THE UNIVERSITY OF CALGARY

REASON AND MORALITY, AGAIN

by

Gordon R. DuVal

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF ARTS

DEPARTMENT OF PHILOSOPHY

CALGARY, ALBERTA

JULY, 1982

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled, "Reason and Morality, Again" submitted by Gordon R. DuVal in partial fulfillment of the requirements for the degree of Master of Arts.

Robert X. Ware

Supervisor, Robert X. Ware
Department of Philosophy

Kai Nielsen

Kai Nielsen
Department of Philosophy

Thomas Hurka

Thomas M. Hurka
Department of Philosophy

Stanley M. Stein

Stanley M. Stein
Faculty of Environmental Design

15 July 1982
(date)

Abstract

The discussion which follows is an examination of some of the most recent attempts to demonstrate a necessary connection between acting morally and acting rationally. I will argue that all of these attempts are unsuccessful. I employ a straightforward means/end conception of rationality, whereby the rational course is that which best promotes the ends of the agent himself. For the sake of simplicity, I have distinguished between moral ends (roughly speaking the desire to do, or see done, the right thing), and non-moral ends (all others). A 'prudent' action is one performed in the promotion of non-moral ends.

The first chapter examines the case where there is a genuine conflict between doing what one ultimately wants to do, and doing that which is morally right. I argue, fairly standardly, that if this is the case, then there can be no necessary connection between reason and morality. This having failed, I turn, in chapter 2, to consider the possibility that there is no genuine conflict (although appearances are to the contrary) between morality and prudence. After briefly canvassing some fairly weak attempts to do this--attempts which make reference to the wrath of God, the efforts of conscience, and the social consequences of immorality; I proceed to initially more plausible views. These views attempt to show that due to the social effects

of cooperation, and the dynamics of human interaction, we do better, even in prudential terms, by being moral. A progression of three views, each increasing in sophistication, is addressed in chapters 2 and 3.

First, I examine a proposal that prudence is self-defeating--that prudence fails by its own standards. A related view is held by David Gauthier, who holds that one who constrains his maximizing (prudent) behaviour will do better than one who maximizes his own utility directly. This constraint is intended to be the restraints which morality places on our behaviour. Two defences of Gauthier's "constrained maximization" are here discussed, the second of which, being more persuasive and sophisticated, is examined in some detail. Toward the end, I offer some very brief and in general speculative remarks, largely concerning the nature of the moral motivation. I suggest that any attempt to reconcile, in a systematic way, reason and morality, which does so by describing the moral motivation in ultimately prudential terms, may have profoundly misconceived the notion of morality.

Table of Contents

> "That's why, I may observe in parenthesis, our social proprieties and conventions are so good. They have profound value, I won't say for morality, but simply for self-preservation, for comfort, which, of course, is even more, since morality is really that same comfort, that is, it's invented simply for the sake of comfort."
>
> Fyodor Dostoevsky,
> The Insulted and Injured

It is in many ways an appealing notion that morality can be rationally grounded. Most people wish to be, ideally at any rate, both rational and moral. It would be too bad if, in doing the morally right thing, we sometimes run afoul of reason. It would be worse still if the dictates of reason sometimes or always counsel a morally evil course. I will nevertheless argue that these are both the case. John Rawls has suggested that "[t]he theory of justice is a part, perhaps the most significant part, of the theory of rational choice."[1] It will be my contention that far from being the most significant part, just action is by no means even a proper subset of any respectable theory of rational choice. That it is outside of the realm

of reason, in some instances, to do the right thing must be, at least for some, disquieting. However, such a conclusion I find unavoidable. This is not to say that the opposite view has not been, or does not continue to be held and vehemently argued. The discussion which follows will address itself to what I take to be the most interesting attempts to show that a necessary connection exists between reason and morality.

> It is not disputed that the man who acts unjustly when it suits his long-term interest is anything but unjust. Of course he is unjust and immoral--but is he irrational? And the man who does not yield to temptation is rightly called a just man. But is he rightly called a rational man?[2]

Certainly our discussion of the relationship between reason and morality will have to include some account of what it is to be rational. For our purposes, a fairly commonplace means/end rationality will be employed. In short, the rational course will be that which best serves a given end. In the context of human action, we will take it that the rational act is the one which most efficiently furthers (which is the best means to achieve) the ends of the individual agent.[3] I will assume also that there is, in this context anyway, no significant distinction between being rational and being reasonable. As a result, I will use 'rational' and 'reasonable', as well as 'rationality' and 'reason', interchangeably. Consistent with this usage, I will further assume a 'reasons' approach to rationality.

That is, the rational course will be the one for which we can provide the best or strongest reasons. In other words, to say that a given course is most rational is to say that there is no alternative course for achieving the same end(s) for which better reasons can be provided. Rationality is a measure of the quality of means to an end. The more efficient is a given means to an end, the more rational it is to employ that means. However, rationality can do no judging of ends themselves. To make judgements about the quality of reasons, one must consult the individual's ends. This is a conceptualization that is common enough. Hume suggests that

> [i]t appears evident that the ultimate ends of human action can never, in any case, be accounted for by reason, but recommend themselves entirely to the sentiments and affections of mankind, without any dependence on the intellectual faculties.[4]

In short, reason cannot provide us with ends. It is the task of reason (rationality) to serve whatever ends we happen to have. David Gauthier (whose work we will examine quite carefully later on) is quite succinct on this point. "Reason takes the ends of our activities as given, and determines the means to those ends."[5] Ends cannot be judged as to rationality except inasmuch as they are also instruments to further ends. Our ultimate goals are neither rational nor irrational. Rationality (reason) has not the tools required to make either such judgement.

It is quite common in the literature on this subject

to contrast moral action, sometimes seen as other-regarding action, with self-interested, self-regarding, or prudent action. While the use of these terms is often appropriate, I think it important to be clear about how these notions interrelate. We should be wary of setting up these contrasts, since it appears that these notions do not, in all cases, pick out different actions. For example, we often contrast moral action with self-interested action. But the morally right thing will often be in the agent's interest. We might in such a case be concerned about the motivation being appropriate. However, it is entirely consistent with all major ethical theories that in many cases the right thing, chosen with the appropriately moral motivation, will be in the interest of the agent.

Another contrast commonly drawn is between other- and self-regarding action, where other-regarding acts are meant to be the morally right ones. But an act performed for the sake of another, even in the absence of any benefit to oneself, need not be morally right--or even permissible. I might aid another by offering to murder one of his enemies, or more tamely, I might help someone in some way to cheat on their income tax. While I may be acting other-regardingly, I am by no means praiseworthy thereby. Contrarily, for similar reasons as I have proposed above, self-regarding action may be moral. For a utilitarian whose own happiness in a given situation is most affected, it will usually be the case that the happiness-maximizing

choice is the one which benefits him most. He is then morally required primarily to regard himself.

The relationship between these notions (prudence, self- and other-regarding action, and self-interest), and the promoting of one's ends, is also less than straightforward. We quite commonly have other-centred ends. In such a case, we serve our own ends by acting in an other-regarding way. Self-interest, like prudence, is generally thought to be exclusively in the service of our ends. But we may not have a corresponding preference for everything which is in our interest. For example, many smoke cigarettes, fully realizing that it is unhealthy and dangerous to do so. We are often compelled to take a large risk--flushed with the excitement of the gamble perhaps--which prudence dictates to be unjustified. If we employ a means/end conception of rationality, we might still be unsure of which end it is rational to serve in cases of conflict. Is it rational to take as an end that which is in our interest, or is it rational to efficiently satisfy our preferences, wants or desires?

On this question, and for the purposes at least of this discussion, I will view ends as nothing more than the ultimate objects of our desires or preferences. I conclude, with J.C. Thornton, that

> it is what a person most wants to do, rather than what is in his own best interests, that provides the ultimate basis for the rational justification of an action.[6]

In the context of means/end rationality, the end is supplied by our wants or preferences. Indeed, it may be argued (although it will not be argued here) that it is difficult to make sense of our ultimate self-interest if not in terms of our basic wants. Whatever the merits of this view, I will assume that our wants are the ends which means/end rationality attempts to advance.

The relations between and among the notions of moral action, self-interest, self- and other-regarding action, and prudence gain in complexity in the face of rigourous examination. It is true that the usual contrasts which are drawn do at least fair service when used (as they are usually used) in the context of a straightforward conflict between the moral, and the personally more attractive immoral course. We generally desire or prefer that which is in our interest. Discussion can proceed tolerably well since we understand that one of the contrasted positions assumes the primacy of moral considerations, while the other side is one in which the agent lacks the requisite moral motivation.

I will however, attempt to clarify and simplify this discussion by, in large part, forsaking the usual terminology which uses "self-interest", "other-regarding", and "self-regarding", at least for the purpose of contrasting moral with non-moral activity. I suggest that we can do the same job by distinguishing between action in service of our moral ends as opposed to that in service of our

non-moral ends. This, instead of contrasting morality with self-interest, or other-regarding with self-regarding action. Very roughly, our moral ends will be those which come from a desire to do or see done, that which we believe to be morally right. Non-moral ends will be all ends which are not moral ends. It may be helpful to see acting for the sake of moral ends as roughly equivalent to acting from the morally disinterested motivation. I will reserve the term 'prudent' to refer to action which promotes our non-moral ends. Moral reasons will be contrasted with prudential reasons, and will be those which appeal primarily or exclusively to moral ends. Prudential reasons will be considerations which appeal exclusively to all other (i.e., non-moral) ends.

While this may be to use 'prudent' in a somewhat different way than usual, I hope this does not prove troublesome. A prudent action will, of necessity, lack moral motivation, even though it may be physically the same act as a moral one. Since we generally think that to truly act morally, one must have the appropriate moral motivation, a prudent act will never be moral. We act morally when we act in promotion of moral ends. It is rational to promote whatever ends we happen to possess, be they moral or non-moral.

I must confess at the outset that a different notion of rationality might well yield quite different results. I have at least two reasons for conceiving rationality in

roughly the way I have chosen. First, it strikes me that this is straightforwardly the most plausible conception. Second, and more importantly, this conception best suits what I take to be interestingly at stake here. What is at stake for the rationality of morality, at least in this discussion, is the possibility of a wholly rational person choosing to do a morally wrong thing some of the time. We want to address the question whether there needs to be a failure of rationality, at least sometimes, in order to act morally. Can a person who acts always reasonably, be persuaded to do the right thing all of the time.

In the interest of clarity, I should like to see the issue in these terms. Can any good reason be supplied to a person to do the right thing, given that he[7] believes it to be the right thing, but does not want to do it; which addresses any consideration apart from the satisfaction of his ends? Presumably, it is <u>not</u> one of his ends that the right thing be done. We might proffer moral reasons, but if he does not wish to be moral, why would such a reason be admitted? We often provide others with reasons to do the right thing which are sometimes heeded. Consider the man who wishes to kick his grandmother down the stairs for the inheritance. While he accepts that it would be wrong, he believes also that it would satisfy many of his ends. It would be fun, and he would enjoy the benefits sooner of a substantial inheritance. We might propose any number of considerations in an attempt to deter him from so

acting. We might point to the likelihood of being caught and imprisoned, thereby losing the inheritance altogether. We might remind him that his brother would be absolutely devastated by the death of a grandmother of whom he is so fond. Or, we might suggest that he could never enjoy the use of the money he would gain due to the resulting oppression of his conscience. Would these, however, be reasons at all if he did not want to avoid capture and incarceration, did not care about the heartache of that particular brother, or felt that the benefits outweigh the admittedly negative effects of his conscience?

In such a case, the act's wrongness would only be a reason to refrain from doing it if the agent wants to avoid doing that which is wrong. We engage in reason-giving, in this instance, when we attempt to convince another either that he has neglected to consider some relevant factor, or that he has some mistaken beliefs about the consequences of his act. It is not unusual to try to convince someone that in fact acting in a certain way will not, contrary to his beliefs, best satisfy his ends--provide him with what he ultimately wants. It is banally true that much of the time, likely most of the time, one's ends are best advanced by acting morally. It appears however that there is little that is enlightening in cases where moral and prudential reasons counsel the same act. The more instructive cases are those in which the satisfaction of one's desires, and the satisfaction of one's

duty, appear to require incompatible courses of action. In this context, moral action would always be rational if (1) contrary to appearances, there is never any such incompatibility. That is, that the counsels of prudence and those of morality always coincide. Or, (2) it would always be rational to be moral if, though morality and goal satisfaction sometimes counsel different courses, better reasons can be found to do the right thing when they diverge.

One might well now ask what it is to be moral--to do the morally right thing. This question too I will leave largely unanswered. It is my hope (and indeed my belief) that this endeavour can proceed without making any decisions about the precise contents of the morality which we discuss. I will assume a view that need not be far different from our ordinary notions of morality. In the general case, killing, raping, stealing, and lying will be, among other things, morally wrong. Conversely, and again generally speaking, sympathy, generosity and compassion will count among the moral virtues. But for now it does not much matter what the contents of the morality are, so long as other-regarding or disinterested considerations play a central part. In addition, while it may or may not be that the only thing good absolutely is a good will[8]; I will take it that the intention to do the right thing is, if not preeminent, then at least morally significant. I expect that these are not implausible

guidelines.

I do accept however the possibility that without a better developed conception of the contents of morality, this endeavour cannot progress. For indeed it might be that judgements about the rationality of accepting moral strictures cannot be made except in the context of a more or less worked out framework of what this acceptance would entail. Needless to say, I think this not the case. J C Thornton, however, cautions that,

> [s]urely, it will be objected, nothing fruitful can be said about the justification of the moral point of view unless it is preceded by a full discussion as to what is meant by this expression...Yet sometimes an examination of the cart can tell us quite a lot about the sort of animal that pulls it.[9]

It seems to me plausible to suppose that we can come to some conclusions about the rational status of disinterested action of a more or less recognizable nature, without being entirely certain exactly what we would be thereby committed to. However, if this is not so, my project might be shown to be seriously misfounded.

With these considerations pointed to, we can proceed. Why would someone think that sometimes the moral act could be irrational? To put this very generally, we are attempting to defend the rationality of moral action, when it is opposed to prudential action. I use the term 'prudential' here because, ex hypothesi, the agent in this case lacks any sufficient moral end(s) to outweigh his desire to

follow the immoral course. Our conception of rationality is tied to how well our ends are furthered by a given action or course of action. How could we go about convincing a person who is moved only by rational considerations to do what is right, when this conflicts with what he, in all, wants to do. As I have suggested, there appear to be two plausible ways. We can show that there is in fact no conflict. We might try to demonstrate that in this case as in all cases, the morally right thing will best promote his (even non-moral) ends. Or, we can show that there exists some as strong or stronger reasons for doing the moral thing as there is for doing what will, in all, advance our ends. It is this second alternative that I will now consider. We will return to the first in due course.

What would it be for there to exist as good or better reasons to do the moral thing than the desired thing? We are familiar enough I expect with having a reason or reasons to do one thing, but having better reasons to do quite another incompatible thing. Reasons are the sorts of things that can be stronger or weaker than one another. Reasons also gang up, in the sense that different factors might be reasons for the same course of action, which can together make a stronger case than a single factor by itself. An analysis of what counts as a stronger or weaker reason for a given course of action will not here be attempted. It is enough that, at least in the context of actions or courses of action, the rational way to proceed

is that for which we have the strongest reasons.

I take it that a reasons approach to rationality is both a fair way to proceed and one which simplifies discussion. We might then wonder the following. Can better (stronger) reasons be found to act morally when to do so would conflict with our desires (ends)? I will for now assume that such conflicts are real and often troublesome. It appears at least initially obvious that we sometimes make a genuine sacrifice in order to act morally. The bulk of our discussion in subsequent chapters will address precisely this question. First however, I should like to examine whether adequate reason can be given to act morally even given a genuine conflict.

