The Resurgence of the Foole

THEODORE S. ZENZINGER
The University of Kansas

Introduction

In Chapter 15 of *Leviathan* Hobbes subjects his Third Law of Nature, "That men perform their covenants made,"[1] to an important test. If Hobbes can not defeat the Foole, who "...questioneth whether injustice...may not sometimes stand with that reason, which declareth to every man his own good...,"[2] the project of *Leviathan* will fail. For, the Foole, in challenging the Third Law of Nature, challenges Hobbes's contention that society can be viewed as the product of a rationally motivated contract between individuals. If the Third Law is invalid, contracts will not be kept, and not being kept, will not be entered into. Meeting the Foole's challenge is thus of critical importance to Hobbes, as well as those with similar contractarian projects.

This essay will be an examination of three attempts to defeat the Foole. In Section 1 I present the Foole's challenge. Section 2 contains an examination of Hobbes's attempt to meet the challenge along with Gauthier's critique of that attempt. Section 3 details Gauthier's reply to the Foole. Section 4 details MacIntosh's critique of Gauthier. Finally, Section 5 critically examines MacIntosh's preferred reply to the Foole.

Section 1: The Foole's Challenge

Hobbes argues that individuals in the State of Nature will rationally choose to form a commonwealth and appoint a sovereign. Each individual, facing the state of war, finds it to be in their own interest to participate in such an agreement. However, that same self-interest would seemingly lead them to violate the agreement when to do so is to their advantage. The tension evident here is nicely modelled by the Prisoners' Dilemma.

Suppose that hobbesian individuals are attempting to decide whether or not to keep the state of war-ending agreement or violate it. It is plausible to suppose that such individuals will most prefer (based on considerations of self-interest) taking advantage of their fellows with impunity; will prefer second a state of peace which places equal restraints

---

1 Hobbes, Thmas. *Leviathan*, . (B. Macpherson, ed.) Harmonds Worth: Penguim Books, 1978, p.203.
2 Hobbes, Thmas. *Leviathan*, . (B. Macpherson, ed.) Harmonds Worth: Penguim Books, 1978, , p.204.

on all; will prefer third the continued state of war; and will least prefer abiding by the agreement-imposed constraints and being taken advantage of by one not so constrained. If we focus on two such individuals, we can capture the situation and the stated preferences in the matrix:

	Keep Agreement	Violate Agreement
Keep Agreement	(2,2)	(4,1)
Violate Agreement	(1,4)	(3,3)

(figure 1)

Although the Foole does not formulate his/her challenge in the vocabulary of rational choice theory, we can use the Prisoners' Dilemma (hereafter "PD") to illustrate its force. To do this, we need to consider a special instance of the PD.

The PD illustrated in figure 1 tacitly assumed that the players would choose simultaneously, but this need not be the case. Suppose, instead, that row chooses first followed by column. Such a situation can be depicted like this:

keep .... keep   (2,2)

C .... violate   (4,1)

R

violate .... keep   (1,4)

C .... violate   (3,3)
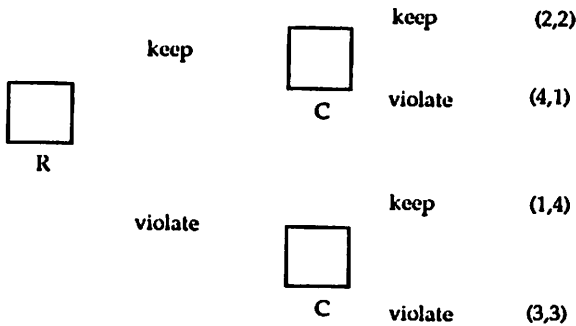
(figure 2)

Suppose that row has chosen "keep". What should column do? The Foole challenges Hobbes to prove that choosing "keep" is the rational choice for column in such a situation as this. It is not obvious that "keep" is the rational choice. Row, knowing column's preferences cannot help believing that if she chooses "keep", column will choose "violate", hence, she will

choose "violate" at the outset. This forces column to choose "violate" in self-defense and thus the players arrive at the outcome (3,3)–the point of non-agreement. (A similar argument demonstrates the rationality of choosing "violate" in the simultaneous move PD.) If row had reason to believe that if she chose "keep" column would choose "keep", the Third Law of Nature would be proven valid.

