

RATIONALITY
AND
POLITICS

Glen Newey

Acknowledgements

I should like to thank the following for their assistance over the three years in which this thesis was written. First, John Horton of York University Politics Department my PhD supervisor: his professionalism - in particular his committed approach to teaching and the breadth of his knowledge both of his own field and related disciplines - was exemplary, as others who have benefited from his tuition can attest. He is not, of course, to be held responsible for any errors which remain in the thesis. Thanks also to members of the York University Political Theory workshop, to Susan Mendus, Peter Nicholson, David Edwards, Chris Megone, David Morland and Barbara McGuinness, who commented on an earlier draft of part of this thesis at one of the workshop's sessions. I should also like to thank two former colleagues of mine on the York Political Philosophy MA course, Neil MacDonald and Maurice Glasman, for their advice and encouragement over the past three years. Finally, my thanks to the States of Jersey Education Committee, who provided a three-year grant to fund the writing of this thesis.

PAGE

NUMBERING

AS ORIGINAL

Introduction

This thesis reviews some recent trends in analytical political philosophy, aiming to identify certain thematic links between traditions of theoretical writing which have hitherto been regarded as distinct and even opposed to each other. The traditions in question both originate from Enlightenment rationalism, but continue to exert a strong influence over contemporary political theory in the Anglophone world. They are Kantianism, enjoying a still-current revival following the publication of John Rawls' A Theory of Justice¹ in 1971; and utilitarianism, which has never entirely gone out of vogue since Bentham's day, and informs much writing in political and ethical theory today, reflected in such works as Derek Parfit's Reasons and Persons².

While Kantianism and utilitarianism both originated with the Enlightenment, they have been taken - in many ways rightly - to embody quite heterogeneous styles of thinking. Their formal differences as varieties of ethical theory are indeed quite marked: notably in Kantianism's stress on the unconditional and categorical nature of morality, as compared with utilitarianism's contingent and empirical bias - the morally right action cannot, for the utilitarian, be determined in advance of a sum weighing its expected costs and benefits (whatever the unit of account). Kantianism's stress on the moral primacy of persons, too, is absent from utilitarian theory, at least in this unconditional form. The interests of the few may have to be sacrificed to promote those of the many. As Rawls has remarked, '[u]tilitarianism does not take seriously the distinction between persons'³.

In spite of these clear differences of ethical approach, however, there are notable similarities in the Enlightenment theories' approach to politics. These similarities I have labelled instrumentalism, a compound theory expounded in Chapters One to Four of this work. In summary form, the principal features of instrumentalism are the following. First, it holds that the ends of political action are determined entirely by pre- or non-

political considerations: these ends being, for example, a certain ordering of civil society, or identifiable forms of non-political (e.g. welfare) benefit. Second, it regards political decision-making as being, at least in principle, amenable to rational calculation; that is, not merely to rational deliberation, but to certain rules or rational algorithms applicable to concrete policy decisions. Third, politics is subordinated, in a hierarchy of deliberative concerns, to non-political (specifically moral) considerations; this is reflected in Rawls' emphasis on the 'priority' of moral over other branches of theory⁴, or in the structuring of political action to achieve utilitarianly desirable goals.

The present study takes issue with instrumentalism, thus defined. It is for the reader to decide, from the characterization presented, whether the affinities between Kantian and utilitarian theories of politics (which very often remain implicit in recent theoretical work, rather than being spelled out in detail) are as strong as I claim, and whether they represent a cohesive view of political action - and whether the alternative sketched in the final chapter is a convincing or appealing substitute for instrumentalist theory. Whatever success the latter enjoys, the present study in no way repudiates the rationalizing aims of the Enlightenment in favour, for example, of some reactive anti-modernism of the form currently in vogue⁵: it comes to praise rationality, not bury it. But (as Kant well appreciated) it is a prime function of reason to determine the limits of rationality. It may be that contemporary political theory is impaired by an unduly rationalistic view of political rationality, and needs (without mystification) a more manifold, less totalizing approach to politics.

This may have to avoid the conceptual methods of recent writings - in particular, the elevation of specific moral concepts, such as that of the person, or citizen, or right, or community, to primordial status in political theory. There may just be no single moral concept which can serve as the repository for all ideas of political value. The pluralistic theories of value with which recent liberal theory has associated itself⁶ appear, at least, to have this implication. A more thoroughgoing pluralism would build this heterogeneity not only into the ordering of civil society, but theory itself, acknowledging that

there can be no single value mediating political conflicts, as some forms of liberal neutrality have supposed⁷. Conflict may be ineliminable from politics. Nor need there be clear rules of engagement in political conflicts. An understanding of this truth requires an appreciation not merely of Enlightenment rationalism, but of Machiavelli.

Notes

1. John Rawls, A Theory of Justice (Oxford 1971).

2. Derek Parfit, Reasons and Persons (Oxford 1984).

3. Rawls, op. cit., p27.

4. John Rawls, 'The Independence of Moral Theory', Proceedings and Addresses of the American Philosophical Association 1975.

5. For an interesting analysis of these trends, see Alex Callinicos, Against Postmodernism (London 1989).

6. E.g. Joseph Raz, The Morality of Freedom (Oxford 1986).

7. E.g. Charles Larmore, Patterns of Moral Complexity (Cambridge 1987).

Chapter One

Utilitarianism and Politics

Utilitarianism and Politics

1. Introduction

This chapter aims to determine whether utilitarianism is able to provide any recognizable account of political institutions and behaviour in its prescriptions for rational agency. The main method by which this inquiry will be pursued will be a conceptual model of politics or, as it will sometimes be called, of political structures, which is intended to capture as far as possible within a treatment of this scope some of the principal criteria used to characterize politics in everyday situations. This term is meant to be general enough to include those features of politics which exhibit some degree of rational coordination, whether in the form of institutions, processes, or ideology. Once the model of political structures has been outlined, it will be used to assess utilitarian treatments of politics. Broadly the conclusion will be that utilitarianism can only take these structures into account by abandoning its most distinctive features as a theory of rational agency. If, on the other hand, these features are retained, utilitarianism is no longer able to provide any recognizable account of politics as described by the model.

Obviously enough, this places a premium on the model's successfully mirroring our conception of politics. Given the form of the argument just sketched, which presents the utilitarian with the choice between political realism and theoretical distinctiveness, it might be objected that the more robustly revisionist brand of utilitarian would wish to claim that utilitarian policies, if consistently pursued, would require political structures so radically different from any now existing that there is little point in evaluating the theory by means of a model drawn from contemporary political practice. This is really a problem concerning the adequacy of the model as a characterization of politics. For if it is adequate, the revisionist has to explain in what respects his own alternative model is identifiable as a description of politics at all. If the conceptual displacements required by the alternative approach are too great, it may be only in the most depleted sense that his is

a theory of politics at all. The argument still takes its starting-point from ordinary usage, even if - indeed, especially if - the objection is that the model does not represent politics as it really is.

Here the difficulty is that the model is meant to capture our thinking about politics, but the only way of checking the adequacy of the model is by reference to this thinking itself. Because of this implicit circularity, the method is open to the criticism that it may bias the conclusions to be drawn about the relationship between utilitarianism and politics. If a selective or distorted picture is given, it is only too easy to show that utilitarianism fails to accommodate political structures as falsely described. Even before adopting the revisionist approach, the utilitarian can claim that the model is distorted, while not denying that it is with reference to some such model that utilitarianism is to be judged. The objections we consider within non-ideal variants of utilitarianism are broadly of this form. This is however not so damaging if it can be shown that in reformulating the model the non-ideal theorist is forced to compromise the principal features of the doctrine: then it will not matter so much whether the original model was accurate or not, since the conclusions about compromising the theory follow even with the utilitarian's own favoured alternative model. But these perplexities are a staple problem in philosophical argument, when theorists seek to sharpen such notions by means of conceptual analysis. The problems here are at least no more acute than in other cases where the method is employed.

We shall, then, be working with a rough and schematic division between forms of utilitarianism, as either ideal or non-ideal versions of the theory. This distinction is not intended to correspond to any of the other distinctions marked in the literature between, for example, act- and rule-utilitarianism, or between either of these and utilitarian generalization.¹ These distinctions will indeed concern us later. But the division of utilitarian theory into ideal and non-ideal forms is based on how corrective an attitude towards human institutions, practices, psychology, etc., a given variant of the theory takes. Non-ideal theories are prepared to accept as part of the decision-making environment the fact that human beings will behave in non-utilitarian ways, and adjust the schedule of available actions

and consequences accordingly. Ideal theories, on the other hand, admit no such 'partial compliance'² possibilities in their body of assumptions. It may be said that this distinction between ideal and non-ideal versions of utilitarianism merely parallels that between act- and rule-utilitarianism. I shall, however, give reasons below for doubting this, and suggest that while the ideal/ non-ideal distinction is real enough, that between act- and rule-utilitarianism is largely illusory.

II

Bearing these comments in mind, I set out below a conceptual model of political structures. There is no attempt to rank the characteristics identified in order of importance, nor is there any pretence of providing a complete analysis of all aspects of these structures, even within the categories distinguished. The model merely aims to make explicit certain intuitively significant aspects of these structures for the heuristic purposes explained above.

(1) The structures must have some public dimension: this means not only that political deliberations must be conducted in some public body, but also, and more importantly, that these deliberations or processes relate to matters of public concern. This is, of course, fairly vague, as concerns both the content and the context of deliberation. By way of comparison, it is assumed that debates over policy in some private concern such as a market corporation are not in the relevant sense public, since the terms of such debate are not usually dictated by considerations of public welfare as (at least on the surface) the terms of political debate are. There may of course be certain concerns, e.g. the environmental costs of production, which obtrude into this domain, of public welfare; the firm may also project a corporate image of itself as motivated by public service or welfare. But in such propaganda the trade-off between these goods and profit is seldom mentioned, and (unlike in at least western politics) the assumption is that financial gain is both the dominant and legitimate purpose of business.

This sort of distinction is, admittedly, harder to sustain if someone takes a thoroughly reductive view of politics and its

rhetoric of concerns as mere rationalizations of self-interest. The fact remains however that even a wholly reductive Namierite view of political motivation has to explain, at the level of appearances, why considerations of welfare bulk so much larger in political rhetoric than in corporate image-building. Apart from this, there is the brute fact that the financial rewards to be had from a career in western politics are markedly lower than its practitioners could reasonably expect from devoting their skills to other professional pursuits. This is reinforced by public controls on corruption in western political cultures, in contrast to those countries where bribery and corruption are accepted features of political life. There is little need to dwell further on this point for our purposes, since utilitarianism would presumably wish to take a broadly non-Namierite view of the conduct of politics: its welfarist concerns preclude endorsement of at least the grosser forms of speculation and political bribery.

This of course leaves it fairly unclear what the matters of public concern referred to above are, and how they are to be identified as such. I take it as a fact about the way politics is conducted that its agenda is neither static nor wholly internally generated. For example, dispute over resource allocation, which now dominates political debate in western democracies, received far less attention in the last century when the public sector played a less prominent role in the national economy. Part of the reason for the shift was the change in the economic infrastructure with its social consequences, which demanded new approaches to policy. But these shifts in the subject-matter of political argument do not mean that it is not reasonably well-defined what is and is not of public concern at any point in time. This does not, of course, require an implausibly naturalistic view of the political agenda. The content of that agenda is determined by what political agents - professional politicians and public observers - think it ought to be. That is, politics is not just about normative claims, but about views of normative claims and their relationship to public action.

(2) The decisions taken as a result of political debate must be backed with force or sanctions of some kind, and more generally the notion of power's being exercised must play some explanatory

role in discussing and understanding political action. This explanatory element is significant, as the existence of political power serves not merely to give effect or coercive backing to decisions already reached, but also to influence the actions of political agents in arriving at those decisions - particularly of those on the receiving end of policy. In other words, it is a familiar feature of political life that the context within which decisions are taken includes the operation of political processes themselves, and the use of power alters the choices, with attendant costs, open to political agents.

This does not mean that political decisions are always directly and overtly supported by the threat of coercion, though this is often true, and some views of political action, such as Machiavelli's and Hobbes's, emphasize strongly the part played by sheer force in political life - so strongly, in fact, that on these views there is little to choose between the notions of monopoly force and political power. One objection to so close an identification as this is that we customarily take power, but not necessarily force, to be intrinsically related to political institutions and processes. It is through this institutional context that we identify the exercise of power, even where, as in military dictatorships, the coercive sanctions which support those in power are immediate and obvious.

(3) Deliberation about political issues must include the possibility of conflict. In its weakest form, one which all utilitarians can endorse, this claim requires only that the fact that deliberation involves making decisions means that there are, at the outset, different possible actions available, and different outcomes which can be produced depending on what decision is taken. It is a necessary feature of acting that the world is (even if the action fails to achieve its aim) altered by it; there is, at the limit, the alternative of doing nothing (leaving the world as it is).³ To the extent, then, that there are different possible end-states open in deliberation, there is a possibility of conflict - that different parties or factions may form in support of rival policies or ideologies.

Of course, the political conflict witnessed in real life goes well beyond this, to take the form of systematic or institutionalized confrontation between groups organized on

ideological, tribal, religious, ethnic, etc., lines. When conflict takes on this institutional character, a stronger form of explanation is required than that offered by the purely semantic considerations about the nature of deliberation mentioned above. Where there is endemic conflict of this form, it is less clear that the utilitarian can accommodate it within his preferred structures of rational explanation. In part the ease with which this can be done depends, as we shall see, on whether it is ideal or non-ideal utilitarian theory which is under review.

(4) Perhaps more contentiously, there must be some sense that politics provides an autonomous area of human concern.⁴ That is, it has to be recognized that much of the content of political activity and argument is self-generated. Political decisions are very often guided or constrained by considerations arising from the activity of politics itself - some political actions are undertaken and have a point only because there already exists a set of political structures which exert an independent force on decision-making.

These are not part of any external goals which politics may aim to promote, since they have no existence outside politics itself. In part this results from the fact that politics, like any other complex collective undertaking, involves instrumental actions in the achievement of its final ends. But it is not merely a matter of performing certain actions for the sake of others, but of being able to explain some actions with reference to political structures themselves rather than merely to the external goal. This can happen because, as already explained, political engagement is normally the scene of conflict, in the strong sense, and therefore demands that politicians devote much of their time to conciliating or overcoming opposition.

Much more could be said on all these matters, and it may be said that there are certain aspects of political structures, such as the role of political ideology,⁵ which would need attention in any full treatment. While it is certainly true that a good deal of explanatory work in political analysis can be done by the notion of ideology, at this level of generality the aspects of these structures which it is important to distinguish concern

power, and conflicts waged in order to secure it. Even apart from the type of reductive analysis already mentioned, which explains apparent ideological conflict as rationalizations by the protagonists of their own interests, it is not clear that systematic ideological conflict is necessary to identify a given set of practices as political. Factional struggles for power may occur outside any ideological context. In this respect the model is likely if anything to err on the side of leniency rather than harshness towards utilitarianism, since it is at best doubtful (as will be argued below) whether much sense can be made of the theory as a political ideology and whether, as a result, the notion of ideology plays any part in its prescriptions for rational agency.⁶

The characterization of political structures given above is meant to apply in the first instance to the paradigm case of national, central governmental or legislative institutions and processes, such as policy-making by the central administration, debates over statutory measures and so on. There are however more marginal applications of the term 'politics' to bodies not usually thought of as political institutions - to decision-making agencies such as private-sector firms, university administrations, etc. Reasons have already been given for doubting that private businesses are best regarded as political agencies. This is not of course to deny that bodies such as those mentioned may bear some resemblance to aspects of political structures. Apart from the considerations of public welfare already mentioned, these bodies usually lack the ideological or factional cohesiveness to mirror party- and ideology-based politics of the kind witnessed in western democracies. But this ambiguity or uncertainty in the extension of the term may be seen, in any case, as an advantage: the model reflects the vagueness of our understanding of the term, rather than imposing any spurious rigour upon it.

III

The question then is whether utilitarianism can accommodate political structures as interpreted in the previous section. As already indicated, the argument will offer utilitarianism the alternatives of either retaining its most distinctive and (at

least prima facie) attractive features, but failing to allow for these structures, or allowing for them at the cost of abandoning those features. It is the latter prong of this fork that the utilitarian will be offered in this section. The focus will be on a notable recent example of (what will be argued to be) non-ideal utilitarian theory, the two-level analysis of moral decision-making presented by Richard Hare in his Moral Thinking.⁷ Hare's is a work of moral, rather than political philosophy, and to this extent it is a matter for speculation what Hare would regard its political implications as being. This is however no more true of Hare than of other theoretical utilitarians now writing.⁸

In Section I non-ideal utilitarianism was informally defined as a version of the theory which allowed for the possibility in calculating schedules of benefits and costs that some agents may fail to do what, ex ante, is utilitarianly rational, and adapts the schedules accordingly. It might be thought from a reading of Hare's book that his theory was not of this form, since it argues repeatedly for the importance of 'critical [ie utilitarian] thinking'⁹ in resolving (what Hare thinks are only apparent) moral dilemmas. While it is certainly true that Hare is not concerned with intricate calculations of the implications for utilitarians of patterns of non-compliance in the behaviour of non-utilitarians (as, for example, Regan is), it is nonetheless the case that Hare, while observing the two-tier stratification of moral thinking, is not prepared, as some utilitarians are, merely to discard wholesale everyday habits of moral judgment in favour of some general aggregative decision-procedure. This is because Hare believes, as these other utilitarians do not, that the habits of judgment are for the most part capable of utilitarian justification - indeed that their ultimate rationale is utilitarian.

Though this makes it unlikely that Hare would wish to be classified as a non-ideal theorist, it remains true that the two-level structure looks like a departure from the standard format of utilitarian calculation. It appears that Hare's more liberal attitude towards the everyday habits of judgment is motivated by the belief that their utilitarian justification rests on the fact that, these habits having already been formed, a good deal of human happiness depends on their continuance. The more corrective brand of utilitarian is apt to dismiss these considerations in

favour of a belief that the calculations in which he is interested can be effected from the ground up, since there is no non-circular argument to show why these particular habits of judgment should enjoy any priority in deliberation. Indeed, a utilitarian of this persuasion could argue that in some circumstances they reduce the benefits available to agents otherwise disposed to act in line with utilitarian standards of rationality.¹⁰

On the other hand, it may be doubted whether the habits of judgment have the rule-utilitarian justification which Hare seems to point towards - that they are as vital to human happiness, even in the world as it is, as Hare claims. The point here is not whether objections like this are well-founded. It is that they mark a difference among utilitarian writers, and show Hare to be one of those largely prepared to accept that an important influence on the utilities available in a given choice situation is the way the world is.

Thus the claim is that Hare counts as a non-ideal theorist because he works without any ex ante notion of utilitarian rationality. This admittedly involves a slight widening of the definition of non-ideal theory to embrace not only the partial-compliance cases originally referred to, but, more broadly, conservative attitudes towards given but in principle alterable features of the decision-making environment (though this still allows for cases of partial compliance). Hare might deny this, on the grounds that the habits of judgment generate, once supplemented with the critical level of moral thinking, the greatest utility available; but even if, as Hare thinks, the ultimate reason why we are educated in these habits in the first place is utilitarian, it is hard to avoid the conclusion that the utilitarian reasons are themselves dependent on the habits' already being in existence, since they provide the raw material from which so much of the calculation is performed.

Hare, as already stated, is not forthcoming on the political implications of his two-level theory. But it is fairly clear how the operation of the dual structure in the moral sphere would transfer to political decision-making. Hare presumably would wish to claim that politicians have, on the whole, been brought up to make similar moral judgments about lying, promise-keeping, and so on, as everybody else, but that, in cases of conflict between

moral judgments, critical thinking must be used to decide the issue. What makes politics distinctive is that the allocation of public resources depend, as is seldom the case with private citizens, on the decisions of politicians. This no doubt makes it more urgent to get the decision right, as the utility numbers involved will be much larger than in private cases; but there is no reason in the basic theory to suppose that the basic format of decision-making is going to be any different in the political case.

In a separate essay, 'Political Obligation',¹¹ Hare has attempted to extend his moral theory to politics (written before his adoption of the two-level formalism, but after his conversion to utilitarianism).¹² Hare's aim is to give reasons - ones which derive from the theory, and are not merely consistent with it - for accepting political obligations as a sub-species of moral obligations. The reasons he presents fall into two categories, prudential and moral. Both of these categories include reasons derived from the existence of the law itself. Hare discusses prudential reasons relating to the penalties attendant on disobedience, and moral reasons dependent on others' likely compliance with the law. (This of course raises the problem of circularity in the justification for our being obliged to obey this law: but Hare gives further reasons for obligation in both categories, and argues that this, like all other legislation, is justified by its utility-maximizing effects.)

This is not the place to consider the merits of Hare's argument as a utilitarian defence of political obligation. The important point for the present purposes is that Hare treats political obligation as merely an outcome of the calculative procedures applicable to moral reasoning in general. He is, furthermore, prepared to allow, as indicated above, for the effect of both compliance and non-compliance on an individual agent's calculations. Thus the fact that others are obeying the hygiene law is an additional reason for my doing so, because the likelihood of my causing an epidemic by not keeping clean is greater than it would be if most people were also disobeying it; and conversely this reason loses its force if most others are also ignoring the law.

At the same time, Hare rules out the possibility of free-riding in other cases by referring to the supposed disutilities

suffered by those cooperating, even when they are unaware that others are failing to cooperate.¹³ This move is justified, as often in Hare's work, by a shift to the critical level of thinking when consideration of personal utilities alone fails to deliver the desired conclusion - in this case, by reference to the universal prescriptions violated by free-riding. It is, however, highly unclear that these prescriptions, at any rate in the form Hare wants them, are consistent with (still less required by) utilitarian theory in its unmodified form.

Such possibilities suggest that for Hare the considerations which inform politicians' thinking at the critical level do not necessarily duplicate those at work in private citizens' practical deliberations. For one thing, the subject-matter of political decision-making is likely to alter the criteria in use, or the relative weights assigned to them in deliberation. Kinship and friendship are obvious examples of the sorts of consideration which Hare regards as having a legitimate bearing on our moral thinking as private persons, but which are less evidently admissible as criteria of political deliberation - not that Hare says as much, but criteriological distinctions of this kind, between public and private forms of practical reasoning, are not ruled out by the dual-level theory itself.

Or are they? The foregoing has assumed that this is what Hare himself would want to say about the political implications of his theory. But it is not clear that the two-level structure, when coupled with a utilitarian account of rational agency, can tolerate such distinctions between private and public spheres of deliberation - which is not to be identified with the clear conceptual distinction between compliance and non-compliance which, as we saw, Hare allows for in his account of political obligation. For utilitarianism seems to recommend a single across-the-board rule of conduct to its adherents: to act in all cases so as to maximize utility.

This being so, it is hard to see why the context of the deliberation, or the identity of the agents undertaking it, should in any way affect the relative weights assigned to the available options. Indeed, if this were true, it would lead to problems of incommensurability - the very problems which utilitarians (including Hare) believe their theory to be best able to solve. For different agents, depending on their position,

would assign different utility measures in some cases to the same action, and there would, as a result, be no unique solution to the question of what ought to be done (no unique pattern of behaviour, involving all causally relevant agents, which maximizes utility). The same pattern of behaviour would have different utility numbers depending on which agent's perspective was taken up, and how that agent was situated.

