

THE PROBLEMS OF DAVID GAUTHIER'S ATTEMPT TO DERIVE MORALITY FROM RATIONALITY



GARETH WILLIAMS



In *Morals by Agreement* (1986), David Gauthier attempted to develop a theory of morality based on mutuality and grounded in individualism. In so doing, he sought to demonstrate the rationality of the development and acceptance of a moral code without dependence on the institution of a coercive leviathan. His theory is of interest to moral philosophers, and to all who take individualism seriously.

This paper will consider how Gauthier, without depending on pre-conceived moral inputs, manages to attempt a construction of moral theory through contractarian methodology. His theory of compliance, bargaining theory, proviso of non-coercion in the initial bargaining position, and the claim that the agreements reached through such bargaining can be considered moral, will all be assessed for coherence and plausibility. It will be argued by this paper that Gauthier's bargaining theory presents a plausible case for how rational individuals would seek to bargain, and that a contractarian theory of the type advanced by Gauthier can have a valid claim to being a system of morality. It will then be asserted that the theory of compliance lacks credibility due to implausible assumptions of translucency of character, the inability of dispositions to assume binding power, and the possibility of the introduction of other dispositions into the choices available. It will be maintained that the rationality of being a constrained maximiser, like that of narrow compliance, can only be justified from a collective perspective and thus is incapable of being reconciled with individual rationality. The proviso of non-coercion in the initial bargaining position will be argued to have only limited rationality, with examples of its irrationality being discussed.

A MORAL THEORY BASED ON THE RATIONAL SELF INTEREST OF INDIVIDUALS AS THEY ARE

Gauthier's objective is to construct a theory capable of answering Nietzsche's question as to where morality comes from; he declares that if he fails in this task then morality is a chimera. Perceiving traditional attempts to establish morality as lacking in foundation, Gauthier sets out to develop a moral theory "not of absolute standards, but of agreed constraints"¹ which can justify moral conduct to the rational self interest of all individuals as they now are. The construction of his moral theory is based firmly upon the rational behaviour of individuals and, as such, rejects any distinction between prudential and moral reasoning, arguing that a person can have no reason for accepting constraints which are independent of their interests. So, for Gauthier, the

constraints which individuals voluntarily impose upon themselves are the basis, and the only basis, upon which morality can be based. In contrast to Rawlsian theory, which attempts to identify the constraints individuals would choose when unaware of their own attributes and tastes, Gauthier's theory must rationally justify constraint to people as they now are.

In a perfectly competitive market, in which all costs and benefits of actions were internalised onto those responsible for them, there would exist a moral free zone. The existence of 'free riding' in a pre moral market place creates a 'prisoners dilemma', which prevents co-operative interaction from taking place and thus creates sub optimal conditions for all. The scope for morality exists in these notions of market failure which suggest that a co-operative surplus could be developed from which everyone could benefit if compliance with agreements made could be ensured. The gains from the co-operative surplus are what entice each individual to desire a co-operative framework of interaction, and thus are what is responsible for the creation of a code of morality.

CONSTRAINED MAXIMISERS AND STRAIGHTFORWARD MAXIMISERS

Gauthier differs from Hobbes by advancing a strong, rather than a weak, theory of compliance. To Hobbes the rationality of agreeing a moral code for each to uphold is dependent upon the presence of external enforcement mechanisms; whereas Gauthier asserts that to reap the benefits of co-operation, one must agree to the development of a morality and develop a disposition to abide by the agreed moral code. The development of a disposition to uphold the moral code is made essential, in order to gain inclusion into co-operative arrangements, by the translucent nature of ones character. Gauthier asserts that one's character is neither opaque nor transparent but translucent, by which he means that others can, in most cases though not with certainty, determine one's disposition to uphold or defect from agreed strategies in a prisoner's dilemma type of co-operative venture. The disposition to comply, held by constrained maximisers (CMs), will enable one to join with others of a similar disposition to reap benefits of co-operative interaction. The disposition to defect from agreements made will enable one to gain from defection when one's disposition goes undetected, but will exclude one from the benefits of co-operative interaction with those who detect one's disposition. So Gauthier's theory of compliance applies only to agreements between constrained maximisers, and does not bind CMs into entering co-operative ventures with those who will defect (labelled straightforward maximisers or SMs). So Gauthier can claim that the disposition to constrained maximisation can enable one to engage in co-operative interaction while managing largely to avoid the problem of free riding by SMs.

