MORALS BY AGREEMENT
DAVID GAUTHIER
Chapter I OVERVIEW OF A THEORY

What theory of morals can ever serve any useful purpose, unless it can show that all the duties it recommends are also the true interest of each individual?1 David Hume, who asked this question, seems mistaken; such a theory would be too useful. Were duty no more than interest, morals would be superfluous. Why appeal to right or wrong, to good or evil, to obligation or to duty, if instead we may appeal to desire or aversion, to benefit or cost, to interest or to advantage? An appeal to morals takes its point from the failure of these latter considerations as sufficient guides to what we ought to. The unphilosophical poet Ogden Nash grasped the assumptions underlying our moral language more clearly than the philosopher Home when he wrote:

'0 Duty!
Why hast thou not the visage of a sweetie or a cutie?,2

We may lament duty's stern visage but we may not deny it. For it is only as we believe that some appeals do, alas, override interest or advantage that morality becomes our concern.

But if the language of morals is not that of interest, it is surely that of reason. What theory of morals, we might better ask, can ever serve any useful purpose, unless it can show that all the duties it recommends are also truly endorsed in each individual's reason? If moral appeals are entitled to some practical effect, some influence on behaviour, it is not because they whisper invitingly to our desires, but because they convince our intellect. Suppose we should find, as Hume himself believes, that reason is impotent in the sphere of action apart from its role in deciding matters of fact.3 Or suppose we should find that reason is no more than the handmaiden of interest, so that in overriding advantage a moral appeal must also contradict reason. In either case we should conclude that the moral enterprise, as traditionally conceived, is impossible.

To say that our moral language assumes a connection with reason is not to argue for the rationality of our moral views, or of any alternative to them. Moral language may rest on a false assumption.4 If moral duties are rationally grounded, then the emotivists, who suppose that moral appeals are no more than persuasive, and the egoists, who suppose that rational appeals are limited by self-interest, are mistaken. 5 But are moral duties rationally grounded? This we shall seek to prove, showing that reason has a practical role related to but transcending individual interest, so that principles of action that prescribe duties overriding advantage may be rationally justified. We shall defend the traditional conception of morality as a rational constraint on the pursuit of individual interest.

Yet Hume's mistake in insisting that moral duties must be the true interest of each individual conceals a fundamental insight. Practical reason is linked to interest, or, as we shall come to say, to individual utility, and rational constraints on the pursuit of interest have themselves a foundation in the interest they constrain. Duty overrides advantage, but the acceptance of duty is truly advantageous. We shall find this seeming paradox embedded in the very structure of interaction. As we come to understand this structure, we shall recognize the need for restraining each person's pursuit of her own utility, and we shall examine its implications for both our principles of action and our conception of practical rationality. Our enquiry will lead us to the rational basis for a morality, not of absolute standards, but of agreed constraints.

2.1 We shall develop a theory of morals. Our concern is to provide a justificatory framework for moral behaviour and principles, not an explanatory framework. Thus we shall develop a normative theory. A complete philosophy of morals would need to explain, and perhaps to defend, the idea of a normative theory. We shall not do this. But we shall exemplify normative theory by sketching the theory of rational choice. Indeed, we shall do more. We shall develop a theory of morals as part of the theory of rational choice. We shall argue that the rational principles for making choices, or decisions among possible actions, include some that constrain the actor pursuing his own interest in an impartial way. These we identify as moral principles.

The study of choice begins from the stipulation of clear conceptions of value and rationality in a form applicable to choice situations.6 The theory then analyses the structure of these situations so that, for each type of structure distinguished, the conception of rationality may be elaborated into a set of determinate conditions on the choice among possible actions. These conditions are then expressed as precise principles of rational behaviour, serving both for prescription and for critical assessment. Derivatively, the principles also have an explanatory role in so far as persons actually act rationally.

The simplest, most familiar, and historically primary part of this study constitutes the core of classical and neo-classical economic theory, which examines rational behaviour in those situations in which the actor knows with certainty the outcome of each of his possible actions. The economist does of course offer to explain behaviour, and much of the interest of her theory depends on its having explanatory applications, but her explanations use a model of ideal interaction which includes the rationality of the actors among its assumptions. Thus economic explanation is set within a normative context. And the role of economics in formulating and evaluating policy alternatives should leave us in no doubt about the deeply prescriptive and critical character of the science.

The economist formulates a simple, maximizing conception of practical rationality, which we shall examine in Chapter II. But the assumption that the outcome of each possible choice can be known with certainty seriously limits the scope of economic analysis and the applicability of its account of reason. Bayesian decision theory relaxes this assumption, examining situations with choices involving risk or uncertainty. The decision theorist is led to extend the economist's account of reason, while preserving its fundamental identification of rationality with maximization.