The upshot of this discussion is not that the very idea of a two-level theory is anti-utilitarian, but that the nature of the judgments which Hare is prepared to include at the intuitive level of moral thinking make it difficult to sustain any consistently utilitarian approach towards agency, since the dual structure does not preclude incommensurabilities in the patterns of moral evaluation which Hare is prepared to accept. On this analysis, Hare fails to take account of the fact that his own permissive attitude towards everyday habits of judgment is likely to allow too much in for the theory to remain utilitarian. It is not immediately clear how the impersonal aggregation required by the utilitarian calculus can allow for the influences on agency which the kinds of considerations Hare appears to have in mind in fact exert on people in acting. One reason for this is the subjectivity implicit in the fact that the reasons we have for acting have ultimately to be our reasons.¹⁴

An obvious area in which this comes out is that of personal relationships, where differential concern for others means that in very many cases deliberation about agency cannot take the universalistic form required by the theory. Thus the kinds of sentiment - such as personal affection or rancour - which motivate people to act have presumably to be included within the utilitarian calculus, even though a committed follower of the theory would have to disregard any such sentiments in himself when calculating consequences. This makes it very doubtful whether utilitarianism can handle these sentiments satisfactorily, while maintaining as its ideal a society all of whose members act on utilitarian principles.

This leaves it open to doubt whether the habits of judgment on which Hare relies can be as readily justified by rule-utilitarian arguments as he assumes. Hare often writes as if unreflective utilitarianism were the common coin of moral judgments nowadays, but the dilemmas arising from clashes between the intuitive-level

principles inculcated by our moral education in childhood are not the only respect in which the ordinary agent departs from the strict canons of utilitarian rationality. Another way in which this can happen is through our subjective standpoints as agents situated in the world.

This can itself give rise to moral dilemmas, of course, but in such cases one limb of the dilemma is not a consideration which a utilitarian can consistently allow the special deliberative weight which agents assign to it. To take a well-worn example,¹⁵ the sea captain faced with deciding whom to save when his vessel is sinking is unlikely to think that whatever weight is carried by the fact that one of the passengers is his wife is merely cancelled or balanced by the further fact that, after all, the other passengers are somebody else's wives or husbands. But asymmetries of this type are not confined to perceived moral dilemmas. They are a pervasive feature of agency.

At this point it may be objected that the original criticism of Hare's theory, that it failed to allow for the difference between the deliberative weights assigned to rival actions in public as opposed to private or personal agency, now seems to be undermined. It may now appear that the personal considerations inadmissible at the public level are precisely what is ruled out by the features of the utilitarian approach just identified, so that the theory turns out to be less disruptive of political action, as conventionally understood, than was argued earlier.

The point however is not so much that this aspect of politics cannot be accommodated within a utilitarian theory of action, but that it is dubious whether the distinction between public and private action is one which has much sense or point for the utilitarian. It is the possibility of differential rankings' depending on the context of decision-making which is alien to utilitarianism. If Hare is conservative enough about the content of the intuitive-level principles, these differential rankings threaten to contravene the utilitarian requirement that any rational decision-procedure must be optimific: if, on the other hand, Hare rules out such rankings, the theory retains its utilitarian credentials (though it remains to be explained in Hare's moral psychology how the rankings got into our thinking at all), but with the result that his approach can no longer reflect the complexity of practical reasoning, and in particular the

special circumstances in which political decisions are taken.

The reply to the objection, then, consists not in denying that utilitarianism can reproduce in its own terms the impartiality of actual political decision-making, but in claiming that this results not from any appreciation on the utilitarian's part of the complexities in our thinking about the relationship between public and private agency, but from a blanket procedure which happens to fit this aspect of political life fairly well. It is its failure to take into account the autonomy of the political which marks the utilitarian approach. If the politician is required to observe impartiality in policy-making, that is not (according to utilitarianism) because of the nature of the politician's role, but because this issues from a more general requirement of impartiality - one which applies equally to those, like the sea captain, in civil society. The reasons underlying the demand that politicians behave impartially as regards personal friends, relatives, etc., in formulating policy, spring from the same grounds which make personal relationships of this kind acceptable grounds for partiality in other circumstances.

This point can be brought out from the opposite direction as well, when we turn to consider the limits on political intervention - particularly in the area of privacy. While it is commonly accepted that there should be fairly strong restrictions on interference with privacy, utilitarianism has difficulty in accepting any hard-and-fast demarcation of public from private concerns. Hare could maintain that the desire for privacy is one which is strongly felt and must therefore be taken into account in any utilitarian calculation. But it looks as if the intuitive level is simply taken at face value when, as here, the first-order sentiments conduce to the desired conclusion, but are subjected to critical-level revision when, as in the free-riding case, reliance on unreflective preferences threatens to upset the calculation.

If enough people want a specified individual's privacy invaded, for example, it looks as though utilitarianism has to sanction this, the offence felt by the individual himself being no more than a debit item to be entered on the overall balance-sheet. What is needed to avoid this outcome is a set of reflective preferences of a rule-utilitarian form against such invasions of privacy. But this will only be effective as a

manoeuvre if, as Hare argues is true in the free-riding case, reflective preferences like this have a first-order foundation. There is little reason to think that this is invariably so.

This issue has been pursued at length because Hare's own views are fairly clear, as is the choice facing a utilitarian who wishes, as he does, to bring everyday habits of moral judgment into the theory. The ambiguities in Hare's thinking lie in the unclarity over how much of the unreflective moral thinking is really justifiable in (rule-) utilitarian terms, and what the effect is if, as has been argued, it is less than Hare supposes. If this is right, there must be some major work of reconstruction rather than marginal tinkering to be done on our unreflective moral attitudes, unless some argument from a distinction between ex ante and ex post utility can be sustained - that is, by building into the calculation some element to cater for the fact that these attitudes are already in existence. It seems plain, however, as already claimed, that Hare has in mind some less pragmatic justification than this for his acceptance of these attitudes.

The discussion has further suggested that utilitarianism cannot make room for any notion of a sui generis political morality, where the moral requirements on agents in other areas of life are either relaxed or strengthened or reformulated to suit political conditions. If utilitarianism cannot make sense of the idea that rationality depends on the agent's situation, there is no reason to think that it will allow for the possibility that its rational requirements can be modified according to context. It is not clear how far Hare is prepared to allow for such contextualization; but if he is, it is hard to see how he can remain a consistent utilitarian.

IV

It is instructive to consider how an apparently quite different brand of utilitarian theory performs when analysed along the same lines as Hare's. In the previous section, non-ideal theory was defined as a version of utilitarianism which does not operate with an ex ante understanding of rationality, i.e. prior to human institutions, practices, and so on, but uses these as raw materials for a utilitarian calculation. It should

now be clear in what respects Hare can be seen as a non-ideal theorist. Regan,¹⁶ by contrast with Hare, is far more concerned with utilitarian responses to partial compliance - with agents' failure to act in what (ex ante) would be the utilitarianly rational way. Regan thus works with a notion of ex ante rationality, but only as a starting-point from which to develop cooperative utilitarianism, his own partial compliance theory.

Regan argues that there are two properties which a utilitarian theory should have, but which cannot be combined in a single theory. In particular, no version of act- or rule-utilitarianism can combine the properties, as each of these, contrary to the extensional equivalence thesis of Lyons, possesses one, and only one, of the desirable properties. Only act-utilitarianism, according to Regan, has PropAU,¹⁷ the property of issuing to each separate agent the instruction to act so as to produce, in the circumstances, the best consequences she can in that situation. Only rule-utilitarianism, on the other hand, has PropCOP,¹⁸ the property of issuing to all agents the instruction collectively to produce by their actions taken together the best consequences achievable by any pattern of action. Because they possess these exclusive properties, Regan argues, act- and rule-utilitarianism are mutually non-entailing. He offers his own cooperative utilitarian theory as an alternative which, it is claimed, comes as close as any theory can to combining PropAU with PropCOP.

Regan uses a matrix of the form shown below¹⁹ in arguing for the claim that act-utilitarianism lack PropCOP.

		Agent 2	
		push	not-push
Agent 1	push	x	z
	not-push	z	y

(x>y>z)

Regan's argument is that if for some reason either of the agents decides to not-push, act-utilitarianism will require the other to follow suit and not-push also, since this is the only way to achieve y, the best consequences available to the latter agent in these circumstances; but rule-utilitarianism requires each agent to push in a situation such as that modelled above, since this pattern of action is the only one which achieves the

greatest benefit attainable by collective action, namely x. Hence, Regan thinks, the two forms of utilitarian theory produce conflicting prescriptions for action in this case, since rule-utilitarianism tells both agents to push, but act-utilitarianism will require its followers in some situations to not-push. So the two theories cannot be equivalent, as Lyons argues,²⁰ since a theory with PropAU will lack PropCOP.

In arguing for the failure of the reverse entailment, Regan introduces the notion of a universal prescription for action (UPA),²¹ which, on his definition, is a feature of rule-utilitarian theories. A UPA is in effect a list of actions assigned to agents: if each agent performs the action assigned under the UPA, the best overall consequences will be achieved. Thus in the above matrix the UPA would assign to both agents the action 'push', though there is no presumption, in general, that the actions assigned to different agents under the UPA must be the same under some non-trivial description.

Regan argues that a rule-utilitarian theory, incorporating a UPA, lacks the flexibility to deal with the possibility of non-compliance by some agents, without collapsing back into act-utilitarianism and so, on the previous argument, lacking PropCOP. If, in the above matrix, either agent should fail to perform the UPA-prescribed action of pushing, the other agent is still required to push even though this action no longer realizes the best consequences in these altered circumstances. Regan argues that the UPA cannot be modified to allow for conditional prescriptions without becoming act-utilitarianism²² - for this is just what the latter is, a set of instructions to all agents prescribing actions contingent on how all other agents behave.

From here Regan develops his own theory of cooperative utilitarianism. The essential idea is that the theory is adaptable,²³ which means that it takes into account the possibility of non-compliance by some agents. Agents act in accordance with cooperative utilitarianism if they combine to produce the best consequences they can, given the behaviour of other agents not following the theory. This might be thought to be circular, given the mention of the theory itself in the definition, but as long as there are univocal utility measures for each possible permutation of actions, it is theoretically possible to define a best pattern of behaviour for any subset of

agents relative to the actions of agents in its complementary set.

Regan's account of the theory emphasizes that the agents who follow it do so not merely by chance, but conscious design. It is important that agents see themselves as cooperators, and recognize reciprocal intentions to cooperate in other agents. This means that agents can fail to act in accordance with the theory even when they produce, adventitiously, the best collective outcome open to them. One possibility this raises is that action could be concerted through institutional means, for example, in order to secure the best available consequences.

Regan is not explicit about how this might take place. But in practice there would be room for provisions of this kind as a means of achieving the best consequences by, for example, disseminating relevant information to cooperators, formulating programmes of action over the longer term, issuing propaganda, and so on. Regan's version of partial-compliance utilitarianism would allow for some form of party organization within a non- or imperfectly utilitarian society, as a means of concerting action to achieve the best available consequences. Since it does not begin from the premiss that the utilitarian need only be concerned with what is rational ex ante, it is, on the definitions already given, a non-ideal theory.

Regan's arguments against the extensional equivalence thesis²⁴ are, however, unconvincing, and as a result his own brand of cooperative utilitarianism lacks the clear-cut advantages over traditional theory which he claims for it. The problem with the arguments using matrices like that above is that they address themselves to different levels of information when applied to act- and rule-utilitarianism. If it really is true, in the matrix, that one of the agents has, for some reason, not-pushed, then it is undeniable that act-utilitarianism will require the other agent to not-push also, despite the fact that before the first agent acted, both were required under rule-utilitarianism to push. But the fact that one agent has failed to do what rule-utilitarianism prescribes cannot be used to show its non-equivalence to act-utilitarianism where the latter is able to adapt to this new information. Unless the theories produce different prescriptions in identical circumstances, nothing can be shown about their non-equivalence. If rule-utilitarianism were

applied to the situation where one of the agents had already not-pushed, it too would require the other to follow suit.

Regan's arguments to prove the impossibility of building conditional prescriptions into the UPAs of rule-utilitarianism are perhaps designed to head off objections of this kind. But it is hard to see why it should be thought impossible for such conditionals to feature in the rule-utilitarian apparatus. It looks as if act-utilitarianism issues, before anyone has acted, a form of UPA to its followers - the same one as rule-utilitarianism; and that once some agents have acted, act-utilitarianism there is just a unique pattern of action by the remaining agents which maximizes utility (or a set of equal best such patterns). This is not a conclusion which Regan is well-placed to resist, given his own exposition of the cooperative theory. For, as was noted above, the latter relies on the assumption that adaptability will enable agents to achieve the best outcomes available to them in the face of non-utilitarian behaviour by others.

Only if it is possible to define sets of actions of some agents which are optimal relative to a given pattern of behaviour by the remaining agents will the notion of adaptability make sense. But these are just the conditional prescriptions for action which we envisaged being added to the UPA of rule-utilitarianism. On closer inspection, it turns out not that Regan has identified rule-, act- and his own cooperative utilitarianism as separate theories, but that they are all one. They all rely, with no more than verbal differences, on relativizing best consequences of actions by some agents to given patterns of action by the rest.

If this is right, there are fewer options for the utilitarian than Regan claims to have identified. This need not mean that the partial-compliance brand of utilitarianism must be seen as the only authentic version of the theory, though if the arguments above are right, it is undeniable that this is of considerable importance in the utilitarian approach to rational agency. There are however forms of utilitarian theory, such as utilitarian generalization, which feature no such property of adaptability or conditional prescriptions. On this showing, the main difference between ideal and non-ideal theory would consist not in radically different theoretical accounts of how to morally justify,

but in different levels of informational input, concerning in particular the actions or intended actions of other agents. It would not be a matter of fundamental incompatibility between competing forms of utilitarianism, but merely of there being, unsurprisingly, different patterns of optimizing action for a given agent or set of agents, depending on the available information.

Much confusion in writings by utilitarians or about utilitarianism arises from a failure to appreciate fully the distinction between optimizing relative to a given set of actions, and, on the other hand, a global view which takes all agents' actions as open to modification. Regan is of course aware of this distinction in the abstract, but blurs it in discussing the supposed difference between act- and rule-utilitarianism. If the important issue is what levels of information, corresponding to given states of action, are taken as the starting-point of the calculation, there is no reason to think that there is any purely theoretical difference between forms of utilitarianism focussing on global calculation and those dealing with partial compliance.

Hare's treatment is marked, as was argued in the previous section, by an ambiguity over the scope of the intuitive-level principles, and their utilitarian justification. This led to uncertainty about how much of what, on the basis of the model presented at the start, we recognize as political structures, could be fitted in with Hare's theory. It is now apparent that a similar ambiguity is latent within Regan's theory, and indeed within utilitarian theory generally. The unclarity in Hare's theory results not from some failing peculiar to him, but from a more general uncertainty in utilitarianism about what to include in the calculation as raw material, and what to treat as potentially subject to rational correction through utilitarian analysis. While on the one hand, it is problematic what, on the global view, should go into the initial calculation, it is also unclear why a consistent utilitarian should refrain from modifying the human institutions and practices which enter into ex post calculations. The political structures which provide a starting-point for the calculations of ideal theory are themselves liable to criticism from within that theory.

These considerations further reinforce the conclusion that there is no distinctively utilitarian theory of political

structures and their role in bringing about or facilitating what, from the utilitarian viewpoint, is rational action. In a theory such as Regan's, which sets out to determine how utilitarians should respond to the prospect of non-utilitarian behaviour by others, there appears to be little account of what the former rationally should do, if anything, apart from devise strategies for adapting to the latter's non-compliance. That is, if a theory like Regan's has anything to say about political structures, even by implication, it seems to be of a purely instrumental nature: to the extent that Regan is prepared to allow for political structures at all, the justification will be that tactical acceptance of such structures is warranted in the name of strategic utility-maximization.

If this is indeed the form of argument which a Regan-style utilitarian advances, it at least has the merit of being clearer about its commitments and their justification than Hare is; but the underlying problem remains that, since there is no independent reason for the existence of political structures, the argument is open to the objection that an ideal state of the world would dispense entirely with these structures. This section has argued that Regan, in his appraisal of rule-utilitarianism, does not escape the force of this objection. Only if there is a well-defined understanding of ex ante utility can it even appear that act- and rule-utilitarianism come into conflict.

Thus similar problems afflict Regan's theory, in its attempts to account for political structures, to those faced by the superficially very different theory of Hare. Just as Regan is left with a view of politics as wholly instrumental, so the equivocations in Hare's position conceal what is either a fundamentally non-utilitarian moral outlook, or else one which authentically utilitarian, but which can no longer provide any independent justification for the existence of the political structures. Because, as has been argued, there are no real points of difference in theoretical terms, as opposed to informational levels and starting-points for decision-making, between the forms of utilitarian theory which we have considered, this pull towards the ideal is bound to assert itself. The next section turns to examine a utilitarian theory which responds to this pull by aiming to provide an account of agency from the ideal perspective.

It should be plain from reading Parfit's book Reasons and Persons²⁵ in what sense his work can be regarded as a contribution towards ideal theory. Parfit uses utilitarianism as a critical tool in his book, as a means of probing and, as he hopes, undermining the assumptions about rationality and personal identity which he sees as motivating non-utilitarian attitudes in everyday moral thinking; he is, as he states at the start of the book, a revisionist philosopher rather than one who is content to analyse and explain already-current beliefs.

This approach raises questions about Parfit's methodology. His usual mode of argument is to ask how intuitively we would respond, presumably equipped with current 'common sense'²⁶ moral beliefs, to the numerous hypothetical examples he devises. If we are not to accept the pre-existing moral beliefs or attitudes at face value, the question arises why these beliefs should be uncritically invoked in response to Parfit's hypothetical examples. These examples may show that our current beliefs cannot be used as a basis from which to extrapolate general principles, since they prove contradictory when any attempt of this sort is made; but that does not show that we were really utilitarians all along, or that the pre-existing beliefs can be selectively used as a basis for utilitarianism as Parfit's methods appear to suppose.

Parfit might respond that this approach is a necessary first step in his broader strategy to expose the inadequacy of 'common sense' morality. If this morality proves to be incoherent, Parfit might argue, there is good reason to abandon it in favour of some alternative theory, such as utilitarianism, which is free of inconsistency. But, whether or not 'common sense' morality, as Parfit interprets it, does contain these inconsistencies, it is extremely doubtful whether his litany of practical examples, and our supposed unreflective responses to them, can establish that this morality is inadequate compared to its utilitarian rival.

As with Rawls' editing of non-philosophical moral beliefs in reflective equilibrium,²⁷ there must always be more than one method of resolving (what may be only prima facie) conflicts in moral reasoning. In Rawls' case the aim is to achieve some rapprochement between first-order moral beliefs and their

theoretical articulations. But Parfit's theory, and the philosophical method he uses to argue for it, leaves - or perhaps exploits - an ambiguity between relying on the 'common sense' beliefs, and showing their incoherence as the prelude to replacing them with something radically different. Utilitarianism may do the latter: Parfit certainly thinks it results in dramatic shifts of (among other things) political priorities. But the wider the gap between the utilitarian destination and the present starting-point, the harder it is to see how current moral attitudes can propel us across it.

In this respect it is instructive to compare Parfit's approach with that of Hare. The difficulties raised by Parfit's style of reasoning interestingly parallel those of Hare discussed in Section 3. There it was noted that Hare has problems in justifying the intuitive-level moral judgments from a utilitarian standpoint. Similarly, Parfit's methods leave us with the question why, on his revisionist brand of utilitarianism, there should be any part for the uncorrected moral attitudes to play, given that much of the argument of his book is aimed at overturning these attitudes. Nor is this similarity accidental, if, as claimed in the last section, the ideal/ non-ideal theory distinction turns less upon theoretically contentious issues than on the stage at which utilitarian analysis is brought to bear. The problem remains for Parfit how to specify, from the standpoint of ideal theory, the content of any moral beliefs purged of utilitarian assumptions. In the same way, Hare's ambiguity over the nature of the intuitive-level habits of judgment could be explained by the critical level's threatening to subsume all moral judgments, while, at this level, it was hard to decide what determined their content.

What really distinguishes Parfit from Hare, then, is at most a difference over how much ordinary moral thinking can be brought uncorrected, or with minimal correction, into the utilitarian framework. Like Hare, Parfit is unforthcoming on the political implications of his work, but it is at least clear that he thinks its consequences for public policy would be fairly radical. For example, Parfit supposes that accepting his philosophical arguments on moral rationality and personal identity entails a commitment to promote demographic and energy policies quite different from those in force anywhere²⁸ - not least because

Parfit's conclusions require a much greater degree of central planning in these policy areas than at present, there being currently no global strategy for determining the use of scarce resources like energy, or for deciding population levels.

It is not possible to discuss at any length the arguments which lead Parfit to these conclusions. Happily, however, the details of Parfit's arguments about rationality and personal identity are tangential to our present concerns, which relate to the political implications of different styles of utilitarian theory. Whereas Hare's broadly conservative attitude towards current moral beliefs suggests that his two-level theory can accommodate recognizable political structures virtually intact, it is far less clear that Parfit's more critical stance is as undisruptive of the existing political order. This becomes apparent when it is asked how we get from the present political arrangements which in Parfit's view results in the pursuit of irrational, because non-utilitarian, policies, to an ordering of the world where the policies he favours are implemented.

Parfit has nothing to say on this in his book, a surprising omission in view of his emphasis on hard-headed practicality. It is not, after all, as if Parfit is not a consistent or thorough enough utilitarian to appreciate that desirable end-states must have offset against their utility the costs incurred in realizing them, or, more generally, that part of the assessment of overall outcomes concerns whatever is involved in securing the final goal - a consideration which may itself be sufficient to alter the goal itself. This is another indication, presumably, that Parfit is operating from the vantage-point of ideal theory, where it is assumed that human institutions, political processes, and so on, are malleable enough to permit the implementation of whatever the theory itself requires. But there is no reason to think even if this is so that the ideal theory can merely absolve itself of all responsibility for explaining how political structures are to be adapted so as to realize the final goals of the theory.

Unless Parfit hopes that world politicians will be swayed by the arguments of Reasons and Persons, therefore, there has to be some account of how the present non-utilitarian state of politics can be brought into line with Parfit's policy prescriptions. It may be that Parfit believes this to be theoretically unproblematic: that there has to be some account of how

politically the policies would be implemented, and the associated costs included in the overall calculation, but that no matters of substance depend on the nature of the political mechanisms by which this is to be achieved. Even from the standpoint of ideal theory, however, there has to be some account of how the favoured policy is not merely implemented, but also of how utilitarians can reach a position where they can implement it. There must, in other words, be some strategy for gaining power in a world where politics is seldom, and never systematically, run on utilitarian lines.

It is doubtful whether ideal theory can maintain the division, which we have conjectured to lie behind Parfit's failure to deal with these issues, between policy and the political means by which it is to be implemented. It is not very realistic to suppose that strategies for gaining power, once put into operation, are likely to leave policy wholly unmodified, not only because these strategies, as already mentioned, themselves incur costs in enforcement, but also because the process of political persuasion induces changes in policy, as a means of canvassing more widespread support. Given Parfit's starting-point, that common-sense morality is non- or imperfectly utilitarian, it is important to make good this gap in the theory. Sometimes utilitarian writers seem to assume that the very act of producing a theory is all that is needed to remedy this deficiency. But (as has been more extensively discussed in socialist writings) the theoretical work has to be supplemented by some programme for gaining power, which does not rely only, or even in some cases at all, on the strength of the purely theoretical arguments themselves. There has to be some explanation, not just of what it is rational to do, but how to deal with (what is seen as being) others' irrationality.²⁹ If this is not to demand some compromise in the content of the utilitarian directives to agents as to how they should act, it becomes pressing to formulate some explanation of how power is to be gained and held.

Another set of questions raised by this latter consideration of how utilitarians are to exercise power once having secured it. Again here there is the possibility of wielding power through existing political channels, by means of the usual legislative and executive institutions, and this is fairly intelligible on at least certain forms of non-ideal theory. But any ideal theory

which eschews the conventional avenues for reaching power is unlikely to rest content with the pre-existing political structures. For example, in democracies like those of western Europe there are usually institutionalized confrontational or non-utilitarian elements built into these structures which the ideal theorist is committed to dismissing on the grounds of their irrationality.