That this is the challenge as Hobbes perceives it is indicated when he writes,

> But either where one of the parties has performed
> already, or where there is a power to make him perform;
> there is the question whether it be against reason, that is,
> against the benefit of the other to perform or not.[3]

Hobbes concludes the preceding thought with: "And I say it is not against reason." It remains to be seen whether or not he is correct in this claim.

Section 2: Hobbes and Gauthier

Hobbes responds to the Foole in paragraph 5, Chapter 15 of *Leviathan*. Briefly he argues:

(1) Individuals act to preserve themselves.
(2) In the State of War the lives of individuals are endangered.
(3) Only the having of confederates and the formation of confederacies make survival possible in the State of War.
(4) Violators of agreements will not be allowed into confederacies and will be denied confederates.
(5) Therefore, those who violate agreements are not acting to preserve themselves.
(6) But such an action violates our initial assumption (1).

In terms of the game depicted in figure 2, Hobbes is arguing that individuals will look beyond its immediate boundaries. Instead of deciding between "keep" and "violate", hobbesian individuals also examine what effect their choice will have on other situations and, ultimately, on their chances of survival. Faced with expulsion from or denial of entry into confederacies, individuals will not have the preferences ascribed to them in figure 2. Instead, the situation will look like this:

---

3 Hobbes, Thmas. *Leviathan*, . (B. Macpherson, ed.) Harmonds Worth: Penguim Books, 1978,, p.204.

(figure 3)

In this situation row need have no qualms about choosing "keep" for if she does, she knows that column will also choose "keep": given his preferences he has no incentive to choose "violate". Thus, hobbesian individuals can reach the agreement necessary to establish the commonwealth; and moreover will have no incentive to violate agreements once there is a commonwealth. The presence of the sovereign guarantees that any violation of an agreement will be subject to punishments which will generate in individuals the same preferences as those generated by fear of the State of War.[4] We can use figure 3, then, to express the situation of hobbesians both within the commonwealth and outside the commonwealth.

---

[4] Although this reading of Hobbes provides an answer to the Foole, it is not without attendant difficulties. If individuals are rational enough to follow this line of reasoning, it might be argued that they need no sovereign to watch over them. Moreover, one might justifiably ask if rational individuals, in a state of nature, will move inexorably to a state of war (for, being rational, will they not be able to maintain a peaceful state of nature?). I think that these are points that need to be addressed, however space does not permit their discussion here. For a discussion of the problem of defining human nature such that we are sufficiently irrational to end up in a state of war and to require a sovereign to extricate us from it; and that we are rational enough to reach a state of war ending agreement, see Hobbes and the Social Contract Tradition, Jean Hampton. New York: Cambridge University Press, 1988.

Let us be sure we understand the moves made by Hobbes:

(1) when individuals are outside the commonwealth (i.e. in the State of War), fear of destruction leads to the keeping of agreements; and

(2) when individuals are in a commonwealth fear of the sovereign leads to the keeping of agreements.

Hobbes, then, "solves" the sequential PD by arguing that individuals will view their choices in the PD within a broader context, from which perspective they will come to modify the dilemma-generating preferences of the PD.

Gauthier's critique of Hobbes (Chapter 6, *Morals By Agreement*[5]) has two threads. First, he denies that there will be compliance without the coercion exercised by the State of War and the Sovereign. Second, he argues that a solution resting on a sovereign is not the best solution. I will consider each in turn.

Gauthier's presentation of the first thread is worth quoting at length:

> But for Hobbes to take full advantage of this response to the Foole, he must revise his conception of rationality, breaking the direct connection between reason and benefit with which he began his reply. Hobbes needs to say that it is rational to perform one's covenant even when performance is not directly to one's benefit, provided that it is to one's benefit to be disposed to perform. But this he never says. And as long as the Foole is allowed to relate reason directly to benefit in performance, rather than to benefit in the dispositions to perform, he can escape refutation.[6]

Take note of three points made by Gauthier in this passage:

(1) Hobbes directly connects reason and benefit,
(2) Hobbes needs to give a disposition based answer to the Foole,
(3) Hobbes fails to argue that it is rational to perform a covenant when it is the result of a rationally chosen disposition.

