

BOOK REVIEW

Gauthier, D., *Morals by Agreement*, Clarendon Press, Oxford, 1986, 367 pp.

D. Gauthier's *Morals by Agreement* is yet another attempt to spell out the relationship between justice and rationality. The structure of such projects is well-known: it is argued that a set of principles of justice are justified, since they can reasonably be taken to be the object of some kind of *rational choice* within some kind of *choice-situation*. In both respects Gauthier defends an extremely thin construct as the groundworks of a theory of justice.

Rational choice is self-interested choice. It may be agreed that a choice is self-interested if and only if the actor chooses to perform some action because the state of affairs she expects as the outcome of such action is of value to her. But what does it mean to say that a state of affairs is of value to a person? Gauthier takes position against any notion of *objective* value. It is meaningless to say that a person's choice is irrational, because the preferences on which she chooses to act do not match a standard of objective value that is independent of her own preferences. Gauthier believes that a person chooses rationally if and only if she *consistently* chooses to act on her own preferences (whatever they may be). In utility theory these consistency requirements are spelled out such that they afford for an interval scale of utility as a measure of preference. Gauthier adds the requirement for rational choice that these preferences must be *considered* preferences. If this condition is fulfilled, Gauthier equates utility and (subjective) value as a measure of considered preference. He then accepts the notion of rational choice which is prevalent in mathematical economics and game-theory, *viz.* that rational choice maximizes utility as a measure of (considered) preference.

Let us now turn to the choice-situation which is the stage for a rational choice of principles of justice. This choice-situation is hypothetical,

Theory and Decision 24 (1988) 289–293.

though it contains actors that are fully aware of their preferences and capacities. Gauthier argues that rational agreement about prospective interaction can solely come about in an initial situation which is free from coercion. He expands on the Lockean proviso: in the initial situation, no person x is allowed to worsen (i.e. decrease the utility of) any other person y 's situation – relative to y 's situation in the absence of x – unless x 's action is the sole action she can undertake to avoid a worsening of her own actual situation.

In this initial situation the prospect of interaction may take two forms, which Gauthier characterizes by means of game-theoretical concepts. It may take the form of a game with an optimal non-cooperative solution (*viz.* some equilibrium outcome) or it may not take such form. In the *former* case rational actors will simply play this non-cooperative game, since any cooperative solution which differs from the non-cooperative solution must afford a smaller payoff to at least one person and would consequently not be accepted. In this case, justice has no place. In the *latter* case it is rational to enter the cooperative game, since there exists a set of outcomes which afford each player an equal or higher payoff than in the non-cooperative game. In this case, the quest for a principle of justice is the quest for a principle which governs the choice of rational players within the cooperative game.

It can be shown that in a non-cooperative game of *perfect competition* the equilibrium outcome which is the solution to the game is also optimal. However, a game of perfect competition can only be played if certain conditions are fulfilled. I will restrict myself to one such condition here, *viz.* the condition that goods on the market should solely be open for private consumption. Consider however a concrete situation in which the condition of private consumption is *not* fulfilled. Imagine a group of shipowners who would each benefit from the construction of a lighthouse. In a non-cooperative game no lighthouse will be built, unless there is a shipowner for whom the benefits of a lighthouse exceed its construction costs. Now, it may well be the case that the *sum* of the benefits of a lighthouse to each shipowner exceed its construction costs. In such case, the equilibrium outcome in the non-cooperative game yields a suboptimal outcome, since no lighthouse can be built. Rational actors must adopt a cooperative strategy (*viz.* sharing the costs) to reach an optimal outcome which is preferred by all to the equilibrium outcome in

the free market. Equilibrium and optimality part from each other due to the presence of goods that are open for public consumption.