Parfit is, again, non-committal on the issue of what political structures would attend the ideal utilitarian society, where power is wielded exclusively by utilitarians (of Parfit's own stamp). This latter qualification is not without importance, as the differences between utilitarian theorists would influence the formation of policy even in a world in which utilitarianism was the dominant political doctrine. Even if the ideal-theory assumption is permitted, that the problems of gaining and retaining power can be ignored, it still remains to be shown how the disputes which rage between utilitarians in present-day academic discussion will disappear or be of no political account in the ideal-world conditions where utilitarians are entrenched in power.

The supposition of much ideal theory such as Parfit's is that those in power will not only be utilitarians, but of his preferred variety. If this is dismissed as grossly implausible, we are left with the prospect, even where utilitarians are in power, of familiar disputes over the policy implications of shared beliefs, and the likelihood that the realities of political power will modify the antecedently rational policy. Once the policy-altering nature of power and political processes in general are taken into account, this pure ideal perspective merely vanishes, and we are pushed back towards the accommodations which the non-ideal theorist has to make, both in allowing for the possibility of dispute between utilitarians, and also, more fundamentally, of having to cope with a world in which few people are utilitarians of any sort.

It is not hard to see why this dimension of indeterminacy should lie beneath ideal theorists' failure to deal with the phenomenon of political power. For utilitarianism's appeal as an action-guiding doctrine is that it is both comprehensive, in applying in principle to all circumstances in which a decision has to be made, and optimific, in specifying a unique best course

of action (or set of equal best actions) as that which it is rational to undertake. The comprehensiveness stipulation demands that the value or values in terms of which the rational calculation is to be conducted are projectible onto all situations where practical deliberation is called for: the optimific condition, that there is at most one value into which all others are translatable in order to decide what it is best to do.

Once these conditions are accepted, there is no room for the idea that values decisive in some areas of practical deliberation may fail to apply to others at all, or that where a value or values do apply, it may not be possible to resolve finally what practical bearing they have on action, in the sense that it may be unclear what actions count, and why, as enacting the values in question. It is clear that practical indeterminacy of this form underlies much political dispute, especially such as occur in political cultures where certain values, such as liberty, are accepted by both parties to the conflict; but this is not the sort of dispute which the ideal utilitarian can countenance, even though internal disputes between utilitarians seem largely to be explainable by reference to these aspects of value.

For these reasons it is highly doubtful whether the telescoping of value required for the utilitarian calculus to retain its appeal is really possible. One of the ways this becomes apparent is in the difficulty of assigning concrete utility numbers to presented courses of action. Some theorists, like Regan, assume the problem away by defining outcomes precisely in terms of utility numbers; while this is an acceptable means of approaching some theoretical issues, it gives no clue as to how practical outcomes are to be assessed in utilitarian terms. Others, like Parfit, assume in examples that outcomes are straightforwardly reducible to numbers (e.g. of beneficiaries of the competing actions).

But even this presupposes a ceteris paribus condition, as otherwise there would be no ground for assuming as Parfit does that the number of beneficiaries is the criterion of assessment. Whatever the Benthamic maxim that each is to count for one and none for more than one may mean, it cannot for the utilitarian just be that a given benefit has the same utility measure regardless of who receives it. But if the above claims about the

nature of value are right, there may be no clear-cut quantification of the sort the utilitarian seeks. As a result, the assumption of ideal theory that pure calculation can resolve political disputes, is at best highly questionable.

This emphasis on calculation as the method of determining policy must exclude the endemic conflict witnessed in the real world of politics. This absence of conflict means that the ideal utilitarian ordering of the world lacks at least one important feature of political structures as defined at the start. What seems to emerge from Parfit's work and that of other ideal theorists is a view of the world where politics has in its essentials been superseded by administration - where the broad distinction between politics and administration, is that the former is concerned with policy, the latter with its implementation. It is far from clear that these can be kept completely distinct, even in thought, as it would be odd if no considerations of practicality entered into discussion of which policy to pursue. An important political consideration is whether a proposed measure can, given the world as it is, be made to work.

It is perhaps this vagueness in the demarcation between the spheres of politics and administration which a utilitarian account of public decision-making can exploit, by reducing, in effect, all questions of policy to those of practical enforcement. But, in spite of the unclarity of the distinction, it is plain that the content of real-world politics is not exhausted by the purely practical issues of implementation. As this section has argued, the balance is largely taken up by controversy over the practical import of values, which utilitarianism is notoriously bad at reproducing in its version of public decision-making. It is not a logical consequence of holding that there is some unitary value in terms of which all values can be expressed, that assessing outcomes in terms of the fundamental value is neutral: there could be two or more (mutually untranslatable) such values, and if so, the fact that each of these could separately express the other values in its own terms would not necessarily yield any means of choosing between them.

Alternatively, it might be held that a sharp distinction exists between facts and values, and value-neutrality can only be

maintained on the factual side of the divide. But utilitarians usually do assert that the calculus is neutral, because they think that utility provides a unique standard of value-reduction, and because any fact/ value distinction in utilitarianism is likely to leave nothing on the evaluative side. If value is allowed to retain the features distinguished above in the discussion of its role in political conflict - if, in other words, it corresponds to our idea of value at all - it will not give the utilitarian what is desired from the calculative procedure.

Sidgwick's writings,³⁰ to which Parfit acknowledges his debt, give an interesting illustration of the reduction of politics to administration. Sidgwick occupies in some respects an intermediate position between ideal and non-ideal theory, since he expressly aims to cater for the existence of non-utilitarian agents, while remaining committed to the view that society can be, and is best, run only by utilitarians. A result of these commitments is that Sidgwick comes up with an ideal society ruled in accordance with utilitarian doctrine, but in which this fact is concealed from those, presumably the majority, who would, if they knew about this, would prefer it not to be.

It is an implication of this position that agents' perception of their interests may be mistaken. Sidgwick's ideal society is one in which no real political conflict can occur, because the structure of the society is such that conflicts over the weights to be assigned to preferences, which for the ideal utilitarian is the sole variety of political conflict, is filtered out in advance. Once the conclusion has been drawn that there is at most one ultimate value, the only political controversy can be over how that value guides action or policy, and the major source of indeterminacy in the particular value which utilitarians favour, whether people's preferences are to be taken at face value, is deliberately excluded from (public) discussion.

Against this it might be objected that even in a society as tightly controlled by utilitarians as Sidgwick's ideal, there could be policy disputes of a kind, just as, for example, there could be disputes over policy within a socialist society. If there is nothing unintelligible in the idea of conflict between, say, the supporters and opponents of a policy of rapid industrialization in some backward socialist economy, then it is

equally possible that a Sidgwickian society could witness conflicts within the ruling group in deciding which policies maximized utility. All that is absent from either of these situations, it may be said, is the deep value-based ideological conflicts familiar in many western democracies; but in the account given in Section 2 above, ideology was not seen as an essential component of political structures.

These considerations are, however, insufficient to show that a Sidgwickian society would have an authentic political culture, and ignore disanalogies between such a society and its socialist counterparts, which arise from the peculiar character of utilitarianism as a theory of action. While it is true that socialist cultures may espouse one supreme value, equality, and utilitarians another, utility, both of which may allow differences of practical interpretation, it is one of the advantages supposedly enjoyed by utilitarianism that it provides clear answers to questions of policy and individual action. If it is admitted that the notion of utility is ambiguous, it becomes much less easy to defend, or even understand, the ostensible rationale of the Sidgwickian dictatorship of utilitarians.

Only if the latter are enlightened as well as despotic is there any ground for keeping everyone else ignorant of how society is being run; if utility proves to be practically ambiguous, it is less likely both that the rulers will have the monopoly of even utilitarian wisdom which Sidgwick assumes, and that there will be an unchallengeable utilitarian mandate for keeping the truth hidden from society at large. Sidgwick's grounds for this form of society were themselves, consistently enough, utilitarian, and if utility is an ambiguous value, that must cast doubt on the governing arrangements themselves.

Here we return once again to the point argued throughout this chapter, that the more consistent a theory is in its utilitarianism, the less capable it is of incorporating the sort of political structures outlined at the start. This section has suggested that a Sidgwickian despotism could only retain its original justification if discussions of policy can be kept free of the conflicts of value which characterize real-life political disputes. If the central feature of utilitarianism is its optimistic decision-procedure, any authentically utilitarian theory must eliminate value-based conflict, and this requires in

effect the end of politics, or at least of one of its most fundamental aspects. The converse point also holds, that any utilitarian theory which accepts the possibility of conflict will be that much better equipped to understand political structures, but at the cost of no longer being a utilitarian theory in the full sense. This claim was argued for in Section 4, where ambiguities in the base-line from which decisions were taken were explained, in Hare's case, by the vagueness in the distinction of levels, itself the product of doubt over the utilitarian justification for the intuitive principles; and in Regan's case, by the pull towards a version of ideal theory in which the need to tailor utilitarian strategies to fit others' non-utilitarianism no longer exists.

Essentially the problem is utilitarianism's open-endedness as a theory of action. This also explains, as already mentioned, the difficulty in maintaining any hard-and-fast distinction between ideal and non-ideal versions of the theory. While on the one hand its aggregative procedure demands some quanta of value if the calculation is to be possible at all, it is, on the other hand, difficult to see what non-arbitrary choices can be made as to what form this value should take - a problem which arises not only with different interpretations of utility itself, but also in the practical significance of these interpretations.

VI

This open-endedness means that there can, for the utilitarian, be no specifically political motives, virtues, or traits of character.³¹ This is not true only of the category of the political in relation to utilitarianism: it is equally true that the doctrine can make no sense, for example, of the correlative moral notions, because utilitarianism sets out a quite general algorithm for decision-making, applicable to all areas of life. Having once ascended to this level of generality, however, it becomes hard to see what is to be put into the calculus. In this, though not in its purporting to provide a decision-procedure for all cases of practical deliberation, utilitarianism resembles Kantian moral theory, which is likewise unable, within the category of the moral, to identify particular contents for actual categorical imperatives, an unsurprising result in view of Kant's

emphasis on the strictly sui generis character of the moral realm. Similarly utilitarianism runs into problems of content once it has been decided that all practical considerations are projectible onto at most one category of value. The result is an empty schematism, which because of its limited perspective cannot avoid arbitrariness in its prescriptions for action.

This schematism has had its parallels in the real world of politics also. In particular, the aggregative and instrumental presuppositions of utilitarianism have found their echo in a commonly-held view of politics as concerned with utility-maximization (both in public expectations and in politicians' rhetoric). Politics is presented as an area of human activity in which disagreements at most concern strategic planning - over how best to realize the benefits which, it is assumed, everyone wishes above all to secure. It is especially striking that the rhetoric of utility-maximization should be so dominant when the political disputes in which it is deployed so often display the value-pluralism which falsifies it.

This tacit utilitarianism has its counterpart in a Weberian style of bureaucratic management, a natural concomitant of the managerial idiom of contemporary politics.³² This should not be surprising given the instrumental assumptions informing much political debate: within this order of assumptions the bureaucrat plays a logistical role, providing the administrative means to promote the ends of policy. In this respect, bureaucracy is the neutral place-filler called for by utilitarian theory, adapting itself to whatever policy objectives are placed before it. But (as will be argued in the following chapters) this neutrality is illusory, and the rhetoric of efficiency and control is unavoidably the expression of a particular ideology.

Like Kantianism in its recent applications to political theory, utilitarianism sets out to deal with the fact of value-pluralism, a recurrent political problem, at least in western democracies. Kantianism tries to handle pluralism by setting out a transcendent category of value, which is capable of overriding or at least mediating between other values when they conflict. Utilitarianism, on the other hand, aims to make conflict appear illusory, by showing that such disputes can be resolved by invoking a single criterion of value.

Both of these approaches, as attempts to deal with pluralism,

eliminate the political as the forum within which conflicts of values occur. A constant problem for liberalism is how to keep a grip on this aspect of the political, while preserving the pluralism which liberalism, whether with enthusiasm or resignation, takes to be a fact of life. It is readily understandable, therefore, that both the Kantian and the utilitarian rejections of politics should have enjoyed support within liberal theory, as a means of preserving value intact, one way or another.

But it is highly doubtful whether either the utilitarian or Kantian approaches do succeed in this, and more doubtful still whether the conflicts of values which the liberal regards as inevitable can be artificially sealed off as part of the private realm, while leaving any recognizable content to politics. Conflict is ineliminable from politics, and it is naive to think that the values about which the liberal wants to be tolerant or pluralistic can be kept out of the political arena. That is why utilitarianism, if it is to keep faith with its utilitarian origins, cannot afford a notion of politics.

Notes

1. In the text, the basic distinction assumed between act- and rule-utilitarianism is the latter's endorsement of rules of conduct which may, on occasion, result in individual actions which are disutile. Numerous questions arise, within this broad distinction, about, for example, the identity and individuation conditions of actions, and what classes of consequences have to be taken into account for the utilitarian calculus. Given the text's acceptance of the extensional equivalence thesis, it is not pressing to pursue these issues here. All that matters is that, in the case of any ostensible such act (such as the case where a computer mailing system occasionally sends out a bill to a customer for a sum below the unit cost of sending the bill itself) the action is only identifiable within the context of a system whose overall operation is or is not itself utile. Given univocal utility measures, and adequate information, there is no reason to think that there is not a single answer to whether or not a given action is utile. Utilitarian generalization is taken

to be the issuing of identical prescriptions (*mutatis mutandis*) to all identically situated agents, regardless of whether this embodies a general rule (which need make no such universalistic prescriptions) or of whether it leads each agent to perform individually utile actions. Here again, it is hard to see how there can be any utilitarian justification for such a system unless it yields outcomes which are utile.

2. The term 'partial compliance' is due to John Rawls: see his A Theory of Justice (Oxford 1971), pp8.

3. Sartre, in his classic existentialist phase, would claim that inaction is still action of a sort, even where it involves no conscious choices: see his Existentialism and Humanism (London 1948). It is doubtful whether mere inactivity manifests the intentionality without which human existence loses the name of action. But, in any case, the point still holds if we take it that realizing this state of inactivity brings about (or perpetuates) a state of affairs in the world.

4. The locus classicus for this position is Machiavelli in the Prince and Discourses, though, more contentiously, its roots may be traced back to Thucydides or Socrates (in his Platonic incarnation in the Republic). To claim this autonomy for politics is not of course to deny that an important part of politics - on many modern views, its totality - is the causal realization of public benefits in, for example, the areas of economic or social policy. What is denied is that this exhausts the content or value of politics as a human concern, a claim which will be taken in the text as a prime assertion of instrumentalism.

5. See e.g. J.B.Thompson, Studies in the Theory of Ideology (Cambridge 1984).

6. Admittedly, there exists a certain ambiguity over the form which a utilitarian political ideology might take, corresponding to the duality discussed in the text between ideal and non-ideal theory. In other words, the non-ideal theorist might appropriate some external ideology (such as socialism), claiming that given the way the world is this is the best method of promoting utility

- even though in conceivable counterfactual circumstances some other ideology, or none, might better serve this end. Although utilitarian assumptions and rhetoric pervade modern political discourse, this option has not been widely canvassed in theoretical writings by utilitarians. The alternative is to regard utilitarianism itself as a political ideology (as presumably Parfit, for example, does), enjoying similar status to, but in competition with, socialism and other political ideologies. This alternative makes it all the more pressing to answer questions of organization and practical implementation.

7. Richard Hare, Moral Thinking (Oxford 1982).

8. But see also Hare's article 'Political Obligation', in T.Honderich (ed.), Social Ends and Political Means (London 1976).

9. See Hare, op. cit. p25f.

10. Parfit is one such, arguing that the 'common sense' judgments presumably delivered by these habits of judgment lack utilitarian justification (and indeed logical coherence). See Section 5 below.

11. See note 7 above.

12. Hare seems to believe that the same metaethical formalism can encompass widely differing moral theories. He assumes in Moral Thinking that the universal prescriptivism for which he argued as a Kantian in The Language of Morals (Oxford 1963) still holds good, despite his conversion from Kantianism to utilitarianism.

13. Hare, op. cit. p9.

14. On this point see Bernard Williams, 'Persons, Character, and Morality', repr. in his Moral Luck (Cambridge 1981), esp. Section III. Williams' target there is mainly Kantianism, but it is striking (as Williams himself would claim) that the central thrust of his criticism applies equally to utilitarianism - the criticism being, broadly, that Kantianism is incapable of handling the structuring of agents' moral life by personal

relationships, and imposes an impersonal universalistic structure of moral obligations on agents. This opens up a gap, as Williams argues, between morality, construed along Kantian lines, as a set of rational prescriptions for action, and agents' subjective reasons for acting as they do. Similarly, utilitarianism only allows these elements to structure the moral life in so far as they emerge anyway from the felicific calculus. In Sidgwick's phrase, the position of the moral agent is not that of an embodied subject, but 'the point of view of the universe'.

15. See Williams, ibid., p18

16. Donald Regan, Utilitarianism and Cooperation (Oxford 1980).

17. For Regan's definition of PropAU, see ibid., pp3-4.

18. See ibid., pp4-5.

19. See, e.g. ibid., p18, p85.

20. David Lyons, Forms and Limits of Utilitarianism (Oxford 1965).

21. See Regan, op. cit., p85.

22. For this argument see ibid., pp87-90.

23. See ibid., p6.

24. Ibid., p6.

25. Derek Parfit, Reasons and Persons (Oxford 1984).

26. Parfit introduces this term (ibid., p99) without any prior definition. As a result, it is a matter for speculation what application it has in the real world - whether, for example, the moral judgments imputed to it by Parfit are indeed the deliverance of common sense. Still more questionable is Parfit's assumption (again unargued) that this morality has objectives, as is implied by his claim that it is in certain situations 'self-

defeating'. This would of course be possible if (as it comes naturally to utilitarians to assume) morality is there to secure external goals (this is implicit in Parfit's parents and children example, ibid. p96-9). This may be true if the moral goals is defined as, for example, the prevention of harms to one's children. But morality is not merely about securing certain states of affairs - or at least not unarguably so. It may also be about given agents' acting so as to secure them. The moral life of the parent involves not merely the securing of benefits for his or her child - for example by hiring a nanny for the purpose - but personally acting so as to secure those benefits. Here again, we find that personal relationships structure moral life in ways unaccounted for by the impersonal aggregation of utilitarianism.

27. Rawls introduces this term in A Theory of Justice, p20f.

28. See Parfit op. cit., Part IV.

29. While the rudiments of a utilitarian non-compliance seem to have been considered by, for example, Henry Sidgwick, modern-day utilitarians - Regan being a notable exception - have either pursued the corrective brand of theory propounded by writers like Parfit, or sought, as does Hare, to produce utilitarian justifications for established patterns of moral reasoning.

30. See Sidgwick's The Methods of Ethics (London 1907).

31. See Bernard Williams, 'Politics and Moral Character', repr. in Moral Luck.

32. As Alasdair MacIntyre has clearly pointed out: see his After Virtue (London 1981), pp26-7.

Chapter Two

Kant, Moral Agency and Politics

I

Recent decades have witnessed a resurgence of interest among Anglo-American political philosophers in Kantian moral theory. Despite their many and obvious differences, the work of Rawls, Dworkin, Gewirth, Nozick and others¹ owes a common debt to Kant, whose most striking manifestation is these writers' shared concern with the moral subject - with, that is, the idea that persons are bearers of special moral claims which cannot merely be negotiated away or used as bargaining chips in 'the calculus of social interests'. Given liberal theory's concern with the moral claims of the individual, this is not surprising, as it is Kantian moral theory which above all exalts the moral subject both as agent and, in the Kingdom of Ends, as the object of moral action. Thus Kant's injunction, in one formulation of the Categorical Imperative, to treat other moral agents never merely as means but as ends in themselves,² has been taken up by liberal theorists as the basis on which to support claims about persons as autonomous or bearers of rights against, for example, aggregative utilitarianism or some other form of welfarism.

While it is not very hard to see why neo-Kantianism should have come into vogue among liberal philosophers, it is less clear that Kant's moral theory can merely be accommodated within liberalism without disrupting some of its other central concerns. One major set of questions relates to the liberal's belief in neutrality, that public institutions and policy should as far as possible remain impartial between different conceptions of the good. Only if Kantian theory can be shown to be neutral in this sense can it enjoy the foundational status within liberalism which philosophers like Rawls have claimed for it; otherwise, the liberal appears to have little defence against the objection that he is advancing a substantive theory of the good which, in violation of neutrality, is preferred to other such theories. We therefore need some argument to show either that Kantianism makes no substantive commitments between conceptions of the good, or at

no substantive commitments between conceptions of the good, or at least that any commitments of this kind which it does make are practically agreed.³

In pursuing this issue, the best starting-point is in Kantian theory itself, and more specifically in its account of the moral subject. This seems the best place to start, as this is the place where (as was asserted above) the concerns of liberalism and Kantianism coincide. In attempting to clarify the relationship between the self and conceptions of value we need to establish both

(1) the identity conditions of the self, and

(2) how, as moral subjects or otherwise, the self as identified in (1) can be seen as bearing value.

The answers to these questions may be connected: it may be held that the conditions which must obtain for a self to continue through time, for example, coincide with the conditions required for a self to bear moral value. There is, on the other hand, no reason on the surface to think that this must be so, and some writers, such as Williams,⁴ give largely distinct answers to these questions. Nevertheless, in Kant's philosophy there is, as will be explained, a tight fit between the metaphysical construction of the subject and his account of moral agency - at least, on one interpretation of the former. There is, moreover, little doubt that this is what Kant himself intended.

This suggests that liberal theory cannot dispense entirely with metaphysical claims, contrary to, for example, Ackerman's assertion that liberalism is committed to no particular metaphysics or epistemology, nor to specific answers to '[b]ig questions of a highly controversial character'.⁵ This need not mean that value-neutrality must be violated. One line of argument, as has already been mentioned, is to invoke the allegedly consensual nature of the (public conception of the) subject, as in Rawls' interpretation of citizenship in terms of a 'moral personality' constructed from shared moral beliefs current in society.⁶ But, as should become apparent in what follows, the

Kantian basis of Rawls' theory is dangerously attenuated. In producing an account of the subject generalized enough to be consensually acceptable, Rawls risks leaving it unclear how the residual subject can bear any moral weight at all.

This is the burden of Sandel's critique of Rawls in his Liberalism and the Limits of Justice.⁷ Sandel argues that Rawls' Kantianism commits him to a 'radically disembodied' self which is too thin to support the strong moral claims, especially those concerning personal desert, required by Rawls' Difference Principle. The present chapter aims to locate these difficulties more precisely in Kantian metaphysics, and to suggest how ambiguities in Kant's conception of the subject of experience threatens to reduce to inconsistency those theories which rely on strongly Kantian assumptions, such as those referred to at the start.⁸ Sandel's approach perhaps understates these difficulties, not least because his own 'radically situated' alternative to the Kantian subject is, as will be argued below, implicit within Kant's metaphysics of experience, as a dubiously coherent counterpart to the noumenal subject.

Although the focus in this chapter will be on Kant himself, the arguments presented here are intended to be more widely applicable, to liberal theory in general. Thus the difficulties identified in Kant's notion of the moral subject are to be taken as symptomatic of a more serious malaise in liberalism, namely an incoherence in its understanding of the individual as the supreme repository of moral value. The argument amounts to a denial that there can be any value-neutral inference from answers to question (1) above, to answers to (2). There is no point from which a consensual notion of value embodied in persons can be used to resolve or mediate political conflict. The response to the miscarriage of Rawls' efforts to establish a consensual standard lies not however in a different answer to (1) as Sandel argues, but in questioning the priority enjoyed by the self in liberal thought. Once the force of this point is grasped, it becomes clearer that the value-neutral politics sought by liberal theorists is unavailable, if they are to retain their pluralistic commitments.