The ability of CMs to detect the disposition of others will increase as they develop their ability to recognise other CMs for co-operation, and to identify SMs so as to avoid being exploited by them. Likewise SMs will learn to detect CMs, whose dispositions leave them vulnerable to exploitation. CMs will also try to increase their own transparency so other CMs will be able to identify them, but this has the negative effect of enabling SMs to exploit them better. Gauthier assigns letters p and q to an equation considering the merits of being a CM or SM. The letter p is the probability that CMs will be able to detect one another and reveal their own sincerity to them, q is the ability of CMs and SMs to conceal their own sincerity and insincerity (respectively) from one another. While any increase in a CM's ability to show their sincerity to other CMs is likely to be offset by their ability to hide it from SMs, CMs will tend to rely on, and develop, their ability to detect the disposition of others. As people improve their ability to detect others dispositions the value of p increases and q can be expected to remain relatively constant. While CMs benefit from increasing p and decreasing q then the increase of p/q must benefit CMs. The more favourable the conditions for CMs the more people will develop the CM disposition, so its dynamic nature makes conditions ever better for constrained maxi-



Philosophical Notes No. 51

ISSN 0267-7091 ISBN 1 85637 426 2

An occasional publication of the Libertarian Alliance,
25 Chapter Chambers, Esterbrooke Street, London SW1P 4NN
www.libertarian.co.uk email: admin@libertarian.co.uk

© 1998: Libertarian Alliance; Gareth Williams.

Gareth Williams is doing an MA in Philosophy at York University.

The views expressed in this publication are those of its author, and not necessarily those of the Libertarian Alliance, its Committee, Advisory Council or subscribers.

Director: Dr Chris R. Tame Editorial Director: Brian Mickelthwait
Webmaster: Dr Sean Gabb

FOR LIFE, LIBERTY AND PROPERTY

misation ensuring that there are more CMs for a CM to co-operate with and fewer SMs for a CM to be exploited by.

With justification for the theory of compliance resting upon the ability of others to ascertain our disposition Gauthier is asserting the rationality of a disposition to comply with agreements irrespective of future opportunities for interaction. Thus Gauthier perceives his theory as offering a rationale for moral behaviour independently of both the possibility of 'tit for tat' repercussions from non-compliance, or of any desire for the acquisition of a reputation for honesty and compliance which may be beneficial to an agent's prospects of social interaction.

MINIMAX RELATIVE CONCESSION

The disposition to comply with agreements reached is dependent not only on the agreements being an improvement upon one's current condition, but also fair. Gauthier draws the distinction between a broadly and narrowly compliant person. The former will comply with any arrangement which improves his condition, but the latter only with those arrangements which improve his condition from a fair baseline. This fair baseline is portrayed as one in which an individual was not made worse off by the other person prior to the bargain, and is exemplified in a hypothetical example of interaction between slaves and their masters. In the example given, Gauthier's hypothetical slaves agree to continue serving their masters if they are not coerced, which spares the slaves the cost of being coerced and the masters the cost of coercing them. Following the removal of coercion the slaves rebel and cease to serve their former masters. It is asserted that a deal made from a position of "coercion is not an agreement, and ... lacks any inherent stability".² Gauthier asserts that people will only recognise as fair deals which they were not forced to accept due to their coerced position, and that upon removal of the coercion they will cease to have reason to comply with the agreements made. Developing a disposition contrary to this would encourage others to coerce one prior to bargaining, in order to strengthen their bargaining positions. Thus, it is asserted, rationality requires one be disposed to comply with agreements only from a non-coercive baseline.

Gauthier constructs his own bargaining theory, called one of minimax relative concession. When bargaining, individuals have no constraints over their utility maximising behaviour, and each, being equally rational, claims the full amount of the co-operative surplus for himself. To claim any more than the full surplus would be to make a claim which worsened the condition of another and thus would drive them away from the bargaining table. To claim any less would be to undermine one's position in the negotiations ahead. Each individual wants the bargaining to succeed and for himself to do as well out of it as possible, and if the bargaining fails they all continue straightforward maximisation. From each claiming the full fruits of co-operation for himself (impossible to fulfil) each makes concessions until an attainable outcome is reached. This process of concession takes the form labelled minimum relative concession, or minimax relative concession (MRC), in which each tries to maximise the sum of what he gets divided by what he had asked for. If one's relative concession is greater than the concession of another then one's ground for complaint is also greater; minimising the maximum relative concession minimises the ground for complaint. The minimisation of relative concession is selected over absolute concession, since, for individuals with different initial claims, concessions of the same absolute value would leave the parties to the bargain with different levels of incentive to reach agreement, and with correspondingly different levels of complaint with the proposed deal.³ Thus Gauthier is able to claim that minimax relative concession is the only solution acceptable to people of equal rationality.