It is fairly natural to think that good reasons can be supplied for doing what we accept to be our duty. For example, that it would be a great evil if we failed to do so. Or, it might be a reason to do what I morally ought that others would be harmed if I failed to do it, or that I will have failed to provide another some benefit. Such reasons are commonly proposed in favour of moral action. But is it rational to accept such reasons? In other words, are the reasons which favour moral action really stronger than those supplied by the opposite, end-serving considerations. To do one's duty in such cases would be ex hypothesi, to act contrary to one's wishes. It is tempting to say that moral reasons, such as those we have suggested, should always outweigh end-serving reasons when they are in

conflict. This is to say that moral reasons count in favour of moral action--and we ought to follow moral reasons. Moral reasons are the ones which would appeal to moral ends if it were the case that the agent had any. In this case of course, he does not.

The trouble is that since reasons are only such in the context of the ends which they serve, and since in our story the agent's ends are not best served by acting morally, it appears that these so-called 'moral reasons' are not reasons at all. If we act rationally only by satisfying our ends, it follows simply that if we lack a given end, then it cannot be that we act rationally by performing acts which are means to that (non-existent) end. In short, it is not rational to be moral if we do not possess some moral ends. In addition, we cannot be branded irrational for failing to have such ends, since it is central to our view of rationality that it cannot judge about ends.

We can now describe the dilemma in terms of moral and non-moral ends. Our case is one where it is one's duty to do that which one does not, in all, want to do. He clearly has, in this case, some non-moral ends, since he wants to do otherwise. He believes that doing other than his duty will best serve his ends. If he has any moral ends, they are either inapplicable to this case, or more likely, are outweighed by the force of other (non-moral) ends which are served by doing other than his duty. A moral end, I presume, would be something like a desire to do the mor-

ally right thing, or to see the right thing done, or to be the sort of person who does the right thing. Doing his duty would be irrational, given the sort of conflict which this story presupposes, simply because his non-moral ends are preeminent. Given that in this instance, his ends are in that way ordered, his moral reasons are (if present at all) weaker than his non-moral ones. Since stronger reasons are better ones, and since rationality requires that we follow that course prescribed by the best reasons, it is rational, in such a case, for an agent to forsake his duty.

A man might face such a situation, that is, one in which his moral duty conflicts with that which will best satisfy his ends. He sincerely believes that the first act is his duty, and that he ought to do it. It has been suggested that the rationality of morality can be salvaged by the following consideration. Can that man, in such a situation, still intelligibly ask why he should not do as he desires, given that he accepts the moral evil of so acting? It has been proposed that such a question would be meaningless. A quick out is taken by those who hold that for one to ask whether he should do X, he must not really believe that X is his duty. To turn this around, if we sincerely and genuinely believe that we are morally obliged to do X, then we cannot ask whether we should do it. This proposed response would certainly expeditiously defuse the problem. For, if there is some necessary

connection between accepting an act as one's duty, and resolving without hesitation to act so, then a reasonable (or any other) man would not (because could not) consider any alternative course. We might be mistaken about what is our duty, but if we cannot consider doing other than what we accept to be our duty, then perhaps the problem disappears. For, if the problem does not theoretically disappear, it does so at least for practical purposes.

The only trouble with this explanation of the view that one cannot meaningfully ask 'why be moral?' is that there seems to be no particularly good reason to think that it is true. The view that this question is unintelligible is for me a mysterious one.[10] It seems that in times of moral weakness (a frequent enough occurrence for most of us) we experience just such ambivalence. However we resolve such issues, it seems odd to claim that we do not know what our 'why be moral?' question means. It appears to suggest either a psychological block about the consideration of this question, or something about the meaning of the words. It is far from clear that the notion of moral duty in some way bars one from considering an opposite course. Further, as a matter of psychology, that people cannot consider doing that which they take to be immoral, seems a proposition which admits of overwhelming contrary empirical evidence.

It seems to me then highly plausible that a person might meaningfully question whether he should do that

which he genuinely and sincerely believes to be his duty. We return to consider what manner of reasons we might supply such an individual. Why should one do one's duty when he can benefit from acting otherwise? It appears that we must agree with Hume that one must be, in such cases, either a fool or a knave.[11] To do what is right must be foolhardy indeed. It is certainly not surprising that others would have you follow morality's dictates--it is others that benefit by your good behaviour. But this does not seem to provide any kind of good reason to be good. The alternative appears to be to live one's life in an iniquitous fashion, uncaring of any moral strictures. Neither alternative is particularly attractive, but what is to be done?

It should be pointed out that even if a person does choose to forsake his duty in a given case, we might propose perfectly respectable reasons, for him to after all do his duty, which might be persuasive to him. If this happens, then one of two thing is going on. First, we might be pointing out additional considerations about the action he has chosen of which he may not have been aware. We might point out that the villainous means that he has chosen will not in fact satisfy his ends. We may argue that he is mistaken in denying that the moral course will yield for him the best results. Second, we might be trying to point out that what he thinks are his ends, in fact are not. This would not be a case of altering, or attempting to alter his ends (no reason can be provided

to do this), rather it would be an attempt to remind him or help him clarify just what his ends are.

In either case however, reasons are supplied in order to attempt to reconcile the counsels of his moral and his non-moral ends. In short, these would be attempts to show that there is in fact no conflict, although appearances are misleadingly to the contrary. I conclude that at times of genuine conflict between the moral course and the course which I genuinely desire to take, no conclusive reason can be provided in favour of the moral course, short of question-begging and inadequate moral ones.

We have considered here a situation where one's non-moral ends are stronger than his moral ones. These are the cases which are, for now, of interest. In future chapters, unless otherwise noted, our cases will continue to be those where non-moral considerations are weightier for the agent. I recognize that if, as is often the case, moral considerations are weightier--that is, the desire to be moral outweighs prudential considerations; then our conclusion would be, in such a case, quite different. For now, however, the case is one of genuine conflict, where one desires to act in a way incompatible with morality. In such cases we are compelled to say that the moral course is irrational. This is, at least initially, a worrisome claim, and I will have more to say about the implications that we might draw from such a conclusion.

But first, we will examine a quite different course.

> Once it is admitted that there are situations in which duty and interest conflict, then it follows that either following duty in such cases is irrational or self-interest is not the only ultimate justifying reason.[12]

I have already suggested that end-serving reasons <u>are</u> the only ultimate justifying ones. It appears then that following duty must be, in such cases of conflict, irrational. It may be however that our intuitions are just false, and that no conflict in fact exists between the serving of our non-moral ends, and the doing of the right thing. For, if it is the case that there is no conflict between the satisfaction of non-moral ends, and that of morality, then the moral action will in every case admit of justification with reference to non-moral ends. If one's moral ends are, in all, weightier, then it will be rational to be moral. More importantly, even if one's non-moral ends are weightier, the most efficient means to their achievement, and hence the rational act, will still be the moral one. This, since the means which advance our moral and non-moral ends always coincide. If such a coincidence can be shown, then a necessary connection between morality and reason, it is suggested, would be assured.

Notes - Chapter 1

1 A Theory of Justice (Cambridge: Harvard University Press, 1971), p.16.

2 J C Thornton, "Can the Moral Point of View Be Justified?", Australasian Journal of Philosophy, vol 42 (1964), p.30.

3 Since we cannot predict with certainty the outcome of our choices, we must see efficiency in end-promotion as a function of the probabilities attached to the various possible results. The difficulties involved in assigning such probabilities will not here concern us. I will assume that we can, in a rough way, decide which will, in all, yield the best expected outcome in terms of the advancement of one's ends.

4 David Hume, An Enquiry Concerning the Principles of Morals (App. I, sec. V), from L A Selby-Bigge and P H Nidditch (eds.), (Oxford: Clarendon Press, 1975), p.293.

5 David Gauthier, "Reason and Maximization", Canadian Journal of Philosophy, vol IV, no 3 (March 1975), p.413.

6 J C Thornton, op. cit., p.34.

7 I do not here mean to imply that only men are capable of such nefarious intentions. Here, as throughout this work, I will employ the masculine pronoun in order to pick out an unspecified individual who might, equally intelligibly, be of the opposite sex.

8 Kant, from Fundamental Principles of the Metaphysic of Morals.

9 J C Thornton, op. cit., p.22.

10 We will see later on that a similar question raised by a different reading of 'why be moral?' might be something like meaningless.

11 Hume, see his Enquiry.

12 J C Thornton, op. cit., p.31.

Perhaps a more fruitful course then is taken by those who hold that the rationality of moral action can be after all rescued by showing that there is no conflict between being moral and serving one's ends. Albeit that appearances are to the contrary, if it could be shown that in all one need never sacrifice one's ends by acting morally, then the rationality of morality might be fairly well secured. It has been supposed that when there exists an apparent conflict between the two, either there are hidden benefits to acting morally, or hidden detriments to acting immorally. When these factors are entered into the calculation, we will always find that being moral and promoting one's ends, come to the same thing.

Various such benefits and liabilities have been proposed. Some have held that the efforts of our conscience, and the social effects of immoral behaviour, remove all conflict. Despite the apparent disadvantages of acting morally, the warm glow that we get from doing our duty is a benefit not to be considered lightly, and one which tells strongly in favour of acting so. Further, one's conscience

can be a cruel tyrant. If we choose to act immorally, in a way that appears to advance our ends, not only do we renounce the feeling of satisfaction which acting morally brings, but we face also the torments of an unforgiving conscience. This combination will always render the moral course rational, since it serves also our non-moral ends.

Many who hold religious beliefs fear the judgement of God when considering any moral indiscretion. If the scriptures are to be believed, then God has horrific plans for any unrepentant sinners. However, for those who attend closely to God's law (which corresponds more or less to the moral law) paradise awaits. Given the infinite nature of both the punishment and the reward, our ends will always be best promoted by the moral course. I hope I may be excused for extending little attention to such a view.

There appears to be little enough reason to believe that God exists, let alone that He is the sort of being we generally suppose Him to be, and that He concurs in our moral views, and that He takes an interest in our affairs, and that He punishes and rewards our actions in the proposed or some similarly extreme manner. I confess that should all this be the case, then this question would admit of re-examination. However, until some more persuasive evidence is introduced in His favour, I cannot consider the efforts of God cogent to this question. As to the effects of one's conscience, we should concede that

these must be considered in the weighing up of benefits and detriments when choosing how we are to act--how our ends are best served. In doing so, we might discover that we after all wish to act in the same way as the moral way more often than we originally expected. However, there appears to be no good reason to think that, in principle, and even considering the effects of our conscience, it could never be a rational choice to act immorally. People seem to be affected to varying degrees by their conscience. It seems not implausible to suppose that there are those who are only slightly affected thereby. We tend to take this view of hardened criminals, unscrupulous business people and politicians. In addition, if the benefits of a given immoral course are great, then one might come out better, in all, by ignoring duty and accepting a certain amount of conscience-inflicted pain. It seems arbitrary to hold that the benefits accrued by acting immorally will never outweigh the pain inflicted by the conscience. In practise as well as in theory, it appears just false that we can never, all things considered, best promote our ends by ill conduct.

A similar sort of thing can be said of another detriment which attends moral action. We might well fear that if discovered, an immoral act will bring many kinds of disagreeable retribution upon the agent. Perhaps most obviously, we have laws against many acts which are immoral. The benefits which come from such an ethical transgression

could be outweighed by the fine or prison sentence which might accompany it. We also face, if our blackguardly behaviour becomes known to others, the chastisement of those around us, the loss of needed companionship, and the loss also of the trust of others which may be helpful to us at future times. These are all detriments which should be considered in deciding whether a given course of action is in fact conducive to the fulfillment of our ends. I will have more to say later on about how our actions affect the way others treat us. However, for now it is sufficient to say again that despite the fact that these considerations might convince us that we will benefit less often than we originally thought by pursuing immoral courses; there is nothing here to show any sort of necessary connection. There is nothing in principle, and apparently nothing in practice to assure us that perhaps secret, perhaps well considered acts might not at times be both immoral and personally expedient. In an often quoted passage, Hume points to some of the benefits which generally come of acting morally. Yet, he observes,

> that honesty is the best policy, may be a good general rule, but is liable to many exceptions; and he, it may perhaps be thought, conducts himself with most wisdom, who observes the general rule, and takes advantage of all the exceptions.1

A somewhat different approach to the reconciling of morality and prudence has been suggested by some, notably

Hobbes[2], and in recent times Kurt Baier[3]. They have proposed that since in the absence of moral institutions everyone would be worse off, everyone is better off adhering to these moral strictures. Hence, it is for all rational to be moral. The major premise of this argument can be afforded ready assent. I take it to be something like trivial that everyone benefits from the existence of, and general adherence to moral strictures. The alternative 'state of nature' would, it is plausible to suppose, inflict on everyone a life, "solitary, poor, nasty, brutish, and short."[4] I will not argue for this claim since, in the first place, it appears to be obviously the case; and anyway, I will argue that from it we cannot infer a good reason always to act morally. This is because it would be consistent for a rational person to accept that everyone (including himself) is better off for the existence of moral institutions; but still demand reasons why he should not, at times, ignore these institutions. He should be glad that such institutions exist, and he may be wary never to act such that they will break down. But infrequent, likely secret acts of immorality, by an isolated agent would do no serious damage to the practice of morality generally. The fact then that moral institutions are valuable for all does not provide a decisive reason against individual acts of immorality. It may, and almost certainly does, limit the range of immoral acts for which conclusive non-moral justification can be found.

Grossly heinous acts, or consistently immoral behaviour might cause or contribute to a deterioration of society's moral standards, and hence might be contrary to the advancement of one's ends. But it does not follow that no instance of immorality can be justified on such grounds.

> Yet, according to the imperfect way in which human affairs are conducted, a sensible knave, in particular incidents, may think that an act of iniquity or infidelity will make a considerable addition to his fortune, without causing any considerable breach in the social union and confederacy.[5]

It strikes us that it is unfair for one to take advantage of others' adherence to moral rules by reaping the benefits of a society with moral principles, while being willing to break such rules oneself. If one were to accept the need for people generally to adhere to moral strictures, but nevertheless break them himself when expedient; we might well be justified in charging him with injustice and hypocrisy. In so doing, however, we will not have provided him with a reason not to do it anyway. Why not be a hypocrite? Why not indulge in injustice? Terrific moral reasons can be produced, but it is precisely the validity of moral reasons that is still at issue. As long as justice and fairness are, for these purposes, moral notions; appeals thereto will not provide any non-circular justification for moral activity.

I have tried, in the above section, to outline some

earlier, and I think less sophisticated attempts to show a necessary connection between reason and morality. I have been, I confess, quite brief in these efforts. I accept what I take to be the generally held view that these attempts are inadequate. At any rate, I am convinced by the standard arguments and see little sense in a lengthy discussion thereof. While I would be glad to hear of any thoughtful attempts at their resurrection, I have nothing particularly novel to contribute to these issues.