---

5 *Morals By Agreement*, David Gauthier. Oxford,: Oxford University Press 1986.
6 Gauthier, p. 162.

Gauthier notes that Hobbes's reply depends on the State of War and the sovereign; that it is fear of these two entities which lead individuals to have the preferences ascribed to them in figure 3. Thus, if either of these entities is removed covenants will not be kept. For, individuals will remain self-interested in such a situation and thus have the preferences ascribed to them in figure 2, with the net result that no covenants will be kept or entered into.

One immediate response to this point asks why we should be concerned that covenant keeping is the result of some form of coercion. As long as covenants are kept, why is it important what motivation for doing so individuals have? Asking this question brings out two points in Gauthier's thought. First, it directs our attention to an important assumption undergirding his project. Second, it introduces the second thread of his critique of Hobbes, to wit, why a sovereign based solution to the compliance problem is not optimal. Let us first consider the assumption made by Gauthier.

Gauthier's reading of Hobbes emerges slowly in Chapter 6; in fact the best indication of how he views Hobbes does not come until after he has finished his overt discussion in sections 1.2 and 1.3. The best indication occurs in section 2.1 where Gauthier tacitly identifies Hobbes with a form of straightforward maximizer. (The Foole is also a sort of straightforward maximizer, hence Hobbes's ultimate inability to defeat him.) In contrasting the straightforward maximizer (hereafter "SM") and the constrained maximizer (hereafter "CM"), Gauthier gives us a clear indication of what he takes to be wrong with Hobbes's reply to the Foole.

The revealing passage is this:

> ...constrained maximization is not straightforward maximization in its most effective disguise. The constrained maximizer is not merely the person who, taking a longer view than her fellows, serves her overall interest by sacrificing the immediate benefits of ignoring joint strategies and violating co-operative arrangements in order to obtain the long-run benefits of being trusted by others. Such a person exhibits no real constraint. The constrained maximizer does not reason more effectively about how to maximize her utility, but reasons in a different way. (Emphasis added.)[7]

This passage contains a fair appraisal of Hobbes's reply to the Foole. It is a form of straightforward maximization. The difference between Hobbes

---

7 Gauthier, pp. 169-170.

and the Foole is merely one of degree: the hobbesian has a broader outlook than does the Foole. The Foole looks only at a particular agreement or situation; the hobbesian at the bearing that cooperation in a given situation has on survival. The sovereign and State of War are effective devices for making individuals take the broad rather than narrow perspective of situations. Gauthier does not, however, consider this a solution to the Foole's challenge.

The key to understanding Gauthier is contained in the bold-faced sentence: "such a person exhibits no real constraint." It is this sentence which reveals an undergirding assumption about what counts as constraint in Gauthier's thought. What Hobbes does, in modern terms, is to argue that individuals' preferences are different when a sovereign or State of War is present than when neither is present. Utility maximization leads, when both are present to the keeping of agreements, and when both are absent, to their violation. Gauthier rejects this reasoning as a solution because it does not make room for constraint. From Gauthier's perspective, the need for such a requirement is easy to see. Morality has traditionally been seen as a constraint on the activities of individuals, and if morality is to be a part of rationality, rationality must lead to adoption of such constraints. Hobbes never asks individuals to constrain the pursuit of their preferences while Gauthier argues that there must be constraint if the actions they perform are to count as moral. Hobbesians, who are coerced into cooperation by fear of the sovereign or state of war, are not exhibiting constraint and are thus not acting morally. Examining what Gauthier means by "constraint" will occupy much of our attention in Section 3.

By itself, this point does not constitute a sufficient reason for rejecting Hobbes's reply to the Foole. Hobbes need only dispute Gauthier's identification of morality and constraint to save his argument. Further, he might contend that his individuals do exhibit constraint, at least insofar as they would prefer not to keep agreements but do so anyway. Lastly, Hobbes might argue that to ask what individuals would do were there no sovereign or state of war makes no sense. Given human nature, there will always be a state of war and a need for a sovereign. What Gauthier has not yet done is to indicate how Hobbes's response is internally flawed.