It seems that Gauthier is able to logically advance his theory of bargaining as the rational option for individuals to use in negotiation, and as an intuitively fair deal when reached. The product of MRC is intuitively fair, as no one is expected to make a con-

cession which another would not be prepared to make from the same situation. As such the theory would be appealing to rational self interested individuals, and is the only bargaining theory that truly embodies the principle of impartiality. MRC also benefits by, unlike the Nash solution, avoiding the problems associated with interpersonal comparisons of utility, by the use of a scale of measurement independent of utility. By separating bargaining from utility Gauthier can also claim to be taking the distinction between individuals seriously, and providing each individual with an impartial bargaining theory suitable to all.

A "MORAL THEORY FOR ADULTS"

The contractarianism advanced by Gauthier does "not afford each individual an inherent moral status in relation to her fellows"⁴ and as such allows "Animals, the unborn, the congenitally handicapped and defective, [to] fall beyond to pale of a morality tied to mutuality".⁵ So Gauthier develops a theory which claims to develop a morality from directly the opposite precepts to those more conventionally used in moral theory. Kymlicka protests that, while it may not be advantageous for the strong to refrain from exploiting the weak, the weak have a prior claim of justice to prevent the strong from doing so. To deny this is "a hollow mockery of the idea of justice — adding insult to injury".⁶ The thrust of Kymlicka and Barry's work is to identify, and provide justification for, transcendent moral claims which each individual can plausibly be entitled to make, these claims being grounded in intuitive moral convictions and impartiality. To understand Gauthier, one must clearly distinguish his project from that of people like Barry and Kymlicka. Gauthier aims to develop a "moral theory for adults"⁷ which can "be justified to people as they now are".⁸ Gauthier sees a failure of conventional moral theory to provide an answer to Nietzsche's question as to where morality comes from, so that it becomes necessary, when starting from no moral convictions, to ground moral theory in the voluntary acceptance of the rational restraint of individuals. The failure of moral theory to provide adequate, non-sentimentalist justification for *a priori* moral claims, which real people can have reason to follow, renders its project unsuccessful. Morality is, this paper asserts, a system of accepted, and impartial rules governing human relations. Gauthier claims to be able to demonstrate how rational individuals would agree upon, and adhere to, a system of rules of conduct which are impartial in nature. So if Gauthier's claim is validated by our investigation then a system of morality can be said to be generated from precepts of mutual advantage. Our investigation will however conclude that flaws exist within Gauthier's thought rendering his theory of compliance implausible, and his proviso of non-coercion in the initial bargaining position as having only limited applicability.

Rationality of compliance is founded on notions of translucency of character and the supposed ability of humans to dispose themselves towards rules of conduct which will then be adhered to irrespective of how one's interest changes during the course of interaction. Translucency of character endows one's potential partners in interaction with the ability to predict what you are likely to do, their ability to predict what you are likely to do being what makes it essential that one's disposition be conducive to interaction. If successful, cooperative interaction produces a cooperative surplus from which all can benefit. The paper will now consider how plausible the notion of a translucent character is, the extent to which a disposition can be developed, and the rationality of developing a disposition to be a constrained maximiser.

HOW COULD DISPOSITIONS BE ASCERTAINED?

In constructing the notion of translucent character Gauthier can rely for support upon our intuition of our ability to be able to detect with reasonable certainty the degree of honesty, or otherwise, of the person we are interacting with. Feelings of trust or distrust are common in everyday life, and one tends to think that these feelings are, in most cases, largely correct.

