

Ken Binmore

Natural Justice: Response to Comments

Abstract: The following responses to the scholars who were kind enough to comment on my *Natural Justice* in this symposium have been kept to a minimum by addressing only issues where I think a misunderstanding may have arisen.

An adequate response to the comments on my book *Natural Justice* offered in this symposium would require writing yet another book. So I plan only to clarify my position where I think it has been misunderstood. However, I should begin with a general apology to those who mistook my book for an attempt at a theory-of-everything in the social realm. My intention was only to follow up a naturalistic explanation of the origins of our sense of fairness all the way through from beginning to end in order to demonstrate that the ideas all fit together in a coherent manner. This task required speculating a good deal about anthropological matters on which we have very little firm data and simplifying a good deal on matters about which we have a great deal of data.

Some of my speculations and my simplifications will doubtless turn out to deserve the criticism directed at them, but the general reaction to my book encourages me in the hope that my basic framework will prove sturdy enough to survive correction on matters where I have gone astray. I certainly do not think that the mathematical modeling of my previous *Game Theory and the Social Contract* that forms a backdrop to the arguments of *Natural Justice* somehow endows my reasoning with an unshakable authority denied to those who use only verbal arguments. The ability to use mathematics when reasoning about the world is a major asset, but the mathematical anecdotes used by game theorists are no less anecdotal because they are expressed in the form of equations.

Anthony de Jasay: *Fairness as Justice*

I had hoped that the fact that I obtain what would normally be regarded as left-wing conclusions from what would normally be regarded as right-wing assumptions might offer a bridge between intellectuals of the left and the right, so it is disappointing to discover that Anthony de Jasay is not open to the idea that fairness norms need not be conceived of as a substitute for the exercise of power but as a means of balancing power.

On his specific comments, let me first say that I do not neglect conventions for bargaining equilibria—whatever the latter may be. In making this criticism, I presume that de Jasay simply means that he does not think the kind of conventions we call fairness norms can usefully be studied using the methods of game theory. His story of ‘After You’ is intended to show that the use of the original position to settle coordination problems is ridiculous, but what then of the convention ‘Ladies First’? As for ownership, he is right that it matters a great deal to us now, but it did not when we were evolving as social animals. It would, in fact, be fascinating to make an anthropological study of how advances in the economic means of production led to the emergence of the convention that some property is to be treated as private. Finally, there is the question of ‘justice as fairness’. Here I leave the field to him, since I am very ready to concede that it may be useful to distinguish the two notions along the lines de Jasay proposes.

Jonathan Riley: *Genes, Memes and Justice*

Jonathan Riley offers a thoughtful summary of the views expressed in my book. He is right to say that the basic argument in *Natural Justice* does not depend on assuming that human beings have an inherently selfish nature. In fact, the argument works with any attribution of consistent personal preferences to the players at all. He is also right that I think the question of what evolution has made of our fundamental nature is an empirical issue to be decided by scientific means. He is right yet again in suggesting that I do not think the available evidence favors a rosy view of human nature. When the chips are down, I think people commonly do give priority to their own well-being and that of their near and dear. I have never made more than passing references to the challenge championed by Gould and Lewontin to the ‘selfish gene’ paradigm, but I think John Alcock’s (1993) *Triumph of Sociobiology* hits the nail on the head when it dismisses their criticism as politically motivated and intellectually dishonest.

I share Jonathan Riley’s enthusiasm for utilitarianism. The argument he offers in its support is another paraphrase of the argument of John Harsanyi that I use in my book. My view is that utilitarianism is what economists call a first-best creed and the egalitarianism to which I give my support in *Natural Justice* is only a second-best creed. A first-best outcome is what you get if you do not have to worry about certain constraints. A second-best outcome is the best you can do taking account of the necessary constraints. In the case of utilitarianism and egalitarianism, I argue that the necessary constraints that preclude the former as a stable social contract for a large society are to be found in a human nature shaped by genes that would not have survived if they had not ‘selfishly’ promoted behavior that gave them better survival chances. I hope that Riley is right in thinking that my Hobbesian view of human nature is mistaken and that it is indeed sufficiently plastic that culture can permanently overwrite what is written in our genes with something more civilized, but I cannot say that the blood-soaked history of the human race offers much grounds for optimism.

Russell Hardin: *The Genetics of Cooperation*

Is my *Natural Justice* a book-length exegesis on Haldane's famous joke that anticipated Hamilton's model of genetic kin selection? I think this is an unfair suggestion, because I repeatedly make it clear that I do *not* think we are genetically programmed to care for the welfare of those outside our inner circle of family and friends. I suspect that Hardin takes me for a member of the new school of behavioral economists of whose views I am a more fierce critic than he is himself.

