

Brian Skyrms

Ken Binmore's *Natural Justice*

Abstract: I raise a few questions about key points in the argument of *Natural Justice*. 1. The pivotal role assigned to the theory of indefinitely repeated games appears to be both implausible and unnecessary. 2. The evolutionary foundations of the Nash bargaining solution are not completely secure, and its role in the account of interpersonal comparisons of utility is questionable. 3. Free renegotiation behind the veil of ignorance appears neither to have an evolutionary rationale nor to be a brute fact about the way men are.

Ken Binmore quotes Spinoza's *Political Treatise* with approval:

“For they (philosophers) conceive of men, not as they are, but as they themselves would like them to be. Whence it has come to pass that, instead of ethics, they have generally written satire ...” (ch. 1)

In *Natural Justice*, and in the two volumes of *Game Theory and the Social Contract* from which it is drawn, he puts forward a theory of justice designed to treat men as they are. The main tool is game theory, with emphasis on the evolutionary and behavioral branches of the subject.

Here is a list of what I take to be some of the main points:

1. The main problem of the social contract is not that there is no good (Pareto efficient) equilibrium, but rather that there are many. The ‘folk theorems’ of the theory of repeated games demonstrate a infinite number of efficient equilibria. This turns the problem of the social contract into a bargaining problem.
2. Morality has evolved as an equilibrium selection device to enable societies to coordinate on one of these equilibria. In particular, ‘Justice as Fairness’—that is to say rational choice behind the veil of ignorance—is the principle notion of justice that we have evolved for this purpose.
3. In a once and for all setting, Harsanyi has shown that rational choice behind the veil of ignorance yields a utilitarian solution.
4. But in a repeated setting, in which the option to renegotiate the contract behind the veil is always freely available, the only viable solution is egalitarian. Otherwise, those who get the short end of the stick would always opt to renegotiate.

5. Egalitarianism makes no sense if individuals cannot make interpersonal comparisons of utility—but, in fact they do. In the long run, evolution inexorably pushes to the Nash bargaining solution and individuals gradually adjust their interpersonal comparisons, so that their sense of distributive justice coincides with Nash. Thus, in the long run, the Nash bargaining solution calibrates the interpersonal comparisons.

In connection with these points, I would like to raise a few questions. To begin with, the folk theorem for *indefinitely* repeated games does not really fit well in an evolutionary setting notwithstanding a tradition starting with Axelrod and Hamilton (1981) that attempts to use it there. Is there a natural rationale for the geometric discounting used in the theory, or is it simply assumed because we can sum the series? If there is some finite upper bound to the number of interactions as there surely is before reproduction or imitation then the folk theorems fail. Binmore finds the theory of *infinitely* repeated games unrealistic (81), but an infinite horizon is also unrealistic.

One might try to use games of incomplete information, along the lines of Kreps, Milgrom, Roberts and Wilson's (1982) explanation of the possibility of cooperation in the finitely repeated Prisoner's Dilemma, but this would require some discussion of the (cultural) evolution of the beliefs needed to support the Bayes-Nash equilibrium. Something more needs to be said.

Difficulties here are not insuperable. I completely agree with Binmore's main point that the general problem is equilibrium selection among many equilibria; but these equilibria can be created in a lot of ways. The folk theorem for repeated games need not bear all the weight.

The Nash bargaining solution is also asked to bear a lot of weight. Many roads lead to the Nash solution, but some lead elsewhere. We are told about Rubenstein's beautiful analysis of bargaining with alternating offers and discounting, but it is not clear how this fits into an evolutionary analysis. Some of the problems are parallel with those attaching to the folk theorem for repeated games. In addition, Rubenstein solves for a unique subgame-perfect Nash equilibrium; there are many equilibria that are not subgame perfect. But as Gale, Binmore and Samuelson (1995) have shown, evolution may not respect subgame perfection. Furthermore, if the problem is discrete—say divide \$ X down to the penny—there are then lots of efficient subgame-perfect equilibria.

There is a genuine evolutionary rationale for the Nash bargaining solution (Young 1993) as a stochastically stable strategy (Foster/Young 1990). There are two questions to be raised about this. First, as Binmore, Samuelson, and Young (2003) show in an extensive analysis of various bargaining models and learning dynamics, the Nash bargaining solution is often but not always the unique stochastically stable strategy. Secondly, there is the question of expected waiting times which, as Binmore knows quite well, may be very long. Perhaps speedup by local interaction as introduced by Ellison (1993; 2000) is an answer, but again more needs to be said.

Binmore is sensitive to questions about his use of the Nash bargaining solution: "What would happen if some sensible alternative were substituted for

the Nash bargaining solution in my theory? The details of the theory would then become more complicated, but I think it unlikely that there would be substantive qualitative changes in the conclusions." (27) This is not clear to me, given the role of Nash in calibrating interpersonal comparisons. Supposing discrete bargaining problems, and a variety of model situations, so many things are evolutionarily possible that the outcome seems quite indeterminate.

It is taken as something of a brute fact in Natural Justice that 'justice as fairness' has evolved as an equilibrium selection device. No special account of that evolution is offered. I'm not complaining about this. The norm is widespread and a rigorous investigation of it is welcome. But I would like to question whether the possibility of always going back behind the veil of ignorance and renegotiating the social contract has a similar status. This is important because it is free renegotiation that pulls the Rawlsian egalitarian rabbit out of Harsanyi's utilitarian hat (168). Its status as a moral norm does not seem to be a brute fact. The panhandler will not get very far with "Lets renegotiate" instead of "Spare change?" The slave does not gain his freedom by simply asking to renegotiate the social contract. The well-to-do citizen who refuses the panhandler, and even the slaveowner do not believe that they are violating a moral norm. Renegotiation behind the veil of ignorance does not seem to be an entitlement in human moral systems. If this is the way that things really are, natural justice may be closer to Harsanyi than to Rawls.

Bibliography

- Axelrod, R./Hamilton, W. D. (1981), The Evolution of Cooperation, in: *Science* 211, 1390–1396
- Binmore, K. (2005), *Natural Justice*, Oxford-New York
- (1994), *Game Theory and the Social Contract I: Playing Fair*, Cambridge/MA
- (1998), *Game Theory and the Social Contract II: Just Playing*, Cambridge/MA
- /Samuelson L. /Young, H. P. (2003), Equilibrium Selection in Bargaining Models, in: *Games and Economic Behavior* 45, 296–328
- Ellison, G. (1993), Learning, Local Interaction and Coordination, in: *Econometrica* 61, 1047–1072
- (2000), Basins of Attraction, Long-Run Stochastic Stability, and the Speed of Step-by-Step Evolution, in: *Review of Economic Studies* 67, 17–46
- Foster, D./Young, H. P. (1990), Stochastic Evolutionary Game Dynamics, in: *Journal of Theoretical Biology* 38, 219–232
- Gale, J./Binmore, K./Samuelson L. (1995), Learning to be Imperfect: The Ultimatum Game, in: *Games and Economic Behavior* 8, 56–90
- Kreps, D./Milgrom, P. /Roberts, J./Wilson, R. (1982), Rational Cooperation in the Finitely Repeated Prisoner's Dilemma, in: *Journal of Economic Theory* 27, 245–252
- Young, H. P. (1993), An Evolutionary Model of Bargaining, in: *Journal of Economic Theory* 590, 145–168
- (1998), Conventional Contracts, in: *Review of Economic Studies* 65, 773–792