Such feelings must however be distinguished from the claim of translucency. The desire which people have to acquire a reputation for honesty, or the desire to avoid negative repercussions from future interaction, must both be distinguished from the development of a disposition which, for Gauthier, stands independent of these. The disposition which Gauthier talks of when referring to constrained maximisation is the tendency to carry out one's agreed plan of action irrespective of the consequences. This disposition is displayed in the case of 'one off' interactions, where one's conduct will not affect one's prospects of future interaction. Given that, as Gauthier accepts, straightforward maximisers will endeavour to portray themselves as CMs, the possibility for gains from free riding will continue to exist. With self-interest being the motivating factor for both SMs and CMs, the difference which distinguishes each is not a difference of benevolence versus egoism, but just a different strategy for self interest. The difference in character between CMs and SMs has to be small for Gauthier to assert the possibility that one can achieve transition from being a SM to being a CM and be able to convince others of one's sincerity. If character difference is so small, it is difficult to see how SMs, who depend for their success upon successful predation, would be unable to develop the ability to portray themselves successfully as CMs.

This being the case it is difficult to see how dispositions could be ascertained by others with high levels of accuracy. If disposition cannot be so ascertained, the only feasible way in which it can be asserted that dispositions can be rationally developed is through iterated interaction or the development of a reputation for certain types of conduct. For Gauthier to adopt either argument would destroy his ability to argue for the rationality of compliance in the case of non-iterated interaction. If this cannot be argued, then the compliance rational individuals adopt with regard to the rules which Gauthier labels moral is purely conditional upon perceptions of their self-interest and thus cannot be relied upon. So it would seem that the arguments Gauthier could use to defend rationality of compliance lack the universal applicability he desires, while the argument of transparency of character lacks the plausibility it needs to provide a convincing theory.

COMPLIANCE WOULD NOT ALWAYS BE THE RATIONAL OPTION

It is equally difficult to see how the development of a disposition is truly possible. It would seem that acquisition of the disposition to pursue Constrained Maximisation would require the individual to undergo hypnosis, or to have their mind in some other way pre-programmed, to only conduct certain types of action. Gauthier's argument expressly rules out the need for externally imposed constraints on action, and nothing in the argument suggests a comprehensive loss of free will to the extent that we no longer control our actions, in short the disposition of which Gauthier writes "is not a precommitment of the kind which bound Ulysses to the mast".⁹ The rational individuals who select constrained maximisation as their chosen lifestyle retain the ability to convert to straightforward maximisation at any time. If they are truly rational it is hard to imagine how in all cases compliance would remain the rational option.

In Gauthier's defence attention could be drawn to the actions of aeroplane hijackers, who, to be successful in their objective, must convince others of their disposition to carry out action detrimental to their own welfare. While it is not disputed that their success is dependent upon others believing in their having this disposition, it is neither clear that it is individually rational to acquire the disposition, or that acquisition of the disposition is actually needed. For hijackers who are seeking a ransom payment for their own use, action which results in their death is highly irrational; what may not be irrational is to convince their bargaining opponents of their irrationality or of their lack of desire to live.

Assuming that (contra this paper) Gauthier's argument concerning transparency of character is valid, it remains the case that at the point at which one comes to blow up the aeroplane, or kill one's

last hostage, one has failed in one's task and one is now acting irrationally. Rational self-interested agents would not act in such a way, although they might pretend until the last moment, possibly even convincingly, that they were not rational and would act in such a manner.

Politically inspired hijacking presents a more interesting case. The infliction of action causing self harm could, in this case, be rational from a collective perspective. The group does not die with the death of a few of its members, and so, providing it will continue to possess the ability to impose similar threats in the future, the aims of the group could conceivably be advanced by such individually irrational behaviour. So it seems that in order to make constrained maximisation a rational option for all cases some element of collective benefit needs to be introduced into the preferences of those who adopt it, in this way the benefits from the upholding of the principle will inspire the individual even when the action causes their own death. Gauthier is however keen to ensure that his theory does not depend upon feelings of benevolence, as he asserts these cannot be relied upon. The whole *raison d'être* of Gauthierian theory is the avoidance of problems relating to free riding and parasitism in the classical prisoner's dilemma situation, if the development of a notion of the common good could be instilled into humans in the real world these problems would not exist and there would be no need for Gauthierian theory.