Both economics and decision theory are limited in their analysis of interaction, since both consider outcomes only in relation to the choices of a single actor, treating the choices of others as aspects of that actor's circumstances. The theory of games overcomes this limitation, analysing outcomes in relation to sets of choices, one for each of the persons involved in bringing about the outcome. It considers the choices of an actor who decides on the basis of expectations about the choices of others, themselves deciding on the basis of expectations about his choice. Since situations involving a single actor may be treated as limiting cases of interaction, game theory aims at an account of rational behaviour in its full generality. Unsurprisingly, achievements are related inversely to aims; as a study of rational behaviour under certainty economic theory is essentially complete, whereas game theory is still being developed. The theory of rational choice is an ongoing enterprise, extending a basic understanding of value and rationality to the formulation of principles of rational behaviour in an ever wider range of situations.

2.2 Rational choice provides an exemplar of normative theory. One might suppose that moral theory and choice theory are related only in possessing similar structures. But as we have said, we shall develop moral theory as part of choice theory. Those acquainted with recent work in moral philosophy may find this a familiar enterprise; John Rawls has insisted that the theory of justice is 'perhaps the most significant part, of the theory of rational choice', and John Harsanyi explicitly treats ethics as part of the theory of rational behaviour.7 But these claims are stronger than their results warrant. Neither Rawls nor Harsanyi develops the deep connection between morals and rational choice that we shall defend. A brief comparison will bring our enterprise into sharper focus.

Our claim is that in certain situations involving interaction with others, an individual chooses rationally only in so far as he constrains his pursuit of his own interest or advantage to conform to principles expressing the impartiality characteristic of morality. To choose rationally, one must choose morally. This is a strong claim. Morality, we shall argue, can be generated as a rational constraint from the non-moral premisses of rational choice. Neither Rawls nor Harsanyi makes such a claim. Neither Rawls nor Harsanyi treats moral principles as a subset of rational principles for choice.

Rawls argues that the principles of justice are the objects of a rational choice--the choice that any person would make, were he called upon to select the basic principles of his society from behind a 'veil of ignorance' concealing any knowledge of his own identity.8 The principles so chosen are not directly related to the making of individual choices.9 Derivatively, acceptance of them must have implications for individual behaviour, but Rawls never claims that these include rational constraints on individual choices. They may be, in Rawls's terminology, reasonable constraints, but what is reasonable is itself a morally substantive matter beyond the bounds of rational choice.[10]

Rawls's idea, that principles of justice are the objects of a rational choice, is indeed one that we shall incorporate into our own theory, although we shall represent the choice as a bargain, or agreement, among persons who need not be unaware of their identities. But this parallel between our theory and Rawls's must not obscure the basic difference; we claim to generate morality as a set of rational principles for choice. We are committed to showing why an individual, reasoning from non-moral premisses, would accept the constraints of morality on his choices.

Harsanyi's theory may seem to differ from Rawls's only in its account of the principles that a person would choose from behind a veil of ignorance; Rawls supposes that persons would choose the well-known two principles of justice, whereas Harsanyi supposes that persons would choose principles of average rule-utilitarianism.[11] But Harsanyi's argument is in some respects closer to our own; he is concerned with principles for moral choice, and with the rational way of arriving at such principles. However, Harsanyi's principles are strictly hypothetical; they govern rational choice from an impartial standpoint or given impartial preferences, and so they are principles only for someone who wants to choose morally or impartially.[12] But Harsanyi does not claim, as we do, that there are situations in which an individual must choose morally in order to choose rationally. For Harsanyi there is a rational way of choosing morally but no rational requirement to choose morally. And so again there is a basic difference between our theory and his.

Putting now to one side the views of Rawls and Harsanyi--views to which we shall often return in later chapters-we may summarize the import of the differences we have sketched. Our theory must generate, strictly as rational principles for choice, and so without introducing prior moral assumptions, constraints on the pursuit of individual interest or advantage that, being impartial, satisfy the traditional understanding of morality. We do not assume that there must be such impartial and rational constraints. We do not even assume that there must be rational constraints, whether impartial or not. We claim to demonstrate that there are rational constraints, and that these constraints are impartial. We then identify morality with these demonstrated constraints, but whether their content corresponds to that of conventional moral principles is a further question, which we shall not examine in detail. No doubt there will be differences, perhaps significant, between the impartial and rational constraints supported by our argument, and the morality learned from parents and peers, priests and teachers. But our concern is to validate the conception of morality as a set of rational, impartial constraints on the pursuit of individual interest, not to defend any particular moral code. And our concern, once again, is to do this without incorporating into the premisses of our argument any of the moral conceptions that emerge in our conclusions.

2.3 To seek to establish the rationality of moral constraints is not in itself a novel enterprise, and its antecedents are more venerable than the endeavour to develop moral theory as part of the theory of rational choice. But those who have engaged in it have typically appealed to a conception of practical rationality, deriving from Kant, quite different from ours.[13] In effect, their understanding of reason already includes the moral dimension of impartiality that we seek to generate.