This section will consider, in a necessarily cursory form, Kant's analysis of the metaphysical subject in the Critique of Pure Reason. As has already been asserted, Kant envisages a close relationship between the answers to questions (1) and (2) - between, that is, the identity or individuation conditions of selves, and the grounds on which we ascribe value to selves. There are, however, reasons for thinking that the attempt to derive this value from answering (1) threatens Kant's account of individuation itself with incoherence. Kant argues in the Critique that

[man,] who knows all the rest of nature through the senses, knows himself also through pure apperception, and this, indeed, in acts and inner determinations which he cannot regard as impressions of the senses. He is thus to himself, on the one hand phenomenon, and on the other hand, in respect of certain faculties which cannot be ascribed to sensibility, a purely intelligible object. We entitle these faculties understanding and reason.⁹

It is Kant's general doctrine in the Critique that we cannot know things as they are in themselves, and that our knowledge of the world is confined to appearances or 'phenomena'. When this doctrine is applied to the human subject, it seems to yield the conclusion that the subject as it is in itself cannot be an object of knowledge: I cannot know myself as I really am. And this looks like a denial of the possibility of self-consciousness, at least if the latter is understood, plausibly, as knowledge by the subject of itself as subject. Yet Kant says in the passage that man is 'to himself...a purely intelligible object'. Is this a contradiction?

If not, we need some explanation of how the supersensible subject is differentiated from the rest of the non-phenomenal world as an object of human knowledge. Kant's claim is that man can know himself 'through apperception': in other words, by means of the faculties of understanding and reason 'which cannot be ascribed to sensibility' rather than the phenomenal route through which, according to Kant, we acquire our other knowledge of the world. It may be that Kant has in mind, as his

use of the phrase 'transcendental unity of apperception' suggests, an application of the general transcendental argument which he uses elsewhere in the Critique, of the following form:

(1*) Conditions C are required for there to be experience; but

(2*) there is experience; so

(3) conditions C are met.

Taken in this way, Kant's argument is that it is a condition of there being experience at all that there is something, namely a subject, which has that experience; but there is experience; so there must be a subject. The positing of the subject is a precondition of experience itself, because the idea that there could be experience which was not predicated of an experiencing subject is unintelligible. From this form of argument there is no warrant for inferring the existence of a self-conscious subject, since there is nothing in the above argument to show that only this kind of subject could have experience. There is no obvious self-contradiction in the notion of a continuing subject of experience, which however lacks knowledge of itself. If this is indeed the most the argument shows, Kant faces problems further on in the argument, when he comes to apply the metaphysical construction on the subject to his moral theory. We will return to this difficulty later.

Even before we reach this stage, however, Kant faces a question about the knowledge of the subject of experience available to us by means of the transcendental argument. According to the argument above, the knowledge we can have of the subject is not given by the normal route, that is, experientially, but rather inferentially, through 'understanding and reason'. This might seem to be the conclusion Kant is seeking, since on his standard distinctions it is by these faculties, rather than by direct experience, that we can acquire knowledge of things as they are in themselves: that is, the subject can gain knowledge of itself as it really is, or self-consciousness.

This however presupposes that intellectual self-knowledge can be identified with self-consciousness. It is debatable whether Kant favoured such a conclusion, given his remark at one point that 'although this thought [that there is a unitary subject] is not itself the consciousness of the synthesis of the representations, it presupposes the possibility of that synthesis'.¹⁰ Kant, then, appears to accept the force of the objection in the last paragraph, that the transcendental argument is insufficient to establish a subject's self-consciousness. But even apart from this textual point, it is not very easy to see how in principle one could unite the representations given in experience as those of a unitary subject, while retaining any coherent distinction between the subject and the world.

Any theory of the nature of the self requires some account of the subject's limits, otherwise there is nothing to tell us how subjects are differentiated from one another. But there appears on the present argument to be nothing within experience which could lead one to that distinction - all that is immediately presented is a manifold of sense-impressions and other sensations, with no indication of what in these representations is properly one's own. There is for example nothing to mark a particular physical sensation, such as a pain, as belonging to one subject rather than another. Hence the line between the subject and the world, from the standpoint of that subject, seems to be pushed back to coincide with that dividing the phenomenal world from that of the faculties.

But if this is so it is no longer clear in what sense we are justified in saying that the separate representations in experience belong to a single subject, rather than any number of different subjects - or at least as many as there are representations. Now it looks as if the transcendental argument relied on the assumption that the unity of the subject was already given in experience, and that it was this which made it senseless to think that there could be subjectless experience. Even if it is granted that there must be some subject to accompany each representation, this does not by itself show that these must be unified within a single centre of consciousness. It

is only if these representations have already been identified as mine that they can be colligated by a unitary subject.

These considerations might prompt us to try to draw the limits of the subject within experience. Most obviously this manoeuvre would involve some attempt to identify a distinguishing feature of certain types of experience which marked them as being of (that is, representing, as well as merely belonging to) a subject. While it is doubtful whether any such attempt corresponds to Kant's intentions, he seems at times to invite this sort of interpretation. It gains currency, for example, from considering the nature of the knowledge which the subject has of itself, as Strawson has noted,¹¹ and in particular the fact that while Kant treats the supersensible subject as existing outside time, it seems undeniable that any self-consciousness must both 'belong to the history of, and must be consciousness of some episode belonging to the history of, a being which has a history and hence is not a supersensible being'.¹²

Kant's problem here is that his construction of the supersensible world as timeless, and hence beyond causal determination (which on his account is necessarily temporal), leaves it unclear how the subject occupying this world is related to the empirical human subject - roughly, the human organism - in the phenomenal world. If we accept that self-consciousness is historically determinate, to the extent both that the appearance of the subject in self-consciousness and that this consciousness itself are part of the world of phenomena, it looks as though '[t]he reference to myself as I (supersensibly) am in myself drops out as superfluous and unjustified', so there is no reason to say that 'in empirical self-consciousness, I appear to myself as other than I really am'.¹³ On this reading, the supersensible subject just disappears from view, as incapable of explaining the historically conditioned nature of human subjects' awareness of themselves.

This is not to say that this alternative phenomenal view is necessarily successful in resolving the problem considered above, concerning the limits to be drawn within experience between the subject and the world; though abandoning the transcendental

argument for the unification of the subject would at least eliminate one source of difficulty in dealing with the problem, namely the argument's inability to explain how the subject, standing outside the world of experience, is distinguished within it. While the alternative approach has not been proven to resolve this problem satisfactorily, it does at least look more promising to take the subject as non-inferentially given within experience. One reason for favouring this approach is that it does not suffer from the same difficulties as the transcendental argument in explaining the fact of self-consciousness.

This survey has unavoidably been cursory, but it should have indicated some areas of difficulty in Kant's theory of the subject, and in particular suggested how his assumptions about the nature of experience lead to problems in his account of the supersensible subject. From this it might be thought that the inference from Kant's metaphysics to his theory of moral agency was less direct than Kant himself supposed - that the fit was less close between the theory of individuation or subject-identity given in answer to (1) above, and the account of value given in response to (2). As will be argued in the next section, one reason for this may be that no answer to (1) which implies any clear answer to (2) can maintain value-neutrality in its account of the human subject.

III

This section will not attempt any full-scale treatment of Kantian theory, any more than the previous one tried to deal comprehensively with Kantian metaphysics. The aim, more modestly, will be to trace the issues pursued so far to Kant's account of moral agency. While it may be true that his theory of the supersensible subject is shaped in part by Kant's conception of what it is to be a moral agent, this section will argue that the latter cannot avoid the metaphysical difficulties raised earlier. As will be explained, the root of the trouble lies in Kant's requirement that the moral agent should be both an object of the understanding as a thing knowable in itself, and when undertaking

moral reflection, a self-conscious author of actions in the world.

The moral agent, for Kant, is capable of rationality only in so far as he is metaphysically free, where to be free in this sense requires that the agent stands outside the determination by causes which characterizes the phenomenal realm. Because of the sharp line drawn by Kant between the phenomenal and the supersensible or noumenal worlds, it is not strictly accurate to think of him as a compatibilist about the question of free will and determinism. That would require Kant to regard the same set of facts (under at least some identity conditions) as admitting simultaneously of two different interpretations or analyses; but, as has been indicated already and will come out again in what follows, it is far from clear that we can identify, for these purposes, some single such set. Without some grasp of identity in this area, it is not evident that we have, as Kant perhaps supposed, a compatibilist theory rather than some form of dualism. Given Kant's assumptions, it is only on this basis that a moral agent can be free, 'for to be independent of determination by causes in the sensible world (and this is what reason must attribute to itself) is to be free'.¹⁴

Kant's absolute denial that empirical considerations can play any part in morality surfaces again in his view that the content of moral principles as rules of conduct must derive from the fact that the principles are self-addressed by agents as purely rational beings.

Empirical principles are always unfitted to serve as a ground for moral laws. The universality with which these laws should hold for all rational beings without exception - the unconditioned practical necessity which they impose - falls away if their basis is taken for the special constitution of human nature or from the accidental circumstances in which it is placed.¹⁵

The categorical nature of morality demands that moral agency be free of contingency. Since this means that empirical facts such as those of human nature can play no part in shaping the content of morality, it is not to human beings as such, but as (potentially) rational agents that moral imperatives are addressed. Only in this way, according to Kant, can morality

retain its autonomy. Thus, by contrast with some noncognitivist theories, such as emotivism, the Kantian account abstracts entirely from motivation in its explanation of our reasons for acting morally; and abstracts, equally, from the actual empirical consequences an action will produce, in assessing its moral character.

In view of Kant's uncompromising rejection of considerations like this, it is important to identify the basis of moral agency in Kantian theory. It is Kant's project to found morality on considerations of rational agency alone. Only in this way, as we have seen, can the categorical status of morality be guaranteed. When it comes to specifying the content of the categorical imperatives, however, which are held to be the sum of morality, it is difficult to see how this content can be generated from a basis as narrow as Kant wants. The more he cuts away from the possible basis of morality, the harder it is to know what content morality is left with in his theory.

It might be thought that Kant's description of the Kingdom of Ends gave a specification both of this content, and an explanation of how it can be produced from the considerations about rationality. Kant's idea is that the subject of the Kingdom are persons qua rational moral agents, and it is their status as ends in themselves which is held to determine other agents' moral actions. Thus the second formulation of the categorical imperative holds that every rational agent should '[a]ct in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end'.¹⁶ Whether a formulation as general and abstract as this can yield some of the rather concrete substantive moral positions Kant favours may be doubted.¹⁷ But even leaving aside the matter of how well this version of the categorical imperative squares with the first-order principles Kant himself endorsed, it is highly dubious, more generally, whether the subject of the Kingdom of Ends has been characterized fully enough to produce first-level principles of any sort.

Kant offers, through the formulation of the categorical imperative, a criterion by which the morality of actions may be assessed. In its most perspicuous form, this criterion is whether the action in question treats other rational agents as ends rather than as mere means. But, bearing the arguments of the previous section in mind, it is highly unclear what the nature of these rational agents is, and how, consequently, their presence in the Kingdom of Ends can have any relevance to action in the phenomenal world. In that section it was argued that on the first, transcendental interpretation of the subject, there was no obvious point of contact between the subject and its experiences of the phenomenal world - or at least not one of which the subject itself could be aware. In particular, it was unclear how a certain kind of experiential content, that of the subject itself, could be distinguished within experience.

But, if this is so, we are left with no clue as to how the self-consciousness required for Kant's highly reflective account of moral agency can be secured. What the subject needs, in applying the universalizability criterion to prospective actions, is an understanding of what, as an embodied agent, it involves for that subject to act in the world; but if, as was argued in the last section, the transcendental form of argument was not strong enough to provide an explanation of the subject's self-awareness, it appears that the degree of reflectiveness required for this criterion to be workable is unattainable. There is no path back, on this reading of Kant's argument, from the concrete phenomenal-world consequences which have to be taken into account in deciding whether an action is morally acceptable, to the position of the purely rational agent - for that is a position which no self-conscious being, in Kant's construction, could occupy.

The same point comes out in the reverse direction, when we consider the route from the supersensible world, where the moral agent is situated, to the world of phenomena. Even if we possessed (as this chapter has argued we do not) a clear conception of how the noumenon or supersensible subject can be self-conscious, and could apply this to moral agency, we are still left with the problem of explaining how decisions taken by

this subject, standing as it does wholly outside the world of phenomenal causation, can bring about events within that world.

Kant's rigorous separation of the noumenon as free moral agent from the world of causes leaves it unintelligible how that agent could do anything to alter the determinism of the phenomenal realm - the more so when the agent is conceived of as lying outside time, in which, Kant held, causal determination can alone occur. On Kant's construction of the noumenon/phenomenon distinction there cannot be any univocal identity conditions linking particulars in the phenomenal world with particulars in the supersensible world; as a result, we cannot identify temporally specific phenomena as the causal products of, or indeed as bearing any special relationship to, the deliberations of the moral agent.

The last few paragraphs have located areas of difficulty in Kant's metaphysics of the subject, as it applies to his theory of moral agency. Kant's problems arose from his insistence on keeping the moral agent wholly outside the world of phenomena. These problems multiply when we consider the dual aspect of the moral subject, as the object, as well as the author, of moral action. The formulation of the categorical imperative given above¹⁸ enjoins agents to treat humanity always as an end, never merely as a means: this is because '[r]ational nature exists as an end in itself'.¹⁹

But, again, in view of the identity problems already discussed, it is not clear how actions undertaken in the phenomenal world can treat noumena as means, or indeed as ends, when Kant has framed his distinctions in such a way that there is a minimal connection between moral subjects and their phenomenal status as human beings. This leaves it very uncertain how decisions as to how to treat the latter can in any way be guided by reference to considerations about the former. We are left with a large gap between the extra-temporal, non-embodied condition of moral subjects, and the fact that what seems to determine the actions of these moral subjects is a rather specific set of causal relations between physical actions and the condition of human bodies.

These difficulties ultimately derive from Kant's presentation of the noumenon as being a free agent only by virtue of being situated outside all causal determination. It is perhaps this, as Strawson surmises,²⁰ which impels Kant to draw so sharp and unconvincing a line around the supersensible subject - it is only in this way, given his assumptions, that he can present that subject, qua moral agent, as being free. If this is so, Kant surely pays an unacceptably high price for his doctrine of metaphysical freedom, a price which includes unintelligibility in his account of reflective moral agency, and a severely incomplete explanation, at best, of how the existence of noumena as rational beings can shape decisions about what human agents are morally required to do.

It is fair to say that these are serious shortcomings in a moral theory. This is not the place to speculate as to how far they might be resolved by adopting the second approach outlined in the previous section, of abandoning the notion of a noumenal realm for a subject given, and limited, phenomenally. This would of course mean jettisoning the idea that subjects, and hence moral agents, are metaphysically free in Kant's sense: the necessary condition of metaphysical freedom, that agents stand outside causal determination, cannot be satisfied if subjects (and hence agents) are purely phenomenal beings.

The important point for this chapter and the argument as a whole is that such a manoeuvre would remove entirely the basis on which Kant saw morality as being founded, for the reason just given. This indicates once again the closeness of fit between his account of subject-identity and the moral theory. This leads us back to the question raised at the start of this chapter, whether the transition from an account of the subject to a theory of its moral value can be effected without compromising neutrality. One of the attractions of Kant's ethical theory, as has been observed, is its seeming even-handedness: it ascribes moral qualities to persons as rational agents, and thus (virtually) everyone qualifies for citizenship of the Kingdom of Ends. And indeed, this is what Kant himself surely intended. Equality of citizenship is only achieved, however, by banishing contingency

from the moral realm - any contingency which threatens to import a moral distinction between persons.

As this section has argued, however, there is a heavy price attached to this moral egalitarianism - a price which has to be paid in more than one currency. The view of the moral agent as noumenon renders persons, as moral agents, inaccessible to reflective self-awareness. It leaves us with a puzzle about the relationship between moral reflection and the causes of action in the world. And it gives us little to go on in deciding what the content of morality should be - in particular, why the substantive imperatives yielded by the theory should be so closely bound up with the fate of rational beings in their purely phenomenal aspect.

The first of these difficulties is perhaps purely internal to Kantian theory itself. But the latter problems directly concern the theme of this chapter, the relationship between the nature of the self and its moral status. Without some resolution of these difficulties we are left with no justification for the centrality of the self in Kant's moral theory. The next chapter will pursue this problem in neo-Kantian liberal theory. As a preliminary to that investigation, it will be instructive to consider how the imponderabilities in Kant's moral theory reappear in his own political philosophy.

IV

Many of the criticisms made of Kantian ethical theory will resurface in the next chapter, when we consider the construction of the self in neo-Kantian liberalism and the neutralist project in contemporary liberal thought. In this section, however, the focus is on Kant's own political doctrines. The aim is to trace the consequences of the views discussed already for Kant's view of the relationship between morality and politics. In particular, the issue is whether Kant's theory of the self can support his belief, echoed in some of his modern followers like Rawls, in the methodological priority of moral over other (and especially political) concerns. The further question this raises, when we

move on to consider modern neo-Kantian theory, is whether the moral priority enjoyed by the self in contemporary liberal thinking is warranted. If the Kantian priorities are arguable, then a particular, self-based, version of those priorities must also come into question.

Kant's most emphatic statement of the subservience of political considerations to morality appears in the Appendix to Perpetual Peace. There Kant asserts that 'all politics must bend the knee before right', and

whatever empirical politics may say to the contrary,...[a] true system of politics cannot...take a single step without first paying tribute to morality. For as soon as the two come into conflict, morality can cut through the knot which politics cannot untie.²¹

It is not of course Kant's point that morality is somehow more effective than politics itself in providing political solutions to problems of policy, but that morality exercises an absolutely overriding claim in situations where moral and political demands conflict. For Kant politics can only occupy the space left over when morality has marked out its territory. As a result, 'there can be no conflict between politics, as an applied branch of right, and morality, as a theoretical branch of right'.²²

Kant's belief in the deliberative priority of moral over political concerns is founded directly on his doctrine of metaphysical freedom. It is the possibility of free agency, in Kant's sense, which both creates a potential conflict between politics and morality, and shows how that conflict is to be resolved.

If, of course, there is neither freedom nor any moral law based on freedom, but only a state in which everything that happens or can happen obeys the mechanical workings of nature, politics would mean the art of utilizing nature for the government of men, and this would constitute the whole of practical wisdom; the concept of right would then be only an empty idea. But if we consider it absolutely necessary to couple the concept of right with politics, or even to make it a limiting condition of politics, it must be conceded that the two are compatible.²³

If, that is, the world could be exhaustively described as a chain of causal interactions, there would be no room for considerations of moral right in political deliberation; the politician would merely be another cog in the midst of a fully determinate physical system. Here, again, we see that Kant is not properly described as a compatibilist about free will and determinism, since he regards the two as logically incompatible - autonomy is only possible if the phenomenal world is excluded entirely from moral deliberation. Moral agency can remain free only if it stands outside the chain of natural causation.

Some commentators, such as Murphy,²⁴ have argued that Kant's moral philosophy is incoherent unless we construe moral freedom as consisting not in (to use Kant's own terminology) free Wille but rather as free Willkur - the former, on Kant's distinctions, being the 'holy will' insusceptible of moral error, and the latter being subject to contingent determination by phenomenal motivations. In an argument which interestingly parallels the theodicy of writers like Plantinga, Murphy maintains that freedom is inapplicable to beings who always act rightly. Since the possibility of error is a condition of freedom, free will can only be ascribed to creatures with Willkur rather than Wille.

It is doubtful, in response to this reconstrual of Kant, whether there is any incoherence in the idea of beings who, of their own free will, always act rightly. As with Plantinga's arguments, if freedom is taken as compatible with universal causal determination at all, there is no greater difficulty in the idea that agents always freely act rightly than in the idea that they sometimes lapse morally; these lapses being, on Murphy's interpretation, expressions of freedom as much as are autonomous moral actions, albeit, in cases of moral transgressions, of free Willkur rather than free Wille.

Such a distinction between types of freedom might be sustainable if Willkur was seen as a departure from autonomy into causal determination of one's actions. But this can hardly be what Kant has in mind. For the dual-aspect theory holds that in some sense all action is causally determined, but that (at least autonomous) action can be redescribed in non-deterministic terms.

Furthermore it is hard to see in what sense, for Kant, action caused by phenomenal motivations can be regarded as free. Hence Willkur, for Kant, must comprise a free and an unfree component, the former being Wille itself, which by definition is free.

But, this being so, it is unclear why there should be thought, even *prima facie*, to be a conflict between moral and political considerations, once the latter have been identified as belonging to the empirical, phenomenal realm. On the dual-aspect theory, there is, as Kant himself remarks in the passage quoted above, just no contest between autonomous agency and phenomenal determination: once an agent appreciates the force of morality, he will not be motivated by phenomenal 'inclination' - not because these motivations will merely cease to exist, but because any fully rational agent will perceive that the grounds of duty lie elsewhere.

There is, as a result, some difficulty in understanding how a Kantian can allow for moral conflict in politics - both in the form of conflict between different moral demands on political agents (since in Kant's view there can never be any clash between moral obligations), and of conflict (as envisaged by the quotation above) between morality and non-moral political concerns. Public policy is conducted in irreducibly phenomenal terms, which gives little ground for thinking that its field of action can ever come into contact - or conflict - with morality. To state the contrast schematically, we are left not with a Platonic republic of politically institutionalized morality, but with Machiavelli.²⁵

The Kantian approach to politics, as interpreted in this chapter, leaves very little scope for the sorts of conflict witnessed in real-life political argument. Once it is concluded that morality bounds the concerns of politics entirely from the outside, any possibility that political conflict might involve difference over the practical implications of value disappears. This is a consequence of Kant's insistence on the categorical nature of duty, that if a situation involves morality at all, it must be in a way which is free of all practical (and a fortiori moral) ambiguity. Kantianism narrows the gap between the

politician and the private moral agent, by collapsing all political decisions into a completely general theory of rational action. What is lost on the way is any sense that moral concerns may apply in different and often conflicting ways to practical decision-making in general, and to political decisions in particular. The neutrality sought by modern liberal theory has, in a sense, been achieved, but only by removing conflicts of moral value from politics.

It was suggested at the start that modern liberal theories based on Kantian ethics risked compromising their commitment to neutrality. The argument was that the Kantian construction of the self has to proceed wholly in abstraction from empirical determinations (such as agents' psychological motivations), whose contingency in Kant's view subverts the categorical status of morality. As a result, both Kant and his modern liberal followers are faced with a dilemma. The neutral self of Kant's argument only retains its neutrality by retreating to a transcendental account of subjectivity from which (at least in the theory of Kant himself) all empirical considerations have been eliminated. This makes it hard to see how the self, thus described, can help to resolve moral disputes - particularly over the practical implications of value.

Of course Kant believed that he had provided, in the Kingdom of Ends, a practical content for the categorical imperative, which would guide action in the world. Hegel and others have reasonably doubted whether the form of universal prescriptivity is as successful in generating a content for morality as Kant himself thought. But, whether or not Hegel was right about that, it is hard to see how the theory can handle value-pluralism as liberalism has traditionally aimed to do. If the theory really does deliver substantive practical content, there will no longer be, from the standpoint of the theory, any way of accounting for the diversity of value which exists in society - or of explaining why the state should be neutral between forms of life embodying these values.

This is not to claim that there is a necessary tension between neutrality and pluralism. Indeed, it might be thought that

neutrality was precisely the liberal response to the existence of plural values: in other words, that the only way of respecting this diversity of value was to remain, as far as possible, impartial between competing conceptions of the good. This is not to claim that neutrality is required by pluralism - the policy implications of plural values are not obvious, and even the endorsement of such diversity in society need not demand the out-and-out impartiality sought by neutrality. At the same time, however, neutrality is clearly one way of attempting to preserve this diversity.

But it is less clear that Kantianism provides a satisfactory basis for neutrality, if at least the latter is designed to support pluralism. Either Kantianism remains purely formalistic, in which case its foundational status in liberal society, as a basis for political institutions and processes, seems to be called into question; or else it embodies specific commitments as Kant thought, and it seems incapable of guaranteeing a pluralistic society: policy, on this alternative, presumably has to be brought into line with the fundamental value, thereby closing off the possibility of entrenching diversity by political means.