GAUTHIER DOES NOT WISH TO DEPEND UPON ALTRUISM

The rationality of the acquisition of the CM disposition is, correctly, described by Gauthier as becoming rational only when a threshold number of CMs exist. Without a sufficient number of CMs, one's conversion to the CM disposition would risk one enduring too much exploitation from SMs who had escaped detection, and would hold little benefit as too few CMs would exist to provide sufficient opportunities for the production of a co-operative surplus. How the threshold number of CMs is attained is not made clear by Gauthier, which is unfortunate as the question is crucial to demonstrating the rationality of the individual acquiring the disposition. It could be argued that such behaviour could develop through bonds of mutual concern within small groups of people, and gradually spread throughout the population.

The problem for such an argument is that although he acknowledges its presence, Gauthier does not wish to depend upon altruism for the construction of his theory. Alternatively a threshold number of CMs could be developed if a sudden intellectual shock jolted a large number of people into a sudden, and simultaneous, realisation as to the demands of rationality; or else, and along similar lines, a select number of people, who were prepared to adopt the CM disposition, decided to form an initially insular community to allow constrained maximisation to work. The latter option is plausible and the community could expand, or weaken its insularity, as SMs on the outside converted to become CMs so as to benefit from the fruits of the co-operative surplus.

Still the rationality of the acquisition of the CM disposition is dependent upon how society is constituted. Holly Smith discusses the possibility of the introduction of other types of actor into Gauthier's model society, which could, if enough of them were present, make the CM disposition untenable or irrational.¹⁰ One of these is labelled KM, and has the disposition to build if you build, or not to build if you do not build, but then doesn't build whatever you do. Such an actor would be impossible for a CM to detect as their initial dispositions are identical, and while in everyday life it is possible to see good intentions going unfulfilled, it is not implausible that such a character could exist. Likewise, if unconditional co-operators (UCs), who would co-operate irrespective of your disposition or actions, were to exist, then it would be irrational not to adopt a policy of straightforward maximisation towards them. UCs possibly do exist in the form of people inspired by religious beliefs, or otherwise altruistic characters.

The other option Smith mentions is one of radical co-operators (RCs) who would only cooperate with those who had chosen the UC disposition. While this disposition is not claimed to be rational, it could, if enough people adopted it, make being anything other than a UC irrational. So it seems that the disposition which it is rational to choose depends upon the characters of those alongside whom one must live, and that the choice of potential dispositions they, or you, could have is greater than that of just being an SM or a CM.

NARROW OR BROAD COMPLIANCE

The disposition to adhere to covenants made, which characterises constrained maximisation, is qualified in Gauthierian theory by a proviso of non-coercion in the initial bargaining position (IBP). The proviso of non-coercion dictates that an agreement should only be respected if it is agreed in an environment in which neither party to the agreement is made worse off by the others' presence. The distinction thus drawn is between a disposition of narrow or broad compliance, the latter will be disposed to respect all agreements which better ones condition whatever the IBP was like, the former only respects those agreements formed from a non-coercive IBP. To be disposed to a constrained maximisation which leaves the fruits of past predatory behaviour untouched is, Gauthier claims, to invite others to coerce you prior to bargaining. The disposition of narrow compliance will make all realise that they cannot benefit, in bargaining, from past predation. Thus it is asserted that narrow compliance is the rational option.

There exists value within this argument, although it seems likely that many instances will exist in which narrow compliance is distinctly irrational. The value of the argument lies in its correct assertion that it often is rational not to agree with a deal from a position of exploitation which may better one's conditions, if so doing will encourage others to exploit one in the future. This is clearly the position taken by governments towards terrorism, and is generally thought to help remove incentives for terrorist activity. The problem with Gauthier's argument with regard broad/narrow compliance includes its failure to acknowledge that sufficient incentive to coerce may continue to exist even when the coercers are deprived of increased strength in the bargaining position, and its failure to clearly distinguish between collective and individual rationality.

A disposition to narrow compliance may prevent others from reaping benefits from our exploitation via the bargaining process, but they can still obtain benefits from our exploitation which may induce them to exploit us anyway. Gauthier's answer to this is to claim an inherent instability to coercive relations, which suggests that coercion cannot become a permanent feature of civilisation because of the possibility of revolution, this argument is also flawed. If the basis of morality is individual rationality, then, for those who cannot wait for a revolution which may take many years or generations to occur, the rational option may well be to negotiate now. Thus in Gauthier's slave/master society¹¹ an argument can be rationally put for the collective rationality of narrow compliance, but it would be difficult to explain the rationality of narrow compliance to the individual who may not live to see the revolution. The disposition to maintain narrow compliance whatever the consequences, could only exist in a person who had a benevolent concern for the common good or/and future generations; in this respect the disposition to narrow compliance suffers from the same flaw as that of constrained maximisation.