Let us suppose it agreed that there is a connection between reason and interest--or advantage, benefit, preference, satisfaction, or individual utility, since the differences among these, important in other contexts, do not affect the present discussion. Let it further be agreed that in so far as the interests of others are not affected, a person acts rationally if and only if she seeks her greatest interest or benefit. This might be denied by some, but we wish here to isolate the essential difference between the opposed conceptions of practical rationality. And this appears when we consider rational action in which the interests of others are involved. Proponents of the maximizing conception of rationality, which we endorse, insist that essentially nothing is changed; the rational person still seeks the greatest satisfaction of her own interests. On the other hand, proponents of what we shall call the universalistic conception of rationality insist that what makes it rational to satisfy an interest does not depend on whose interest it is. Thus the rational person seeks to satisfy all interests. Whether she is a utilitarian, aiming at the greatest happiness of the greatest number, or whether she takes into independent consideration the fair distribution of benefit among persons, is of no importance to the present discussion.

To avoid possible misunderstanding, note that neither conception of rationality requires that practical reasons be self-interested. On the maximizing conception it is not interests in the self, that take oneself as object, but interests of the self, held by oneself as subject, that provide the basis for rational choice and action. On the universalistic conception it is not interests in anyone, that take any person as object, but interests of anyone, held by some person as subject, that provide the basis for rational choice and action. If I have a direct interest in your welfare, then on either conception I have reason to promote your welfare. But your interest in your welfare affords me such reason only given the universalistic conception.

Morality, we have insisted, is traditionally understood to involve an impartial constraint on the pursuit of individual interest. The justification of such a constraint poses no problem to the proponents of universalistic rationality. The rational requirement that all interests be satisfied to the fullest extent possible directly constrains each person in the pursuit of her own interests. The precise formulation of the constraint will of course depend on the way in which interests are to be satisfied, but the basic rationale is sufficiently clear.

The main task of our moral theory--the generation of moral constraints as rational--is thus easily accomplished by proponents of the universalistic conception of practical reason. For them the relation between reason and morals is clear. Their task is to defend their conception of rationality, since the maximizing and universalistic conceptions do not rest on equal footings. The maximizing conception possesses the virtue, among conceptions, of weakness. Any consideration affording one a reason for acting on the maximizing conception, also affords one such a reason on the universalistic conception. But the converse does not hold. On the universalistic conception all persons have in effect the same basis for rational choice-the interests of all-and this assumption, of the impersonality or impartiality of reason, demands defence.

Furthermore, and perhaps of greater importance, the maximizing conception of rationality is almost universally accepted and employed in the social sciences. 14 As we have noted, it lies at the core of economic theory, and is generalized in decision and game theory. Its lesser prominence in political, sociological, and psychological theory reflects more the lesser concern with rationality among many practitioners of those disciplines, than adherence to an alternative conception. Social scientists may no doubt be mistaken, but we take the onus of proof to fall on those who would defend universalistic rationality.

In developing moral theory within rational choice we thus embrace the weaker and more widely accepted of the two conceptions of rationality that we have distinguished. Of course, we must not suppose that the moral principles we generate will be identical with those that would be derived on the universalistic conception. Its proponents may insist that their account of the connection between reason and morals is correct, even if they come to agree that a form of morality may be grounded in maximizing rationality. But we may suggest, without here defending our suggestion, that few persons would embrace the universalistic conception of practical reason did they not think it necessary to the defence of any form of rational morality. Hence the most effective rebuttal of their position may be, not to seek to undermine their elaborate and ingenious arguments, but to construct an alternative account of a rational morality grounded in the weaker assumptions of the theory of rational choice.

3.1 Morals by agreement begin from an initial presumption against morality, as a constraint on each person's pursuit of his own interest. A person is conceived as an independent centre of activity, endeavouring to direct his capacities and resources to the fulfilment of his interests. He considers what he can do, but initially draws no distinction between what he may and may not do. How then does he come to acknowledge the distinction? How does a person come to recognize a moral dimension to choice, if morality is not initially present?

Morals by agreement offer a contractarian rationale for distinguishing what one may and may not do. Moral principles are introduced as the objects of fully voluntary ex ante agreement among rational persons. Such agreement is hypothetical, in supposing a pre-moral context for the adoption of moral rules and practices. But the parties to agreement are real, determinate individuals, distinguished by their capacities, situations, and concerns. In so far as they would agree to constraints on their choices, restraining their pursuit of their own interests, they acknowledge a distinction between what they may and may not do. As rational persons understanding the structure of their interaction, they recognize a place for mutual constraint, and so for a moral dimension in their affairs. That there is a contractarian rationale for morality must of course be shown. That is the task of our theory. Here our immediate concern is to relate the idea of such a rationale to the introduction of fundamental moral distinctions. This is not a magical process. Morality does not emerge as the rabbit from the empty hat. Rather, as we shall argue, it emerges quite simply from the application of the maximizing conception of rationality to certain structures of interaction: Agreed mutual constraint is the rational response to these structures. Reason overrides the presumption against morality.