This is not of course the end of the argument. It remains open to claim (as for example Larmore does) that there is a value so fundamental that it can allow for the diversity of other values characteristic of a pluralistic society. A fuller evaluation of this response will appear in the next chapter. For the time being, it may be doubted whether any one value really is universal and fundamental enough to perform this role. Pluralism may mean not merely that there is, at a level other than the fundamental one, diversity of values, but that this diversity extends also to beliefs about the nature of the fundamental value itself.

Perhaps liberalism, contrary to what some recent writers such as Ackerman have claimed, can dispense with neutrality without undue discomfort.²⁶ As is perhaps implicit in what has already been said, this partly depends on how wide a range of applicability is claimed for neutralist principles. There may be

areas of policy in which different conceptions of the good or, less grandiloquently, different items which people value, are treated unequally or at least differently by public policy, but which involve no violation of neutrality.

Certainly such violations will be all too frequent if what is required is that different objects of value always receive equal treatment, because this is impossible. There is no detraction from liberalism's commitment to neutrality if it is restricted to operating that principle only when it can. It is perhaps difficult for liberalism to abandon neutrality entirely, given its epistemological underpinnings. Even the political pragmatism which is often invoked against neutrality has to make epistemological claims about what, given the limits on our vision, it is expedient to do, and that brings back the old worry for liberalism, about why even this claim to knowledge is privileged. And this dilemma is fully reflected in Kantian moral theory, which either holds up as the supreme object of value a self so replete as to subvert value-pluralism, or consigns the moral subject to a limbo of vacuity. An important aspect of that vacuity is the self's exclusion from any role within political life.

Notes

1. John Rawls, A Theory of Justice (Oxford 1972), and 'Kantian Constructivism in Moral Theory', Journal of Philosophy 1980; Alan Gewirth, Human Rights (Chicago 1982); Robert Nozick, Anarchy, State and Utopia (Oxford 1974); Ronald Dworkin, Taking Rights Seriously (London 1977).

2. Immanuel Kant, Groundwork of the Metaphysics of Morals ed. H.J.Paton (London 1948), p96.

3. The device of the Original Position in A Theory of Justice seems to indicate that Rawls has the first of these alternatives in mind in his book, whereas the emphasis in 'Kantian Constructivism in Moral Theory' on a socially constructed consensual morality points to the second alternative.

4. Compare Williams' arguments for bodily criteria of personal identity in his Problems of the Self (Cambridge 1973) with the essays 'Ought and Moral Obligation' and 'Internal and External Reasons' in Moral Luck (Cambridge 1981), where Williams stresses the individualized moral claims of a self seen largely in terms of a continuous psychological history.

5. See Bruce Ackerman, Social Justice and the Liberal State (Yale 1980), pp356-7, p361.

6. For the notion of moral personality, see Kantian Constructivism in Moral Theory, p521f; for the application of this notion to citizenship, see Rawls' 'Justice as Fairness: Political not Metaphysical', Philosophy and Public Affairs 1985.

7. Michael Sandel, Liberalism and the Limits of Justice (Cambridge 1982).

8. Cf note 1 above.

9. Immanuel Kant, Critique of Pure Reason tr. N.Kemp Smith (London 1929), p472.

10. Ibid., p154.

11. P.F.Strawson, The Bounds of Sense (London 1966), p248.

12. Ibid., p154.

13. Ibid.

14. Kant, Groundwork, p120.

15. Ibid., p109.

16. Ibid., p96.

17. This is not to deny that the formulation quoted of the categorical imperative could, given a suitable interpretation, be thought to lead to these consequences, but to point out the

oddity of a doctrine which ends in such definite physical injunctions, having begun by excluding the phenomenal world as being irrelevant to moral agency.

18. Cf note 16 above.

19. Kant, Groundwork, p96.

20. Strawson, op. cit. pp247-8.

21. Immanuel Kant, Political Writings, tr. and ed. H.Reiss (Cambridge 1970), p125.

22. Ibid., p116.

23. Ibid., pp117-8.

24. Murphy, op. cit. p95.

25. The reference to Machiavelli here is made with the familiar interpretation in mind of The Prince as advocating the total exclusion of morality from political conduct. While this has a good deal of plausibility as applied to Christian morality (specifically to the Christian moralists writing in the mirror-for-princes genre of which The Prince is an example), it is less convincing to claim that Machiavelli regards politics as ideally being uninformed by any moral considerations at all. The Prince is, after all laden with the vocabulary of moral evaluation - but a classically based, and non-Christian morality.

26. Ackerman, op. cit.

Chapter Three

Kantianism and Contemporary Liberalism

Kantianism and Contemporary Liberalism

In the previous chapter it was argued that Kantian moral theory is in important respects incoherent. While Kant requires his metaphysical doctrine of transcendental selfhood to support the claim that rational (and therefore moral) agency is beyond phenomenal determination, the moral subject's nature as noumenon leaves it incapable, on this account, of causally acting in the world. Nor is it clear how the status of human agents as phenomenal beings is supposed to decide which actions we are morally required to perform. As a result the causal antecedents of moral agency are obscure, as are Kant's reasons for thinking that the presence of other rational agents determines the content of morality.

The present chapter aims to extend this criticism by examining neo-Kantian moral theory, and in particular its applications in contemporary political philosophy. Here the treatment must be selective, since Kant has exercised a pervasive influence, both directly and indirectly, on much recent theory, especially in the United States, and full consideration of all work of broadly Kantian inspiration produced in the past few decades would demand far more space than is available in this work. By confining the analysis to some of the most notable and influential examples of neo-Kantian theory, it should be possible to gain some grasp both of its affinities with Kant's own writings and of its wider impact on contemporary liberal philosophy. This task is made easier by the fact that some of the most prominent work in the field owes a fairly obvious debt to Kant's own writings.¹

On the whole Kant's influence has been channelled through his moral rather than his political theory.² The first part of this chapter will give a brief indication of this influence, with the aim of suggesting something of its nature and extent in modern-day political philosophy. There will however be no attempt to decide whether the interpretations of Kant implicit in the writings of his modern followers are valid, though it is of course assumed by the format adopted in this and the previous chapter that elements of Kant's theory are traceable in the work of contemporary theorists.

Most of the chapter will, however, be devoted to a critical assessment of this body of work as an attempt to ground liberalism in Kantian moral theory. In particular, it aims to establish whether these writings avoid the problems identified in the previous chapter. The ultimate concern will be whether neo-Kantian liberal theory can perform its foundational role while preserving any understanding of the distinctiveness and autonomy of politics as an argumentative and decision-making process. In this discussion it will be suggested that the problems in Kantian moral theory also afflict the work of his modern followers, and that these emerge most clearly in an inadequate understanding of the relationship between moral value and political dispute. There is good reason - good liberal reason - to think that value-based conflict is ineliminable. Part of what this means, as liberals themselves have often remarked, is that conflict may itself be a dimension of value, one which cannot in all cases be eradicated without moral cost. Politics could not play the part it does in human life in the absence of such conflict. An important consequence of neo-Kantian theory is that political dispute can no longer help to resolve disputes about values and their practical implications since such disputes in effect no longer exist - the theory redefines politics in such a way as to preclude them.

I

As has already been remarked, this chapter cannot attempt a comprehensive survey of neo-Kantian moral and political theory. The detailed analysis will accordingly be confined to two notable recent writers working, as will be argued, within Kantian assumptions. Whereas the first of these, John Rawls, is explicit about the Kantian roots of his theory, Alan Gewirth, the second, has disclaimed Kant as a precursor, maintaining that his moral theory is without historical precedent,³ being derived solely from reflection on the nature of agency.

This disclaimer will make necessary a somewhat lengthier treatment of Gewirth's Kantianism than Rawls'. The conclusion will be that the emphasis on agency does not mark any difference of substance between Gewirth's theory and Kant's, since the argument from the nature of action does not, after all, depend on

considerations of rational agency alone, but on a moralized view of the agent. At the other end of the argument, Gewirth is less explicit than Rawls about the political implications of his theory; in his book Reason and Morality Gewirth is more concerned to prove the general thesis that rights are derivable from generic features of action than to specify the content of particular rights.⁴ It is, nonetheless, fairly clear that Gewirth envisages the first-order rights as dealing, familiarly, with such matters as privacy, personal liberty, non-interference, and so on - familiar, that is, as the content of rights within liberal theory. This is not surprising, given the role of the moral subject in underpinning the theory. It is only given a prior view of the subject as being, in some sense, the supreme repository of value in the world that it becomes important to specify conditions on legitimate action.

We begin, however, with the explicitly Kantian theory of John Rawls, as expounded in his book A Theory of Justice and numerous subsequent articles.⁵ This discussion will focus on two features of Rawls' theory where his debt to Kant is particularly evident - the device of the Original Position and the methodological apparatus of reflective equilibrium, the latter having been more recently amplified or replaced by a 'constructivist' procedure.⁶ This seems a good place to start, since the metaethical standpoint of constructivism, and of reflective equilibrium, is meant to formalize the intuitive moral judgments embodied in the Original Position: the latter's appeal is meant to derive from its encapsulating moral notions or beliefs which we already hold, and its delivering conclusions (i.e. the principles of justice) conforming with our prephilosophical beliefs about what is fair. This is the formal basis of the theory, and the source, as Rawls hopes, of its appeal, that it gives theoretical effect to moral convictions which we already hold.

It may be said that there is nothing particularly Kantian about the method of reflective equilibrium, or of constructivism as such; all these require is that we arrive at sets of moral judgments capable of being reduced to consistency when the 'ill-considered' judgments among them are discarded. Thus, at the minimum, all that the methodology of reflective equilibrium or constructivism demand is consistency in our (dispositions to assent to) moral judgments.

In response to this objection it is worth saying, first, that although there is indeed nothing uniquely Kantian in this method, reduction to consistency along lines similar to this is at least a necessary feature of Kant's moral system. Because Kant regarded moral imperatives as being exclusively categorical, there could be no circumstances in which an agent was confronted with two or more incompatible moral claims upon him. For in such situations one imperative would have to override the other, and this would mean that the second was, after all, not categorical, but conditional upon the first's not applying.

Similarly, the deontological metaethical structure of Rawls' theory of justice requires that the principles derived from the theory are categorical, in the sense that they 'are not subject to political bargaining or to the calculus of social interests'.⁷ Since the principles of justice have this non-negotiable status, any argument for them must conform to 'considered' moral judgments which share this status. In the process of reflective adjustment, recognizing the validity of the principles of justice may itself induce us to revise the substantive judgments to which we assent. This process of adjustment is meant to be reciprocal - we are meant to derive general moral principles from philosophical reflection, while discarding any such principles which entail intuitively unacceptable first-level judgments. But by the end of the process, we are supposed to be left with a set of principles which are 'in accordance' with the 'everyday judgments we...make'.⁸

It can also reasonably be maintained that Rawls' more recent constructivist procedure is a reworking, in a more explicitly Kantian vein, of the reflective equilibrium method of A Theory of Justice. In Rawls' Dewey lectures, the emphasis shifts from an individual's reaching an equilibrium between his moral principles and first-order judgments, to a conception of moral theory in which '[o]bjectivity is to be understood by reference to a suitably constructed social point of view'.⁹ Here the idea is that the principles of justice codify moral beliefs already current in society, and which (for example) inform in part the content of political debate. Rawls seeks to construct, from the materials provided by the current moral notions, a consensual standard, that of the moral person, which we can all accept after reflection. It is not claimed that the moral standard is trans-

historically valid, merely that it encapsulates the shared standards which Rawls believes to underlie familiar disputes such as those where alternative policies appear to offer a choice between promoting freedom as against equality.

This dimension of historical contingency might be thought to be at odds with the Kantian view of morality as being unconditionally binding and free from (e.g. cultural or historical) contingency; but Rawls sees Kant's theory as 'the leading historical example of a constructivist doctrine'.¹⁰ The Kantian element in Rawls' constructivist position is held to consist in the priority accorded to the notion of moral personhood, in contrast, for example, with what Rawls labels the 'rational intuitionism' of Sidgwick, Moore and Ross;¹¹ the latter, in contrast with Kantianism, hold that there is 'a moral order that is prior to and independent from our conception of the person'.¹²

It is questionable whether there is anything particularly Kantian about the constructivist methodology itself, as opposed to the specific notion of moral personhood which provides the materials from which the construction is made. Kant himself would surely have regarded the categorical imperative, and the actions which it enjoins, as gaining its force from considerations of rational agency alone, rather than from an assessment of current moral beliefs.

Similarly, the publicity requirement, which Rawls presents in the Dewey lectures and elsewhere¹³ as a constructivist notion, while being an authentically Kantian requirement upon agents' principles of moral action, is not obviously derivable from the idea of a constructed morality alone. It is at least conceivable that the moral notions at large in society might not accord such centrality to the person as a moral notion, even if it is hard to see how constructivism could result in social and political institutions which violated the publicity condition.¹⁴ However this may be, Rawls' theory is clearly Kantian both in its central 'model conception' of moral personhood, if this is identified with the rational agent of Kantian theory, and in the deontological priority of the 'right' over the 'good'.

It is because the notion of the moral person enjoys this centrality in our moral thinking that Rawls believes it to be justifiable to introduce the theoretical device of the 'Original

Position' as a means of deriving the two principles of justice. If we accept the ground rules and conditions of the Original Position as equitable because they conform to or formalize the moral beliefs we already hold, then we will be committed to accepting the outcome of the deliberations which go on behind the 'veil of ignorance' as also being just. Hence the Original Position works in the theory as a means of spelling out the consequences of our prior moral convictions. The veil of ignorance restricts the information available to the parties in the Original Position so as to conceal from each person 'his place in society, his class position or social status' and 'his fortune in the distribution of natural assets and abilities, his intelligence and strength, and the like'.¹⁵

The conditions of the Original Position are so arranged as to eliminate the natural and social contingencies which are held to be irrelevant to persons' status as rational agents: the limitations on the parties' knowledge as decision-makers in the Original Position are imposed because the rational capacities, which are unaffected by these restrictions, are the essence of the parties' moral personhood. Seen in this way, the Original Position turns out to be a modernized version of the Kantian Kingdom of Ends. What is left when the personal attributes excluded by the veil of ignorance are subtracted, in the amplified presentation of the theory given in the Dewey lectures, are the two moral 'powers', the 'Reasonable' and 'Rational' capacities of agents. What emerges, then, is a fundamentally Kantian conception of the person, and this conception can be seen, Rawls thinks, as an elucidation of our actual moral thinking about persons.

Rawls has latterly denied¹⁶ that the view of the person implicit in the construction of the Original Position relies on any controversial metaphysical claims. The theory rather involves, in Rawls' view, a 'political conception of the person',¹⁷ according to which the 'moral powers' are given effect in the values enshrined in public institutions and procedures. As this is meant to be the end of the theory, however, it is hard to accept that it can also, at least without circularity, be its beginning: the theory was meant to provide some theoretical rationale for adopting a certain set of political structures, and it can hardly do this if the theoretical construction is made

from the very structures in question. Presumably Rawls would regard the non-metaphysical status of his conception of the person as arising also from the method of its construction: since the conception is built directly out of the moral beliefs which people actually hold, there is no need to introduce any questionable metaphysical doctrines in its support. It is the use of these beliefs as a basis for constructing the public consensus, rather than the ultimate foundation or justification of the beliefs, that Rawls wishes to explain. What is important is the fact that the beliefs are held, not their philosophical grounding.

It is, however, doubtful whether Rawls can evade so easily all obligation to justify the metaphysics of the Original Position. Even if it is true, in some sense, that the final conception of citizenship which results from the theory codifies or formalizes moral notions already held by real people, this does not mean that no metaphysical assumptions have been used in the process of construction. The very process of constructing a set of shared public values out of the disparate and often inconsistent moral judgments made in everyday situations may demand such assumptions.

This is already implicit in the fact that, if the original judgments really are inconsistent, there must be alternative ways of reducing them to consistency, and it is hard to rule out the possibility that some metaphysical apparatus will be needed in the reduction - certainly some consideration will have to be brought in to decide why one form of reduction should be chosen rather than another. For example there must in the Original Position be some minimal conditions of identity upon the contracting parties, so that they can rationally calculate their best self-interested strategy in the light of the possible outcomes: only if these outcomes are identifiable, when the veil of ignorance is lifted, as ones which they have chosen, will Rawls' argument have any force. As the veil of ignorance excludes a wide range of personal attributes from the knowledge of the parties in the Original Position, the identity conditions are contestable. As Sandel has argued,¹⁸ Rawls is relying on a 'radically disembodied' self, one bereft of (what are thought of as being) its contingent attributes.

This is not necessarily to claim (as Sandel's use of the label

'disembodied' might be taken to suggest) that Rawls ends up with an unconditionally Cartesian view of the self as contingently embodied but essentially mental. But the restrictions imposed on the self in the Original Position, and Rawls' extended account of moral personality in his writings since A Theory of Justice, suggest that he does indeed rely on controversial metaphysics both in his assumptions about personal identity, and in his understanding of the relationship between this and the moral value of persons. This does not show by itself that Rawls must run into the same problems as Kant encounters in explaining moral agency. This issue will be examined in the next section. There must at least be some explanation of how, given Rawls' abstract conception of the self, the existence of other rational agents operates as a constraint on what actions may morally be performed, or what public principles of justice may be adopted in Rawls' 'well-ordered society'.

It appears, then, that Rawls' constructivist method has to rely at some stage on metaphysical claims of one form or another, even if his task is confined to developing publicly applicable moral notions out of (philosophically unexamined) original beliefs. Since the initial situation is one of conflict between conceptions of public institutions and policy, it is at best doubtful whether there can be any direct progression from shared moral beliefs, even if they exist in the first place, to the resolution of such conflict. What may be at issue, after all, is precisely what metaphysical notions underlying these shared moral beliefs.¹⁹ Even if people are unanimous in the values which they profess, this does not show that similar unanimity must prevail about the metaphysical claims which these values license. As the argument so far has tried to show, Rawls' theory relies on assumptions about the identity of persons and how this bears upon our understanding of persons' moral worth.

In his book Reason and Morality Alan Gewirth believes himself to have shown, without reliance on controversial metaphysical claims, how reflection purely on the nature of rational agency entails moral commitments, and specifically to the assignment to all agents of rights, as individualized moral claims against other rational agents. In Gewirth's view, his argument depends solely on a consideration of the so-called generic features of action - those features which must be present in order that a

piece of behaviour counts as action at all. Hence the argument is supposed to depend only on logical or semantic premisses about agency. From these premisses Gewirth hopes to prove that anyone is committed, purely in virtue of being an agent, to accepting that they are morally obliged to perform certain actions and refrain from others in the appropriate circumstances.

The core of Gewirth's argument is that '[s]ince the agent regards as necessary goods the freedom and well-being that constitute the generic features of his successful action, he logically must also hold that he has rights to these generic features and he implicitly makes a corresponding rights-claim'.²⁰ That is, there are certain necessary conditions which must be satisfied if action is to be possible, among which are (a certain amount of) freedom and well-being. Assuming that it is impossible rationally to will any action without at the same time willing its necessary conditions, it follows that each rational agent must will a certain amount of these goods.

From this Gewirth concludes that each agent is logically bound to accept that these goods are likewise necessary to all other rational agents if they are to act. The conditions of non-interference with my freedom and well-being, which I am logically obliged to will in my own case as necessary to my acting, must by parity of reasoning apply to my fellow agents as well. This being so, I acknowledge that I have logically to accept that other agents must receive a measure of the generically necessary goods. This establishes, Gewirth believes, valid rights-claims by each agent in respect of all other agents: the rights-claims set out the formal logical requirements of agency. These rights-claims are intended to embody the acceptance by each agent that the generic features of action impose the obligation to respect the generically necessary goods of all other agents. Hence Gewirth claims to have arrived at substantive moral conclusions from considerations of rational agency alone.

An assessment of this argument is deferred until the next section. The present task is to expose the latent Kantian assumptions in Gewirth's argument. As has already been observed, Gewirth regards the agency-based derivation of morality as wholly without precedent, since '[a]lthough the importance of action for moral philosophy has been recognized since the ancient Greeks, it has not hitherto been noted that the nature of action enters into

the very content and justification of the supreme principle of morality'.²¹ It might be wondered whether, for example, the existence of any such supreme principle was a claim to be argued for rather than 'noted'; this assumption apart, Gewirth dispenses with any explicit metaphysical foundation for his attempted derivation of morality from rational agency.

The question, however, is whether his argument proves to depend, apart from tacit assumptions about the nature of morality, on further claims about the status and nature of agents, as Kantian theory explicitly does. If this does turn out to be the case, Gewirth seems to be threatened with parallel objections to those made against Kant in the previous chapter, where the problems in Kantian moral theory were traced to his account of the metaphysical subject. This is not to say that the full apparatus of Kantian metaphysics lies beneath the surface of Gewirth's argument. But he is likely to run into similar problems to those of Kant if his argument presupposes a moral subject parallel to that which Kant creates from his metaphysics of experience.

As has been said, Gewirth starts from the premiss that there are certain generic goods necessary for action, and that these at least include freedom and well-being. It might be thought that this claim was almost vacuously true, since anything recognizable as action in the world will require such goods if 'freedom' is just explicated as 'freedom of action', and 'well-being' as 'capacity to act', or something similar. It looks, indeed, as though Gewirth has in mind some formulation of this level of generality, so that 'the agent's well-being is to be identified primarily even if not exclusively with the general abilities and conditions required for attaining any of his purposes'.²² Although agents must pursue determinate goods, or conceptions of well-being, in their lives, it is the general capacity to pursue, in some specific form, such a conception which constitutes the generic feature of action: 'it is not the particular purposes or outcomes but rather the generic abilities and conditions that for the agent primarily constitute his well-being, since they are the necessary conditions of all his pursuits of his purposes'.²³ This requires that agents are identifiable separately from the particular ends which, at any point, they are pursuing. Agents are not, in Sandel's terminology,²⁴ 'radically situated' selves,

identified with the particular ends they pursue, but more like the 'radically disembodied' selves which, in Sandel's view, lies behind Rawls' presentation of the Original Position.

This of course does not show that Gewirth is working with a Kantian conception of agency, but it does suggest that he cannot avoid making some commitments about the nature and identity of agents in his argument. This to some extent lessens the force of his claim to rely only on considerations about action itself. One way in which this becomes apparent is not merely in the thought that particular actions are irreducibly attached to individual agents, but also as Gewirth acknowledges that actions acquire their meaning and content from their playing a role in wider projects of the same agent; this is true even of the 'radically situated' agent whose identity consists in the unity of the project which he pursues.

Gewirth may say in reply that he is not considering the question at this level, but is rather concerned with entirely general features of agency shared by all individual actions, whatever their content and whoever performs them. This is of course a quite coherent line of response, but it moves Gewirth far closer to the neo-Kantian position of Rawls, for whom '[t]here is no need to identify individuals with their existing conception of the good, with their actual plans of life'.²⁵ What comes out of this stage of the argument, then, is the assumption both of (temporally continuous) agents as authors of action, and the conceptual separability of these agents from the particular goods or projects they pursue.

Gewirth's second premiss is that it is impossible rationally to will an action without also willing its necessary conditions. It might be thought that the necessity referred to in this premiss was that of logical identity - that anyone who willed an action must will its conditions because this is what it means to will an action. From the argument set out above, however, it appears that Gewirth has in mind a more substantial notion of necessity than this, concerning the part played by action in implementing conceptions of the good. On this reading Gewirth is claiming that it is a necessary condition of acting that an agent possesses a determinate conception of his own good, which must comprise or include freedom and well-being; as a result, the agent is rationally committed to willing these latter goods.

This presumably introduces a closer relationship between action and its necessary conditions than that of cause and effect, as it is unlikely that Gewirth is advancing the implausible thesis that an agent is committed, by willing an end, to willing also whatever is causally necessary to produce it. At the same time, the fact that Gewirth is concerned with goods defined generally enough to serve as the conditions of any actions whatsoever suggests that he is operating with an entirely abstract and general conception of agency. It is as being acting on quite general considerations of rational agency that the conditions hold.