I think that Gauthier does indicate a failure point in Hobbes's response to the Foole. Gauthier argues that the sovereign's existence requires a maintenance cost and thus the sovereign solution is sub-optimal: all would do better were they able to keep agreements without paying the sovereign.[8] If individuals are rational, they will not enter into

---

8 Let us not forget that for Hobbes "paying" the sovereign means giving up almost all rights.

sub-optimal agreements; hence no sovereign will be appointed. The argument here is not that a sovereign will fail to compel the keeping of agreements, rather it is that no sovereign will be appointed.[9] Gauthier's position is that it is possible to reach and keep agreements where no sovereign is present.

I would like to see Gauthier make a stronger case against Hobbes. The preceding argument assumes that individuals will not enter into sub-optimal agreements, but this will be true only where optimal agreements are possible; Hobbes might well argue that no better agreement than one creating a sovereign is possible for individuals in the

What I do not find in Gauthier is an argument proving that a non-disposition based solution will fail. Such an argument might go something like this. Individuals who do not have fixed dispositions to comply with made agreements are not to be trusted. Even if there is a sovereign in residence, individuals will still, at bottom, prefer to violate agreements and will do so whenever they believe that they can get away with it. I may, then, fear that you believe you can get away with violating our agreement, to which the rational response on my part would be to not enter into the agreement in the first place.

We might extend this argument to include the sovereign-generating agreement. In Hobbes's vision of the social contract, all turn over all powers (save self-defense) to the sovereign. How can anyone be sure, without some sort of disposition which prevents it, that someone will not renege on this agreement? If this fear becomes pervasive enough, there will be no social contract.

This is, of course, a mere sketch of an argument. I include it to show one argument Gauthier might make against any reply to the Foole which is not disposition-based. Gauthier might not embrace this argument, but he needs to provide one to prove that non-disposition based replies to the Foole will fail. Otherwise, Hobbes has it open to him to deny the assumptions about constraint, morality, the nature of individuals and the situation in which individuals find themselves upon which Gauthier's rejection of Hobbes depends.[10]

---

9 Gauthier does not himself make this argument. He does claim that the sovereign solution is sub-optimal. The use of the argument to indicate a problem with Hobbes's solution is my own.

10 I am indebted to Dr. A. Cudd for pointing out both the need for such an argument and one way one might make such an argument. She might not, of course, endorse the argument as I have formulated it.

Section 3: Gauthier

I have already introduced Gauthier's idea of constrained maximization.
To defeat the Foole (a straightforward maximizer) Gauthier must prove
that rational individuals will choose to become constrained maximizers.
This means proving two propositions:

> (1) that individuals will choose to base their actions on a joint
strategy when they expect others to and to act on an individual strategy
when they expect others to.[11]
> (2) that individuals will choose at one time on principles of utility
maximization to forego further choices by that standard and will instead
act from a disposition of constrained maximization.[12]

Both conditions are important and each must be looked at carefully.
Condition (1) is the more discussed by Gauthier[13], but condition (2) might
be even more important. Gauthier's method is to compare the expected
utility of disposing oneself to maximize utility given other's expected
strategy choices with the utility of disposing oneself to co-operate with
others in bringing about nearly fair and optimal outcomes.[14]

At first glance the case for choosing to dispose oneself to act as a
straightforward maximizer seems strong. We can use a familiar matrix to
examine the choice:

	CM	SM
CM	(U^{**},U^{**})	(U,U^{***})
SM	(U^{***},U)	(U^{*},U^{*})

(Where: $U^{***}>U^{**}>U^{*}>U$)

(figure 5)[15]

One might reason as follows. No matter what column chooses, row does
better by choosing SM than by choosing CM (for $U^{***}>U^{**}$ and $U^{*}>U$). A
similar argument can be made for column. SM thus dominates CM and

---

[11] Gauthier, p.167.

[12] Gauthier, p.158.

[13] See section 2.2 of Chapter 6 of *Morals By Agreement*.

[14] Gauthier, p. 171.

[15] This formulation of Gauthier's argument is from Campbell's "Moral
Justification and Freedom", *J. of Philosophy*, Volume 84 (April 1988): 192-
213.   See also his "Gauthier's Theory of Morals By Agreement",
Philosophical Quarterly, Volume 38 (No 152): 343-364.

will be chosen as the disposition to form by rational agents. Gauthier
rejects this argument.