There is however, one exception. I discussed, at no great length, the view that since everyone benefits from the existence of moral institutions, we are thereby provided a non-moral reason to adhere to these institutions. This view had its most forceful original presentation in Hobbes, and has since been re-worked in Baier's The Moral Point of View and elsewhere. I argued, (fairly standardly) that while the rationality of moral institutions was demonstrated, we could not from that infer the rationality of each individual act prescribed by such institutions. While in this fairly crude form the argument has little chance of success, it is at the root of other arguments which bear closer scrutiny. Through the remainder of this chapter and the next, we will examine, in broad terms, three such arguments. In a sense, they are all extensions of a single argument, subsequent attempts progressing in sophistication. Each assumes a roughly contractarian model of morality, and all attempt to show

that the subtleties of social interaction are such that we can maximize the satisfaction of our ends only by being moral. This includes all of our ends, even non-moral ones. The assumption which underlies each of these views is that in cooperative society, the ends of all, in a sense, stand together. That is, that one furthers one's own ends by promoting (with some qualifications) the ends of others.

It has been suggested, most recently by Baier and David Gauthier, that our relatively simple and straightforward notion of prudence fails to do justice to the complexities of social interaction. I would like to examine the view that in some circumstances, the actions of others can render the prudent course of action less advantageous than some other course. A class of such special circumstances are proposed to be found in so-called prisoner's dilemmas. The classic prisoner' dilemma (hereafter 'PD') is as follows.[6]

Two men are arrested for some fairly serious crime, and are called upon to make a full confession. They are separated and told that the following table of jail terms will be applied. Notice that the only factor which affects the length of their sentence is whether either or both of them confess to the commission of the crime.

		B	
		confess	not confess
A	confess	8,8	0,10
	not confess	10,0	2,2

The decision grid is straightforward enough. Whichever combination of A and B, either confessing or not confessing obtains, A will receive the first jail term listed (in years), B the second. Presumably if both confess, then the prosecution can gain a conviction on a fairly serious charge for both of them. If one confesses and the other does not, then the one who keeps silent will be implicated, and will have perjured himself. He is therefore dealt a stiffer sentence still. As a reward, the one who confesses goes free. If neither confess, then the prosecution will be able to gain a conviction for each, but due to the lack of confession, only on a lesser charge.

The story we tell about the values is not important. It is important however, to see how the prosecution has set this up to encourage confession. To see why this is so, consider A. If B confesses, then by not confessing, A gets 10 years. However, if he confesses as well, he will get only 8 years. If B does not confess, and A does likewise then he (A) will get 2 years. But, by confessing while B does not confess, A is set free. Therefore, whether B confesses or not, A serves two fewer years by confessing. It apparently most satisfactorily serves his ends to confess. The situation is precisely the same for B. For exactly the same reasons, B always does better by two years if he confesses, whichever course A takes. Given that it is in the interest of each to confess, the

punishments will be those found in the top left outcome, that is, each will serve 8 years.

In this situation there are, for each, four possible results. They could go free, or they could serve 2, 8, or 10 years. Notice that if they both make what appears to be the most desirable choice (the one which will apparently best promote the individual's ends) each ends up serving only his third favoured jail term. While they have guarded against the worst outcome, their choice has left them in a not much better position. We might think this simply an unfortunate feature of the way the dilemma is constructed--and perhaps it is. However, consider the result if we keep the same table of punishments, but change the attitude of the people facing the choice. Both A and B could choose the course that would be optimal for the <u>other</u>. That is, A would not confess, since by doing so, whichever course B chose, B would be better off than if A confesses. The case is, once again, the same for B in respect to A's situation. So by concerning themselves with the wants of the other, and ignoring what they themselves want, both A and B will choose to not confess. This being the case however, the punishments would be those in the bottom right corner--each serve only 2 years. Remember that by considering only his own ends, each received an 8 year term.

This is a surprising result. For each prisoner does better in this case by ignoring his own ends than by con-

sidering only his own ends. Presumably the ends of each are best promoted by serving as little time as possible. So the prudent course of action does not seem to yield the more favourable result. Yet, we judge the prudence of an action in terms of how well the agent's ends are satisfied. This then is the dilemma. The point of adopting the prudent course (that which advances best one's non-moral ends) is to achieve maximal good for oneself. However, the PD suggests a situation in which the point of being prudent is undermined. It fails to achieve for the agent the greatest possible advantage. Good here is of course seen in terms of non-moral end satisfaction. Indeed, it appears that a greater good is realized by ignoring the prudential course, and adopting a policy of concern not for one's own ends, but for those of others. The coherence of employing prudence as a basis for rational action is challenged, It might be that prudence undermines itself. That is, it is imprudent, in some cases, to follow the dictates of prudence.

Proponents of a view such as we are now discussing could be read to be saying something like this. This dilemma (and similar ones) can be seen as symbolic of activity within a society. It is inescapable that we interact with other members of our community. Further, the promoting of one's ends is a more complicated sort of thing in the context of social interaction. On a personal level, as well as to the community in general,

cooperation yields a better result for all than hostile competition. We might see the 2,2 outcome as the result of a cooperative way of proceeding. Attempts to take advantage of others, as does the prisoner who confesses, banking on the good nature of the other to not confess; could be a strategy which within society is ultimately pernicious--even for the agent himself. It could be that in civil society, the ends of its members stand and fall, as it were, together. I do not wish to extend the analogy between PD and cooperative society too painfully far. While the PD has only two participants, society has thousands or millions. I take it, however, that there is some plausibility to the suggestion that social interaction makes a difference to the dynamics of a situation. There is enough such plausibility, at any rate, I suggest, to proceed. While I will argue that this attempt ultimately fails, we will see that the roots of more sophisticated attempts are here found.

At this point we are concerned that a chain of reasoning has suggested that prudence might, at least in some instances, undermine itself. This would leave moral practices on a much stronger footing rationality-wise, since we appear to best serve our ends by being moral. If this reasoning is something like correct, we might be justified in concluding that it is not prudent to follow prudence--or more baldly, that in this case anyway, the prudent act is not prudent. Being apparently a contradiction,

this conclusion is disquieting indeed. For if we have in fact derived a contradiction, we have derived a conclusion which is necessarily false.

On the other hand, this conclusion might, as many have suggested, point to the self-defeating nature of prudent action, at least in the context of cooperative society. In short, that the notion of prudence in this context is somehow paradoxical. If this is the case, the argument might suggest that the serving of one's non-moral ends is a futile effort. It might further show that we have good non-moral reasons for having moral ends. In this way we attempt to show that it is rational to have moral ends--that it is rational to be moral.

I take it that to hold that prudence is self-defeating is to say that in order to achieve the aims of prudence, that is to maximize what is desirable for oneself, it is not the case that one should act prudently. It appears that this argument runs roughly parallel to the so-called paradox of hedonism.[7] We sometimes hear proposed the following objection to various hedonistic theories. It cannot be the case that we ought to strive to maximize happiness, since human experience convinces us that striving for happiness will not render us happy--certainly not maximally so. A quest for happiness cannot be successful. This is because we can achieve happiness only by directing our energies in some other direction. Happiness will come (if it comes at all) in a

somewhat more mysterious way. Happiness then, or so the argument goes, is self-defeating in this sense. It cannot be achieved by attempts to be happy.

Whether this argument succeeds or not, I trust that the parallel between it and the troubles we are having with the notion of prudence is tolerably clear. In the same way that happiness is frustrated by specific attempts at its achievement, so the aims of prudence are frustrated by acting prudently. This argument, if successful, would be I suspect, very persuasive. If prudence is in fact self-defeating, we might be thereby provided a (non-moral) reason to take the view that moral reasons override prudential ones. If prudential reasons fail to succeed even in their own terms--if even prudential considerations give us reason to be moral, then prudence would be undermined as a rational justification for action. Notice that this argument does not attempt to undermine prudence by showing that there are moral considerations which tell against it. Rather, it attempts to show that the notion of prudence as a justification for action is inadequate by its own standards--that it defeats itself, and favours morality.

It seems clear that there <u>is</u> a sense in which prudence is self-defeating. This is the straightforward sense in which our taking the prudent course (in this case to confess) will yield a result less favourable for both, than if both chose the altruistic[8] course. Derek Parfit calls

this the sense in which prudence is "collectively self-defeating"[9]. So if A chooses not to confess while B does the same, they each achieve a better result than if both confess. It certainly is the case that less time is served by the two. However, for A this is not the optimum outcome. For, given that B does not confess, it is still in A's interest to confess. In this way, he gets off serving no time. As always, the situation is precisely the same for B. While it is preferable to serve 2 years instead of 8 years, it is best of all to serve no time at all. We cannot escape the feature of the dilemma, that whatever B does, A is best served by confessing, and likewise for B. Prudence, taken as a theory of individual action, judging individual benefit and detriment, is not self-defeating. Prudence counsels that we confess, and <u>individually</u> we do better by confessing. Prudence must be an individual notion. This is because, as Gauthier points out, we should

> take for granted that the primary subject of action, or activity, is the individual human person. And I shall presuppose that it is primarily to the individual that we ascribe rationality.[10]

There is a fairly straightforward sense then, in which prudence is self-defeating--it is collectively self-defeating. "If we were choosing a collective code, something that we will all follow, prudence would here tell us to reject itself."[11] Given that this is so, does pru-

dence condemn itself as a justification for action, or as an alternative to following the dictates of morality? I think it quite clear that it does not. For as we have seen, prudence is not individually self-defeating. In PD whatever course one agent chooses, the other is better off choosing the prudential course. Therefore, the individual does best by acting prudently. It cannot be an indictment of prudence as a reason for acting that it will not yield what we take to be the 'happiest' outcome, i.e., each serving two years. The purely prudential man will not consider it the best result. He (A for example) would find the happiest result the one where B makes the altruistic choice (confesses), and he makes the prudential choice (does not confess). Instead of serving two years, he serves none at all. That the other serves 10 years can hardly be considered a prudential consideration. The prudent man need not be at all troubled.

This points, I think, to a feature of PD which occurs in interaction between people in the real world. A person appears to benefit from his own prudence, but is harmed even more by the prudence of others. In PD, each benefits from his own prudential action, but is harmed to a greater extent by the prudence of the other. This might often be the case in civil society. Is this a reason to reject prudence? I think not. For even if we assume that prudence provides a benefit to the agent, and a greater

harm to the other; and that altruism provides a benefit to the other and a harm to oneself, we still do better by acting prudently. If the other acts prudently, then by acting altruistically we take both harms upon ourself (we serve 10 years, for example). At least by acting prudently, we get some benefit along with the harm we anyway incur at the hands of the other. Conversely, if the other acts altruistically, then we can claim both benefits by acting prudently (in our case by being set free). This is better (in end-serving terms) than the one benefit and one harm that would come of choosing altruistically when the other chooses altruism.[12]

Prudence, I take it, is collectively self-defeating. However, it is defeated only in a way that should not concern the prudent actor. As we have seen, prudence is not, nor should it claim to be, a code of conduct meant for the good of all. Rather it is a strategy whereby, if successful, one maximizes benefit for oneself. That prudence does not succeed in providing the best outcome for all should be by no means surprising.

We can see the resolution of the apparent contradiction within prudence in another way. When we concluded that it is not prudent to follow prudence, we had in mind two differing senses of 'prudent'. The two senses of 'prudent' are just the two that we have isolated in our search for a self-defeating prudence. When we say "the prudent course is not prudent" (as we had

concluded some pages earlier) we are using the notion of individual prudence in the first instance of 'prudent' and a collective notion which we misleadingly termed 'prudence' in its second instance. We seem to have been saying that the individually prudent course is not (in this case) collectively "prudent"--a far less remarkable claim. Indeed it strikes me as altogether unremarkable that a course which will maximize benefit for an individual might not maximize benefit for all.

I conclude then that prudence does not defeat itself as a rational justification for acting. Recall that we have characterized prudence as necessarily an individual notion, since it is the promotion of one's <u>own</u> ends that is prudent. As a consequence, it cannot be that individual prudence is self-defeating. That we might think of prudence as paradoxical is a result, it appears, of a confusion between the notion of individual prudence, and a collective notion which indeed is not prudence at all. If one acts so to be collectively "prudent", we might well conclude that he is accepting a morality--a morality of sorts at any rate. It is not this sort of prudence that we contrast with morality. It seems that we cannot yet rely on prudence to demonstrate the rationality of being moral.

We might think however, that our treatment of PD is a bit simplistic. For it may seem that we have failed to

do justice to this program by ignoring the effects that our choice between morality and the promotion of our non-moral ends, has on others and their choices.

There is a fairly common fallacy that we can describe in the following way. Consider a person (say myself) who wonders whether or not he should quit smoking. I might reason in this way. Smoking gives me pleasure, but getting cancer brings much pain. Indeed, cancer brings much more pain than smoking does pleasure. We might arbitrarily assign a value of +10 to the utility of smoking, and a value of -1000 to that of getting cancer. It is tautologically true that I will either get cancer or not get cancer. I can choose either to smoke or to stop smoking. I might set out a decision table as follows:

	cancer	no cancer
smoke	-990	+10
quit	-1000	---

All other things being equal, cancer will cause a utility of -1000 whether I smoke or not. Smoking will provide a utility of +10 whether I get cancer or not. Since it is surely the case that (like everyone) I will will either get cancer or not, and since ex hypothesi I am able to choose to continue smoking or quit, and since both of these pairs are mutually exclusive and exhaustive; we can be certain that one of the states of affairs described (one of the above results) will obtain. Straight-

forward game theory counsels that I should smoke. This, because whether I get cancer or not, I am +10 better off smoking. If I get cancer, my suffering is alleviated (albeit in a small way) by the pleasure of smoking. If I do not get cancer, there seems no reason to forsake the enjoyment of smoking.

The fallacy committed here is, I suspect, obvious. The decision procedure is faulty inasmuch as no account is taken of the causal influence of smoking on the probability of getting cancer. I presume that the decision grid could be made to reflect this, likely by providing the various probabilities, and having these range over the possible outcomes. But I am not so much concerned with making the argument work, as with pointing to the fallacy. Could Gauthier or Baier complain that our reasoning about the PD admits of the same fallacy?

Our conclusions about PD might be skewed because, like the above dilemma, they do not take account of the effects that the actions of one chooser has on the choice of the other. This is not very plausible in the one shot decision of our case. However, if we examine the way that the findings of PD might relate to every day moral situations, this suggestion acquires additional plausibility. As members of a given community, it could be that we are in some sense either in good standing or not in good standing as moral agents. It might further be that this standing is in some substantial way a result of the way

we make such social choices, and affects the way we are, by others, treated. It is almost certainly the case that the causal link between smoking and getting lung cancer is responsible for the flaw in reasoning to which I have pointed. It could be that in much the same way, a link between the effects that other's choices have on our own, might have skewed the results of our PD discussion unfairly in favour of the egoist. If this is so, then a different way of examining interdependent social relationships might be called for. David Gauthier in "Reason and Maximization" and elsewhere has attempted to do just that. His attempt is to reconcile morality with rationality using a roughly Hobbesian contractarian approach.

Gauthier accepts that rationality can be measured only in terms of benefits to the agent himself, and sees benefit in terms of preference satisfaction. One acts rationally by doing that which he believes will yield for him the greatest utility.

> The rationality which may be exhibited in choice is conceived in maximizing terms. A numerical measure is applied to the alternative possibilities, and choice among them is rational if and only if one endeavours to realize the possibility which has been assigned the greatest number. The measure is associated with preference; the alternative possible states of affairs are ordered preferentially, and the numerical measure, which is termed utility, is so established that the greater utility indicates greater preference. The complications of this procedure need not concern us here.[13]

This is consistent with the view of rationality which we

have been using. We can see maximizing utility or benefit as roughly equivalent to best promoting our non-moral ends. The attempt again, is to show that since it is by acting morally that we maximize the serving of our ends, we have even non-moral justification for having moral ends. So, it would be rational in terms either of moral or non-moral ends to be moral, because all ends are served by moral action.