This can be brought out further by examining the nature of the goods which in Gewirth's view must be willed by the rational agent. On the face of it, there is no need for any agent to consider, on the level of generality assumed by Gewirth's argument, what conditions must be satisfied if he is to act at all. All that is required, it appears, is that there are some determinate goods which he aims to pursue in his actions, and this need not involve any general thoughts about agency. If so, the generalization step of the argument is threatened, since it relies on each agent's being committed, purely through identity of inference from identical situations, to accepting that the generic goods are needed for all other agents to act. Unless the conditions are set at the most general level, it looks as if all that an agent will be required to recognize is that anyone else identically situated to himself will need the same range of particular goods to fulfil the purposes which he in fact has. This certainly yields a weaker conclusion than Gewirth is seeking: it is far from clear that this leads us from reflection on agency into morality, as opposed to the recognition of a trivial logical truth. That is, the considerations about the necessary conditions of agency threaten to permit only a limited universalization, from the particular circumstances of a given agent, and this must be a less strong conclusion than Gewirth is seeking.

This perhaps explains why Gewirth has to resort from the outset to a wholly general conception of agency. In his treatment agents are divorced conceptually from whatever ends they happen to pursue at any given time, and the conception of agency with which they are credited has, for the universalization step of

Gewirth's argument, to be set at the most general level. There is, moreover, a further respect in which the argument relies on a substantive, and quasi-Kantian, view of agency. For the universalization step to go through, there must be some explanation of why other agents' purposes are to be taken into account in the first place - of why it is a relevant consideration in analyzing the rationality of an agent's action whether it deprives others of their generic goods. It is not obvious that any agent who fails to take account of the latter in practical deliberation is convicted of logical inconsistency, as Gewirth claims. For that to follow, it would be at least a necessary condition of the agent's deliberating that he was aware of other agents standing in an analogous relationship to their purposes as he does to his. But it is unclear why this must figure in an agent's deliberations: surely it is logically possible for a person to engage in practical reasoning without this awareness of others as authors of action.

If this is right, it looks as though Gewirth has both to provide an account of agency, and to show that this account must appear in the deliberations of any rational agent. Gewirth also has to show, even once this account is to hand, why the considerations must have practical force for the agent, in guiding his projected actions. In response to this second point, an argument might be constructed from the allegedly prescriptive nature of desires.²⁶

That is, it is held that whatever is desired is desired as something good; and some measure of freedom in acting is necessary for the pursuit of whatever is desired. Furthermore, on the least objectionable version of ethical naturalism, identical evaluative properties must supervene on the instantiation by two objects of identical sets of non-evaluative properties.²⁷ Any agent desiring freedom as necessary to the satisfaction of his other desires, and therefore regarding it as a good, must in the light of this also accept that such freedom is good, because instrumentally desired, for other agents as well. So any process of deliberation which yields the conclusion for a given agent that his purposes are good must by the same token force him to accept that others' purposes are good also. One way of giving moral effect to this conclusion is to ascribe rights to each agent against all others in respect of the generic goods of

action.

On this interpretation, Gewirth's argument is that any agent must grant practical force in his deliberations to the purposes of others, since otherwise his own purposes would lack the prescriptive or evaluative force they have for him. In this form the argument still relies, of course, on the agent's reasoning, with a generalized conception of agency, back to his own case, in deciding what he can rationally will. The importance of this revised version of the argument, however, is that it attempts to show how any rational agent is compelled to give force to others' purposes in his practical deliberations.

Whatever else may be said about the argument in this form, it is doubtful whether the sense of 'good' needed for the naturalistic entailment step is strong enough to produce the conclusion Gewirth wants. While it follows from the premiss that whoever desires something regards it as good, that anyone who desires something (i.e. freedom or a measure of well-being) as necessary to satisfying his other desires must see that thing as good, it does not follow that a person must regard the things desired instrumentally by others as good. This is so even if it is also held that anyone who thinks something good because he desires it must regard others as thinking the objects of their desires good. Even this does not require an agent to think of others' purposes as being, in any unqualified sense, good, nor for that matter good for the persons seeking them. Agents only have to recognize others' purposes as seeming good to those pursuing them.

Gewirth's argument cannot work by attempting to endow others' purposes with the motivational force which an agent's own purposes have for himself. It is only if the agent is already disposed to regard the existence of other purposive agents as founding some claim - for example a moral claim - on his action, that these purposes will acquire any deliberative weight for him. The argument from the symmetrical deliberative positions of each agent with respect to all the others gains little leverage against, for instance, a Stirnerian egoist who uses this symmetry to support an argument for moral individualism.

What is needed, then, is some grounds for thinking that agents must regard it as relevant to their deliberations what others similarly situated would be motivated to do. But it is hard to

see how this can happen unless the agent concerned has made some prior decision that others' ends are to be given deliberative weight: the point is not so much that there is no transition from non-moral to moral deliberation in practical reasoning, but that the purely formalistic considerations advanced by Gewirth are insufficient to explain how any agent unmotivated originally to grant them weight can come to do so merely by reflection on the nature of action.

It is this, perhaps, which pushes Gewirth further in the direction of Kantian moral theory than he is prepared to acknowledge. Once the idea is abandoned that Gewirth's argument rests on the motivations an optimally rational agent could come to acquire through reflection on agency, we are left with the position, more Kantian in appearance, that there is a rational constraint, or set of constraints, which hold irrespective of agents' motivations, on our actions as rational beings. In other words, Gewirth's assignment of rights to the generic goods of agency looks like a paraphrased version of the Kantian claim that there are moral imperatives which hold unconditionally (specifically, in this case, regardless of agents' subjective motivations) on rational agents as such.

This becomes clearer once it is appreciated that Gewirth's argument cannot rely for its force on the actual motivations of agents. If it did, the argument would have to take account of the fact that many agents ignore others in their deliberations. Consistency alone is not enough to compel agents motivated as they are to accept the right of others to the generic goods of action. Gewirth's argument relies solely on considerations about the nature of action in general, not on any reconstruction of deliberative processes which an agent might undertake from an original set of motivations.

While Gewirth is not as explicitly Kantian as Rawls in his moral theory, he nonetheless is forced to rely on something very similar to Kant's conception of moral agency if his argument is to succeed. Without some such conception there is no reason to think that agents are committed, purely by their being agents, to accepting Gewirth's moral conclusions. It is not sufficient to point to the existence of other agents in the world in arguing for Gewirth's principle of generic consistency, since it need not enter into an agent's deliberations how others' projects will be

affected by some course of action which he wishes to pursue.

Such considerations will only have any bearing on the agent's practical reasoning if it can be shown that there is some antecedent requirement of an 'external reasons'²⁸ form that agents acknowledge the existence of other agents and adapt their actions accordingly. The most obvious form of external reasons claims are those mounted by (a cognitivistically interpreted) morality, which makes it a criterion of rationality whether agents take into account and act upon the fact of other moral agents' existence. This narrows the distance between the Kantian Kingdom of Ends, where agents act out of respect for each other as rational beings, and Gewirth's account of the generic conditions of agency.

Gewirth's neo-Kantianism is thus apparent in an implicitly moralized view of agency. It emerges also in the separation, to which, as has been argued, Gewirth is committed, of agents' motivation from the grounds of moral obligation, which is a notable feature of Kant's theory of moral agency.²⁹ The mutual rights Gewirth derives cannot be contingent on particular motivations which agents may happen to have, but rather depend on quite general conditions of agency. This generality comes out in the restriction of Gewirth's argument to those conditions which apply to all action. The argument is concerned with general conditions of compossibility in actions, that is with sets of actions (of different agents) which can be performed simultaneously.

This, again, bears a strong resemblance to the Kantian requirement that moral action must be founded on universalizable maxims, the content of which is to be determined, in Kant's view, by asking what actions could consistently be willed by a society of purely rational agents. Similarly, for Gewirth the generic features of action have to be so defined that the rights which emerge at the end of the argument sanction compossible sets of actions. This is after all the point of the generic consistency requirement, that it generate sets of actions all of which can be performed. Thus while universalizability need not enter into the content of the agent's reflective thoughts about his own actions, it operates as a tacit condition on action.

The generality of Gewirth's argument is also apparent in his interpretation of the generic conditions themselves. While it

would be misleading to claim that his understanding of the conditions of action in any way presupposes the full apparatus of Kantian metaphysics, it remains true that Gewirth's emphasis on freedom as a generic condition has to be taken in a highly abstract sense. As we have already seen, it is not enough for his argument to view freedom merely, for example, in terms of what agents subjectively happen to want: there has to be some wholly general formulation of the notion, which enables it to serve as a fundamental condition of all action. It is unclear how this notion of freedom, set at this level of generality, can play much part in circumscribing the actions which agents can perform, unless it is implausibly supposed that the sorts of actions Gewirth wishes to proscribe by means of the PGC are such as to prevent some agents acting at all. If this is not so, there is no reason to think that an argument based on the necessary conditions of agency will be powerful enough to preclude (what Gewirth would regard as) gross violations of persons' rights. As was argued in the last chapter, similar problems beset Kant's attempts to derive substantive moral conclusions from his noumenal conception of freedom.

This section has tried to bring out some salient points of contact between Kantian theory and the work of Rawls and Gewirth. As Rawls himself has remarked,³⁰ this relationship is one of analogy rather than identity, but, as has been argued above, the similarities are strong enough to identify both Rawls and Gewirth as Kantian writers. The next section asks whether they accordingly fall prey to the criticisms made of Kantian moral theory in the previous chapter.

II

It will be helpful to begin by presenting a brief comparison between the main features of Rawls' argument and that of Gewirth. This comparison is intended to trace (what will be argued are) the shortcomings of these theories to the common inspiration in Kantian moral theory. As the previous section tried to show, despite the substantial differences of approach between Rawls' and Gewirth's work, the underlying Kantian similarities, whether acknowledged or not, are clear enough. If the problems afflicting these theories are indeed of Kantian origin, the question arises

whether the Kantian project was doomed to failure from the start. This question will be considered in the next chapter.

The exposition of Gewirth's neo-Kantianism in section I of this chapter emphasized the centrality of the moral agent in both theories, though it may be said that for Kant this notion forms the starting-point of the theory rather than, as for Gewirth, its conclusion. As has already been noted, the notion of moral personality is likewise a fundamental idea in Rawls' theory, and particularly in the writings he has produced since the publication of A Theory of Justice.

Rawls' theory of moral personality extends the abstract conception of agency first presented in his account of the Original Position. For Rawls, the fundamental notions informing moral personality are the so-called 'moral powers', namely 'the capacity for a sense of right and justice' (the Reasonable) and 'the capacity for a conception of the good' (the Rational).³¹ As has already been observed, the latter capacity does not require that individuals are identified with their current conceptions of the good, and indeed other aspects of Rawls' theory, such as the account of primary goods, demand that these conceptions are revisable.

Similarly, Gewirth's argument in Reason and Morality assumes, as was mentioned, that agents are not to be identified with the specific ends which they pursue; this was held to be essential to Gewirth's case, that the understanding of agency was set at the most general level. As with the parties in Rawls' Original Position, the assumption underlying Gewirth's argument is that agents are 'radically disembodied'³² to the extent that their identity-conditions are set independently of their particular conceptions of the good. This already involves certain metaphysical commitments regarding the nature of persons, and about our reasons for valuing the ends they pursue.

Rawls is certainly more explicit than Gewirth about the moral content of his model of personality. The participants in the Original Position are explicitly defined as 'self-originating source[s] of claims'³³ upon the provision by society of scarce goods. It is the capacity of moral persons reflectively to form a conception of the good and then mount claims upon society as a means to its fulfilment, rather than the content of that conception, which imposes moral requirements on the other members

of society; conversely, the presence of other agents pursuing their own conceptions of the good constrains the possible goods which an agent can pursue while acting on his own conception. The decisions taken behind the veil of ignorance occur within an already moralized framework - one whose moral content Rawls has emphasized in his more recent writings, in contrast with his earlier stress on rational decision-making mechanisms in A Theory of Justice. In the revised version of the theory, this means not only that circumstances in which the principles of justice are decided are so arranged that morally arbitrary attributes of persons are excluded from consideration, but also that the factors to be taken into account by the parties in determining the principles include an awareness of others' status as moral persons.

While Gewirth's approach does not build moral considerations as directly as this into the argument, there is reason to think, as we have seen, that the argument depends on the tacit assumption that persons bear certain moral claims against one another. How far are these to be identified with the 'moral powers' of Rawls' theory? In so far as Gewirth is thinking mainly of persons as setting for themselves projects and concerns which shape their lives, it might be supposed that his argument laid emphasis on the claims of the 'Rational' rather than the 'Reasonable' - in other words with self-interested calculation rather than with any moral or altruistic concern for others. This is indeed what Gewirth takes himself to be doing: the argument is meant to lead from self-interest into morality. But once the idea is rejected that it is the intrinsic content of the ends which persons pursue which gives those ends value, in favour of the view that they acquire value vicariously, as the ends of persons, it is much harder to resist the conclusion that the conception of agency is already moralized, and in a way which accords primacy to the moral claims of persons.

Gewirth would of course reject this interpretation of his argument, as it denies that the rights-based conclusions which he draws can be reached solely by means of logical or semantic considerations about agency. But once the possibility of the rational egoist is introduced, Gewirth seems to be faced with a choice between accepting that such a person has rational grounds for flouting (what Gewirth regards as being) the rights of

others, or imputing to the agent some reason or set of reasons lying outside his 'subjective motivational set',³⁴ which purport to present the agent with reasons for acting in ways in which he is currently unmotivated to act. Only if some argument supporting the latter conclusion is given can Gewirth maintain that agents are required by pure considerations of rationality to accept the limitations on their actions imposed by the PGC.

There is certainly no element of the 'Reasonable' in Gewirth's argument in the sense of an express acknowledgement by agents of right and justice as moral values. This does not however mean that the argument dispenses entirely with any assumptions about agents' moral capacities. Apart from anything else, if the argument is to have any practical effect, agents must be capable of understanding the moral constraints imposed upon them by the presence of other agents and of acting upon them. Beyond this, each agent must be thought of as able to respect others' purposes because they are the purposes of other agents. This requirement could only be avoided if each agent were already subjectively motivated to respect others' purposes. But the assumption of the argument is that this is not so, and indeed it would render the argument unnecessary if it were already true. Unless, then, agents are thought of as having the motivational capacity to act on moral reasons, and the ability to recognize such reasons in the first place, Gewirth's argument will not establish any unconditional practical necessity to observe persons' rights from the generic features alone.

While it is true that the methodology both of reflective equilibrium and of constructivism, its replacement in Rawls' more recent writings, have little in common with Gewirth's argument from the generic features of agency, both Rawls and Gewirth work with a highly formalized conception of agency. This emerges partly through a conception of persons' moral status as independent of their particular ends, but also in the assumption that it is possible to draw substantive conclusions about the rational requirements on agents from considerations of this degree of generality.

This is not to question the possibility of making any philosophical statement about rational agency; but it does not follow from the truth of this claim that a philosophical theory will be able to deliver substantive conclusions about which

actions rational agents will or will not perform. Ultimately this demands that there must be a set or disjunction of goods which any being who is rational must want. It is doubtful whether there is even a disjunction of such goods - at the limit, it is possible rationally to reject the (causally necessary) conditions of all agency. Suicide is a rational possibility, and one which Gewirth seems to recognize as causing problems for his theory.³⁵

Apart from this denial of the existence of intrinsically rational goods, there is a peculiarity in any attempt, within the broadly liberal assumptions of Rawls and Gewirth, to specify certain goods as intrinsically rational, in the sense that any rational agent will pursue or desire them, whatever else he pursues or desires. For the picture which both Rawls and Gewirth present is of agents as 'radically disembodied' choosers of their ends. If their accounts of rational agency are cast in fully general terms, it is hard to see, given the plurality of possible goods, how any can be earmarked to all rational agents as such.

It is an avowed aim of both writers to produce a theory which achieves maximum generality by accounting for the diversity of goods (with minor exceptions)³⁶ which agents in fact pursue. Agents are depicted as bare choosers confronted with an array of possible goods with which to structure their lives. Since these goods are plural, incommensurable and, given scarcity, often exclusive, there has to be some decision taken, for the purposes of public policy, between these goods; but as far as possible³⁷ public provision should refrain from biasing the choice between the competing goods by differential funding or otherwise favouring some of these goods over others.

Once this degree of neutrality is thought of as desirable, however, there no longer appears to be any basis on which agents can choose between the competing alternatives. Once agents, viewed either as parties behind the veil of ignorance in the Original Position, or as the locus of capacities for rational action, it is difficult to see what considerations, at this level, can inform their choices and consequently to understand in what sense the latter are to be thought of as rational. This reflects a wider problem confronting a certain strand of liberal theory, that its metaphysical view of the agent as an untrammelled chooser leaves it obscure how any choice can be made by a self this thin. In its quest for a self free from contingent

determinations of its choices liberal theory is left, in the end, with a self too attenuated to be capable of deciding anything.

This problem seems to be a direct result of neo-Kantian attempts to derive a value-neutral account of the self and its choices. It may be said in response to the criticisms voiced above that for the philosophical purposes of Rawls and Gewirth we can just divide through by the common factor shared by each rational chooser, and this leaves us with something like the participants in the Original Position, or Gewirth's model of agency. What the argument of Rawls and Gewirth are concerned with, it may be said, are not real-life choices, but a general analysis of rational commitments on all agents.

One line of response to this is to ask, as Sandel does, whether this possibility is as coherent as it may appear at first sight. It may be that selves are so identified with certain of their projects that this abstraction makes no sense of their status as agents. They may also be strongly characterized by certain (moralized or non-moralized) traits of character, the virtues and vices being an obvious example. This possibility will be considered more fully in the final chapter. For the present it may be noted that the very notion of virtues of character poses intractable problems for a Kantian view of morality. It is plain that such virtues may be unequally distributed among agents, and therefore that agents' moral capacities may vary from one person to another, rather than depending exclusively on their status as rational beings.

Similar difficulties arise when we move from considering the relationship between the self and its ends to the self's status as a bearer of moral worth. As Rawls explicitly admits and Gewirth, as has been argued, also assumes, there has to be some explanation of how the self, as a moral subject, comes to be seen as making categorical claims on the actions of others: there must be some attributes of the self in respect of which these moral claims hold. But the more the conception of the self is narrowed to accommodate the liberal notion of a metaphysically free chooser, the harder it becomes to understand how the residual subject can be thought of as mounting any significant moral claims, or, in practical terms, how projected courses of action could infringe upon the territory occupied by such a self. The corollary of interpreting metaphysical freedom as the absence of

any specific objectives is a self whose existence in the world seems to make little difference to the factors influencing agents' practical thinking about what ought to be done.

This is, of course, the reverse of the conclusion which both Rawls and Gewirth want, as it is the existence of moral persons or rational agents which is supposed to give rise to the principles of justice or rights to free action. If we define persons or agents at the most general level as bearers of capacities for rational or moral agency, it may be said, this does not leave us without an explanation of how the existence of these individuals can have a practical influence on our concerns: it is precisely the fact that they have these capacities which entitles them to consideration in the practical deliberations of others.

The criticism, however, relates not to the derivation of these entitlements from persons moral or rational capacities, but to the difficulty of understanding how the mere presence of these capacities could have practical effect. At the same time, it may reasonably be said that rational agency, as conceived of by Gewirth, does place certain minimal obligations upon the public authorities, to provide at least a subsistence income to citizens under their jurisdiction - without some such minimal provision, all agency will, for some individuals, be impossible. This is not to say, of course, that Gewirth has demonstrated that any rational agent must accept the need for at least this level of welfare, but merely that the capacities are causally dependent upon it.

Once the subject's freedom is abstracted from substantive pursuits, it is hard to see what possible courses of action could impede or restrain the capacities themselves - all the actions do is limit the available ends, which we have been given no reason for believing to be valuable in themselves. Against this, it may be objected that the exercise of the rational or moral capacities will be severely restricted if others' behaviour is not itself subject to appropriate constraints; it is not the existence of these bare capacities which provides the basis for agents' moral entitlements (it may be added), but their practical exercise. If the entitlements consist not in the capacities themselves but in their exercise, there may be grounds for guaranteeing the maximization of the latter by prohibiting certain forms of

interference with agents' freedom of action.

This line of response is all right as far as it goes. If no limitations at all are placed on how agents may act, it is obvious that the ability of some to pursue their ends may be impaired. Nor is the point answered by the consideration that, after all, the imposition of such limitations is similarly going to restrict agents' ability to pursue their ends. It is possible, as in Prisoner's Dilemma cases, to imagine situations in which the unrestricted pursuit of self-interest leads to worse self-interested outcomes for all agents than if some control or at least coordination had been imposed. This does not of course establish what limitations are to be placed on agents' freedom of action, but it would provide some argument from rational considerations alone for the principle that agents were committed to accepting some restrictions on their actions.

This would not fully answer the original criticism, that Rawls and Gewirth fail to show how their view of the rational or moral agent can produce the particular action-prescriptions for which they argue. But it would retain some hold on the idea of agents' moral worth consisting in a rational capacity, while not leaving the agent beyond the reach of all possible actions in the world. Some further argument would be required to specify the nature of the limitations: this chapter has argued that they cannot be provided by considerations of practical rationality alone. What we might be left with are sets of compossible actions, each one of which can, when taken by itself, consistently be pursued by agents. But consistency by itself is not sufficient to decide between these sets; though it might be possible to define from these some core set of actions or goods, analogous to Rawls' primary goods, or Gewirth's generic goods of action, which was common to all these sets.

It should readily be observed that the line of criticism pursued so far in this section strongly resembles the objections raised against Kantian theory in the last chapter, and it will be helpful to examine the respects in which Gewirth's and Rawls' work both reflect the difficulties in Kantian theory, and depart from it in certain respects. In the last chapter it was argued that Kant's construction of the moral subject leaves questions unanswered both about how the moral subject can act in the phenomenal world, and about the factors which can influence the

subject's decisions about which actions to perform. The objection was the Kantian theory preserved the metaphysical freedom of the moral subject by drawing a sharp line between the noumenal and phenomenal worlds, but at the cost of leaving inexplicable the practical dimension of moral agency. This was traced to Kant's account of the metaphysical subject of experience as lying wholly beyond phenomenal determination, and it might be supposed that, once this metaphysical background was removed, its associated problems might disappear with it.

Kant's difficulties arise, as the last chapter argued, from his insistence that free agency can only occur outside the phenomenal world: we cannot then explain how the moral subject can effect any actions in that world. Despite their avoidance and indeed explicit disavowal of Kantian metaphysics, however, both Rawls and Gewirth may be unable to avoid similar problems in their own theories. We have seen that these theories present a picture of the agent as being wholly separate from any contingent attributes when deciding on the principles of justice, or upon structures of rights. It is only on this basis that agents can be regarded as free choosers of their ends - as standing outside those ends.

Once this has been imposed as a condition of the agent's freedom, it is unclear how he is to make any decisions about what particular ends to pursue, as the notion of agency has been so narrowed as to remove any basis on which a choice might have been made. While this difficulty is not formally identical with the problem about moral agency in Kant's theory, Rawls and Gewirth seem also to be unable to explain how the agent can move from the purely formalistic account of the rational subject to the pursuit of concrete ends. In the same way as Kantian theory is unable to explain how the noumenal subject can act in the phenomenal world, modern liberal theories which seek to free the individual from determination by factors external to his or her condition as a rational agent have difficulties explaining how free agents can arrive at practical decisions.

There is also, as should now be clear, a common problem in both Kantian and neo-Kantian theory regarding the agent's ability to take decisions about action as a result of rational deliberation. Even leaving aside the metaphysical questions concerning agents' freedom, it still has to be shown how agents

could arrive at any decisions with practical effect about their rational obligations. In Kantian theory, this difficulty appeared in the form of an indefinite number of sets of consistently universalizable prescriptions for action: as Hegel pointed out, despite Kant's reliance on the character of institutions such as promising to derive the content of the categorical imperatives, and whatever may be the case in this example, it is at best highly doubtful whether Kant shows that universalizability is a sufficient condition for a proposed action to be morally acceptable.