Gauthier writes,

> ...this argument would be valid only if the probability of
> other's acting co-operatively were, as the argument
> assumes, independent of one's own disposition. And this
> is not the case. Since persons disposed to co-operation
> only act co-operatively with those whom they suppose to
> be similarly disposed, a straightforward maximizer does
> not have the opportunities to benefit which present
> themselves to the constrained maximizer.[16]

Gauthier here argues that the reasoning supporting the rationality of
choosing SM is flawed owing to its failure to appreciate the true nature of
the CM disposition. Because of this failure, the outcomes represented in
figure 5 are incorrect.

The CM disposition, we noted earlier, is one which:

(a) leads to cooperation with other CMers, and
(b) leads to non-cooperation with SMers.

Thus, a more correct matrix would be (using the same payoffs as specified
for figure 5):

|	CM	SM
CM	(U^{**},U^{**})	(U^*,U^*)
SM	(U^*,U^*)	(U^*,U^*)

(figure 6)

In this matrix CM weakly dominates SM and is thus the rational choice.
Given transparency of dispositions, no one will be able to take advantage
of a constrained maximizer, and only constrained maximizers will gain the
benefits of cooperation.

The reader will note (and Gauthier agrees) that this argument
goes through only if we assume that the dispositions of all are open to
inspection.   In the real world this hardly seems to be the case.
Consequently, it is important for Gauthier to establish that CM is the
rational choice in a world (more like this one) where dispositions are
translucent. For our purposes, however, we need only concern ourselves
with the ideal situation where transparency of dispositions holds.

---

[16] Gauthier, p.172.

The second condition adumbrated earlier was that the choice of CM or SM be a one time choice; a disposition, once chosen, is not subject to alteration. Campbell brings out the importance of this condition neatly.[17] Suppose that dispositions are transparent but that they can be changed. With this supposition in mind, consider a sequential representation of the game depicted in figure 5:

(figure 7)
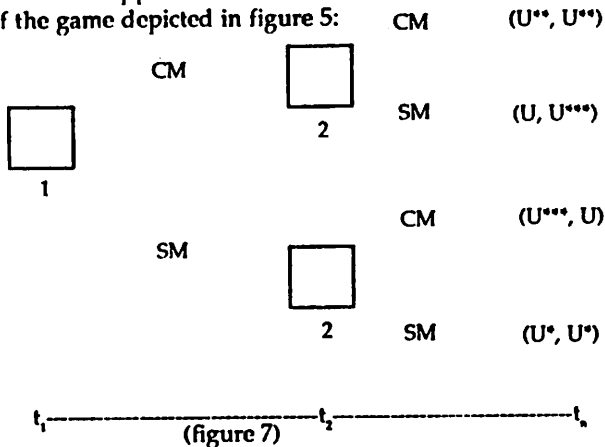
Player 1 is attempting to decide between CM and SM. At t1 player 2 has the disposition of CM, thus leading player 1 to choose CM (the expected outcome being (U^{**},U^{**})). But, if player 2 can change dispositions, supposing player 1 has chosen CM at t1, what is player 2 likely to do at t2? Her preferences have not changed by adopting the disposition of CM, so preferring U^{***} to U^{**}, player 2 will choose to change dispositions and take advantage of player 1. Player 1, aware of this danger, will not choose the disposition of CM. Once more, there is no cooperation. To prevent this unravelling of cooperation, Gauthier must argue for the non-alterability of dispositions after they are adopted.

Gauthier never explicitly makes such an argument. It would, I believe, have to be something like the following:

(1) It is rational to choose a disposition to CM.
(2) If it is rational to choose a disposition, it is rational to do that which is necessary to implement it.
(3) One only gains the benefits of a disposition to CM if one renounces the ability to change dispositions.
(4) Therefore it is rational to renounce the ability to change dispositions.

---

17 See either of the articles by Campbell listed in note 13.

The critical step is (2). If one can accept it, the rest of the argument follows quickly.

This completes our sketch of Gauthier's response to the Foole. The reply, we noted rested on two claims:

(1) It is rational to choose a disposition to CM, and
(2) It is rational to renounce the ability to change dispositions.

Both are needed if the Foole is to be defeated. Both have been questioned.

Section 4: Gauthier Critiqued

MacIntosh presents his critique of Gauthier in two papers: "Two Gauthier's?"[18] and "Libertarian Agency and Rational Morality"[19]. I cannot here hope to discuss the large number of issues that MacIntosh raises in these papers. Instead, I will draw attention to what I take to be the largest problem he raises for Gauthier.