He wants to draw a distinction however, between straightforward maximization and a new notion which he terms 'constrained maximization'. Straightforward maximization (hereafter 'SM') is the familiar policy employed by the egoist which counsels that one should, in each individual instance of acting, choose that course that will maximize expected utility for oneself. Constrained maximization (hereafter 'CM') on the other hand, is a policy which counsels the forming of, and adherence to, mutually beneficial agreements. These agreements include our broad societal moral institutions and conventions, as well as agreements and understandings between and among smaller groups. For example, we have seen that each of the prisoners in PD do better when both keep silent than when both confess. It would seem then, that a more rational course than the prudent one would be for the prisoners to agree not to confess. Both benefit if there is no confession. It might appear that such agreement is simply a good strategy, in some cases, for one who wishes

to maximize straightforwardly. To see how a policy of CM diverges with that of SM, we refer to Gauthier's condition of rational action. This condition, he proposes, is necessary. Sufficient condition(s) is(are) not provided.

> ...a person acts rationally only if the expected outcome of his action affords each person with whom his action is interdependent a utility such that there is no combination of possible actions, one for each person acting interdependently, with an expected outcome which affords each person other than himself at least as great a utility, and himself a greater utility.[14]

When the choices of others will affect the utility of one's own action, this action is said (by Gauthier) to be 'interdependent'. When one's action will give the same utility no matter what others do, his action is, by contrast, 'independent'. It is Gauthier's view that in civil society our actions are largely interdependent, since the actions of others affect the utility (for us) of our actions. It is in interdependent action that the strategy of CM is proposed to be more rational than that of SM, and as such provides the proper framework for justice.

> ...it enjoins each individual to agree, that when others are also willing, he will refrain from behaviour which would directly maximize his own utility, when the effect of everyone refraining is to bring about a state of affairs with greater utility for each person.[15]

CM and SM differ in just this way. The CM will take part in mutually beneficial agreements, to which he

commits himself to adhere. Since he adheres to such agreements, he is a credible moral agent. On the other hand, the SM will not have the opportunity to take part in such agreements, since he cannot be trusted not to violate the agreement. The CM is 'constrained' because he renounces the option of violating an agreement to which he is party, even if (as in the case of our PD) he would benefit still more from violating while others adhere. The SM is not so constrained, but pays the price in lost credibility. When agreement would not be beneficial, both CM and SM counsel that one not take part. In this case, the CM strategy requires that one maximize straightforwardly. The SM will likely wish to take part in such beneficial agreements, but presumably will not be accepted since the other parties are provided no guarantee of his adherence. His untrustworthiness will make others understandably wary of including him in agreements, so he will be left out.

CM then is a strategy which

> ...is clearly intended to maximize the agent's overall expected utility, by enabling him to participate in agreements intended to secure optimal outcomes, when maximizing actions performed in the absence of agreement would lead to non-optimal outcomes.[16]

On this basis, it is held to be more rational than SM, since

> If we compare the effects of holding the con-

> dition of SM with the effects of holding the condition of CM, we find that in all those situations in which...[SM]...leads to an optimal outcome, the expected utility of each is the same.

This is presumably because in such a case agreement is no more beneficial than no agreement, so the CM will act just like the SM.

> [B]ut in those situations in which...[SM]...does not lead to an optimal outcome, the expected utility of the SM is less. In these latter situations, a CM but not an SM can enter rationally into an agreement to act to bring about an optimal outcome which affords each party to the agreement a utility greater than he would attain acting independently.[17]

For example, in our PD, that each prisoner is given 2 years in jail is an optimal outcome for the prisoners. SM yields an 8 year term for each. If, however, the prisoners are CM's, they will receive the preferable 2 year term. Presumably if a CM is dealing with an SM, then the CM will not make any such agreement, since he is not able to trust the SM to adhere. Of course, that other parties to the agreement benefit is immaterial except inasmuch as their participation is required for the success of the agreement; and such participation could not be secured unless they too benefit thereby.

In short,

> ...since the CM has in some circumstances some probability of being able to enter into, and carry out, an agreement, whereas the SM has no such probability, the expected utility of

the CM is greater.[18]

It bears pointing out that this assumes both that the CM will always be able to identify other adherers, and that SM's will never be permitted inclusion in such agreements. Both of these assumptions are at best doubtful. Presumably, Gauthier holds that no agreement is possible among SM's since it will often best serve their ends to break any agreement made. For this reason, no agreement will ever be sincerely struck, and each party knows this. It is not the case that agreement will always be possible, even among CM's. For each party to an agreement, it must be that entering into, and abiding by the agreement will yield a greater expected utility than no agreement. Otherwise, it would not be rational to take part. Hence the condition of optimality even for others in the agreement. Circumstances, and the nature of other potential participants to an agreement will in some cases prevent agreement. Such a case would be, for example, when one has reason to suspect that others will fail to abide by the terms of an agreement.

I take it that society's moral conventions can be seen as large-scale agreements, to which society's members have extended at least tacit consent. It is morality as a framework of mutually beneficial agreements between and among society's members that Gauthier wishes to rationally justify. I take it that the following is at

the heart of Gauthier's argument. The CM can achieve all of the benefits (in terms of expected utility) as can the SM, and some additional ones. That is, he can in some cases benefit from rational agreement with others--a thing the SM cannot do. Our notions of justice are here based. Justice can in this way be derived from an economic[19] view of man, and is rational because it allows one to maximize benefit for oneself.

But the difference between a CM and an SM must be a real one. In order to retain the distinction between the two, the CM strategy will have to require the sacrifice of some benefit in abiding by an agreement, which the SM can claim. Otherwise, the CM is nothing more than an SM with a more enlightened strategy for personal utility maximization. It seems clear that he sacrifices (except perhaps in extraordinary circumstances) the option of breaking an agreement to which he is party. In terms of our PD, it appears that the CM gives up the opportunity of achieving his most preferred result (that of suffering no punishment) in order that he may claim, by cooperation, a two year term instead of an eight year one. This is the 'constraint' which the constrained maximizer accepts.

Gauthier must hold (and does hold) that the difference between an SM and a CM is not simply the difference between a naive, short-sighted egoist and an enlightened, far-sighted egoist. It cannot be simply the case that the CM is one who gives up certain present advantages that he

may achieve greater benefits at a later time. For this would mean that a CM is just an SM who is cleverer and more efficient. It is surely not inconsistent with the SM strategy to give up certain short-term goods in favour of greater future profit. We must be shown that CM is a strategy different from SM in some substantive way.

But can Gauthier uphold the substance of this distinction? I suspect that he cannot. Keep in mind that for Gauthier, rationality is seen only in terms of maximizing one's own utility--in rough terms, efficiency in providing for one's wants or preferences. I will claim that Gauthier succeeds only in showing one of the following. Either a person's willingness to enter into and abide by agreements (be a constrained maximizer) simply makes him a more efficient straightforward maximizer (but an SM nonetheless) or that the policy of CM is not rational inasmuch as it does not permit one to maximize his own utility.

To see why this is so, we will examine what a rational person does in the context of the formation of an agreement. It is clear that it will often be promotional of an agent's ends to strike an agreement, or be party to an agreement. In our case to agree with the other prisoner not to confess, on condition that the other does not confess. The constrained maximizer will abide by the agreement, and if the other does likewise, each will get two years in jail. This is apparently preferable to the out-

come if two SM's follow their strategy. We recall that they would each get eight years. We are meant to see that the CM can achieve all of the benefits that an SM can achieve, since his strategy is the same when his action is independent, and he can achieve more as well in cases like this one. But has the CM, in this case, acted rationally? Apparently not, for given that he is facing someone who will abide by the agreement, he is most rational to break the agreement. More rational, because if he breaks while the other adheres, he will get off entirely free, saving himself two years in jail. In addition, we must take seriously the possibility that the other will break the agreement. He may have been duped into thinking that the other is a CM. If this happens, then he ends up in prison for ten years--the least desirable result. It appears to be yet the case that he should break the agreement by confessing. Unless it is in some way guaranteed that the choices will both be the same, it might be rational to break any agreement we make.

One might well object, at this point, that by breaking agreements, we will in the longer term do ourselves much greater harm. Since we will lose credibility as an agreement-keeper, agreements will not be made with us. Since we benefit by participation in agreements, we will be, in fact, worse off for not abiding by the terms of those we make. I think that the thrust of this objection is basically right, but that it does not show what it is intended

to show. A good case is made that the CM is not so irrational as we might have originally thought. On the other hand, it is just at the point that this objection acquires some force, that the relevant distinction between SM and CM begins to break down.

If the CM is rational to adhere to at least some agreements, it appears to be because it is important that, at least among those with whom we regularly deal, we be trusted as an agreement-keeper. This credibility is beneficial because it is beneficial to be party to many agreements, and we will not be invited to participate if we are known to be untrustworthy. In this way, the rationality of abiding by the terms of some agreements is fairly well shown. However, this rationality is only in the context of concern for one's reputation as being trustworthy. It does not follow that one is rational to be in fact trustworthy.

Consider a case where the violation of an agreement will be beneficial, but could also remain undetected--or has some high probability of remaining undetected. It appears that it would be irrational to abide by such an agreement, since to do so would be to fail to achieve the full benefit possible. This since no credibility or reputation is lost by violation. Further, if the benefit realized by breaking a particular agreement is very great, it might be worthwhile to sacrifice a certain amount of credibility in order to achieve this extraordinary advan-

tage. Or, if we are dealing with a person or group of people upon whom we are unlikely to rely in future, there appears to be little need to be viewed as trustworthy. It could be that in any of these cases, agreement and cooperation will yield greater expected utility than no agreement at all. Hence it is prescribed by the doctrine of CM that we should adhere to agreements made in these contexts. But, it seems also to be the case that convincing others to adhere, but violating oneself, will yield <u>even greater</u> utility. We can only conclude that in these sorts of cases, CM gives a result that is not optimal for himself--hence not rational.

Remember the attempt here is to show that Gauthier's notion of CM either is irrational (in his terms) or comes to the same thing as SM. I have tried to show that in cases where (a) detection of agreement violation is unlikely, (b) the benefits of violation outweigh the detriments of lost credibility, and (c) violation is among a group with which we deal infrequently or only once (when we are far from home for example--say, in a strange city or country); adherence, even to a mutually beneficial agreement would be irrational, so long as violation would be more beneficial still. Were the CM to agree that we should, in such cases, violate an agreement, he would run seriously afoul of his own doctrine. At any rate, to do so would leave CM indistinguishable from SM.

But what of the cases (doubtless common enough)

where it is rational to adhere to an agreement. These would be cases in which the benefits that would accrue from violating an agreement are outweighed by the detriments that come of being seen as untrustworthy. These disadvantages are not insignificant. But this is not to say that the disadvantages that come of being untrustworthy are not insignificant. It is important, apparently, in seeing to our ends (being rational) that we seem to be trustworthy. If we do, others will include us in agreements which are beneficial. However, one who appears to be trustworthy, but is in fact not so, will be included in agreements; while those who are trustworthy, but are not seen to be so, will be left out. It appears then, that one who is rational will strive to appear to others as trustworthy as possible, but violate agreements whenever the advantages of violation outweigh the disadvantages brought about by lost credibility (by the loss of appearance of trustworthiness).

Seen in this light, however, the issue becomes somewhat simpler. It seems that it is not always rational to abide by agreements into which we have entered. It is sometimes rational to do so, and the way we go about deciding whether it is rational is vital. It appears that we should decide whether it is rational to violate or not on the basis of a straightforward utility calculation. It is not clear that an SM will never enter into and adhere to a mutually beneficial agreement. He might decide

that on balance, it is better to renounce the benefits of violating, in order that he can gain the greater benefits of being seen as trustworthy. But it seems implausible to suppose that such a calculation of benefits (expected utilities) will always dictate adherence. The CM no less than the SM is subject to the dictates of rational prudence. If we assume that we can balance the utilities and disutilities of agreement formation and adherence with those of agreement formation and violation; it will sometimes be the case that we act rationally to do one, sometimes that we act rationally to do the other, and sometimes that we act rationally not to enter the agreement at all. If the CM adheres all of the time to even beneficial agreements, he is guilty sometimes of failing in rationality. In other cases, his action, and its only rational justification is precisely that of the SM.

Constrained maximization is meant to be a policy of action that is rational and which provides some basis for rejecting purely prudential reasons for acting. If my argument is roughly correct, Gauthier's notion fails to meet one or other of these objectives. However, more recently, Gauthier has proposed a new defence of his constrained maximization. This is the third step in the progression we have been considering, and since I take it to be the most sophisticated and challenging, will now examine it in some detail.

Notes - Chapter 2

1 Hume, Enquiry, op.cit., sec. IX, pt. II, pp. 282-283.

2 see especially Leviathan, chs. 13, 14, 15, 17.

3 see especially The Moral Point of View (Ithaca: Cornell Univ. Press, 1958), chs. 1, 8, 11, 12.

4 Thomas Hobbes, from the Leviathan, ch. 13.

5 Hume, op. cit., p.283.

6 The classic prisoner's dilemma is attributed to A.W. Tucker, by R.D. Luce and H. Raiffa in Games and Decisions (New York: John Wiley & Sons, 1957).

7 see, for example, John Mackie, in Philosophical Quarterly, vol. 23 (Oct. 1973) p.290.

8 Here 'altruism' names a policy of promoting what one takes to be the ends of another, while not considering one's own ends. It will be associated here with acting morally. While I recognize that as a generality this would be troublesome, for the purposes of this limited context, this use will serve.

9 Derek Parfit, "Against the Importance of the Person", unpublished manuscript, p.8.

10 "Reason and Maximization", op.cit., p.412.

11 Parfit, "Prudence, Morality, and the Prisoner's Dilemma", Proceedings of the British Academy, vol. 65 (1979), p.563.

12 Parfit suggests something similar in his manuscript, op. cit., p.9.

13 Gauthier, "Bargaining Our Way to Morality", Philosophic Exchange, vol. II, p.15.

14 "Reason and Maximization", op. cit., p.427.

15 Gauthier, "Economic Rationality and Moral Constraints", Minnesota Studies in Philosophy, vol. III (1978) p.91.

16 "Reason and Maximization", op. cit., p.428.

17 This passage and the one above are from "Reason and Maximization", op. cit., p.429.

18 ibid., p.430.

19 The 'economic' man is a maximizing man. Gauthier, in "Reason and Maximization", op. cit., p.411, describes him as follows. "Economic man seeks to maximize utility. The rationality of economic man is assumed, and is identified with the aim of utility-maximization."

I think that the similarity between our discussion of prisoner's dilemma, and that of Gauthier's constrained maximization is tolerably clear. Each relies on facts about the dynamics of social interaction in an attempt to show that even our non-moral ends are best promoted by following the morally right course. Reasons are being proposed, it seems, for one to adopt, if possible, moral ends. If we adopt moral ends, then it will be rational to be moral in order to promote those ends. These moral ends, it must be granted, will be instrumental to further non-moral ends. But by serving these moral ends we will always act rationally. It is rational to be moral in all cases, because morality best promotes our moral ends and our non-moral ends--all of our ends. Being prudent (directly promoting only one's non-moral ends) on the other hand, fails to optimally satisfy either our moral or our non-moral ends. This at least if constrained maximization is indeed the supremely rational strategy. I have argued that, as yet, this has not been shown.

More recently, Gauthier has proposed that we can

resurrect the rationality of CM. His new defence picks up roughly where my objections (and those of others) leave off. Let us say (he proposes) that one is 'transparent' if his being a CM or an SM is easily recognized by all. Further, that one is 'opaque' if precisely the opposite is true, that is, it is impossible for another to discern whether one is moral or is purely self-serving.