The genuinely problematic element in a contractarian theory is not the introduction of the idea of morality, but the step from hypothetical agreement to actual moral constraint. Suppose that each person recognizes himself as one of the parties to agreement. The principles forming the object of agreement are those that he would have accepted ex ante in bargaining with his fellows, had he found himself among them in a context initially devoid of moral constraint. Why need he accept, ex post in his actual situation, these principles as constraining his choices? A theory of morals by agreement must answer this question.

Historically, moral contractarianism seems to have originated among the Greek Sophists. Glaucon sketched a contractarian account of the origin of justice in Plato's Republic but significantly, he offered this view for Socrates to refute, not to defend.15 Our theory of morals falls in an unpopular tradition, as the identity of its greatest advocate, Thomas Hobbes, will confirm. Hobbes transformed the laws of nature, which lay at the core of Stoic and medieval Christian moral thought, into precepts of reason that require each person, acting in his own interest, to give up some portion of the liberty with which he seeks his own survival and well-being, provided others do likewise.16 But this agreement gives rise to actual constraint only through the efficacy of the political sovereign; from the standpoint of moral theory, the crucial step requires the intervention of a deus ex machina. Nevertheless, in Hobbes we find the true ancestor of the theory of morality that we shall present. Only recently has his position begun to acquire a significant following. G. R. Grice has developed an explicitly contractarian theory, and Kurt Baier has acknowledged the Hobbesian roots of his central thesis, that 'The very raison d'etre of a morality is to yield reasons which overrule the reasons of self-interest in those cases when everyone's following self-interest would be harmful to everyone.'17

To the conceptual underpinning that may be found in Hobbes, Grice, and Baier, we seek to add the rigour of rational choice. Of course the resulting moral theory need not be one that they would endorse. But the appeal to rational choice enables us to state, with new clarity and precision, why rational persons would agree ex ante to constraining principles, what general characteristics these principles must have as objects of rational agreement, and why rational persons would comply ex post with the agreed constraints.

3.2 A useful vantage point for appreciating the rationale of constraint results from juxtaposing two ideas formulated by John Rawls. A contractarian views society as 'a cooperative venture for mutual advantage' among persons 'conceived as not taking an interest in one another's interests' .18 The contractarian does not claim that all actual societies are co-operative ventures; he need not claim that all afford the expectation of mutual advantage. Rather, he supposes that it is in general possible for a society, analysed as a set of institutions, practices, and relationships, to afford each person greater benefit than she could expect in a non-social 'state of nature', and that only such a society could command the willing allegiance of every rational individual. The contractarian need not claim that actual persons take no interest in their fellows; indeed, we suppose that some degree of sociability is characteristic of human beings. But the contractarian sees sociability as enriching human life; for him, it becomes a source of exploitation if it induces persons to acquiesce in institutions and practices that but for their fellow-feelings would be costly to them. Feminist thought has surely made this, perhaps the core form of human exploitation, clear to us. Thus the contractarian insists that a society could not command the willing allegiance of a rational person if, without appealing to her feelings for others, it afforded her no expectation of net benefit.

If social institutions and practices can benefit all, then some set of social arrangements should be acceptable to all as a co-operative venture. Each person's concern to fulfill her own interests should ensure her willingness to join her fellows in a venture assuring her an expectation of increased fulfilment. She may of course reject some proposed venture as insufficiently advantageous to her when she considers both the distribution of benefits that it affords, and the availability of alternatives. Affording mutual advantage is a necessary condition for the acceptability of a set of social arrangements as a co-operative venture, not a sufficient condition. But we suppose that some set affording mutual advantage will also be mutually acceptable; a contractarian theory must set out conditions for sufficiency.

The rationale for agreement on society as a co-operative venture may seem unproblematic. The step from hypothetical agreement ex ante on a set of social arrangements to ex post adherence to those arrangements may seem straightforward. If one would willingly have joined the venture, why would one not now continue with it? Why is there need for constraint?

The institutions and practices of society play a co-ordinative role. Let us say, without attempting a precise definition, that a practice is co-ordinative if each person prefers to conform to it provided (most) others do, but prefers not to conform to it provided (most) others do not.19 And let us say that a practice is beneficially co-ordinative if each person prefers that others conform to it rather than conform to no practice, and does not (strongly) prefer that others conform to some alternative practice. Hume's example, of two persons rowing a boat that neither can row alone, is a very simple example of a beneficially co-ordinative practice.20 Each prefers to row if the other rows, and not to row if the other does not. And each prefers the other to row than to act in some alternative way.

It is worth noting that a co-ordinative practice need not be beneficial. Among peaceable persons, who regard weapons only as instruments of defence, each may prefer to be armed provided (most) others are, and not armed provided (most) others are not. Being armed is a co-ordinative practice but not a beneficial one; each prefers others not to be armed.