We noted earlier that one reason Gauthier rejected Hobbes's reply to the Foole was because of its failure to include the notion of constraint. Moral actions are moral only to the extent that they are constrained actions; only to the extent that individuals act contrary to their preferences:

> Our claim is that in certain situations involving
> interaction with others, an individual chooses rationally
> only in so far as he constrains his pursuit of his own
> interest or advantage to conform to principles expressing
> the impartial characteristics of morality.[20]

Earlier we looked at a passage which explicitly denied that we pursue such constraints on grounds of self-interest. Let us, for the moment, accept that moral actions are constrained actions.

MacIntosh wishes to argue for an additional feature of moral actions. If moral actions are to be identified with a subset of rational actions, then since rational actions, to be rational, must be freely chosen, moral actions must be freely chosen: "If people only do something because of causally stabilized deterministic dispositions, they have not

---

[18] "Two Gauthiers", *Dialogue*, Volume 28: 43-61.
[19] "Libertarian Agency and Rational Morality: Action-Theoretic Objections to Gauthier's Dispositional Solution of the Compliance Problem", *Southern J. of Philosophy*, Volume 26: 499-525.
[20] Gauthier, page 4.

chosen to co-operate, and so do not act rationally in co-operating."[21] Let us then accept that moral actions must be freely chosen actions.

With these two conditions in hand, MacIntosh's point can be presented in the form of a dilemma (though he himself does not make his point this way).

[1] If the disposition to CM is non-alterable, then the actions resulting from it are not freely chosen (note that given such a disposition it is impossible to choose certain actions, this is the basis for Gauthier's rejection of the argument which demonstrated the rationality of adopting SM over CM). Nor would they be chosen else constraint would have no meaning. If moral acts are freely chosen acts, then a non-alterable disposition does not lead to the performance of moral acts. And, not being freely chosen acts, the acts cannot be rational. Thus, a non-alterable disposition gives rise to irrational acts. It seems that, to avoid this result, individuals would not choose to adopt such a disposition, or if they did, they would no longer be rational or moral agents.

[2] If the disposition to CM is alterable then one of two things is the case:

> (a) preferences must be altered so that cooperation will be the rational choice for individuals moving second in the sequential PD where the first player has chosen cooperation.[22] Changing preferences, however, has as a consequence the abandonment of constraint as a part of morality; or
> (b) if preferences are not altered, the compliance problem will remain unsolved.

Briefly, then, we must choose between giving up freedom and giving up constraint as part of morality. Or, must we?

Gauthier could reject [1] as presented, for he enjoins us to look at the rationality of dispositions, not the rationality of actions. Bearing this in mind, actions qua emanating from a rational disposition are rational and given that individuals freely chose to adopt a disposition, they are free actions. Such a disposition, chosen on the basis of utility maximization, is not subject to change: we choose on "utility maximizing grounds, not to make further choices on those grounds."[23] Since our preferences remain unchanged (we would still prefer the betrayal payoff), we are constrained for, when moving second in the sequential PD, we do not do what we would prefer to do.

---

21 MacIntosh, "Libertarian", p.509.
22 This is the test case suggested by Hobbes and the one which we have employed throughout this investigation.
23 Gauthier, page 158.

MacIntosh and McClennen[24] argue that this move will not save Gauthier. Although MacIntosh presents a seemingly wide variety of objections to Gauthier's disposition-based solution to the compliance problem, they can, I believe, be reduced to just one. The chief point of contention is the requirement that a disposition be chosen once and for always. There are a number of ways to make this point. Here are two:

> If it is expected utility calculated from the ex ante point of deliberation over policies--that judges a CM policy to be superior to the SM policy, it is also expected utility-calculated from the ex post point of choice of a course of action--that will judge implementation of SM superior to implementation of CM.[25]

> The CM disposition qua standing preference for conditional co-operation is thus, at the time of the behavior in question, one that it is irrational to have. It is thus, then, an irrational disposition, issuing in irrational behavior.[26]

The sentiment of both authors is the same. If it is appropriate to base adoption of a CM disposition on considerations of utility maximization, then why is it not appropriate to continue to employ utility maximization? If there is some flaw in utility maximization, how can it be useful for proving that the adoption of the CM disposition is rational? If there is no flaw, how is its discontinuation to be justified? Is Gauthier merely saying be a utility maximizer when it is rational to be one and not be one when it is not rational to be one? If so, how is he giving meaning to "rational"?