However, it is true that I believe genetics is important. In particular, I think that all human fairness norms have a common deep structure with which we are genetically programmed, just as we are genetically programmed with the deep structure of all human languages, but this is not to deny the importance of culture. Culture matters no less to the actual form that a fairness norm takes in a particular society than it does to the form that a language takes. Nor have I neglected the views of my hero David Hume on what Hume calls mirroring and I call empathetic identification. On the contrary, most of my modeling is concerned with the problem of actively incorporating the cultural evolution of a person's empathetic preferences into my theory of fairness norms, because I believe that this is where we have to look to understand how we make the interpersonal comparisons of welfare that are intrinsic to any fairness judgment.

I guess I must have pressed the wrong button somewhere to make Russell Hardin think that he and I are not singing from the same hymn sheet. In particular, I am at one with him in endorsing the line of thought he promotes in his paper. The investigation of 'mirror neurons' is particularly interesting, although I found time for only a passing mention of this subject in my book.

Dieter Birnbacher: *Binmore's Humeanism*

I much appreciate Dieter Birnbacher's recognition that my attempts to contribute to social and moral philosophy are in the empirical tradition of the great David Hume. I even try to emulate Hume's writing style. However, it should be noted that I do not slavishly follow Hume on all subjects. In particular, there is a tension in Hume that is evident in the two quotations with which Birnbacher ends his paper.

Is reciprocal altruism between largely self-interested individuals the rock on which human cooperation is founded, or do we have built-in sympathetic preferences that incorporate a substantial concern for the welfare of others outside our inner circle of family and friends? In his *Enquiries*, Hume emphasizes the latter explanation, which is currently being vigorously canvassed by the new school of behavioral economists. However, without denying that nearly everyone is willing to pay a small cost to alleviate the suffering of strangers and a few saints are willing to pay a very large cost, I think that the evidence comes down in favor of the reciprocal altruism that Hume was the first to expound in his earlier *Treatise*. Indeed, although Hume insists that the *Enquiries* are to supersede all

his previous philosophical work, I think it is the *Treatise*—published when he was only 23—in which his true genius shines most brightly.

As for my skeptical comments on metaphysical theories of morality, I understand that they may seem shrill to the European ear, but things are different in America, where the Kantian orthodoxy is so firmly rooted that the mere suggestion that morality might have a relative component is typically greeted with scornful contempt by moral philosophers. Birnbacher thinks it likely that religious folk would be most hostile to a naturalistic approach, but that is not at all my experience. If you have religion, you do not need to lean on the spurious authority of the metaphysical arguments I make fun of in my Chapter 3. It is the metaphysicians without religious faith who are the most outspoken enemies of the kind of naturalism I espouse. My book would doubtless have been better received if I had preserved a dignified silence on their lines of thought, but I feel it would have been intellectually dishonest not to have made it clear what I think of their ramshackle arguments.

Bernd Lahno: *Making Sense of Categorical Imperatives*

I agree with Bernd Lahno's excellent account of why people behave as though there were categorical imperatives. For example, they are socially conditioned to regard keeping promises as a good thing in itself. If this is all it means for a principle of action to be a categorical imperative, then of course categorical imperatives make sense. Nor do I doubt that the real reasons that Immanuel Kant and his followers believed that categorical imperatives exist is because they had themselves been socially conditioned in the manner that Lahno describes so well. However, when I deny that categorical imperatives make sense, I mean only that the way in which they were conceived by Immanuel Kant makes no sense. If rationality is understood as in game theory, it is simply false that any rational being will necessarily act on the maxim that he would will to be a universal law. If rationality is to be construed in some wider sense, what are the foundational principles that are to be applied?

Lahno refers to Rawls' idea of a reflective equilibrium as a way round the lack of any firm foundational principles from which one might seek to justify the claims of rationalist philosophers like Kant. I agree; but the results of such a reflective equilibration cannot evade being conditioned by the cultural biases of those who do the reflection. Any 'categorical imperatives' they may come up with are therefore not categorical in the Kantian sense.

In summary, I think the apparent difference of opinion between Lahno and myself is linguistic. The categorical imperatives that I deny are not the categorical imperatives he defends.

Fabienne Peter: *Justice: Political Not Natural*

I am not sure that I disagree very much with Fabienne Peter—except perhaps on how much we can rely on the reports of heterodox economists like Gintis and Sen

about the extent to which experimental and field data support the hypothesis that people are intrinsically nice.

Peter would presumably agree that the fact that we all belong to the same species implies that we share a certain irreducible commonality. And I agree with her that history shows that this commonality allows an enormous amount of cultural divergence in the social contracts that evolve at different times and places. Peter takes for granted that when such social contracts emerge, we are not always the helpless victims of uncontrollable cultural forces. This also seems obvious to me, and so I agree that the fairness norms which operate in modern societies are sometimes partly political constructs.