The co-ordinative advantages of society are not to be underestimated. But not all beneficial social practices are co-ordinative. Let us say that a practice is beneficial if each person prefers that (almost) everyone conform to it rather than that (most) persons conform to no practice, and does not (strongly) prefer that (almost) everyone conform to some alternative practice. Yet it may be the case that each person prefers not to conform to the practice if (most) others do. In a community in which tax funds are spent reasonably wisely, each person may prefer that almost everyone pay taxes rather than not, and yet may prefer not to pay taxes herself whatever others do. For the payments each person makes contribute negligibly to the benefits she receives. In such a community persons will pay taxes voluntarily only if each accepts some constraint on her pursuit of individual interest; otherwise, each will pay taxes only if coerced, whether by public opinion or by public authority.

The rationale for agreement on society as a co-operative venture may still seem unproblematic. But the step from hypothetical agreement ex ante on a set of social arrangements to ex post adherence may no longer seem straightforward. We see why one might willingly join the venture, yet not willingly continue with it. Each joins in the hope of benefiting from the adherence of others, but fails to adhere in the hope of benefiting from her own defection. In the next two chapters we shall offer an account of value, rationality, and interaction, that will give us a precise formulation of the issue just identified. Prior to reflection, we might suppose that were each person to choose her best course of action, the outcome would be mutually as advantageous as possible. As we fill in our tax forms we may be reminded, inter alia, that individual benefit and mutual advantage frequently prove at odds. Our theory develops the implications of this reminder, beginning by locating the conflict between individual benefit and mutual advantage within the framework of rational choice.

3.3 Although a successful contractarian theory defeats the presumption against morality arising from its conception of rational, independent individuals, yet it should take the presumption seriously. The first conception central to our theory is therefore that of a morally free zone, a context within which the constraints of morality would have no place. 21 The free zone proves to be that habitat familiar to economists, the perfectly competitive market. Such a market is of course an idealization; how far it can be realized in human society is an empirical question beyond the scope of our enquiry. Our argument is that in a perfectly competitive market, mutual advantage is assured by the unconstrained activity of each individual in pursuit of her own greatest satisfaction, so that there is no place, rationally, for constraint. Furthermore, since in the market each person enjoys the same freedom in her choices and actions that she would have in isolation from her fellows, and since the market outcome reflects the exercise of each person's freedom, there is no basis for finding any partiality in the market's operations. Thus there is also no place, morally, for constraint. The market exemplifies an ideal of interaction among persons who, taking no interest in each other's interests, need only follow the dictates of their own individual interests to participate effectively in a venture for mutual advantage. We do not speak of a co-operative venture, reserving that label for enterprises that lack the natural harmony of each with all assured by the structure of market interaction.

The perfectly competitive market is thus a foil against which morality appears more clearly. Were the world such a market, morals would be unnecessary. But this is not to denigrate the value of morality, which makes possible an artificial harmony where natural harmony is not to be had. Market and morals share the noncoercive reconciliation of individual interest with mutual benefit.

Where mutual benefit requires individual constraint, this reconciliation is achieved through rational agreement. As we have noted, a necessary condition of such agreement is that its outcome be mutually advantageous; our task is to provide a sufficient condition. This problem is addressed in a part of the theory of games, the theory of rational bargaining, and divides into two issues.22 The first is the bargaining problem proper, which in its general form is to select a specific outcome, given a range of mutually advantageous possibilities, and an initial bargaining position. The second is then to determine the initial bargaining position. Treatment of these issues has yet to reach consensus, so that we shall develop our own theory of bargaining.

Solving the bargaining problem yields a principle that governs both the process and the content of rationale agreement. We shall address this in Chapter V, where we introduce a measure of each, person's stake in a bargain--the difference between the least he might accept in place of no agreement, and the most he might receive in place of being excluded by others from agreement. And we shall argue that the equal rationality of the bargainers leads to the requirement that the greatest concession, measured as a proportion of the conceder's stake, be as small as possible. We formulate this as the principle of minimax relative concession. And this is equivalent to the requirement that the least relative benefit, measured again as a proportion of one's stake, be as great as possible. So we formulate an equivalent principle of maximin relative benefit, which we claim captures the ideas of fairness and impartiality in a bargaining situation, and so serves as the basis of justice. Minimax relative concession, or maximin relative benefit, is thus the second conception central to our theory.

If society is to be a co-operative venture for mutual advantage, then its institutions and practices must satisfy, or nearly satisfy, this principle. For if our theory of bargaining is correct, then minimax relative concession governs the ex ante agreement that underlies a fair and rationale co-operative venture. But in so far as the social arrangements constrain our actual ex post choices, the question of compliance demands attention. Let it be ever so rational to agree to practices that ensure maximin relative benefit; yet is it not also rational to ignore these practices should it serve one's interest to do so? Is it rational to internalize moral principles in one's choices, or only to acquiesce in them in so far as one's interests are held in check by external, coercive constraints? The weakness of traditional contractarian theory has been its inability to show the rationality of compliance.