Central to my primary objection to the notion of CM is the distinction between being a CM and seeming to be one. I have argued that Gauthier's argument goes some distance in demonstrating the rationality of the latter, considerably less in the case of the former. The issue, as I see it, is one of credibility. It must be rational, at least in the general case, to act so that we will be invited to participate in beneficial societal or personal agreements. It seems plausible that others will be less disposed to accept us if we appear to be of an untrustworthy nature. This is not to say that others will be less disposed to accept us into the moral community if we <u>are</u> of an untrustworthy nature.

However, if it is a fact about human nature or psychology that people are transparent, then the two (being trustworthy and seeming trustworthy) for practical purposes come to the same thing. If this were the case, it appears that CM would be the more rational strategy. For, given that any potential agreement-breaker could be expeditiously identified, none would be permitted to partici-

pate in beneficial agreements, and could never take advantage of CM's. CM's on the other hand would be able to identify other CM's and, at least sometimes, benefit from cooperation among themselves. Gauthier accepts, however, what is patent. That is, be it happy or sad to say, persons are not so constituted. It is far from true that the moral nature of people is transparent to others.

Further, were it the case that people are opaque, that is, precisely the opposite of transparent, SM would clearly be the more rational strategy. For in this case, SM's no less that CM's would be permitted participation in beneficial agreements. From this position, an SM could, with impunity, break any agreement to the detriment of agreement-keepers, whenever it was to his benefit.

For both the CM and the SM the worst outcome obtains when dealing with an SM. But the CM would come out worst of all. The CM, being deceived as to the nature of the SM, would be taken advantage of in an agreement. In our story, this corresponds to the CM getting 10 years in prison. The SM, however, would find himself in a situation where both break any agreement made. The result is much like that of no agreement whatever. This result roughly corresponds to the top left PD outcome where each SM gets 8 years. Better outcomes are realized when either is dealing with a CM. But while better for each, it is still the SM that does best in this case if people are opaque. While the CM would be able to make an agreement

and have it kept, resulting in a jail term (in our dilemma) of only 2 years; the SM would be able to take advantage and get off free. Clearly then, opacity would undermine the rationality of CM. However, I think that we can agree with Gauthier that it is no more reasonable to assume opacity in people than transparency.

Gauthier then argues as follows. It appears that transparency favours the rationality of CM, and opacity favours that of SM. However, it seems that we are neither transparent or opaque. So, if we can somehow objectify the degree of transparency or opacity that people have, we might be able to discover the point at which rationality ceases to favour SM. To do this, he is obliged to posit values for the benefits resulting from the four possible outcomes. He suggests the following table:

keeping to an agreement while other(s) break	- 0
no agreement or all break agreement	- 1/3
all keep agreement	- 2/3
breaking agreement while other(s) adhere	- 1

Of course, 1 is the most preferable outcome, 0 the least preferable. His argument is, as it might be expected, largely mathematical. Using the above pay-offs as constants, and the following variables, the calculation can proceed.

These variables are necessary to attempt to objectify the degree of transparency or opacity. Here 'strategy'

refers to one's choice either of CM or SM.

p - probability of identifying another's strategy

q - probability of being identified oneself

r - proportion of CM's in the population, i.e., the probability that a given person will be a CM

The idea is to determine which values for p, q, and r will, in combination, produce cross-over points between transparency and opacity, above which CM becomes the rational strategy. If realistic values for a given moral community reach or exceed such a point, then CM, it is proposed, is truly rational.

The utility calculations can proceed in the following way:

> If one is a CM, then one's average expected utility from interaction is 1/3, plus the gain from successful cooperation, [and] minus the loss from being taken advantage of. Successful cooperation yields a gain of 2/3 - 1/3 with probability rpq (the probability that one is interacting with a CM and that mutual identification successfully occurs). Being taken advantage of yields a loss of 1/3 with probability (1-r)(1-p)q (the probability that one is interacting with an SM who one misidentifies while being correctly identified oneself).[1]

So, for the CM, the average expected utility is:

(A) $1/3 + rpq(1/3) - (1-r)(1-p)q(1/3)$

> If one is an SM, then one's average expected utility from interaction is 1/3 plus the gain from taking advantage of a CM. Taking advantage (i.e. successful violation) yields a gain of 1 - 1/3 with probability rp(1-q) (the probability

that one is interacting with a CM whom one successfully identifies while being misidentified oneself).

In this case, the average expected utility is:

(B) $1/3 + rp(1-q)(2/3)$

For Gauthier, it pays to adopt CM if and only if (A) is greater than (B). Simplifying, we find equivalent expressions to be:

(A) $q(p+r+1)$

(B) $2rp(1-q)$

Employing these formulae, we find that CM becomes rational at (for example) the following points:

p	q	r
.7	.7	.75
.7	.8	.462 (6/13)
.8	.7	.636 (7/11)
.8	.8	.333

While further value combinations are producible, this abbreviated table serves to give one a rough idea of Gauthier's way of proceeding. At any rate, it is not the calculations or the mathematics to which I will primarily object. As an example though, if I am able to identify the strategy of another 7 times out of 10, and I will be correctly identified 7 times out of 10, and the population is 75% CM's then it will be rational for me to be a CM. If Gauthier's framework is well-founded, and if the correct

values for p, q, and r in our society give a larger value for (A) than for (B), then in our society CM is the rational strategy. Agreements are here meant to include society's moral conventions and institutions. Since to adhere to one's agreements is to adhere to morality's strictures, the CM is meant to be the moral individual. Hence if CM is rational, then the rationality of morality is also demonstrated.

Before I proceed to the heart of my objection, I should like first to point to the entirely arbitrary nature of his value assignments. It appears to be well-nigh impossible to tell whether the figures which Gauthier proposes as outcome utilities are in any sense reasonably related to the satisfaction of one's ends. Even if we did have some tolerable indication that the figures are realistic (and I am quite at a loss to speculate what such evidence would be like) we have no way of knowing whether the corresponding necessary values for p, q, and r are achieved or achievable for individuals and/or for our social organization. It is disturbing that Gauthier's utility constants are so unashamedly arbitrary. Further, his presumption that these figures are in any way related to the situation within a community appears to be at best an article of faith.

This consideration, I feel, leaves us with little reason to think that CM has been, or can be, justified

rationally. However, my objections to this framework run much deeper. Even if it is the case that Gauthier's values are realistic; and that realistic values do, using the method that Gauthier suggests, yield for everyone the rationality of CM, its case would be by no means made. I fear that the project is, at a basic level, seriously misfounded.

First, it should be pointed out that Gauthier might have some difficulty explaining how, in the first place, there got to be CM's in the population. In order for it to be rational to adopt CM oneself, there must be a certain relatively large number of CM's already present. How did already existing CM's come to adopt this strategy? It cannot be that there was a simultaneous choice by vast numbers of people to adopt CM. This because, at the time of the choice, there were no CM's and no way of knowing how many would choose to adopt this strategy. As a result, the choice of CM would not have been rational for any of the population. If CM was chosen, as it were, by people one at a time, the choice to adopt CM would have been irrational for everyone who chose CM before the requisite proportion of the population had done likewise. For example, if an r value of .5 is required in a society of 1000 to rationally justify CM, then it would be rational to choose CM only after 499 others had <u>irrationally</u> chosen CM. Either way, we must assume that for there to be CM's in the population at the time we make a choice, a profound

number of these must have themselves chosen irrationally. If the point of CM is to allow one to be supremely rational, it seems odd that it was, at least by many of those who came before us, chosen despite that it was, at the time of choosing, irrational.

In addition, his use of average values for p and q are liable to objection on the following level. Gauthier appears to assume that the values for p and q are the same for all. That is, the probability of successfully identifying the strategy of another, and that of being successfully identified oneself are values which are constant for all members of a given moral community. But is this really so? Both CM's and SM's benefit from being thought of as CM's. Therefore, if rational, all people will do their best to appear to follow CM. I suppose that different people are, to varying degrees, capable of discerning the moral strategy of others. This suggests the possibility that while a good predicter might be more rational to choose CM, one who is easily deceived (the good-natured, trusting sort) might be better off, depending on other values, being an SM. This is because, as a CM, he would be deceived and taken advantage of more often than others. For such a person, the value of p would be relatively small. It could be that for him, the utility of cooperation which he would be afforded as a CM is outweighed by the disutility that would come of being more often deceived.

It is further reasonable to assume that different

people will be, again with varying degrees of success, able to deceive others as to his strategy. For a person who is particularly adept at making others believe that he is an CM, it might be rational (again depending on other values) to do so while following the SM strategy. Even if for many others it is rational to be a CM, for this person, the benefits derived from taking advantage of others would be greater, since he would be in a position to take advantage more often. It might be that these benefits outweigh the detriments that would come of a somewhat reduced opportunity to benefit from cooperation. It seems that unless it is in principle impossible for one to be so easily deceived, or such a good deceiver as to justify the SM strategy; it is the case that SM could be rational for some, while CM is rational for others. As long as each wishes to find out what is personally rational, the values for p and q which apply to different individuals might result in different rational strategies. It would be mistaken to take cross-societal average values for p and q, since this theory intends to rationally justify for each individual the strategy of CM. That this framework might counsel different strategies for different people must be troubling for Gauthier. In "Reason and Maximization" he points out that

> [i]t is the individual who is rational, but qua rational, one individual is the same as another. Hence any answer to this question must be the same for all persons: What one person must do

> in virtue of being rational, is to be characterized in the same way as what any other person must do, in virtue of being rational.[2]

We might also wonder about the values which Gauthier has chosen as utilities for the various outcomes. Aside from the obviously arbitrary nature of the value assignments, we might question the values on the following level. His assumption appears to be that the loss which is felt by those taken advantage of, is the same as the gain enjoyed by the person taking advantage. But there seems to be no good reason to think that this is so. If a beggar steals a hundred dollars from a millionaire, surely the gain which the thief enjoys is greater than the loss which the victim suffers. Different goods are more and less important to different people. The loss of the same benefit may be less distressing for one person than for another. Indeed it seems that if there is an equality at all between the utility of the loss, and that of the gain from a given crime, it is simply the result of coincidence. In the context of Gauthier's framework, it is not altogether clear in which direction this difficulty skews the results. It is clear however, that the results are not quite right because of this.

Let us attempt to put this discussion back into some more concrete perspective. The strategies of CM and SM diverge in just this way. While both see the virtue of making beneficial agreements, the CM will never break an

agreement (at least not one with another CM) even if doing so will result in greater benefit still. Conversely, the SM will always break such an agreement if the benefit so achieved will be greater than that of adhering. In rough terms, the CM corresponds to the morally just man, inasmuch as moral strictures can be seen as agreements between and among his society's members which are for all parties beneficial. There will be times when to contravene a moral requirement will be (it appears initially) of benefit to an individual. The CM (moral man) will refrain from such an indiscretion and abide by society's moral conventions. The SM will violate any moral stricture when doing so will result in some greater personal gain.

The SM strategy is straightforward enough. In every case, do that which will maximize one's own expected utility. The CM strategy is a bit more complicated. When dealing with one who he takes to be a CM, the CM will first calculate to decide whether agreement is, for him, more beneficial than no agreement. If it is not, then no agreement is struck. If agreement is beneficial, then he will join in and will not violate the agreement. The CM, when dealing with one he takes to be an SM, will not make an agreement. The SM will take part in beneficial agreements if permitted, but presumably this could only happen if he is misidentified. It is tempting to suggest that the CM might enter an individual pact with an SM, but be willing to violate if beneficial. Gauthier is not clear

on this point, but it seems that, unless it is for both more beneficial to keep than to break such an agreement, there is little in it for the CM. Since the utility of both violating is assumed to be the same as that of no agreement, the CM can do no better than keeping out of the agreement. It also seems that being prepared to violate an agreement is evidence of an insincerity which is not compatible with our view of the CM as a good or moral man. However, I will have more to say about the treatment of SM's by CM's later.

The relevant difference then, between the SM and the CM (moral person) comes when breaking an agreement to which all have agreed or would agree is beneficial to the agent. The only way that CM could benefit the agent in that case is by the effects which his strategy has or will have on the way others treat him. I can think of at least two ways in which this might be so. While suggestive, Gauthier's account does not make clear which is meant.

It might be the case that, in abiding by the agreement, one will get a reputation as a good or moral person. He will be trusted with participation in further agreements. Other members of the moral community will become aware of who the agreement-keepers, and who the agreement-breakers are. Being a CM (an agreement-keeper) is more rational because we will be invited to participate in enough additional beneficial agreements to outweigh the negative

effects of occasionally being deceived and taken advantage of. While the SM gains some from taking advantage of others, this circumstance will become increasingly rare since his agreement-breaking will result in decreasing participation in such agreements. Again, the presumption is clear. It is that the SM's reduced opportunity to take advantage of others will decrease the benefits therefrom to the point where he comes out, in all, worse than the CM who can benefit from cooperation.

But if this is the conception that Gauthier proposes, then our earlier argument against "Reason and Maximization" will once again tell. For it appears that the SM strategy is being misrepresented. The SM strategy is just this. The SM will choose the course which will, of all possible courses, yield for him the best expected outcome. Here best refers to that which will maximize the satisfaction of his wants or preferences. However, if the framework is as I have suggested, and the SM's agreement-breaking will result in diminished future benefits (due to the effect it will have on the behaviour of others toward him); then it seems that such an instance of agreement-breaking would not truly be in keeping with the SM strategy. As long as the future consequences of a given instance of agreement-breaking are more dire than the advantage gained by doing so, the agent has not straightforwardly maximized his own utility.

If the reason that CM is more rational is that the

SM prejudices his own treatment by other of society's members with his agreement-breaking behaviour; then the comparison has been unfairly made. It is hardly a remarkable observation that those who indiscriminately break society's moral conventions will be generally looked down upon, and will be in all worse off for it. It is something short of contentious that the moral man will be better off than the barefacedly wicked man. It has yet to be shown that he will do better than the selectively wicked individual.

A more plausible reading of this view might be that the SM does in fact do what will best promote his ends. It is not simply his past actions that prejudice others against him. In addition, and in some significant way, there must be something about his amoral nature which can be, to some extent, discerned by others. It is vital to this view of the question that the reason the SM is not trusted, is that others are able to see his untrustworthy nature by means of considerations which include some that are <u>not</u> past instances of immorality. For, if it is purely past acts that have caused the individual to lose benefits, we can conclude that those acts were not those of an SM. Rather, such a person would be a short-sighted, unenlightened and somewhat foolish egoist.

It might be that persons who have the disposition to take advantage of others in social contractual situations can be with some accuracy identified. At any rate, it

appears that Gauthier needs to show something like this in order to salvage the rationality of CM. I take it then that the most reasonable position open to Gauthier is this.

Anyone who is not a CM (hence not a moral man) will lack a roughly identifiable disposition toward moral (agreement-keeping) action. Conversely, moral persons (CM's) will be, again in a rough way, identifiable by some clues which point to this disposition. The immoral man is distinguished not just by his immoral behaviour, but also, and to a significant extent, by his immoral disposition or nature, the evidence for which comes in the form of some factors other than his intentional behaviour.

We can, at this point, say more about the CM. First, he cannot violate an agreement to which he is party, at least when dealing with other CM's. It is at the core of Gauthier's argument that always cooperating (constraining one's maximizing behaviour) is most rational. We will later discuss the treatment of SM's by CM's. As a result, the CM's adherence to an agreement is guaranteed by his participation in that agreement.