Similarly, the procedural treatment of rationality presented by Rawls and Gewirth is unable, without introducing further assumptions, to deliver any concrete conclusions about moral requirements on action - the additional input being, in Rawls' case, the arguments based on decision-theoretic reasoning under uncertainty in the Original Position, and in Gewirth's case, a far more substantial and moralized view of the agent than Gewirth himself acknowledges.

Part of this indeterminacy in Kantian theory issues from uncertainty over the practical implications of the the moral reasons which are meant to guide agents when they are deciding what to do. In the formulation of the categorical imperative which enjoins agents to treat others never merely as a means but also as an end, it remains unclear what practical force this general injunction has, without further argument about what practically it involves to treat someone as an end in themselves. We saw in the last chapter that Kant's rigid separation of the noumenal and phenomenal worlds failed to explain how morality, in the form of categorical imperatives, could have any bearing on events in the phenomenal world, even though, on a common sense view, the content of any acceptable set of moral beliefs would be involved, to a very considerable degree, with the treatment of phenomenal objects, such as human bodies. It was therefore uncertain how the status of persons as noumena could be thought to have any bearing on the actions which moral agents were required to perform.

It should now be clear that the construction of the moral subject in Rawls' and Gewirth's writings runs into similar difficulties. For in these modern works the moral person or agent is too thinly characterized not only to be capable of practical

decision-making, as already argued, but also to be the critical factor influencing others' moral deliberations. Once again, in their efforts to present a subject free from contingent attributes, as an unconditioned pursuer of its ends, liberal theorists are unable to explain how a subject drawn so narrowly can be of such importance to moral and political decisions.

This is true even if, as Larmore³⁸ has recently insisted, we should regard liberalism as a purely political philosophy rather than some all-inclusive philosophy of man. Whether or not liberalism is best described in such terms, an awkward question remains about how the subject, thus defined, can appear as the principle factor influencing public policy and the structure of political institutions. Even if the so-called 'expressivist' view of the state is rejected,³⁹ as a contravention of the liberal theorist's sought-after moral neutrality, there still needs to be some explanation of how such pared-down political institutions as do exist are there for the sake of an individual defined in such a way that all the subjective sources of value which provide structure and meaning to persons in their lives are eliminated. This highlights one of the curious anomalies of liberal theory, that it holds up the individual as its supreme value, yet denies any part in deciding political arrangements to the conceptions of value which give individuals' lives meaning.

This is not to say that it is logically inconsistent to hold both that in some sense the individual is the main consideration in devising political arrangements, and that the values which individuals pursue or enact in their lives may be diverse and plural. But it is a peculiarity of neutralist theories like Larmore's that they remove all subjective sources of value from the political arena, while also claiming that there is a supreme value, the individual, which exists regardless of whether it does, or even can, figure in the practical concerns of actual persons.

This suggests that the relationship between moral pluralism and the neutrality which is seen by many modern liberal theorists as both a necessary and desirable feature of liberalism may not be as harmonious as often supposed. On standard theory, neutrality is invoked as the appropriate response to the presence of diverse conceptions of the human good. Since there is in fact a plurality of mutually incompatible beliefs about value held by

the members of society, with no correct or agreed mechanism for resolving conflicts about values, the best that can be hoped for is that the public authorities act as umpire between these values. Even if full impartiality proves impossible, the argument runs, neutrality is nonetheless the best available means of dealing with the heterogeneity of value.

The problem with this position is not merely that, as has already been suggested in previous chapters, neutrality is compromised if a particular moral ideal, that of the person, is adduced in support of neutrality - the particular moral values being respected as those held by persons. Beyond this, and the implicit restrictions on the scope within which neutrality can operate, there is the problem that neutrality itself looks like a particular value or ideal, or at least its invocation as the best means of arbitrating between diverse values has this appearance.

Here a defender of political neutrality might reach for a distinction of levels, and claim that a line must be drawn between the area within which first-order conceptions of value operate (the private), and on the other hand the public principles which are meant to regulate the competition between them. But it is not clear that the scope of pluralism can be delimited this easily. It is reasonable to think that such a distinction between public and private spheres is part of the conceptual apparatus of a particular value-system, namely liberalism, and one which other systems may reject. This is all the more obvious if the argument for neutrality has itself to rely on claims about value.

For these reasons, it seems questionable whether neutrality is sanctioned by a belief in moral pluralism. In the case of Rawls and Gewirth, this doubt emerges in the failure of the supposedly consensual ideal of the moral person, or of a purely descriptive notion of agency, to yield the moral conclusions which these writers seek. There is no route leading from a morally neutral (either non-moral, or unanimously approved) starting-point into a conclusion which defends moral pluralism. For either this starting-point into a conclusion which supports moral pluralism. For either this starting-point is moralized in advance, in which case neutrality stands in danger of being violated at the outset, or else it is genuinely free of moral commitments, and then seems to be incapable of delivering any moral conclusions with

practical force. This is what we found in transferring the criticisms made of Kantian moral theory to his modern successors, that the non-moralized self, while arguably neutral in its moral content, was too insubstantial to issue in principles of justice, or sets of rights, with significant content.

The previous chapter concluded by examining the impact of Kant's moral theory on his political philosophy. It gave reasons for doubting whether, in view of Kant's overriding emphasis on the primacy of moral over political considerations in action, there was any autonomy left to the activity of politics in his theory. Since politics is defined as a sphere of action whose imperatives are hypothetical rather than categorical, any conclusion of political reasoning must in principle be defeasible if there is some conflicting moral requirement which bears on the circumstances of the decision. The conclusion was, as this chapter has tried to re-emphasize, that neutrality can only figure in liberal theory at the cost of compromising some of its other commitments, notably its bias towards a sceptical epistemology: this is another version of the problem already considered in this chapter, that it is hard to construct a neutral justification for neutrality itself, which does not undermine moral pluralism.⁴⁰

Thus if there is no commonly-accepted method of resolving conflicts between values, it is not obvious that it is open, even given an accepted understanding of 'neutrality', to appeal to this as a pragmatic solution to the dispute. Rawls and Gewirth both seem to be attracted towards a broadly neutralist position, but are unable to show how it can be combined with a full commitment to pluralism. One of the ways in which this comes out, as the remainder of this chapter will argue, is in an attempt, for which they provide little argument, to limit the scope within which pluralism operates.

This emerges most obviously in a limiting of the scope of politics in these theories. For Rawls the political institutions of his well-ordered society are designed to embody the shared conceptions of value which his constructivist method uses in formulating the principles of justice. Rawls sees value-based conflicts, such as those over the implications of value for public policy, as the fundamental feature of political life in modern-day democracies; he takes his task as being, accordingly,

to devise a set of political arrangements which can as far as possible obviate these conflicts. In this connection Rawls identifies what he takes to be the fundamental shared moral notion held in contemporary western societies, that of the person, and elaborates from it an ideal of citizenship, or 'political conception of the person'.⁴¹ The point is that citizenship embodies a publicly disseminated ideal of the moral claims of persons, and helps to define the space within which individuals' private conceptions of the good may be pursued. Thus Rawls presents a picture of a diversity of individual notions of the good, mediated by the consensual moral values enshrined in the public institutions.

One question which this approach raises is whether in the face of plural moral beliefs the consensus Rawls seeks is broad enough to produce any significant principles of justice, or institutions to implement them. If no sufficient wide basis exists on which to found these institutions, there will be no justification for distinguishing between the neutral public sphere, embodying the shared ideals, and the private realm in which individuals are free to pursue their favoured conceptions of the good life. For in this case one of the points at issue between different values, or systems of value, may be precisely where, if anywhere, the line is to be drawn between private and public spheres or, more fundamentally, whether such a distinction is desirable at all. Given the initial diversity of beliefs, there is nothing to prevent its taking the form of disagreement over the extent to which public institutions can justifiably intervene in persons' lives, and in what ways.

These are, or can easily become, political questions. It is, for example, a familiar point of disagreement between advocates of capitalist and socialist systems of production to what extent the state should intervene in the market to redress inequalities of income distribution brought about by its uncontrolled operation. Rawls may claim, in response, that the principles of justice arising from his argument are general enough to accommodate disputes of this nature. But if they are, it is hard to maintain that the principles model our shared values accurately enough to be able to mediate disputes over freedom and equality, as enacted over a particular issue of public policy. One of the ways in which value-based conflict can emerge, as has

already been observed, is in disagreement over the practical courses of action sanctioned by values which in themselves are shared. These facts weaken the ability of the agreed principles to settle political disputes when they arise. If, on the other hand, the principles are sharpened so as to be capable of giving concrete answers to questions of policy, it then becomes doubtful whether they really embody shared values.

It is persuasive to think of real-life political debate as providing a forum in which these disputes over value are fought: political argument is partly about the consequences of policy for moral values, and the rational interminability of such argument testifies to the insoluble nature of conflicts between values invoked during its course. At least part of the time political debate implicitly or explicitly is about the significance which politics itself should be given in human concerns. This alone is enough to cast doubt on the neat separation of institutional public consensus from the diversity of values permitted at the private level.

None of this need be taken to imply some adversarial model of political debate, rather than a consensual model. Political argument can occur within a context of shared values. But even if, at a general level, values such as freedom and equality are agreed by political adversaries, much scope remains for dispute over the relative weights to be assigned them in particular policy issues and the extent to which a proposed course of action will involve favouring one action at the expense of another. While political debate is not, much of the time, about the single-minded promotion by one side of some value no less single-mindedly rejected by the other, a lot of room is still left for political manoeuvre within a broadly shared set of values. Some of these dimensions of political engagement seem inadequately served by Rawls' theory, which assumes, on the whole, that since argument in politics is concerned with questions of value, it is only necessary to arrive at a consensual set of values in order to render such arguments obsolete.

These gaps in Rawls' treatment of politics can perhaps be traced to his Kantian model of moral theory. In part they can be accounted for by Kant's rigid subordination of political to moral concerns in practical reasoning. Rawls' theory reproduces this order of priority both in Rawls' belief in the methodological

primacy of moral theory over other branches of philosophy⁴² and in the related assumption that once theoretical modelling has been carried out in line with the constructivist procedure, there will be little left for political processes themselves to contribute towards deciding questions of social distribution.

In part, also, Rawls' Kantianism emerges in the supposition that certain kinds of consideration operate in practical reasoning so as to override or eliminate entirely others with which they might come into conflict. This is a distinct point from the one about the priority of morality over politics, since it applies within, as well as between, these deliberative spheres: what is in question here is the idea that certain sorts of reason can enjoy an absolute priority with respect to others even within some such sphere. The most obvious example of this in Rawls' theory is the lexical ordering of the two principles of justice, and between the two parts of the second principle. But it seems to be an endemic feature of political engagement that disagreement can occur over just such questions of priority. The notion of a hierarchy of concerns is internal to a particular theory of what human beings are or should be, and cannot be produced merely by constructing a model of people's shared beliefs about value.

A similar desire to reduce practical concerns to a schematic deliberative hierarchy is apparent in some of Gewirth's work, notably his essay 'Are there any absolute rights?'.⁴³ There Gewirth tries to show that certain rights are absolute in the sense that they can in no circumstances be justifiably overridden whatever other costs are incurred by respecting them. Despite Gewirth's distinction between 'abstract' and 'concrete' forms of moral absolutism,⁴⁴ perhaps made with Kant in mind, it is unclear how far this distinction can be maintained in practice, as 'the concrete absolutist is concerned with consequences and empirical connections, but always within the limits of the right he upholds as absolute'.⁴⁵

Rights are one way of assigning special deliberative weight to certain forms of concern. Gewirth seems to assume, however, that these rights must be absolute if the moral claims about persons which they embody are to be defensible against an aggregative moral theory such as utilitarianism. It does not follow from saying that there are no rights which cannot be

overridden in some circumstances that all deliberative concerns are projectible onto some unitary scale of value, such as utility. It may be that the moral concerns which give rise to these rights conflict in certain situations with other such concerns, and there is no neutral method of resolving the conflict. In cases like this, it cannot be true that the concerns that the rights claim expresses must 'trump'⁴⁶ the others, since it is a feature of these situations that the conflicts which they involve are, in some sense, insoluble. This is not utilitarianism, as it denies that the conflict can be settled by reference to some single standard, whether of rights or utility. Gewirth seems to be drawn to something very like a Kantian hierarchy of concerns, his denials notwithstanding.

While the reduction of politics to moral theory is less obvious in Gewirth than in Rawls, there remains a fairly clear reductivist bias in the attempted derivation of rights from the generic features of agency - where, again, the influence of Kantian thinking is detectable. We find, in Gewirth, a similar choice to that which Rawls faces in spelling out the political implications of his theory. Either the rights are framed at such a general level as to be nearly vacuous, in which case the theory fails to deliver conclusions with any practical force, the content of the rights being specified by political decision; or else there is practical content in the rights, but this pre-empts any debate over their practical significance.

Gewirth does not perhaps wish to produce a very general theory for tackling issues of distribution in the way Rawls seeks to do, and it could be argued from this that the area within which the rights generated by the PGC operate is supposed to be closely circumscribed. But this does not detract from the attempt to place absolute limits on what may in practical and hence political terms be done, while if the generic goods are indeed about the causal conditions of agency, there will be important political questions to be answered about what agents need in order to satisfy these conditions, and about what political measures are justifiable to ensure that agents receive the relevant goods. These are questions to which liberal theory returns a notoriously uncertain answer.

Rawls' and Gewirth's Kantian debts are most obvious in their efforts to use Kant's moral theory to ground liberal

individualism. But their attempts to clear a space within which individuals may pursue their private conceptions of the good life leaves it unclear why, given that the aim is to allow this diversity of value, the value of individualism itself should be privileged in this way. It is not through an enforced restriction of its field of operation, but in recognizing the conflict of values which pervades politics, that pluralism is most fully and coherently expressed.

Notes

1. See John Rawls, A Theory of Justice (Oxford 1971) and, more recently, his Dewey lectures entitled 'Kantian Constructivism in Moral Theory', Journal of Philosophy 1980, pp515-572; Robert Nozick, Anarchy, State, and Utopia (Oxford 1974); Charles Fried, An Anatomy of Values (Harvard 1970); Thomas Nagel, The possibility of Altruism (Oxford 1970).

2. But see Rawls' article 'Justice as Fairness: Political not Metaphysical' in Philosophy and Public Affairs 1985, pp223-251.

3. See Gewirth's Reason and Morality (Chicago 1978), eg p26.

4. Gewirth does this in his collection of essays Human Rights: Essays on Justification and Applications (Chicago 1983).

5. See the references cited in note 1, and the development of the theory in: 'A Well-Ordered Society', in P. Laslett and J. Fishkin (eds.), Philosophy, Politics and Society, Fifth Series; 'Social Utility and the Primary Goods', in B. Williams and A. Sen (eds.), Utilitarianism and Beyond (Cambridge 1982); 'The Basic Liberties and Their Priority', in S. McMurrin (ed.), The Tanner Lectures on Human Values 1982.

6. See 'Kantian Constructivism in Moral Theory', op. cit.

7. Op. cit., p4.

8. Ibid., p46.
9. Op. cit., p570.
10. Ibid., p556.
11. Ibid., p560.
12. Ibid.
13. Notably in 'Justice as Fairness', op.cit.; see also 'Kantian Constructivism', op. cit., p561.
14. Though the 'Government House' utilitarianism to which Sidgwick perhaps subscribed, and which was considered in the first chapter, offers an example of institutions whose operations are non-public and yet (the argument goes) in some sense are so arranged as to satisfy citizens' desires.
15. Op. cit., p137.
16. In 'Justice as Fairness', op. cit.
17. Ibid., p244.
18. In his Liberalism and the Limits of Justice (Cambridge 1982).
19. As, eg, Derek Parfit has sought to spell out in his utilitarian critique of common-sense theories of personal identity in Reasons and Persons (Oxford 1984).
20. Op. cit., p63.
21. Ibid., p26.
22. Ibid., p61.
23. Ibid.
24. Sandel op. cit., p21.

25. See Rawls' 'The Independence of Moral Theory', Proceedings and Addresses of the American Philosophical Association 1975, for a statement of this view.
26. Bernard Williams takes this course in his critique of Kantianism in Ethics and the Limits of Philosophy (London 1985), Chapter 4.
27. Disregarding complications introduced by the presence of a referentially opaque context in practical reasoning.
28. For this term see Bernard Williams' 'Internal and External Reasons', repr. in his Moral Luck (Cambridge 1981).
29. As Charles Larmore points out in his Patterns of Moral Complexity (Cambridge 1987), p80.
30. Op. cit., p517.
31. Op. cit., p16.
32. Sandel, op. cit., p21.
33. Op. cit., p544.
34. Williams' term, from 'Internal and External Reasons', op. cit.
35. Reason and Morality, pp136-7; p264; p324.
36. But, at any rate in Rawls' case, the exceptions are so trivial as to include among the primary goods the conditions of all the kinds of good rational agents are likely to wish to pursue.
37. It is compatible with a neutralist position to accept that this may not be very far.
38. Op. cit.

39. Ibid., p73.

40. Larmore, ibid., pp50-5, considers this objection, finding the solution in a version of the 'neutral dialogue' approach favoured by Bruce Ackerman in his Social Justice and the Liberal State (Chicago 1980).

41. 'Justice as Fairness', p244.

42. See Rawls' 'The Independence of Moral Theory'.

43. Repr. in J. Waldron (ed.), Theories of Rights (Oxford 1984).

44. Ibid., p106.

45. Ibid.

46. For this term see Dworkin in ibid., 'Rights as Trumps'.

Chapter Four

Instrumentalism: Some Antecedents

I

Instrumentalism, the view of politics constructed from the materials assembled in the foregoing chapters, has little to say about politics as an autonomous activity. Where utilitarianism and neo-Kantianism, the instrumentalist theories considered in previous chapters, have found anything to say about politics at all, their attention has been directed mainly to institutional analysis - for example, in Rawls' writings,¹ to specifying the political institutions of the perfectly just or well-ordered society. Utilitarian work, on the other hand, divides itself between technical discussions, in welfare economics, of the conditions within which supposedly optimific decision-procedures for public choice can operate, and, more recently, in the work of Parfit, Singer, Harris and others,² concentration on the details of policy, with little consideration of the processes by which policy is decided.

What is lacking in both the utilitarian and neo-Kantian approaches is an appreciation of how political processes can modify policy by altering the options available for practical action and by altering agents' understanding of the relationship between policy and value-based commitments. The present chapter sets out to expose some of the central assumptions underlying instrumentalist theory and, by suggesting alternatives to these assumptions, to point towards a different, non-instrumental view of politics.

Traditionally liberal theory has seen politics as occupying a confined - in some recent treatments, vanishingly small - space within human concerns, in an area circumscribed by other, private sources of value and obligation. In the theory recently expounded, for example, by Larmore,³ politics is ideally a neutral residuum left over when the partisan commitments permissible in private life alone have been subtracted. For Larmore, political neutrality is the corollary of value-pluralism: it is because values are irreducibly heterogeneous that public institutions and policy must, as far as possible, remain neutral between private conceptions of the good - or, at any rate, departures from

neutrality must be justified by reference to public procedures.

In direct contradiction to this, the present chapter will argue that accepting the heterogeneity of value demands the rejection of political neutrality as conceived by modern liberalism. One aspect of this heterogeneity is to be found in the indeterminate relationship between different values, and between values and public policy. Concentration, in liberal pluralist philosophy, on the grander cases of abstract confrontation between rival sets of values has obscured the more humdrum possibility that such indeterminacy may arise in questions of practical agency also. Arbitrating between private conceptions of value becomes far more difficult if there is no value-free or agreed method of regulating the competition between values in practical action. Moreover, the entire notion of a hard-and-fast division between public neutrality and private partisanship can only be sustained if these realms can be kept apart. But it is unclear that pluralism is compatible with such a demarcation.

This claim, that there can be no plausible neutrality in the presence of a commitment to pluralism, does not mean, as has sometimes been supposed, that all human concerns must become politicized (nor that any hitherto private concern may not in conceivable circumstances become politicized). To reject any monolithic distinction of public from private concerns is not to claim that everything is therefore reducible to the former - quite the contrary, since an acceptance of pluralism requires (as will be argued) that any single dimension of value cannot serve as that onto which all others are projected.

Pluralism is precisely the claim that value is not so projectible, and that different categories of value cannot stand to one another in any fixed or hierarchical relationship. To take a familiar example, it is not clear that there can be any determinate relationship within a pluralist account of value, between moral and political concerns, as separate categories of reason for action; in particular, there may not be any definitive means of resolving the conflicting claims, in practical deliberation, of moral as against political considerations. If morality enjoyed some form of 'lexical'⁴ priority over other categories of practical reasons, pluralism could not be true, as the latter is just the denial that there is this sort of hierarchical relationship between different kinds of reasons for

action.

At this point it is advisable to draw a distinction, which will be of some importance in the following discussion, between different forms of pluralism. It is fair to say that most philosophical discussions of pluralism have worked with the assumption that pluralism is a notion applicable within moral theory: that is, it has been assumed that pluralism identifies a conflict, or incommensurability, or at least a multiplicity, of ultimate goods - usually, goods which some persons in society actually pursue, or might reasonably be expected to pursue. This definition obviously passes over numerous issues relevant to a philosophical understanding of pluralism, concerning the nature of the pluralistic relationship between these goods - in what circumstances, for example, pluralistic goods are involved in practical conflicts, and whether incommensurability between goods is a necessary condition of pluralism.⁵

On this issue, of the relationship between the pluralistic goods, it is worth noting that questions of practical policy arise even without the more dramatic forms of confrontation envisaged by some discussions of the subject. It is quite possible, for example, that there may exist in a society (a unitary sovereign state, say) a number of discrete communities each of which encapsulates relatively coherent and homogeneous sets of values, with little interaction between these communities. In such a case, political questions may well arise concerning, for instance, funding levels for cultural activities pursued by these different groups, without the values themselves coming into direct conflict. Of course, the scarcity of public resources itself may give rise to conflict, but this is not because of some intrinsic incompatibility between the values themselves - rather the conflict arises from the contingent fact that groups espousing them are in competition for scarce public resources.

Obviously matters like this require much fuller investigation. For the present purposes, the point is that politics can generate this possibility of conflict, and it is this that gives rise to the broader sense of pluralism to be examined in the present chapter. On this broader understanding, both the weaker claims of pluralism in the more restricted sense explained above, concerning the multiplicity of goods, and the stronger claims, concerning conflict, are accepted. But the present account differs from

neutralist theory in denying that politics provides a forum within which these conflicts can be resolved - rather the very existence of politics testifies to their rational insolubility. Accordingly, the broader understanding of pluralism sees the political process as an expression of pluralistic conflict rather than a means of resolving it. The conflict between these rival conceptions is manifested by their engaging in competition for power. There is thus no neutralist solution to value-based conflict, as envisaged by neutralist theory.

To this distinction of forms of pluralism corresponds a further distinction, of a methodological nature, drawn by Rawls in his paper 'The Independence of Moral Theory'.⁶ There Rawls argues for the philosophical subordination of non-moral and in particular political concerns to moral ones. At least part of the inspiration for this claim comes from the thought that pluralism exists because persons are diverse in their moral outlooks, and any political attempt to impose a particular moral outlook on persons across this diversity is likely to violate their moral autonomy as agents. As a result, the moral claims of persons are regarded as fundamental (Gewirth's rights-based position parallels Rawls in this regard⁷), and politics can only legitimately operate within these moral constraints.

In contrast with this methodological claim, the present chapter will argue that there is no justification for this priority, and that it threatens to violate the very neutrality which writers like Rawls have aimed for in their political theory. For on the broader variant of pluralism, the form of consensual value sought by Rawls in his efforts to provide a moral framework for political institutions either does not exist or else is insufficient to yield the political solutions which he wants from his theory. Part of what the broader version of pluralism means is that there can be no a priori ordering of this kind, based on subscription to an allegedly foundational moral principle. By contrast, it is part of the function of political activity to enlighten agents about their rational commitments - including what they have good reason to value. This being so, the present chapter will argue that there can be no pre-political ordering of values of the kind Rawls is aiming for.