Here we introduce the third conception central to our theory, constrained maximization. We distinguish the person who is disposed straightforwardly to maximize her satisfaction, or fulfil her interest, in the particular choices she makes, from the person who is disposed to comply with mutually advantageous moral constraints, provided he expects similar compliance from others. The latter is a constrained maximizer. And constrained maximizers, interacting one with another, enjoy opportunities for co-operation which others lack. Of course, constrained maximizers sometimes lose by being disposed to compliance, for they may act co-operatively in the mistaken expectation of reciprocity from others who instead benefit at their expense. Nevertheless, we shall show that. under plausible conditions, the net advantage that constrained maximizers reap from co-operation exceeds the exploitative benefits that others may expect. From this we conclude that it is rational to be disposed to constrain maximizing behaviour by internalizing moral principles to govern one's choices. The contractarian is able to show that it is irrational to admit appeals to interest against compliance with those duties founded on mutual advantage.23

But compliance is rationally grounded only within the framework of a fully co-operative venture, in which each participant willingly interacts with her fellows. And this leads us back to the second issue addressed in bargaining theory--the initial bargaining position. If persons are willingly to comply with the agreement that determines what each takes from the bargaining table, then they must find initially acceptable what each brings to the table. And if what some bring to the table includes the fruits of prior interaction forced on their fellows, then this initial acceptability will be lacking. If you seize the products of my labour and then say 'Let's make a deal', I may be compelled to accept, but I will not voluntarily comply.

We are therefore led to constrain the initial bargaining position, through a proviso that prohibits bettering one's position through interaction worsening the position of another.24 No person should be worse off in the initial bargaining position than she would be in a non-social context of no interaction.

The proviso thus constrains the base from which each person's stake in agreement, and so her relative concession and benefit, are measured. We shall show that it induces a structure of personal and property rights, which are basic to rationally and morally acceptable social arrangements. The proviso is the fourth of the core conceptions of our theory. Although a part of morals by agreement, it is not the product of rational agreement. Rather, it is a condition that must be accepted by each person for such agreement to be possible. Among beings, however rational, who may not hope to engage one another in a cooperative venture for mutual advantage, the proviso would have no force. Our theory denies any place to rational constraint, and so to morality, outside the context of mutual benefit. A contractarian account of morals has no place for duties that are strictly redistributive in their effects, transferring but not increasing benefits, or duties that do not assume reciprocity from other persons. Such duties would be neither rationally based, nor supported by considerations of impartiality.

To the four core conceptions whose role we have sketched, we add a fifth--the Archimedean point, from which an individual can move the moral world.2s To confer this moral power, the Archimedean point must be one of assured impartiality--the position sought by John Rawls behind the 'veil of ignorance'. We shall conclude the exposition of our moral theory in Chapter VIII by relating the choice of a person occupying the Archimedean point to the other core ideas. We shall show that Archimedean choice is properly conceived, not as a limiting case of individual decision under uncertainty, but rather as a limiting case of bargaining. And we shall then show how each of our core ideas--the proviso against bettering oneself through worsening others, the morally free zone afforded by the perfectly competitive market, the principle of minimax relative concession, and the disposition to constrained maximization--may be related, directly or indirectly, to Archimedean choice. In embracing these other conceptions central to our theory, the Archimedean point reveals the coherence of morals by agreement.

4. A contractarian theory of morals, developed as part of the theory of rational choice, has evident strengths. It enables us to demonstrate the rationality of impartial constraints on the pursuit of individual interest to persons who may take no interest in others' interests. Morality is thus given a sure grounding in a weak and widely accepted conception of practical rationality. No alternative account of morality accomplishes this. Those who claim that moral principles are objects of rational choice in special circumstances fail to establish the rationality of actual compliance with these principles. Those who claim to establish the rationality of such compliance appeal to a strong and controversial conception of reason that seems to incorporate prior moral suppositions. No alternative account generates morals, as a rational constraint on choice and action, from a non-moral, or morally neutral, base.

But the strengths of a contractarian theory may seem to be accompanied by grave weaknesses. We have already noted that for a contractarian, morality requires a context of mutual benefit. John Locke held that 'an Hobbist . . . will not easily admit a great many plain duties of morality' .26 And this may seem equally to apply to the Hobbist's modern-day successor. Our theory does not assume any fundamental concern with impartiality, but only a concern derivative from the benefits of agreement, and those benefits are determined by the effect that each person can have on the interests of her fellows. Only beings whose physical and mental capacities are either roughly equal or mutually complementary can expect to find cooperation beneficial to all. Humans benefit from their interaction with horses, but they do not co-operate with horses and may not benefit them. Among unequals, one party may benefit most by coercing the other, and on our theory would have no reason to refrain. We may condemn all coercive relationships, but only within the context of mutual benefit can our condemnation appeal to a rationally grounded morality.