MacIntosh and McClennen have asked a good question, certainly one that Gauthier must address. Gauthier could respond by arguing that there is nothing wrong with utility maximization. Each time the disposition needs to be renewed, the rational individual, for the reasons adumbrated earlier, will choose to renew it.[27] On this reading Gauthier gives up the choosing of a disposition once and for all time, but contends instead that each time a choice of dispositions is made, individuals will select dispositions of the appropriate type, to wit, a disposition to choose as a constrained maximizer. Knowing this,

---

24   E. McClennen, "Constrained Maximization and Resolute Choice" in *The New Social Contract: Essays on Gauthier*, eds. Paul, Miller, Paul and Ahrens. Oxford: Basil Blackwell, 1988.

25 McClennen, p. 105.

26 MacIntosh, "Libertarian", p.514.

27 I am indebted to Dr. A. Cudd for pointing out this response to me.

entering into agreement with one another.

The problem with this response, apart from the fact that we must now be transparent to one another in more than our dispositions, is that it will not provide a guarantee that agreements will be kept. For, individuals will still have the option of changing preferences and even if we are fully transparent to one another (i.e. able to observe everyone's reasoning process and preferences), one might enter into an agreement and then face the possibility that the other person will change dispositions; what disposition and preferences they had in the past is no sure guide to those they will have in the future. Once again, there is room to doubt that agreements will be kept and thus a disincentive to enter into them.

There is another difficulty in Gauthier's reply to the Foole. Let us look at figure 6 once more:

	CM	SM
CM	(U^{**},U^{**})	(U^{*},U^{*})
SM	(U^{*}, U^{*})	(U^{*},U^{*})

Here CM weakly dominates SM. Are the utilities correct as presented? I suggest that they are not. If one chooses CM one must also choose a non-alterable disposition; such a disposition carries an extreme penalty with it—the inability to take advantage of a change in circumstance.[28] How much of a penalty? If U^{**} is only a bit larger than U^{*}, then appreciation of this penalty could lower U^{**} enough so that it was equal to or less than U^{*}. Were this to occur, no one would choose a disposition to CM and we would once again be left without a solution to the compliance problem. In a nutshell, I do not think that Gauthier takes sufficient notice of the impact the type of disposition required to be an effective CM will have on the preferences of agents faced with a choice between CM and SM.

Serious questions, then, remain to be addressed by Gauthier. Both of the questions I have raised here involve the preferences of individuals. Perhaps we must, as Hobbes suggested, change our preferences if the Foole's challenge is to be met. Consideration of the preferences of individuals is at the heart of MacIntosh's reply to the Foole.

---

28 The penalty is even more severe if we consider non-transparent individuals. For example, imagine being locked into a disposition to co-operate only to find that the other person is not as you supposed them to be. Would one really risk such a trap? It would be much like asking a hobbesian to give up their right to self-defense.

Section 5: MacIntosh

Let me say at the outset that I believe that MacIntosh points a finger at a rewarding approach; I do not, however, believe that his approach, as presently formulated, is a viable solution to the compliance problem.
MacIntosh's basic idea is that:

> ...we construe the CM disposition as just a set of preferences for justice for its own sake, practical act-rationality then justifying co-operative choices as maximizing of utility-of individual preference satisfaction-given the new preferences.[29]

The basic idea is that individuals will alter their preferences to value justice for its own sake: "I say one must revise the overall balance of one's preferences so that one prefers, above all else, to have co-operated with those who prefer to co-operate with conditional co-operators."[30]
In game form:

(figure 8)

Given these preferences, it is rational for utility maximizing individuals to choose a CM disposition: considerations of justice here play the role for MacIntosh that the State of War and the sovereign play for Hobbes. (Thus the similarity between figures 8 and 3.)
There are two aspects of this account which require attention:
(a) Preference Alteration. In order to preserve the freedom of choice individuals are allowed to alter preferences at will. According to what rule will they change preferences? That preferences can change should alarm the player choosing first in a sequential PD. Knowing that preferences can change, will the player moving first adopt a CM