Second, it must be that the CM disposition is one which is not easily renounced. This is because the existence of this disposition to moral action will not be a particularly credible indication of the trustworthiness of an agent if it can come and go with relative ease. It appears that unless this disposition is to some degree unshakeable, the rationality of adopting it is lost. If

it were possible to easily alter one's disposition, it would be rational, in some cases to do so. It would be to one's advantage, in at least isolated instances, to renounce one's CM disposition, commit an immoral act--violate an agreement; and then re-adopt CM. But it is Gauthier's attempt to rationally justify refraining from immorality. If it is both possible and more rational to act immorally (in an agreement-keeping way) then this attempt has failed.

It is also important that the CM not be able easily to alter his disposition for the following reason. If it were the case that one could alter one's disposition from CM to SM and back again with anything like ease, this would presumably be a state of affairs of which all would be aware. This being so, it would be generally known that while a CM will refrain from breaking society's agreements, any CM might (particularly if he is rational) adopt SM in isolated instances and exploit the adherence of others. It appears that if this were the case, then not even a CM could be or would be trusted. Without the greater opportunities afforded by being trusted, the CM loses his advantage over the SM. Indeed, the SM would have the superior strategy if he were no less credible. Once again, in order to defuse complaints about irrationality, it appears that we must assume minimally that it is an involved sort of process to alter one's disposition in this way.[3] It must, however, be possible to become a

CM, if one is not such already, and this decision must be freely taken. Otherwise, there can be little point in showing the rationality of CM. I will assume that all are able to adopt CM, but are able also to choose SM.

Given the nature that this disposition would seem to have to give the CM, we can wonder first whether such a thing is present or possible in people. It appears that we do sometimes identify a moral disposition in others. Some people are good-natured, attentive when spoken to, and adopt a concerned and sympathetic expression when hearing of another's misfortunes. Others seem to go through life wearing a scowl, and looking hesitatingly or with suspicion upon others. We often (at least in literature) learn of the villain having 'shifty' eyes. While I presume that we can rule out the evidence of physiognomy as indication of a moral or an immoral disposition, I suppose it possible that bearing, facial expression, tone of voice, various bodily movements and other clues might provide some evidence of an individual's either having or lacking a moral disposition. We must assume that one acquires these traits as he adopts the moral disposition, and that he carries whatever traits are involved in a generally more sincere fashion than can be feigned by an SM.

Before we can proceed to my substantive objections to Gauthier's view, it will be necessary to say something more about the CM and his disposition. It is central to Gauthier's theory that it is rational to adopt CM. Given

a choice between CM and any other strategy, it is CM that will result in maximal benefit for the chooser. I have tried to show that it cannot be the case that an SM can consistently choose to act such that future advantages will be prejudiced by the bad opinion that such acts would engender in others. Doing so would be to fail to be an SM. In order then to keep the distinction clear between the CM and the SM, it is important to show that the CM benefits not simply by acting in a certain way (an agreement-keeping way) but rather that he benefits by adopting this disposition. He benefits from this adoption because he will be able thereby to participate in more of society's mutually advantageous contracts. Therefore, the choice to adopt CM comes to the same thing as a choice to program oneself to moral action. It cannot be a choice, case by case, always to act morally.

Although Gauthier is not clear about this, it seems that we might conceive the difference between the CM and the SM in the following way. The SM will examine the utilities which face him on a case by case basis. In each individual instance, the course he chooses will be the one which will yield for him the greatest expected utility. The CM, however, does not, at least when party to an agreement, consider the individual circumstance. While he will weigh the utilities when deciding whether to take part in an agreement, once he has decided that the agreement is beneficial, and he has agreed to its terms, then he will

not consider the utility of violating or adhering.

> One cannot be disposed both to cooperate with one's fellows, and to maximize one's utility given one's expectations of their behaviour. For the latter involves being disposed to take advantage of one's fellows rather than genuinely to cooperate with them.[4]

It appears that Gauthier sees an inconsistency between cooperating behaviour and calculating behaviour. I get the impression from this passage that with his strategy, the CM also adopts a certain cooperative spirit. Perhaps this happens necessarily, due to the nature of CM itself. One cannot, in some sense, both have this cooperative spirit, and weigh the utilities of violation and adherence. Notice however, that the CM <u>is</u> capable of calculating utilities when judging whether to take part in an agreement. Gauthier is very careful to talk about the conditions which must hold before participation in an agreement is justified. Undoubtedly he does not expect that the CM will ally himself with every agreement offered him. So, it is clear that some weighing up must go on at that point. It is odd that CM appears to demand that one calculate utilities prior to agreement, but forbids one to do so after agreement is struck.

We have now drawn, I think, an adequate picture of the notion of constrained maximization, and the nature of the constrained maximizer. We can now return to my criticisms. I earlier attempted to point to the limitations

which this view has. I pointed to the fact that, if there were fewer than enough CM's in the population at a given time, it would not be rational for anyone to adopt that disposition. This means that, unless people in large numbers act, or have acted irrationally; it will never anyway be rational to adopt Gauthier's strategy. In addition, it appears that since values for p and q will not be the same for all, that the rationality of CM might only apply to some, but not all people. These considerations indicate that the applicability of this view is by no means universal. It is not a strategy that can be rationally held by anyone, or at all times. For a moral theory, I take this to be a profound drawback.

The problem of the limited scope of the theory, coupled with the intolerably arbitrary character of the value assignments leaves a theory with serious initial troubles. Gauthier is left with difficulty relating his formulae to any practical view. It will be my contention however, that this framework admits of rather more serious problems than these. We have discussed briefly the nature of the CM disposition. It is the acquisition of this disposition that is meant to be rational.

I think it necessary however, to look more deeply at the nature of the cooperative spirit which comes from the disposition to CM. That the doctrine of CM forbids (at least in agreement with other CM's) the calculating of benefits is clear enough. It is also clear that, being a

disposition, this comes to more than simply a decision, case by case not to violate. It seems that there is a sense in which it does not occur to the CM to calculate--that his adherence is, we might say, automatic. But what exactly does this come to? It is worth examining just what effects the CM disposition has on the CM himself. We are tempted to say that while the CM disposition leaves one unwilling to calculate about, or violate an agreement, we wish perhaps to stop short of holding that he is unable to do so. Surely there is some free choice remaining for the CM, despite the fact that he will not exercise it. I am not convinced, however, that this free choice does not rob him of rationality.

Is it possible for a CM to violate an agreement a small number of times (or once) without loss of credibility-without losing any opportunities to participate in future agreements? If he could, then CM would be irrational in its demand that one always adhere. This is because it would be more rational to be just like a CM (have the CM disposition), but take advantage of extraordinary circumstances to violate. It must be then, that the CM loses a certain amount of credibility (actually reduces his agreement forming opportunities) if he ever violates. It must further be that the loss of benefit from this loss of opportunity is as great or greater than the gain resulting from the violation. In order for it to be always rational to adhere, it must be shown that it is in principle impos-

sible for the gain from violation ever to exceed the loss from reduced opportunity to make agreements and benefit thereby. But what if the benefit from violation is extraordinarily large? How could adherence be justified in this case? We might suggest that the larger the benefit, the more the effects of our conscience will prey upon us. Others will be able to recognize this state to the degree that we have benefitted, that is, to the extent of our crime. This appears, however, a dubious assumption. While it is usually the case that a larger crime will result in stronger feelings of guilt, it is not at all clear that the guilt feelings are in proportion to the benefit which we claim. This, particularly since the blameworthiness of an act (the source of our conscience troubles) is not always proportional to the benefit which one claims from the act. It is still less clear that the physical manifestation of that guilt will be apprehended by others in proportion to the benefit claimed.

In all, I see no reason to suppose that it might not be more rational for the CM to occasionally violate. Our only alternative is to hold that the CM is not irrational for not violating because he is <u>incapable</u> of violating his agreements by the force of his CM disposition. Such a strongly deterministic conclusion is, I take it, unacceptable. But is is avoidable? I am left to conclude that unless it is in principle impossible for the benefits of violating to outweigh the detriments of lost credibility

of the sort which manifests itself physically (and there seems no reason to think that this is in principle impossible) then CM is not the supremely rational strategy. If there is a failure in rationality by the CM because of his disposition, then it cannot be that we have successfully grounded morality (as constrained maximization) on reason.

It might be objected that the choice of CM is rational, because, over all, we do best by its adoption. One who considers violation, or is willing to violate, is not a CM, and hence would not be in this favourable position in the first place. This assumes that anyone who can break an agreement, or who can consider breaking an agreement, will not have the roughly identifiable disposition to CM. It is true that someone who violates an agreement is not a CM. I argue that such a person is more rational than a CM. This being the case, CM cannot be the supremely rational strategy. But there may be something to the complaint that this person would not be invited to participate in beneficial agreements so often, and so will not be in the position to benefit from violation.

This point recalls our discussion of the cooperative spirit which the CM disposition instills. Remember that we concluded that the CM does not calculate utilities after an agreement has been made. Due to his cooperating nature, he is unwilling to question whether to break his agreements. Central to this issue, though, is whether the CM

is unable to violate or consider violating, as well as being unwilling to do so. What is at stake here is the nature of the disposition. Can one have the CM disposition and yet be able (if unwilling) to calculate or violate, or is it the case that if we are able to violate or consider violating, then we do not have the disposition? One or the other of these clearly must be true. I will address the former possibility first.

If it is the case that the CM is able to consider violating an agreement, but is unwilling; then he is being consistent with his CM strategy. However, if this is the case, it must also be that one can violate or consider violation if one chooses, and still be roughly recognizable by others as a CM. It seems that an ability to violate does not remove whatever attributes identify one as a CM in the eyes of others. Given that one can still be a CM (with the advantage of credibility perhaps somewhat lessened) while being able to violate, we can only conclude that our objection above is after all valid. Since there seems to be no reason why, in principle, the detriments of lost credibility will always outweigh the benefits of violation, it appears that it will be rational sometimes to do so. If a CM is able to violate (his unwillingness notwithstanding) he is irrational not to do so in such a case. But ex hypothesi, the CM will not violate. Therefore, given that a CM is able to violate or consider violation (even though he is unwilling to violate) without

losing his CM disposition, the CM is, in cases as above, irrational. For Gauthier this will not do at all. It is vital that the CM be at all times rational. It must be then that the opposite is true, and that the CM is unable to violate or consider violating.

If it is the case that the CM is unable to consider violation (since if one did so he would not be a CM and would not be identified as such), then he would be in a position to keep agreements and benefit by so doing. The ability to violate or consider violation will be apparent to others, and would bar one from the trust of the community. Having the CM disposition is incompatible with even considering breaking an agreement. If this were the case, we could not charge the CM with being irrational, since reason cannot insist we do that of which we are incapable. This would mean that the CM disposition is a powerful force indeed, in that it renders people unable to break, or consider breaking their agreements. For if it does not so compel them to adherence, then it would be possible both to have the CM disposition (and be recognized as such) and be able to violate. If one is able to violate, then it is rational to do so, at least sometimes. Since the CM will not violate, he is irrational, because he fails to do that which is rational. Anyone who can consider breaking would presumably be identified as lacking the CM disposition, and would be treated accordingly. Unfortunately, if this is the case, then I am

afraid that there are few if any CM's in existence. For the CM is a person whose disposition in some sense compels him to moral (agreement-keeping) action. The problem is that there does not seem to be any such people around. If everyone who (however unwilling) is able to calculate and violate is seen as not a CM, then probably everyone will be seen as not a CM. We can agree that a willingness to violate is inconsistent with CM, but it appears now that an ability to violate is also in this way inconsistent. Unless he cannot but will to adhere, then the CM's unwillingness to violate, except in extraordinary circumstances, must be irrational. While I suppose it possible that the CM disposition does not cause this inability, it is enough for our purposes that there is an incompatibility between having the CM disposition and being able to break or consider breaking an agreement.

If this is the case, then the CM is a humanly improbable specimen. There appears no way that there can ever be enough of such individuals (a high enough value for r) to satisfy Gauthier's formulae. But if the CM is not this sort of person, then he is quite simply irrational at times. I take it that, to be rational, the CM must have such a nature, and that this constitutes an absurd consequence of Gauthier's framework. I do not suppose that anyone can choose to be so determined or compelled to moral action as the CM apparently must do. Anyone who has such a mental block must be a human oddity--a psychologically fascinating

study. Lacking this remarkable quality, however, the CM is irrational.

But there is another, perhaps more serious difficulty with the way that Gauthier has set up his formulae. It seems obvious enough that in some cases, it would be rational even for a CM, to violate an agreement to which he is party. At least, I have argued, this would be if he were capable of so doing. The above proof depends, for its force, on there being circumstances of this sort, even for the CM. Such a case would be an instance of immorality which is nevertheless personally beneficial. It is such cases, I have suggested, that render CM irrational. We might wonder, however, how the existence of such rational acts (apparently so damning to the notion of CM) is reconcilable with Gauthier's arithmetical framework. Recall that his figures seemed to indicate that CM is the optimally rational strategy. I believe that the required explanation will also serve to show where Gauthier went fundamentally wrong in his calculations, and provides perhaps the most persuasive objection to his notion of constrained maximization.

How can it be that there are cases in which some larger benefit can be gained by agreement violation than by adherence, even for the CM. These are cases where it is a failure in rationality to adhere to an agreement. But if there is a failure in rationality, how can the mathematical formulae have shown the supreme rationality

of CM? Surely the existence of such circumstances is at least compatible with there being adequately large values for p, q, and r. The source of the problem, I will argue, comes in the use of average values for the assigned utility outcomes. We have already seen that the values for p and q could vary from person to person.

In addition, it is surely the case that the various payoffs will also vary from case to case. For example, take the utility accrued by violating an agreement to which others adhere. Gauthier has as a "simplifying assumption" that violating in such a case provides on average 4/3 the expected utility of adhering to that agreement. Even if this were so, it seems clear that the benefit from some cases of violation will be much higher. For such cases it might be rational to violate (given requisite values for p, q, and r) even if the average benefit from violation were not so, due to the detriments involved in the loss of credibility. If the assumption is that no payoff from violation will be greater than the average, (and hence none less) then the figures might be of some value. However, it is plain that the payoff for some violations will be above, on occasion well above, the average value used. We can see how this same difficulty infects each of the other three 'simplified' utility outcomes.

The problem I see here is that Gauthier has justified CM as a moral strategy by using average values for all of

the utility constants. However, it appears that the policy of CM is not defensible on these grounds, since in individual cases, probabilities and utilities which are above or below the average might show the rationality of violation. It appears that one will do better than a CM, by adopting the CM disposition, acting like a CM most of the time, but exploiting exceptional circumstances. Even if it were the case that correct average values yield the rationality of CM, it does not follow that CM is counselled in all cases, or as an unshakeable strategy. It seems that at best his formulae might serve to test the rationality of a particular choice between adherence and violation. If this is so, it would likely be more helpful to the SM who judges utility on a case by case basis. Unfortunately, due to the unjustifiably arbitrary character of the figures employed, it is difficult to see how this framework could be of use to anyone.

Once again it may be objected that the average values are valid and that the CM does best after all because it is only by giving up such extraordinary advantages that the CM can claim the greater advantage of being included in such agreements in the first place. He would not be included in these if he were willing ever to violate them. But, as I argued earlier, this assumes an incompatibility between having the CM disposition (and being seen as having it), and being able to calculate or violate. If there is not such an incompatibility, then the CM will be able to

violate agreements, although he is unwilling to do so. Admittedly, there may be some loss of trustworthiness involved in violation or even in considering violation, but there seems no reason to think that the benefits which come of violating will never outweigh these. To fail to take advantage of such opportunities (as the CM must) is to fail to be rational.