Moral relationships among the participants in a co-operative venture for mutual advantage have a firm basis in the rationality of the participants. And it has been plausible to represent the society that has emerged in western Europe and America in recent centuries as such a venture, For Western society has discovered how to harness the efforts of the individual, working for his own good, in the cause of ever-increasing mutual benefit. 27 Not only an explosion in the quantity of material goods and in the numbers of persons, but, more important, an unprecedented rise in the average life span, and a previously unimaginable broadening of the range of occupations and activities effectively accessible to most individuals on the basis of their desires and talents, have resulted from this discovery.28 With personal gain linked to social advance, the individual has been progressively freed from the coercive bonds, mediated through custom and education, law and religion, that have characterized earlier societies. But in unleashing the individual, perhaps too much credit has been given to the efficacy of market-like institutions, and too little attention paid to the need for co-operative interaction requiring limited but real constraint. 29 Morals by agreement then express the real concern each of us has in maintaining the conditions in which society can be a co-operative venture.

But if Locke's criticism of the scope of contractarian morality has been bypassed by circumstances that have enabled persons to regard one another as contributing partners to a joint enterprise, changed circumstances may bring it once more to the fore. From a technology that made it possible for an ever-increasing proportion of persons to increase the average level of well-being, our society is passing to a technology, best exemplified by developments in medicine, that make possible an ever-increasing transfer of benefits to persons who decrease that average.30 Such persons are not party to the moral relationships grounded by a contractarian theory.

Beyond concern about the scope of moral relationships is the question of their place in an ideal human life. Glaucon asked Socrates to refute a contractarian account of justice, because he believed that such an account must treat justice as instrumentally valuable for persons who are mutually dependent, but intrinsically disvaluable, so that it 'seems to belong to the form of drudgery.' Co-operation is a second-best form of interaction, requiring concessions and constraints that each person would prefer to avoid. Indeed, each has the secret hope that she can be successfully unjust, and easily falls prey to that most dangerous vanity that persuades her that she is truly superior to her fellows, and so can safely ignore their interests in pursuing her own. As Glaucon said, he who 'is truly a man' would reject moral constraints.32

A contractarian theory does not contradict this view, since it leaves altogether open the content of human desires, but equally it does not require it. May we not rather suppose that human beings depend for their fulfilment on a network of social relationships whose very structure constantly tempts them to misuse it? The constraints of morality then serve to regulate valued social relationships that fail to be self-regulating. They constrain us in the interests of a shared ideal of sociability. Co-operation may then seem a second-best form of interaction, not because it runs counter to our desires, but because each person would prefer a natural harmony in which she could fulfill herself without constraint. But a natural harmony could exist only if our preferences and capabilities dovetailed in ways that would preclude their free development. Natural harmony would require a higher level of artifice, a shaping of our natures in ways that, at least until genetic engineering is perfected, are not possible, and were they possible, would surely not be desirable. If human individuality is to bloom, then we must expect some degree of conflict among the aims and interests of persons rather than natural harmony. Market and morals tame this conflict, reconciling individuality with mutual benefit.

We shall consider, in the last chapters of our enquiry, what can be said for this interpretation of the place of moral relationships in human life. To do so we shall remark on speculative matters that lead beyond and beneath the theory of rational choice. And we may find ourselves with an alternative reading of what we present as a theory of morals.33 We seek to forge a link between the rationality of individual maximization and the morality of impartial constraint. Suppose that we have indeed found such a link. How shall we interpret this finding? Are our conceptions of rationality and morality, and so of the contractarian link between them, as we should like them to be, fixed points in the development of the conceptual framework that enables us to formulate permanent practical truths? Or are we contributing to the history of ideas of a particular society, in which peculiar circumstances have fostered an ideology of individuality and interaction that coheres with morals by agreement? Are we telling a story about ideas that will seem as strange to our descendants, as the Form of the Good and the Unmoved Mover do to us?

NOTES

[1] See David Hume, *An Enquiry concerning the Principles of Morals*, sect. ix, pt. ii, in L Selby-Bigge (ed.), *Enquiries concerning Human Understanding and concerning the principles of Morals,* 3rd edn. (Oxford, 1975), p. 280.

[2] Ogden Nash, 'Kind of an Ode to Duty', *I Wouldn't Have Missed It: Selected Poems of Ogden Nash* (Boston, 1975), p. 141.

[3] See David Hume, *A Treatise of Human Nature,* bk. ii, pt. iii, sect. iii, ed. L. A. (Oxford, 1888), pp. 413-18.

[4] Thus one might propose an error theory of moral language; for the idea of an error theory, see J. L. Mackie, *Ethics: Inventing Right and Wrong* (Harmondsworth, Middx., 1977), ch. I, eSp. pp. 35,48-9. 5 The idea that moral appeals are persuasive is developed by C. L. Stevenson; see *Ethics and Language* (New Haven, 1944), esp. chs. vi, ix.

[5] For the idea of the proviso, see note I to ch. VII, below. For the idea of an Archimedean point, see Rawls, *A Theory of Justice*, pp. 260--5.