---

[29] MacIntosh, "Two Gauthiers?", p.53
[30] MacIntosh, "Two Gauthiers?", p.55.

disposition? I think that MacIntosh's emphasis on freedom gets him into trouble. If individuals are too free to change preferences, then there is no sure way to know what they will do. Fearing the worst, a prudent first-mover will choose SM.[31]

(b) Constraint. Are we to give up the idea that moral actions exhibit constraint? MacIntosh has allowed for freedom of choice, but since we are acting to maximize our preferences, this account is subject to the same criticism raised against Hobbes by Gauthier. MacIntosh must present either a case for not viewing constraint as integral to moral acts, or explain how his account incorporates constraint.

Conclusion

Despite the efforts of Hobbes, Gauthier and MacIntosh, the Foole's challenge has not been met. If we are ever to defeat the Foole, we must try to appreciate why each effort failed.

Hobbes suggested that individuals would prefer to uphold the Third Law of Nature when fearful of the State of War or the Sovereign. We argued that this fear was not a sufficient guarantee that agreements would be kept; that individuals would, at bottom, prefer to violate agreements; and that the fear-generated preferences were ephemeral and hence not to be relied upon. One could not know when entering into an agreement whether, when it came time to fulfill it, the other party would be sufficiently motivated by fear to keep the agreement. Without such knowledge, rational individuals will not enter into agreements.

Gauthier attempted to assure compliance with made agreements by arguing that individuals would choose dispositions which would dictate the keeping of made agreements. The problem with this approach is that Gauthier supposes that while our overall preference is to adopt such a disposition, some of the actions it requires (keeping agreements, for example) will be contrary to some of our discrete preferences (this allows Gauthier to demonstrate that his theory includes constraint). With respect to our discrete preferences, we thus behave irrationally. Additionally, for the disposition to ensure compliance, it must not be subject to change;

---

31 MacIntosh might respond that individuals will make probability calculations of how likely it is that free individuals will alter their preferences in unfavorable ways. This response will avoid the objection, but only if MacIntosh can develop an account which demonstrates that prudent first movers will not use disaster avoidance reasoning and that they will have sufficient faith in their calculations MacIntosh will also have to provide a formula to act upon them.

however, this would seem to imply that we are not free and we noted that free choice is a characteristic of rational choice. If we can change our dispositions, then we face the same uncertainty in our dealings with others that hobbesians face. For, who can be sure that someone's overall preference will not fluctuate with the situation they find themselves in? (Especially given that our discrete preferences, if acted upon would lead to a change of disposition.) Gauthier, to, then, fails to provide a guarantee that the Third Law of Nature will be upheld and thus that a social contract is possible.

MacIntosh attempted to ensure compliance by arguing that individuals would prefer a disposition which once adopted, would alter their preferences such that one preferred each and every action dictated by that disposition. Since our preference is to keep agreements (because the disposition we choose dictates the keeping of agreements), there is nothing to fear--agreements will be kept if entered into. The difficulty here, we noticed, is that individuals remain free to revise their preferences and thus their dispositions. We cannot be sure that someone will not alter their preferences and thus their dispositions in the space of time between the making of an agreement and its fulfillment. Like the responses of Hobbes and Gauthier, MacIntosh's leaves room for the Foole to insert doubt sufficient to undermine the making of agreements.

Hobbes, Gauthier and MacIntosh recognize that to defeat the Foole they must provide a sufficient guarantee that agreements will be kept. Hobbes attempts to ground his guarantee in the preferences of individuals. Gauthier and MacIntosh attempt to ground their guarantees in dispositions. The problem is that preferences and dispositions are subject to alteration; hence there is no guarantee that agreements will be kept. In fact, as the Foole points out, there is every reason to believe that they will not be kept. For, if we take as our domain self-interested individuals, the factors that lead them to choose to cooperate, will also dictate their non-cooperation when it is to their advantage to do so. The Foole's challenge thus makes it doubtful that we can view society as the product of a rationally motivated contract between individuals; rational individuals do not enter into contracts which they do not believe will be kept. Unless the Foole's challenge can be met, the prospects for social contract theory are dim. Despite 350 years of effort, the Foole continues to foil our attempts to meet his challenge.[32]