On the other hand, if there is this incompatibility then it truly will be the case that the CM will be the only one to benefit from cooperation in agreements. However, this would render the CM a strange character indeed. For he would be the sort of person who is unable to consider violating an agreement, or any moral stricture. If there are any such people, it is clear that there are very few. I conclude then that CM is a strategy which is not, as Gauthier had hoped to prove, rational. This because the only rational conception of a CM is either plainly false (as a conception of human behaviour) or which restricts the class of CM's to include, most likely, no one.

While I have proposed a number of objections to Gauthier's framework, there are other aspects to his view to which one might object. It seems clear, for example, that the notion of justifying moral action on only personal end-promoting grounds, leaves a morality which ill-corresponds to our ordinary reactions. While I am not convinced that I am here saying anything decisive, it

could be that an attempt of Gauthier's sort misses the point. It seems to be just false that when we do what we think is right, we always believe that it will ultimately be to our advantage. Whether we in fact ever sacrifice to be moral, it appears that we anyway believe that we do and accept the sacrifice. Gauthier provides self-seeking reasons to be moral. However, it has been observed by many that this is a motivation which is plainly inappropriate. It is the wrong kind of motivation, and not just the morally wrong kind. It may be that as a motivation, it misdescribes the agent's reasons when coming to a decision about acting or failing to act morally. I suppose it could be the case that people often act morally, and with a suitably moral motivation; and that morality is nevertheless, in his egoistic sense, rationally defensible. It strikes me as untenable, however, that we could be so systematically confused as to think that we are truly sacrificing to act morally, and that morality comes to something more than serving one's own ends; when in fact the dictates of morality and the serving of one's ends come to the same thing.

Another, but perhaps related objection has to do with the range of activity over which Gauthier's view extends. If it might be that this view poorly describes our everyday notion of why we act morally, it appears that it further fails to do justice to the complete range of moral judgements we commonly make. I wonder, for example,

whether a CM should be moral to a drowning SM. Except in extreme cases, it is our usual view (it is my view at any rate) that we should save a drowning person, and wonder about his nature later. Even if he is not altogether a moral person, we nevertheless, it is thought, have a moral obligation to save his life. How could Gauthier account for this presumed obligation? I assume that if the CM in question were to judge that it was likely an SM that was drowning, to save him would be to be taken advantage of. It is not entirely clear how this story accords with the kind of description we want. However, it would probably run something like the following. In order for CM to be the most rational strategy, it is vital that SM's be, at least whenever they are identified as such, excluded from beneficial social agreements. If they are not excluded, then it would be more rational, it appears, to be an SM. This is because as an SM, one would gain not only the advantage of being party to such beneficial agreements, but also the advantage of breaking such agreements when expedient. I suppose that a save-another-when-drowning agreement would be beneficial to anyone who took part. By saving the drowning SM, thereby allowing that he is party to this agreement, the CM is forsaking the superior rationality of his own strategy. For if SM's are permitted to reap the advantages of such contracts (by being saved, for example), while being prepared not to adhere himself if his ends lie elsewhere; it would clearly be more beneficial

to be an SM. The only way that CM can truly be more rational, is if CM's conscientiously exclude those they judge to be SM's from the benefits gained by cooperation. It appears that the only consistent CM course is to let the SM drown. If Gauthier is right, and the rational course is the moral course, then we are left with the moral rightness of allowing another to drown, when we have the wherewithal to save him. It hardly needs pointing out that a similar argument can be proposed for any (to speak roughly) charitable act toward those we take to be SM's.

These are damning enough considerations on their own. However, the same sort of case might be made for other examples of interaction between CM's and SM's. Is there any good reason why a CM should refrain from lying to an SM--or stealing from an SM? If, as Gauthier suggests, moral injunctions thereagainst arise from social agreements or contracts which are beneficial to all participants; and if SM is a less rational strategy because it denies its followers the same access as CM's to such agreements, then there appears little reason to encourage the SM by permitting him to take part. The disadvantage under which the SM labours (that which makes SM less rational than CM) is that he is excluded more often from such beneficial social institutions. Once again, if the CM is to retain his rational edge over the SM, it is necessary, for the reasons I have provided above, that those who are judged to be SM's not be permitted to be party to

these agreements. But what does denying the SM access to social agreements mean? I can only presume that it would mean that the CM need not feel obliged (except insofar as it is in his interest) to treat the SM in a moral way. If an injunction against stealing is a relatively large scale, cooperative agreement, whereby all participants agree to refrain from stealing from each other, then the CM should be part of this agreement and refrain from stealing from other participants. But these are apparently only other CM's. For if the SM is to be excluded from such beneficial social contracts then there is no reason in Gauthier's morality to refrain from stealing from an SM. Collaterally, there is no particular reason to expect an SM to refrain from stealing from him. Of course, fear of getting caught, the effects of one's conscience, and other such considerations might provide reason not to steal, lie, or kill for either the SM or the CM. When agreement formation and adherence is most rational for both, then they might both agree and adhere. But there seems to be nothing in morality so characterized that prohibits such villainous practices by the CM against the SM. It is an obvious understatement to point out that this consequence of Gauthier's view is grossly inconsistent with our usual moral notions.

Gauthier, in later chapters of Morals By Agreement, recognizes that not all of our moral concepts are rationally justified by the framework provided by constrained maximization. He apparently feels that even if there are

some aspects of the morality which we accept that are apparently not rationally justified, we can at least salvage the rationality of some, perhaps most. But if there are some fairly broad categories of our commonly held moral judgements that cannot be made to fit, it seems an equally reasonable conclusion that it is his theory that is inadequate. This particularly since there appears to be no clear demarcation between those classes of morality that can be rationally justified, and those which cannot. In "Reason and Maximization", Gauthier claims that "[w]e must not expect that an account of morality, based on agreed optimization, will necessarily resemble our existing conception of morality." This, we are told, is because there "is little reason to suppose that our present conception has developed to correspond to rationality, conceived as identified in any way with utility maximization."[5]

If this is Gauthier's view, then we may have some grounds to complain that he is unfairly connecting rationality with morality by developing a framework for rational action, and labelling the resulting strategy 'moral'. I think we have the right to demand of Gauthier that the morality he deems rational also bear some significant relation to our everyday moral notions. However, Gauthier's ground seems to have shifted at least somewhat by the time of Morals By Agreement, where he remarks,

> Actual moral principles are not, in general, those to which we should have agreed in a fully rational

> bargain....But it is reasonable to adhere to these principles insofar as the outcomes they yield approximate to those which would have been achieved by fully rational bargainers. We may defend the principles by reference to this ideal bargain, and the closer the principles fit, the stronger the defence. We do not suppose that our actual moral principles derive historically from agreement, but insofar as the constraints they impose are acceptable to a rational constrained maximizer, we may fit them into the framework of a morality rationalized by agreement.6

It appears however that there is a large volume of our "actual moral principles" which are not "acceptable to a rational constrained maximizer". I here speak of moral obligations toward non-constrainers. I am not sure what to say about Gauthier's claim that morality can be rationally justified, when it is conjoined to the admission that there is a significant and respectable body of moral judgements which do not conform. It is with some hesitation that I suggest the oddness of justifying an inconsistency by pointing out that it is half correct.

Our discussion of the moral responsibilities which the CM has toward unfortunate SM's points once again to the issue of motivation. For if the CM is the moral man--the man who cooperates socially by adhering to societal agreements, then presumably CM's would, in fact, often perform such charitable acts. This being the case, it appears again that Gauthier may have missed something very basic about doing the right thing. I rather suspect that even if the CM were apprised that there was no benefit to be derived from (for example) saving the drowning SM,

he would do it if possible anyway. We might be able to convince him that he was acting irrationally--even foolishly. We might even be able to convince him that he was betraying his CM credo--and to some degree his fellow CM's. However, I suggest that it would be a rare individual who would be deterred from so acting by such arguments. The reason for this should be clear to anyone for whom doing the right thing is is some sense important. It is because in the last analysis, being moral is not about doing for oneself. To look for ways whereby morality is in fact in service of one's ends is to badly misread the notion. When the CM dives in to attempt to save the thief or liar or scoundrel, perhaps at some personal peril, he is aware that it is somehow beside the point to wonder first about how well his ends will be, in so doing, advanced. To brand him irrational seems a bit like name-calling--and a bit absurd, because it merely points to a failure of comprehension on his part about something essential to morality. What would be the substantive difference in motivation between cases of a CM saving an SM, and saving another CM? I suggest it would likely be small. At least in this case, it seems that the rightness of his act of person-saving is not primarily, for him, a function of his expected utility.

Could it be the case that while we think that there is something more to morality than the serving of one's ends, there in fact is nothing. Perhaps I am wrong about

people's motives when they sacrifice to do the right thing. I am tolerably well convinced that the consistent, rational CM must exclude SM's from their agreements wherever possible, in order to ensure the supreme rationality of the CM strategy. Yet it would be foolish to deny that brigands are every day included in truth-telling, save-from-drowning, do-not-steal and other such agreements. If CM's do this for 'moral' reasons, Gauthier would have to say that these people are mistaken. Perhaps he would hold that psychological egoism is true. That is, whatever we think motivates us, in fact our own ends are the only factors. Or, it might be that while we can act in the perceived interest of others, such action is always beyond the realm of our ethical duty. In either case, the CM is mistaken if he believes that there is moral praiseworthiness in aiding or cooperating with SM's. Indeed, if doing so compromises the opportunity for himself and other CM's to benefit (and it seems that it always will), then one might be positively blameworthy.

I take this to be a radically objectionable view. If it is a consequence of Gauthier's view that CM's have no moral obligations whatever to any but other CM's, I think it can be generally agreed that a profound difficulty exists. It hardly needs be said that a morality which fails to include such duties is one sufficiently removed from our usual conception as to render it unacceptable.

Notes - Chapter 3

1 This passage and the one following by David Gauthier, Morals By Agreement, unpublished manuscript, ch. 8, p.27.

2 "Reason and Maximization", op. cit., p.412.

3 This might not be the case if it were possible that when a CM renounces his disposition, it is still guaranteed that he will keep to all agreements made while he was a CM. I mention this only parenthetically because I take it to be a humanly implausible assumption.

4 Morals By Agreement, ch. 8, p.8.

5 This passage and the one above by Gauthier, "Reason and Maximization", p.433.

6 Morals By Agreement, ch. 8, pp.19-20.

The notion of constrained maximization, it appears, fails as a conception of rationality which can justify the practise of morality. Gauthier's most sophisticated defence made CM a complex and difficult beast, but it seems that at bottom, its problem is roughly the same as faces other such attempts. In short, the CM must either fail to be an individual utility maximizer--and hence not be rational, or must act in a way which is indistinguishable from that of an everyday egoist. The complexity of this most recent defence seems to have bred a complexity of objections.

I have tried to show that his use of average values in the mathematical formulae rendered the results unusable as a universal justification. In the first place, since different people will be better and worse at identifying others, and better or worse at masking their strategy; the values for p and q will vary from person to person. This means that, contrary to our usual notions about morality, this framework might counsel different moralities (strategies) for different people. In addition,

the use of average values for the utility outcomes, fails to reflect differences of circumstance between instances of choice between agreement violation and adherence. The utility payoffs and the probabilities of being identified and identifying others can vary independently not only between different agents, but also among different cases involving the same agent. Even if the averages for these factors counsel constraint (no agreement violation) it does not follow that in each individual case, with each individual agent, constraint is counselled. Indeed it seems virtually certain that there will be some circumstances in which, if these formulae are applied to the given situation, violation will be shown to be the rational course. It is of course the rationality of <u>not</u> violating which this notion is meant to demonstrate.

This consideration by itself constitutes, I expect, a crippling objection to Gauthier's constrained maximization. However, even if there were not this difficulty, or if this problem could be somehow rectified, there are troubles enough with this view. It is worth pointing out again how unsupportably arbitrary Gauthier's figures are. Let it be the case (what I think is clearly not the case) that Gauthier's formulae, using his figures, would demonstrate the rationality of CM. We are still provided no good reason to think that the figures which he uses bear any relation at all to human endeavours. He offers as a 'simplifying' assumption that the utility out-

comes are as he suggests. He does not, however, suggest any reason why we should accept them. Neither does he explain in what sense the procedure is 'simplified' by providing entirely unjustified arbitrary values. But let us accept what strikes me as eminently unacceptable, and assume that his utility outcomes are reasonably reflective of our social interactions. We are still left with nothing that even resembles evidence that sufficient values for p, q, and r are achieved or achievable in our or any other social community.

Quite the contrary, I suggest thet there is much reason to think that that these values are not achievable, and certainly not achieved. I have argued that one's strategy cannot be discerned by others (for the purposes of p and q) with reference only to past behaviour. I am not altogether sure what sort of considerations remain on which to judge of another's strategy. However, it seems that physical clues such as bearing, facial expression, and tone of voice provide at best meagre and misleading corroborative evidence of one's disposition. This means that the values for p and q are likely low. As a consequence, the value for r (the proportion of CM's in the population) will need to be high to justify CM even in Gauthier's terms. I do not suppose that there is any way of knowing with accuracy how many CM's there are in the population. I have suggested, however, that the best guess would be that there are <u>no</u> CM's.

This is because, assuming there to be no strict incompatibility between having the CM disposition and being able to consider violation, it must be that the CM is sometimes irrational. For, if there is this incompatibility, then in cases where the benefits of violation outweigh its detriments, it is rational to violate. To say that the CM is unwilling to violate is not enough to rescue his rationality. If he wills to do that which is irrational, then he is irrational. Since the CM strategy forbids one to violate, it must be, in such cases, that CM is irrational. We can only assume then that there is some incompatibility between having the CM disposition and being able to violate.

It appears then that if this is the case, then the CM is an individual who is unable freely to choose about moral matters. The CM's (likely psychological) block might just be an inability to calculate utilities (consider benefits) of violation, due to the cooperative spirit or attitude which the CM necessarily has toward other cooperators. In order to protect the rationality of CM in exceptional cases, it is necessary to see him as being compelled in some respect by his disposition always to adhere. While there may be those who are, at least usually unwilling to calculate or violate, it would be remarkable indeed to find someone who is unable to do these. It would be more remarkable still, since that person, despite his inability to calculate benefits in the context of

an agreement, is perfectly able (indeed is required) to calculate the benefits of an agreement prior to taking part.

We should also point out that if there are not enough CM's in the population at a given time to justify CM, there likely never will be. This is because it will never be rational for anyone to adopt CM until enough others do so. If there are not any, or not enough CM's now, it will never be rational (except if a great many people are irrational first) to adopt CM.

So it appears that we have some reason to believe that even if Gauthier's method of calculation were of value (it almost certainly is not), the values requisite to counsel CM do not anyway hold. But if they did hold, and even if all of the difficulties to which I have pointed could be in some way resolved, I am by no means convinced that CM is morally defensible as a rational strategy. We have addressed the question of the treatment of SM's by CM's. It appears that the CM has no moral responsibilities to the SM whatever. What gives the CM his rational edge over the SM is presumably the fact that SM's are less often participant in beneficial social or individual agreements. But,

> even if agreement is reached, a CM is committed to carrying it out only in the context of mutual expectations on the part of all parties to the agreement that it will be carried out.[1]

If a CM is dealing with an SM, he has no expectation that

the SM will necessarily adhere to the agreement. The agent then is not, it seems, committed to carry out the agreement, at least when dealing with the SM. Indeed, it seems that since it is his ability to take part in more beneficial agreements that is meant to give the CM his advantage over the SM, it would be positively irrational to accept moral obligations toward him, thereby tacitly including the SM in a beneficial agreement. The CM would simply strengthen the SM's case by permitting him to take part in such agreements, knowing that he might take advantage when expedient. To be rational, the CM must at all costs exclude the SM from moral agreements. This means that he need not trouble himself with any tedious moral responsibility toward the SM. Indeed it may be that he has a positive obligation to refrain from acting in a moral way with the SM unless he is directly benefitted thereby.