[6] Our sketch of rational choice owes much to J. C. Harsanyi; see 'Advances in Understanding Rational Behavior', in *Essays on Ethics, Social Behavior, and Scientific Explanation* (Dordrecht, 1976), pp. 89-98, and 'Morality and the theory of rational behaviour', in A. Sen and B. Williams (eds.), *Utilitarianism and beyond* (Cambridge, 1982), pp. 42-4.

[7] J. Rawls, *A Theory of Justice* (Cambridge, Mass., 1971), p. 16; J. Harsanyi, 'Morality and the theory of rational behaviour', p. 42.

[8] See Rawls, p. 12.

[9] See ibid., p. 11; 'the principles. . . are to assign basic rights and duties and to determine the division of social benefits.' Principles for individuals are distinguished from the principles of justice; see p. 108.

[10] See Rawls's distinction of 'the Reasonable' and 'the Rational', in 'Kantian Constructivism in Moral Theory', *Journal of Philosophy* 77 (1980), pp. 528-30.

[11] See Rawls, *A Theory of Justice*, pp. 14-15, and Harsanyi, 'Morality and the theory of rational behaviour', pp. 44-6, 56-60.

[12] See Harsanyi, 'Morality and the theory of rational behaviour', p. 62.

[13] This conception of practical rationality appears with particular clarity in T. Nagel, *The Possibility of Altruism* (Oxford, 1970), esp. ch. x. It can also be found in the moral theory of R. M. Hare; see *Moral Thinking* (Oxford, 1981), esp. chs. 5 and 6.

[14] See Harsanyi, 'Advances in Understanding Rational Behavior', p. 89; also J. Elster, *Ulysses and the Sirens: Studies in rationality and irrationality* (Cambridge, 1979); 'The "rational-choice" approach to human behaviour is without much doubt the best available model. . .', p. 112.

[15] See Plato, *Republic*, 358b--359b.

[16] Thomas Hobbes, *Leviathan* (London, 1651), ch. 14, pp. 64--5.

[17] See G. R. Grice, *The Grounds of Moral Judgement* (Cambridge, 1967), and K. Baier, *The Moral Point of View: A Rational Basis of Ethics* (Ithaca, NY, 1958); the quotation is from p. 309.

[18] Rawls, *A Theory of Justice*, pp. 4, 13.

19. The discussion here is related to my characterization of a convention in 'David Hume, Contractarian', *Philosophical Review* 88 (1979), pp. 5-8.

20. See Hume, *Treatise*, iii. ii. ii, p. 490.

21. This is the theme of ch. IV, below. See also my earlier discussion in 'No Need for Morality: The Case of the Competitive Market', *Philosophic Exchange* 3, no. 3 (1982), pp.41-54.

22. For references to the literature on rational bargaining, see notes 12-14 to ch. V, below.

23 This conclusion rests on a reinterpretation of the maximizing conception of rationality, which we develop in ch. VI, below; see especially the opening paragraph of 3. 1.

24. For the idea of th proviso, see note 1 to Ch. VII, below.

25. For the idea of an Archimedean point, see Rawls, *A Theory pf Justice*, pp. 260-5.

26 Locke MS, quoted in J. Dunn, *The Political Thought of John Locke* (Cambridge, 1%9), pp. 218-19.

27 We offer no explanation of this discovery. There seems no reason to suppose that it resulted from deliberate search.

28 For the increase in average life span, see N. Eberstadt, 'The Health Crisis in the U.S.S.R.', *New York Review of Books* 28, no, 2 (1981), p. 23. For the broadening in the range of accessible occupations, note that 'As late as 1815 three-quarters of its [Europe's] population were employed on the land. . .', *The Times Concise Atlas of World History*, ed, G. Barraclough (London, 1982), p. 82.

29. Thus the idea of economic man as an unlimited appropriator comes to dominate social thought. The effects of this conception are one of the themes of my 'The Social Contract as Ideology', *Philosophy and Public Affairs* 6 (1977), pp. 130--M.

30. The problem here is not care of the aged, who have paid for their benefits by earlier productive activity. Life-extending therapies do, however, have an ominous redistributive potential. The primary problem is care for the handicapped, Speaking euphemistically of enabling them to live productive lives, when the services required exceed any possible products, conceals an issue which, understandably, no one wants to face, Without focusing primarily on these issues, I endeavour to begin a a contractarian treatment of certain health care issues in 'Unequal Need: A Problem of Equity in Access to Health Care', *Securing Access to Health Care: The Ethical Implications of Differences in the Availability of Health Services*, 3 vols., President's Commission for the Study of Ethical Problems in Medicine and. Biomedical and Behavioral Research (Washington, 1983), vol. 2, pp. 179-205.

31. *Republic*, 358a, trans. A. Bloom (New York, 1968), p. 36. 32 ibid., 359b, p. 37.

33. The thoughts in this paragraph have been influenced by R. Rorty; see esp. 'Method and Morality', in Norma Haan, R. N. Bellah, P. Rabinow, and W. M. Sullivan (eds.), *Social Science as Moral Inquiry* (New York, 1983), pp. 155-76.