In claiming the power of each individual to select dispositions, and thus asserting that these dispositions are not innate to human character, Gauthier highlights a further problem with his argument for narrow compliance. Persons who have not yet acquired a disposition of narrow compliance (e.g. young children) have no ability to deter enslavement from a rational actor who wishes to attain the master servant rather than master slave relationship through the bargaining which Gauthier discusses. Having not selected a disposition prior to enslavement the enslaved person is faced with the option as to which disposition to obtain, if the

assumption of translucent character is correct, those who select broad compliance will enjoy a better life as a servant, those who select narrow compliance will continue to suffer as slaves. It is difficult to see how, in this circumstance, narrow compliance can remain, for these people, a rational disposition to select.

Gauthier claims that a disposition of narrow compliance would avoid others engaging in predatory behaviour towards one. This assumes, contrary to this paper, that dispositions can be given absolute power and can be detected with a high level of certainty by others. Taking these assumptions to be true, there can still be real cases in which one would suffer considerably by this disposition. As Sugden points out,¹² it is in the nature of war that the most efficient way of dealing with prisoners of war is to kill them. If one is a prisoner disposed to broad compliance, and one's character as a CM can be detected by one's captors, then one could negotiate release and a safe passage home in return for cash to be given later to one's captor. A narrow complier will be shot.

PROBLEMS HAVE BEEN FOUND

To conclude, this paper has found that Gauthier's assertion of the existence of no distinction between moral and prudential reasoning is correct, and thus that an attempt to develop moral theory from precepts of mutual advantage is a valid project. Likewise Gauthier's bargaining theory has been argued to be both acceptable to the rational individual and an acceptable account of the real bargaining process.

Problems have however been found in the theory of compliance and proviso of non-coercion in the IBP. The theory of compliance is critiqued due to the implausible assumption of translucency of character, the impossibility of the acquisition of a disposition that actually binds one to comply with agreements made, the individual irrationality of complying with agreements as soon as the other party has fulfilled their side of the bargain, and the apparent irrationality of the CM disposition if other dispositions were introduced into the equation. It was argued that the CM disposition, like that of narrow compliance, could only achieve the rationality Gauthier ascribes to it if an element of collective concern were injected into it. Gauthier does not wish to depend upon altruism for his theory as if it could be depended upon externalities would not be a problem, a moral free zone would exist, and contractarian theory would not be needed.

Likewise a noncoercive IBP was argued to lack universal individual rationality with the adoption of narrow compliance only being individually rational in certain cases. It was argued that the incentive to exploit is not removed by the adoption of narrow compliance, and that the instability of coercive relations may not produce a revolution in time for all of the coerced to benefit from it. As such it can, for many people, be rational to adopt a disposition of broad compliance, and thus no reason can be said to exist to morally condemn compliance with agreements made from a coercive IBP.

NOTES

1. Gauthier, D., *Morals By Agreement*, Oxford University Press, New York, 1986, p. 2.
2. Gauthier, D., in Vallentyne, P., *Contractarianism and Rational Choice*, Cambridge University Press, New York, 1991, p. 29.
3. See p. 138 in Gauthier, D., *Morals By Agreement*.
4. Gauthier, D., as cited in Kymlicka, W., *Contemporary Political Philosophy*, Oxford University Press, New York, 1990 p. 222.
5. Gauthier, D., *Morals By Agreement*, p. 268.
6. Barry, B., as cited in Kymlicka, W., *Contemporary Political Philosophy*, p. 131.
7. Gauthier, D. and Sugden, R. eds., *Rationality, Justice and the Social Contract*, Harvester Wheatsheaf, Hemel Hempstead, 1993, p. 2.
8. *Ibid.*, p. 11.
9. *Ibid.*, p. 45.
10. Gauthier, D., in Vallentyne, P., *Contractarianism and Rational Choice*, Cambridge University Press, New York, 1991, pp. 235-243.
11. See Gauthier, D., *Morals By Agreement*, pp. 190-2.
12. Gauthier, D. and Sugden, R. eds., *Rationality, Justice and the Social Contract*, p. 21.