Needless to say, this is a view which clashes violently with our usual moral notions. That there is no reason in Gauthier's morality (although there may be prudential ones) for one to refrain from killing, robbing, or lying to an SM, or to help a drowning SM; is a result which I take to be unavoidable for the rational CM. It is also one which is undeniably and unacceptably repugnant. A view which countenances such behaviour is not, in any meaningful sense, a moral one. This final objection points, I think, to what is really wrong about the use

of CM to rationally justify morality. It is the same problem that we discussed much earlier. The attempt to justify morality in terms of benefits to the agent will, of necessity perhaps, be a frustrating one. This since it involves finding a prudential justification to choose to be moral instead of prudent. It appears however trivially true that for everyone, prudence and morality sometimes conflict.

Even at times when there is no conflict (even if there is never any conflict) there is a difference between doing something prudently, and doing it morally. It is hardly an original notion, but it bears pointing out at this stage, that it is a part of acting morally that we do the right thing regardless of its effects on ourself--beneficial or harmful. To be motivated in another way is natural enough, but it is not consistent with acting morally. That the CM's motivation is ultimately prudential, I feel, renders it unserviceable as a moral strategy.

To this point we have discussed some attempts to demonstrate a strong, perhaps necessary connection between acting rationally and acting morally. I have tried to show that there has developed a rough progression which has its roots in Hobbes' view that everyone benefits from social cooperation. It is out of such community cooperation and the dynamics of social interaction that the prudence of morality is thought to come. Each of

these attempts, however, has run against the same barrier. While it is easy enough to show that we usually do better by acting morally, all arguments have failed to show why we must not exploit what appear to be inevitable exceptions. In the context of such exceptional circumstances, it seems we face either a failure of rationality (if we do the morally right thing), or a failure in morality (if we take the rational course). But there is nothing new here. This is the same conclusion which Plato was at such pains to dispute, and at which Hume arrived in the 18th century.

It has been my conclusion that Hume was basically right, and that the major philosophic attempts to ground moral action on reason have been uniformly unsuccessful. In this final section, I intend to offer some very brief and largely speculative remarks, premised on the assumption that the reason that attempts to show a rational base for morality have failed is the obvious one. That is, that there <u>is</u> no base in reason for being moral. This means there is no necessary connection between being rational and being moral.

In chapter one, we briefly discussed the view that the question 'Why should I be moral?' is a meaningless one. We argued that there is a straightforward sense in which this question is not meaningless. However, it might be interpreted in a quite different way. It could be requesting some reason for one to take the moral point of view, or to have moral ends in the first place. Viewed in

this way, the question does not seem to admit of any answer. For, if we interpret 'should' here as a moral notion, then the question appears to misunderstand its own terms. It requests a moral justification for being moral and is, as a result circular in much the same way as is 'Why are bachelors unmarried?'. If, however, 'should' is interpreted prudentially, it is clear that the question cannot be answered, since there can be no reason to forsake one's ends when they conflict with morality. In either case, we are confounded in our attempt to answer this question. I will not enter the debate as to whether a question is, in virtue of being unanswerable, meaningless. Suffice it to say that when viewed in this light, the question 'Why should I be moral?' fails to admit of an answer.

As we saw in the first chapter, however, this question can be viewed in a different way. If one who asks 'Why should I be moral?' is looking for reasons to adopt moral ends--for reasons to accept that doing the right thing is important, then the question is indeed unanswerable. But if, in an individual situation, one asks this question; he might be wondering why he should be moral in <u>this</u> case--why he should <u>now</u> do the morally right thing. To this question, a perfectly good reason can be provided, at least to some. It is undeniable that many want to do the right thing, or at least want usually to do the right thing. We commonly enough have moral ends, as I have earlier described

them. For many of us, the doing of the right thing can be numbered among our ends. In some cases, we might answer the question 'Why should I be moral?' by pointing out that being moral is what the asker wants to do. Perhaps we wish to remind him of a resolution to be moral which he had, at some earlier time, formed.

Compare this question with 'Why should I take turkish delight?'. While we cannot, in one sense, give reasons why a person should like or crave turkish delight; in another sense, he might be asking about a specific bit of candy. An answer might, in this case be supplied, by reminding the asker that he enjoyed it the last time he ate it (perhaps he has forgotten), or by pointing to elements to the flavour of turkish delight which are in accord with the asker's tastes. I confess that there is a certain tension here due to the fact that we generally know better than another what our tastes or ends are, and whether a particular choice will best serve them. In addition, our asking of such questions often assumes that, in terms of our ends, the object of the question (being moral or turkish delight) is unacceptable. Nevertheless, there are straightforward cases in which both 'Why should I be moral?' and 'Why should I take turkish delight?' admit of answers which take the form of reason-giving. It is sometimes the case that others can predict the consequences of our actions as well as can we. Also, others are often more aware, or at least as aware of our wants

than are we.

It appears now that our conclusion concerning the relationship between morality and reason can be made more complete. For while there is no necessary connection between the two, neither are they inconsistent notions. If it is a person's primary goal to do the right thing--if he, as it were, holds the moral point of view above all other ends; then doing the right thing will always be rational. This since his ends or preferences will be best satisfied by always acting morally. Hume suggests that

> in all ingenuous natures, the antipathy to treachery and roguery is far too strong to be counterbalanced by any views of profit or pecuniary advantage. Inward peace of mind, consciousness of integrity, a satisfactory review of our own conduct; these are circumstances, very requisite to happiness, and will be cherished and cultivated by every honest man, who feels the importance of them.[2]

Since rationality has not the tools to judge of ends, the goal of being a morally good man is no less reasonable than any other. That being one's goal, the rational course is the one which most satisfactorily allows one to be such. This course can be none other than striving, in all things, to do what one takes to be morally right. Perhaps more realistically, to the degree, and at those times when one wishes to be morally good; it is rational to act so.

I earlier complained that Gauthier's view was, at a

basic level misfounded, due to the fact that it assumes an ultimately prudential basis for our motivation to moral behaviour. While I think that this objection is still good, we might say something more about the nature of this motivation. For it is to be admitted that the CM's motivation is not blatantly self-seeking. Once the CM disposition is chosen, and an agreement struck, the CM ceases to calculate. At this point it appears that he adheres in the morally appropriate, disinterested way. It could be too that to demand the 'right' motivation at any deeper or more ultimate level is naive and perhaps a bit syrupy. A view such as Gauthier's moral contractarian one might also claim an explanatory role. Not only does this moral framework, it could be suggested, provide a way of deciding what <u>is</u> moral, but it also explains how the practice of morality could have arisen in the first place. In very broad and rough terms, moral institutions arose in a Hobbesian sort of way. That is, the pernicious character of life in the absence of such conventions gave way to an obviously happier state of affairs which included them. The reason that people ever acquired a desire to act morally stems from the fact that everyone is better off (hence <u>I</u> am better off) for the existence of, and general adherence to, moral conventions. Whatever <u>our</u> reason is for being moral, the genesis of the individual practice of morality is best seen in something like this light.

As a consequence, we might continue, the moral motiv-

ation of which CM admits is as deeply disinterested as we can reasonably expect moral motivation to be. While the CM is meant to be the moral individual, his morality is at best instrumental, and in the final analysis, in service of himself. But what more is to be expected? We may have the desire to do the morally right thing, or to see the right thing done; but this could be the result of social conditioning. There is nothing wrong in this, but the reason that we have to some degree lost track of the ultimately egoistic motivation for morality could be found in our training, and not in the nature of morality.

I have argued that the doctrine of CM might be misconceived at its foundation. That is, any view of morality which is ultimately (if not instrumentally) egoistically motivated, has in some serious way missed the point. The considerations which I briefly point to above attempt to show that an instrumentally indifferent motivation is, morally speaking, the best one to be found. At an ultimate level, we can find no better than an egoistic motivation, since it was ultimately egoistic reasons that led originally to the development of moral notions and practices. The most satisfactory way of applying morality's contractarian roots to a present day moral framework, is by seeing the moral man as one who constrains his maximizing behaviour, in order to maximize more efficiently.

But even if the practice of morality did arise in something like this way, it does not appear to follow that

we are obliged to construct our morality on some compatible framework. That moral institutions developed on a roughly contractarian model, based on mutual benefit, need say nothing about the sort of moral framework that we should employ. If reason is, as I have suggested, silent about the validity of ends, then a moral end is no less rational than a non-moral (prudential) one. This being the case, we have no reason to prefer prudence, tout court, to morality. Indeed, that our intuitions are quite the opposite seems to say something about the way that our ends are, in general, ordered.

Do we have any ultimate moral ends? That is, is it ever the case that we desire to do the right thing not as an instrument to some further non-moral good, but only for the sake of doing the right thing? This is, of course, the challenge of the psychological egoist. While this is likely a question best answered by a psychologist, it appears to me not entirely implausible that we do have moral ends which are, in this sense, ultimate. Assuming that we do, we must also assume that these have been acquired in some non-rational way. Since there can be no reason to alter one's (at least ultimate) ends, (reason being the sort of thing which cannot address the worth of ends), we must acquire and alter ends in some other way. The most likely explanation will, I expect, be strongly related to our emotions and sentiments; and likely includes, in a profound way, the considerable forces of family and

social conditioning. However moral ends may have arisen, the point is that if we have them, then we are not obliged to develop a system of moral beliefs which is premised on the assumption that ultimately, the only rational goals are prudential ones.

Gauthier has developed his notion of CM with a view to establishing the rationality of our moral behaviour. But for one to whom doing the right thing is, in an ultimate sort of way, important; the project is unnecessary, and perhaps misconceived. This is because the rationality of moral behaviour is established, at least to the degree that people wish to be moral. For those who wish to be moral, or for a person who wishes, in a given case, to be moral, the belief that a course is moral provides adequate reason to take it. Such a person, or a person in such a circumstance, has no need of the prudential persuasion which constrained maximization purports to provide. It is already rational to be moral.

There is however, no necessary connection between reason and morality. If an individual does not desire to do the right thing, or if this desire is sometimes or often outweighed by opposing prudential considerations; we cannot charge him with irrationality when he acts immorally (although we may wish to brand him with any number of other uncomplimentary epithets). Reason dictates that he should act immorally. While this may be too bad, I am not convinced that we need be overly concerned. At any rate,

I suspect that this points to something that we have known all along. That is, if one does not wish to be moral--if one has no overriding moral end in a given case; reason, even that in the form of constrained maximization, is anyway not competent to bring one around to virtuous behaviour. He will, in any case, do as he pleases.

Notes - Chapter 4

1 "Reason and Maximization", op. cit., p.429.

2 Hume, Enquiry, op. cit., sec. IX, pt. II, p.283.

Bibliography

Baier, Kurt, "Good Reasons", Philosophical Studies 4 (1953).

_____, The Moral Point of View (Ithaca: Cornell Univ. Press, 1958), esp. chs. 1, 8, 11, 12.

_____, "The Conceptual Link Between Morality and Rationality", unpublished manuscript (1981).

Beehler, Roger, "Reasons for Being Moral", Analysis 33, no. 1 (1972).

_____, "Morals and Reasons", Analysis 33, no. 1 (1972).

_____, Moral Life (Oxford: Blackwell, 1978), esp. ch. 4.

Blake, R.M., "The Ground of Moral Obligation", Ethics 38 (1968).

Bradley, F.H., "Why Should I Be Moral?", from Ethical Studies: Selected Essays (New York: Humanities Press, 1951).

Broad, C.D., "Egoism as a Theory of Human Motives" (1950), from Cheney (ed.), Broad's Critical Essays in Moral Philosophy (New York: Liberal Arts Press, 1971).

Falk, W.D., "'Ought' and Motivation", Proceedings of the Aristotelian Society 48 (1947-48).

_____, "Morality, Self, and Others", from Casteneda and Nakhnikian (eds.), Morality and the Language of Conduct (Detroit: Wayne State Univ. Press, 1963).

Flew, Antony, "Must Morality Pay?", from Carter (ed.), Skepticism and Moral Principles (Evanston, Ill.: New University Press, 1973).

Foot, Phillipa, "Moral Beliefs", Proceedings of the Aristotelian Society 59 (1958-59).

_____, "Reasons for Action and Desires", Proceedings of the Aristotelian Society (Sup.) (1972).

_____, "Morality as a System of Hypothetical Imperatives", Philosophical Review 81 (1972).

Frankena, W.K., "Obligation and Motivation in Recent Moral Philosophy", from Goodpaster (ed.), Perspectives on Morality (Univ. of Notre Dame Press, 1976).

Gauthier, David, "Morality and Advantage", Philosophical Review 76 (Oct., 1967).

_____, "Bargaining Our Way to Morality", Philosophic Exchange 2 (1972).

_____, "The Impossibility of Rational Egoism", Journal of Philosophy 71 (15 Aug. 1974).

_____, "Reason and Maximization", Canadian Journal of Philosophy 4 (March, 1975).

_____, "Economic Rationality and Moral Constraints", Midwest Studies in Philosophy 3 (1978).

_____, Morals By Agreement, unpublished manuscript (1981).

Grice, G.R., The Grounds of Moral Judgement (Cambridge Univ. Press, 1967), esp. Introduction, ch. 1.

Harman, Gilbert, The Nature of Morality (New York: Oxford Univ. Press, 1977) esp. pt. IV.

Hobbes, Thomas, Leviathan, for example, Oakeshott (ed.), (London: Collier MacMillan, 1978) esp. chs. 13, 14, 15, 17.

Hospers, John, Human Conduct (New York: Harcourt, Brace, Jovanovich, 1972) pp. 174-195.

Hume, David, An Enquiry Concerning the Principles of Morals, from, for example, Nidditch and Selby-Bigge (eds.) (Oxford: Clarendon Press, 1975).

_____, "Of the Original Contract" (1748), from Glickman (ed.), Moral Philosophy: An Introduction (New York: St. Martin's Press, 1976).

Munro, H.D., Review of Baier's The Moral Point of View, Australasian Journal of Philosophy 37 (May, 1959).

Nagel, Thomas, The Possibility of Altruism (Oxford: Clarendon Press, 1970).

Nielsen, Kai, "Is 'Why Should I Be Moral?' an Absurdity?", Australasian Journal of Philosophy 36, no. 1 (1958).

_____, "Why Should I Be Moral?", Methodos 15 (1963).

_____, "On Giving Reasons for Being Moral", Analysis 33, no. 1 (1972).

Parfit, Derek, "Prudence, Morality, and the Prisoner's Dilemma", Proceedings of the British Academy 65 (1979).

_____, Against Prudence, unpublished manuscript.

Phillips, D.Z., "Does it Pay to be Good?", Proceedings of the Aristotelian Society 65 (1964-65).

Plato, The Republic, for example, Adam (ed.), (Cambridge University Press, 1969) see esp. bks. I, IX.

Prichard, H.A., "Does Moral Philosophy Rest on a Mistake?", Mind 21 (1912).

_____, Duty and Interest (Oxford: Clarendon Press, 1928), esp. pp. 3-29.

Thornton, J.C., "Can the Moral Point of View Be Justified?", Australasian Journal of Philosophy 42 (1964).

Wadia, P.S., "Why Should I Be Moral?", Australasian Journal of Philosophy 42 (1964).