How will the costs for the construction of the lighthouse be shared between rational actors? Clearly no player is willing to pay a share larger than the benefits she expects to derive from the lighthouse, and the sum of all shares should equal the costs of the lighthouse. But these conditions are not sufficient to set a solution to the cooperative game, since it may leave us with many feasible ways to share the costs. Which of these feasible outcomes will rational actors agree upon? Each player has a worst and a best possible outcome. The worst possible outcome is set by the proviso, i.e. no lighthouse is constructed and there are no costs. Furthermore I am indifferent between this outcome and the outcome on which I pay a share for the construction of the lighthouse such that my costs equal my benefits. On the best possible outcome the lighthouse is constructed at minimal costs to myself. Between the worst and the best possible outcome is a set of outcomes on which I pay a progressively smaller share of the costs for the construction of the lighthouse. To each of these outcomes I then assign utilities that are defined up to an interval scale. The Kalai–Smorodinsky solution to the bargaining game predicts that rational bargainers will agree upon the Pareto-optimal point which affords each player an *equal* proportion of the utility difference between their worst and their best possible outcome. (Gauthier slightly modifies this solution to a *maximin* proportion, since A. Roth has shown that the Kalai–Smorodinsky solution does not work for the *n*-person game). Rational actors thus come to agree to a solution in a cooperative game in accordance with what Gauthier names the *Principle of Maximin Relative Benefit*. Any cooperative decision in society which satisfies this principle is thus a just decision, since it is the solution which rational bargainers would come to agree to in a fair initial choice-situation. And thus the principle of maximin relative benefit is the central principle of justice for society at large.

So far I have spelled out what I take to be the core of Gauthier's theory of justice. After this charitable reading I will now turn to some critical remarks. Traditionally, the justification for bargaining models is couched either in strategic terms (i.e. in reference to self-interest) or in normative terms (i.e. in terms of some conception of fairness). Gauthier tries to bridge this gap in claiming that fair bargaining is nothing but strategic

bargaining starting from a rationally acceptable initial situation. I am sceptical about this project for the following reasons. I believe that neither the initial situation, nor the bargaining solution in Gauthier's construct can be justified on the basis of strategic considerations. Furthermore, it seems to me that this construct, due to its reliance on subjective value or utility (as a measure of considered preference), yields results that clearly do not match common intuition, concerning fairness.

Consider the initial situation. It seems to me that the demand for a non-coercive initial situation cannot be justified on strategic grounds, but requires some appeal to our intuitions concerning fairness. A coercive situation is not acceptable, solely because rational bargainers have some *independent* notion of what counts as a fair initial situation. However, I believe that a justification from fairness cannot hold as well, because Gauthier understands the notion of 'worsening a person's situation' in terms of subjective value. It does not seem to match our intuitions concerning fairness that I may search for an imprudent bargaining partner to afford a better initial situation for myself, or that, in bargaining about a policy on visitors, I ought to take into account the harm I cause to my racist roommate by asking about foreigners.

Concerning the bargaining solution, I do not see why in strategic bargaining, a person would take any interest in the maximal claim of her opponent, since this claim is most often a wildly unreasonable demand. (I believe Nash's classical solution and Kalai's more recent egalitarian solution yield a much more plausible account of strategic bargaining than the Kalai-Smorodinsky solution which Gauthier appeals to) (Bovens, 1987). Furthermore, Gauthier's bargaining solution does not match our intuitions concerning fairness. Let us imagine Rich and Poor are bargaining from some fair initial situation about \$ 100. If we set Rich's utilities linear with money, it is reasonable to assume that Poor's preferences are expressed by a utility-function with decreasing marginal utility. Now does it match our intuitions concerning fairness that Rich runs off with the largest share as is the case on Gauthier's bargaining solution? Indeed, in *strategic* bargaining a person who is strongly in need of a good may be coerced into settling for a bad deal. Gauthier attempts to rule out coercion in the initial situation, but is it not troublesome that his bargaining solution precisely captures the coercion which is due to the relative structure of the utility-functions of the bargainers?

There is both empirical work (Yaari and Bar-Hillel, 1984) and normative models of fair bargaining (Barry, 1979 and Roemer, 1986) which put into question that the preferences of the bargainers can be a sufficiently rich informational basis for setting a fair bargaining solution. I believe a confrontation of such views with Gauthier's theory may prove rewarding.

REFERENCES

- Barry, B.: 1979, 'Don't Shoot the Trumpeter – He's Doing His Best', *Theory and Decision*, **11**, 153–180.
- Bovens, L.: 1987, 'On Arguments from Self-Interest for the Nash Solution and the Kalai Egalitarian Solution to the Bargaining Problem', *Theory and Decision*, **23**, 231–260.
- Roemer, J. E.: 1986, 'The Mismatch of Bargaining Theory and Distributive Justice', *Ethics*, **97**, 88–110.
- Yaari, M. and Bar-Hillel, M.: 1984, 'On Dividing Justly', *Social Choice and Welfare*, **1**, 1–24.

*Research Assistant in the Belgian
National Fund for Scientific Research
University of Minnesota,
Department of Philosophy,
Ford Hall 356,
Minneapolis, MN 55455,
U.S.A.*

LUC BOVENS