I am therefore not hostile to the political perspective on justice in a modern state that Rawls (1993) took up in his *Political Liberalism*. Where I perhaps differ from Peter and Rawls is in thinking that the irreducible commonality we derive from our genetic heritage offers both opportunities and handicaps to those who would like to use political means to reform our current social contract. The handicaps derive from the fact that a social contract that is incompatible with human nature cannot survive in the long run—and hence the demise of all planned human utopias of which I am aware. The opportunities lie in the fact that we can perhaps make political use of the social tools that biological evolution has cast up as flotsam on our social beach—and hence my interest in the possibility that all fairness norms may share a common deep structure that was written into our genes before we became recognizably human and so capable of political activity at all.

When Peter argues that naturalist theories are not written by nature, but are scholarly attempts to reflect on a select set of data about social life, my agreement is therefore uneasy. I think that a successful naturalist theory should be able to survive a scientific confrontation with genuine data, but who would want to claim the same of the typical scholarly attempt to reflect on social life?

Christoph Schmidt-Petri: *Binmore's Egalitarianism*

In his sympathetic review, Christoph Schmidt-Petri points out that my 'utilitarianism' and my 'egalitarianism' differ in major respects from the various creeds discussed under these headings by traditional political philosophers. In fact, even welfare economists are not always happy with my using these terms in what they regard as a positive sense rather than a normative sense. My use of the term "social contract" for a consensus within a society to operate a particular equilibrium from the many available in our game of life excites similar unease in other quarters. My own view is that such differences over the use of language are inevitable if advances are to be made. It is not reasonable to expect innovators to strap themselves into the intellectual straitjackets that traditionalists so often make of the words available to us in the dictionary.

I am grateful to Schmidt-Petri for also pointing out the complexity of the social and psychological considerations that are relevant to the manner in which people make interpersonal comparisons of utility. Even if the general principles

of my theory of how interpersonal comparison works turn out to be correct, he is right to say that there would remain an enormous need for socio-psychological research before the principles could reliably be applied. But I think it is unreasonably optimistic to expect more than general principles at a stage in the intellectual development of the subject when the mere possibility of meaningful interpersonal comparison continues to be denied by many scholars.

Cushman, Young, and Hauser: *The Psychology of Justice*

I am pleased that this symposium has allowed Marc Hauser and his coworkers the opportunity to put their fine research before a wider audience. Their insistence that *process* matters to what we count as just is well taken. In my theory, process becomes important when one asks how we should interpret the phantom coin that is tossed in the hypothetical circumstances of the original position to decide who will occupy what role in society.

If Adam and Eve both require a heart transplant, but only one heart is available, who should get the heart if both are equally worthy? No woman is likely to agree that it is fair that the heart should always go to the man on the grounds that Eve had an equal chance of being male when the egg from which she grew was fertilized. Everybody understands that a new coin must be tossed independently every time the problem arises. It is, in fact, this simple observation that is crucial in deriving the egalitarian bargaining solution from the circumstances of Rawls' original position.

Brian Skyrms: *Ken Binmore's Natural Justice*

Brian Skyrms' summary of the main points of my theory and the weaknesses of some of my assumptions are all very much to the point. The need for further research is apparent. Perhaps it will lead to conclusions substantially different from those I defend in my book. If so, I will still claim some credit for having inspired the new research.

Douglass C. North: *On Kenneth Binmore's Natural Justice*

What can I say to criticism that endorses my objectives but expresses misgivings about my methodology that I share myself? But I do not feel particularly apologetic, since I do not see what else one can do but to make what progress one can with the tools to hand. Without people being prepared to take on this kind of endeavor, it is difficult to see what incentive there would be to develop better tools. I think it particularly important that progress be made with versions of the folk theorem that apply not just in classical repeated games but in the kind of environments that North calls 'non-ergodic'.

Ahlert and Kliemt: *Binmore, Boundedly Rational*

Ahlert and Kliemt argue that my conclusions are speculative, which is a view that I hasten to endorse. In my final chapter, I point to a number of weaknesses in my approach, of which Ahlert and Kliemt are especially concerned with the problem of large societies. I wish I knew how to handle both the coalitional and the monitoring issues that arise in large societies, but, in spite of a large literature, I think these problems remain unsolved.

I am more confident of my approach to bargaining issues. It is true that my use of the Nash bargaining solution is open to question, but how different would the conclusions be if some other paradigm were used? It is very unlikely that the elegance of the conclusions one is able to obtain with the help of the Nash bargaining solution would survive, but it is hard to see how the qualitative nature of the conclusions would change. But perhaps Ahlert and Kliemt are already planning to explore the consequences of replacing the Nash bargaining theory by the theory they prefer—although I hope it is better founded than the theory of Gauthier which they mention in passing (Gauthier/Sugden 1993).

Bibliography

- Alcock, J. (2001), *The Triumph of Sociobiology*, Oxford
Binmore, K. (2005), *Natural Justice*, New York
— (2006), Why do People Cooperate? In: *Politics, Philosophy, and Economics* 5(1), 5–32
Gauthier, D./R. Sugden (1993), *Rationality, Justice and the Social Contract*, Hemel Hempstead
Rawls, J. (1993), *Political Liberalism*, New York