There is then little reason, when we accept the wider understanding of pluralism, to assume that morality must enjoy

'lexical' priority over other practical concerns. Pluralism can only be combined with this priority ordering if it is taken in its narrower sense, as a notion operative only within moral theory: in that case, it can be reconciled with the lexical structuring, as Rawls' own theory demonstrates. But once we move to the alternative version of pluralism, which questions any a priori hierarchy among deliberative concerns, it is hard to see what justification there could be for Rawls' claims. Pluralism, on the wider definition, just is the rejection of a hierarchy of concerns, and the structuring of public institutions by reference to it.

It is of course true that lexical orderings can be a mark of incommensurability between goods: one way of indicating this relationship is by claiming that no amount of one good will compensate for, or balance, any amount of the other. Money and love, for example, have sometimes been assumed to be such, though the direction in which the lexical ordering is taken in this case is doubtless a matter of personal preference. But the significant point is that this, as in other cases of incommensurability, indicates not so much the possibility of lexical orderings between the goods, as that neither may be translatable into the currency of the other; this comes out also in the fact that a person may accept both as goods without accepting any such ordering between them. While incommensurability between goods is, as has been suggested above, a mark of pluralism, it does not follow that the pluralistic expression of this relationship must taken the form of a lexical ordering between the goods - nor between categories of goods.

Given at least the possibility of the broader interpretation of pluralism, it is hard to see how political neutrality can survive if the notion of a deliberative hierarchy is challenged. For in the absence of any such ordering of concerns, the rationale disappears for thinking that the fundamental political task is to arbitrate between different conceptions of the good. That thought assumes that these conceptions have value, and are of political account, as the ideals of life held by persons. Since persons are the supreme bearers of moral worth, all other concerns must be subordinated to that of facilitating persons' pursuit of these ideals. Neutrality, then, is relativized to a particular conception of the supreme good, which itself is non-neutral; in

other words, neutrality cannot justify its own foundational principle.

The search for a neutral justification for political decisions has dominated much recent liberal theory. One prominent form which this has taken, in the work of Ackerman,⁸ Rawls, and others, has been the search for a mode of justifying political arrangements whose neutrality is guaranteed by the neutral procedures giving rise to them, and specifically by neutral conditions of dialogue between participants in the political decisions which determine those arrangements. If the conditions in which the protagonists confront one another are neutral between their separate conceptions of the good, it is claimed, the resulting political dispensation must be mutually acceptable to them so long as they adhere to rational criteria of justification.

The neutralist approach thus presents a picture of political debate as issuing in rational agreement, at least when conducted in ideal conditions. In the next section this position is located within the context of the Enlightenment philosophical ideas which inform much utilitarian and neo-Kantian thinking, while examining instrumentalism's view of political argument. After that I will outline an alternative account which questions both the neutralist assumptions underlying instrumentalism and, more specifically, the psychological premisses on which (as it will be argued) its characteristic positions are based. By the end we should have a better grasp both of why the instrumentalist project of providing neutral grounds for political debate is doomed to failure, and the general shape of an alternative view which stresses the rational interminability of political argument.

II

We have already seen that the instrumental view of politics holds, both in its utilitarian and neo-Kantian versions, that the ends of political action are given extra-politically. In the case of utilitarianism, this takes the form of aggregating preferences for the purposes of public policy, while the Kantian brand of instrumentalism assumes, as has been indicated, that political agents operate within a system of morality and pursue, at best, only contingent objectives of their own. In both cases, the format

of public decision-making is held to be isomorphic with that undertaken by the individual agent. On neither Kantian nor utilitarian theory is there any distinctive contribution made to the process by political debate as a collective deliberative process: the assumption is that the position of an individual agent is paradigmatic for practical reasoning in general, as similar structures apply at both the individual and collective levels. Instrumentalism, then, maintains that the ends of political action are given extra-politically: political processes make no independent contribution towards deciding the final goals of political action. All that these processes do is to play an executive role in deciding how the given ends are best achieved.

It is helpful, in understanding the implications of instrumentalism, to consider the Enlightenment philosophical psychology which lies behind much of its thinking. One root of the instrumentalist assumption that the ends of politics are given extra-politically is to be found in Humean or quasi-Humean philosophical psychology: on the Humean view, all that 'reason' can contribute to practical deliberation is a decision about how best to satisfy 'passions' which are themselves regarded as incorrigible data of experience:

a passion is an original existence or, if you will, modification of existence, and contains not any representative quality, which renders it a copy of any other existence or modification. When I am angry, I am actually possessed with the passion, and in that emotion have no more a reference to any other object, than when I am thirsty, or sick, or more than five feet high. 'Tis impossible, therefore, that this passion can be opposed by, or be contradictory to truth and reason: since this contradiction consists in the disagreement of ideas, considered⁹ as copies, with those objects, which they represent.

Hume's position is, then, that only reason, being the faculty concerned with the comparison of ideas, can be representational; and thus the passions, as brute 'existences' in the world, cannot be opposed to reason, since they are incapable of representing (and so by the same token of misrepresenting) states of affairs. It follows that any attempt to persuade someone on rational grounds to abandon a preference which they hold, or to adopt another which they do not, must be futile. The notion of rational persuasion has no place in Humean philosophical discourse.

On this view, the ineffectuality of political argument merely

results from the claim that any attempt at rational argument must be ineffectual, where it involves an effort by the speaker to persuade his audience to abandon their preferences. This is not necessarily because the Humean is committed to thinking that preferences can never change, though the theory leaves it unclear how this is supposed to happen. Nor is it that Hume is prevented from giving some explanation of how preferences can be changed (there can be a rational account, within Humean theory, of how to go about brainwashing somebody, given that this is what, on other grounds, one wants to do). Rather the unclarity arises when we ask why the Humean would want to go about modifying preferences in the first place: since preferences are rationally unfounded, it is hard to see on what basis any policy could be formulated with such a goal in view, or how the grounds on which such a policy was formulated could bear any rational relationship to the grounds on which its political proponents set out to persuade an audience of its merits.

To this analysis it could be objected that there are variants of modern utilitarian theory which draw on Humean premisses for their psychological basis, but allow that preferences may, after all, undergo rational modification. Rather than seeing preferences as static and given, some more sophisticated brands of utilitarianism, such as Regan's,¹⁰ allow for the possibility that preferences can be modified to meet external changes in an agent's decision-making environment. In Regan's cooperative utilitarian theory, for example, agents are presented with a schedule of possible outcomes and are supposed to adapt their own behaviour to maximize utility in the given conditions.

But this kind of adaptive behaviour does not show that there is any rational modification of preferences at work here. All we are presented with in Regan's theory is a table of conditional outcomes. But even apart from the fact that the outcomes are not different preferences, but merely utility numbers, there is no ground for inferring that preferences are modifiable, rather than just being conditional upon external information. To move to a different point in a given schedule of preferences is not the same as to regard the schedule itself as rationally negotiable.

Admittedly, this does raise the question how Humean theory copes even with the less dramatic possibility of conditional preferences. It is Hume's view that the most which reason can

contribute to practical deliberation is the cognitive servicing of desires - the latter being, in his description, 'original existences'¹¹ which are both incorrigible by reason alone and which provide the motivational force required to induce an agent to act. All that reason can do is inform the agent of which objects to secure in order to satisfy given passions, and how to go about securing those objects. Without the impulse to action given by desires, agency would be impossible. Thus, as Hume says, any deliberative chain must ultimately lead back to some passion or desire which cannot be accounted for by reason:

Ask a man why he uses exercise; he will answer, because he desires to keep his health. If you then enquire, why he desires health, he will readily reply, because sickness is painful. If you push your enquiries farther, and desire a reason why he hates pain, it is impossible he can ever give any. This is an ultimate end, and is never referred to any object...It is impossible there can be a progress in infinitum; and that one thing can always be a reason why another is desired. Something must be desirable on its own account, and because of its immediate accord or disaccord with human sentiment and affection.¹²

From this it is clear that Hume thought reason was restricted to matters of purely instrumental calculation. In view of this restriction, there can be no question of any rational disputes over ultimate aims, of the form which may be plausibly thought to characterize real-life political conflicts. For, on the Humean-inspired instrumentalist view, conflicts of this kind are impossible. The only room for disagreement - in principle resolvable given accepted procedures for gathering and processing information - is over how best to achieve goals which do not themselves admit of rational dispute.

One clear limitation imposed by Humean analysis on the scope of political argument is the latter's inability to give any convincing account of political persuasion. It is a matter of common observation that politicians frequently do draw upon a variety of techniques to persuade their opponents, or some wider audience, that they should support a proposed policy. The Humean gloss on such possibilities has, presumably, to be that these efforts are either pointless or else aim to inform their audience that the policy in question really is that which best fulfils an objective agreed in advance. There are occasions when something like this picture is true: there can, for example, be genuine

political disagreement over which policy most effectively furthers some agreed aim, such as cutting crime or increasing employment. But political dispute, where it does not involve conflicts over final ends - for example over what form of society we should live in, with what productive relationships, and what values that society should express or promote - is very often about balancing conflicting final objectives, and this, according to the Humean view, is a task in which reason can offer us no assistance.

It is important to bear in mind the impact of Humean assumptions on the form as well as the aims of political argument. As the passage quoted above from the Treatise indicates, Hume explicitly denied that the passions have a representational or intentional component: a consequence of this claim is that emotional states cannot be seen as rational (or irrational) in Hume's view, since to be in such a state is to 'have no more reference to any other object, than when I am thirsty, or sick, or more than five feet high'.¹³ The passions, being non-representational, are incapable of truth and falsity, and so cannot be modified by rational argument.

As a result, the role of rhetoric in political argument, which familiarly relies in its repertoire of techniques on an appeal to the audience's emotions, has been discounted in Humean-inspired instrumentalist writings. The emotions are seen, from this perspective, as fundamentally non-rational states, and so their function in any purportedly rational structure of argument becomes obscured. The most that the rhetorician can hope to achieve in the way of emotional persuasion is to by-pass the audience's rational faculties and appeal directly to their non-rational passions or 'affections'.

This is not to say that such appeals will not be successful - though, as has already been remarked, it remains unclear in the theory how these 'original existences' can be modified. But even where it succeeds in its aims, political rhetoric cannot, in its nature, conform to the canons of rational argument which apply, for example, in philosophical discourse, because its appeal is made to a psychological faculty which is essentially non-rational. This helps to explain the neglect of rhetoric within political persuasion in recent theorizing, and the obloquy in which it is held in academic discussion. Whatever there may be to say about rhetorical argument as an appeal to the emotions, there can be

nothing to say, according to Humean philosophical psychology, about it as a rational form of persuasion.

While it may be superficially less obvious that Kantianism, which the present work has identified as the other main inspiration for instrumentalist theory, leads to similarly dismissive consequences for political argument, and an equally limited view of the scope of political action, the applications of Kantian moral theory in recent political philosophy lead to conclusions similar to those already attributed to modern Humean-inspired writing. This is to be traced to common features of Hume's and Kant's theories, and particularly to shared psychological assumptions, the great and evident differences between them notwithstanding. It will be helpful to trace in more detail the common ground between Humean and Kantian theory, and to indicate the latter's influence in modern instrumentalist writings.

As has already been argued in the Introduction to this chapter, Kantianism asserts a rigid priority between categories of reason for action. Because of this priority, non-moral categories of reasons, including, in Kant's own demarcation, political reasons, occupy a necessarily subordinate position with respect to the moral category. For Kant, moral reasons must apply categorically if they are to apply at all. This means that there can be no permissible trade-off between moral and non-moral considerations in practical reasoning (or, for that matter, between different moral imperatives with conflicting practical consequences). To allow the possibility that moral imperatives could be overridden in certain circumstances would in Kant's view remove the sole basis on which morality can apply, that is, unconditionally.

The result of Kant's insistence on the categorical status of the moral within deliberation is that there is in his theory a stark division between the grounds on which moral obligations hold, and agents' motivations for acting morally. These cannot, except fortuitously, coincide in Kantian theory. To make an agent's reasons for acting morally contingent upon his subjective motivation for doing so would be to subvert the categorical nature of the moral imperatives. For Kant, understanding the character of moral reasons for action is precisely to see those reasons as holding independently of any agent's subjective inclination to act on them.¹⁴

In the work of Kant's modern followers, such as Rawls, this emphasis on the categorical grounds of moral imperatives emerges in the methodological subordination of political to moral theory, as already mentioned. In Rawls' work, the priority of the moral is seen in his efforts to derive political institutions from supposedly consensual moral notions. These notions are meant to structure the processes by which the parties to the Original Position deliberate about the ideal form of society and the social and political arrangements which result from these deliberations. They operate, for example, so as to exclude from the discussion any information which could lead the parties to conclusions incompatible with the inviolable status of persons, which Rawls sees as the prime consensual moral notion held in contemporary (western) society. When the deliberations are over, the resulting principles of justice and the institutions embodying them are intended to express this fundamental moral concern for persons.

For the present purposes the question raised by Rawls' and other neo-Kantian theory is how far either the procedure by which the institutions of the just society are determined or the society itself allows for the possibility of political dispute, and hence for rational persuasion. From the neutralist perspective of Rawls' theory, it is unclear that either the process or its outcome leaves any room for persistent value-based conflict of the form witnessed in actual political argument. Since the theory is meant to express shared moral convictions, such conflict can at most occur within a framework of given values and aims - to act within the values, the priorities between values, embodied in Rawls' two principles of justice.

Thus political action is circumscribed by shared goals, much as in Kant's own theory the content of political action is defined residually, as what remains when the claims of morality have been taken into account. Beyond purely executive matters, such as deciding on appropriate allocations of primary goods in accordance with the two principles, it is difficult to see what part endemic political disagreement can play in Rawls' ideal society. The fundamental conflicts of values present in actual political argument have been banished by institutionalizing the common beliefs thought to underlie these conflicts.

It might be said in response that there could still be limited political disagreement in a Rawlsian society, concerning non-

consensual values (i.e. those not relating to moral personality), and its limitations imposed by an acceptance that discussion was framed by the principles embodying the supreme value. As already argued in the Introduction, the primacy of a particular value does not preclude incommensurability between other values. Such questions as the indexing of the primary goods could give rise to political disagreement over basic values, or at least over the practical implications of value in policy-making. Given this possibility, it may be said, there is, after all, room for political disputes in the well-ordered society, and of a familiar kind.

One problem this answer faces is that it is hard to recover any very substantial shared moral content from political dispute as it occurs in modern democratic societies. This is not to say that such dispute often involves the wholesale repudiation by one side of the values espoused by the other: conflicts over the practical weight to be assigned to considerations of freedom as against those of equality, for example, seldom take the form of rejection by one party of freedom as a value, balanced against the no less wholehearted rejection by the other of equality. The lexical ordering of the two principles, and the very idea that there must be some shared nexus of value in our notion of personhood, is difficult to square with the fact that political disagreements, and the grounds on which arguments are made in the course of them, so often belie any claim that certain values enjoy the supremacy which Rawls accords them. The existence of shared values does not mean that there is a fixed hierarchy between values, to be observed by the participants in political disputes. Rather these disputes often concern the relative practical weights to be given to these values in concrete questions of policy.

Rawls' theory faces a further problem in showing how the well-ordered society could witness recognizable political conflict, a problem shared by other neutralist theories such as Larmore's. Rawls and Larmore hope to combine a theory of political institutions as encapsulating certain agreed moral values, while maintaining a pluralistic theory of value within civil society; though in Larmore's theory the state is meant to be less expressive of substantial moral notions than its Rawlsian counterpart, he still finds himself promulgating the ideal of equal respect as being universal or unexceptionable enough to

stand as his basic constitutional principle. Similar remarks apply to Nozick's minimal state theory.¹⁵ The problem is that once the line has been drawn between the values embodied in public institutions and those pursued by private citizens, it is hard to see how the conflicts which, on any realistic pluralist theory, these private pursuits must generate, can be arbitrated by the state without compromising its neutrality.

It is persuasive, by contrast, to abandon as artificial any hard-and-fast division of public from private conceptions of value, and assume that political dispute can often take the form of confrontation between values which neutralist pluralists like Larmore see as irreducibly private. If this is right, it looks as though any consistent neutralist theory with a rigid public/private distinction will have to exclude from debate large areas of discussion which feature in actual political argument. These problems are multiplied if it is acknowledged that debate may proceed not only over the private values themselves, but also over the empirically necessary conditions for pursuing them.

These problems arise in Rawls' and Larmore's theories from the attempt to combine public neutrality with private pluralism. The separation between public and private spheres of action can only postpone a decision about where, and more pressingly on what grounds, the line is to be drawn. In contrast with his modern-day followers, Kant's own political and moral theory has the merit of confining all (non-aesthetic)¹⁶ value within the realm of morality, thereby avoiding the perplexities which beset his contemporary disciples in their efforts to describe the relationship between public and private values. This at least escapes the difficulty already mentioned above, that there is a broader interpretation of pluralism according to which there is no fixed ordering between classes of reasons for action. Kantian theory only achieves this, however, by cutting back still further the ground within which politics operates, and leaves it obscure how the commitment to different forms of value can inform political disputes. While it shares this problem with its historical precursor, neo-Kantian theory compounds its difficulties by coupling a pluralist account of value with an attempt to set once and for all the boundaries within which different kinds of value can operate.

It should now be clearer how, despite large and obvious

differences between them, the philosophical psychology of Hume and Kant both contribute towards instrumentalism's dismissal of political argument. That psychology works, in Kant's as in Hume's writings, with a schematic division of the soul into ratiocinative and motivational elements. Just as for Hume reason can provide no grounding, as we have seen, for motivation, so in Kant's moral theory the status of the rational agent as noumenon can only be preserved by rigorously abstracting from its motivated, and therefore phenomenal, characteristics. Looked at in this way, Kant's and Hume's moral theories emerge as opposed answers to the same question - what is the foundation of value - within a shared set of psychological assumptions.

In both cases the starting-point is this faculty division between reason and motivation, and the foundational task is to assign the source of evaluation to one or other of these faculties. Hume offers a credible account of how we can act from moral motivations, but his noncognitivist metaethics leaves unexplained the phenomenology of moral experience - in particular, the appearance, to those operating within its value-system, as a set of categorically binding requirements on conduct. By contrast, Kantianism, which offers a rational account of this aspect of the moral life, is less able, given its dual-aspect theory of the agent, to explain how reflection on the formal attributes of morality can motivate agents to act morally.

This section has tried to indicate the origins of some of instrumentalism's characteristic assumptions in Enlightenment philosophical psychology. Instrumentalism shows its debt to Kant and Hume in its assumption that the concerns present in the formation of policy generally are determined by extra-political considerations. Neo-Kantianism, particularly in its modern neutralist form, places value outside or prior to politics, so that the most that political processes can concern themselves with is how to work within an external set of values: values which, on the neutralist position, negatively define the space within which politics can operate, or which condition the objectives which politicians may aim at in their actions and the procedures by which these may be secured.

This approach pays little attention to the continuities between politics and other human concerns, and tells us little about the part which political argument can play in modifying our perception

of our interests. The Humean-inspired utilitarian, on the other hand, can readily explain how interests, in the form of preferences, may shape the goals of policy, but can offer no account of how agents' perception of their interests can change rationally through political persuasion. On either account, there is little scope left for politics to make any independent contribution towards setting its ends. What is needed to remedy this situation is an alternative to the Enlightenment psychology underlying the instrumentalist view and the theories of evaluation to which that psychology gives rise. This task is undertaken in the next section.

III

Instrumentalism thus combines a distinctive philosophical psychology with an associated account of evaluation, and it is this which the present section aims to question. We have already seen that Hume believed moral distinctions to be rationally unfounded, since if they were, we could not explain how agents are motivated to act morally: for Hume, reason can only represent, and as a result can give no explanation of how it can motivate agents to act morally. As Williams puts it in his own neo-Humean theory, if all that a statement about, for example, moral obligation does is to tell us 'one more fact about the universe, one needs some further explanation of why one should take any notice of that particular fact'.¹⁷ Any satisfactory non-Humean theory must, then, account for the motivational capacity of morality.

One prominent example of a non-Humean theory, which Williams has in his sights in the above quotation, is cognitivism: the claim, roughly, that there are moral truths, in the same way as (on a realist view) there are, for example, scientific truths about the world, and we can come to know these moral truths in a way analogous if not identical to that in which we can know facts about the physical world. Thus defined, cognitivism opposes the Humean view to the extent that it allows reason to play a part in discovering truths about morality. Its major advantage is that it gives theoretical effect to what looks like a pervasive feature of everyday moral thinking, that to be under a moral obligation appears to the person concerned to be something more than a projection, as some varieties of noncognitivism have claimed, of

his subjective or affective states. Morality seems to involve something more than the pseudo-objectification of a person's feelings of approbation or personal preferences. Cognitivism's strength lies in its ability to explain this aspect of the phenomenology of moral experience.

Against this, however, cognitivism suffers from telling disadvantages. One problem is that something has to underwrite its ontology of moral truths: to pursue the analogy made above, we have a fair understanding of what it would be for the world to exist, without our existing to represent it in our scientific theories, but it is, to say the least, less clear what it would be for us never to have existed, but for moral truths relating to such matters as promise-keeping, truth-telling, virtues of character, and so on, to exist regardless of this fact. Secondly, the cognitivist has to give some account of how we come to know these truths. Again, in the scientific case, we have a well-formed understanding of the role of research programmes in developing theoretical models of aspects of the natural world, but it is less obvious how one goes about discovering truths about morality - particularly how this is achieved by examining aspects of the natural world. According to cognitivism, there are moral truths waiting to be known, but in disanalogy with the scientific example, there is no clear experiential route which leads to these truths by a systematic method of investigation.¹⁸

For the present purposes, however, the major problem with cognitivism is that it fails to provide a satisfactory alternative to the Humean position. It works within the bipartite psychology of Humean theory, transferring its explanation of the foundation of value from the motivational to the ratiocinative faculty. This move opens up a gap in explanation which the Humean can readily exploit, a gap to be filled by an account of how we are motivated to act morally. In developing a non-Humean alternative to cognitivism, we should distinguish the claim that there are moral truths from the claim that we can arrive at knowledge of these truths by ratiocination. The significance of this distinction should become clear if we examine another important shared feature of cognitivism and Humean theory, the distinction of fact and value.¹⁹

It is fairly obvious how the bipartition of faculties and the fact/value distinction are mutually supportive. If there is this

ontic distinction, it is natural to suppose that there must correspond to it some distinction of the faculties by which we arrive at factual as opposed to moral judgments: while, on the other hand, the distinction of faculties invites a distinction of the world into items which can be represented, and those which have the causal power to issue in action. Cognitivism errs in accepting the fact/value distinction, thereby encumbering itself with the task of explaining the nature of moral facts, and how they can be known; it then has also to explain how morality, assimilated to the class of facts, can gain any practical hold on us as agents.

Once we reject, however, any rigid distinction between neutral facts and non-neutral values, it is easier to see how cognitivism could be replaced by a more serviceable alternative to Humean moral theory. We need to distinguish between two claims conflated within cognitivism: first, that there is an objective realm of moral facts; secondly, that these facts can be discovered by reason. We have already seen some of the problems encountered by the first claim, concerning the status of this realm of moral facts and the means by which we are supposed to gain knowledge of it. In so far as the second claim presupposes the first, by assuming that there is an object of rational discovery, it fares no better, but it will be suggested below that if we abandon the fact/value dichotomy, it has better prospects of being worked up into something more satisfactory. This move aims to avoid the problematic ontic claims made by cognitivism, and replace them with a general account of evaluation which does not entirely exclude reason from the process of forming moral judgments and acting upon them.

It is important to keep in view the wider issues in the discussion. Instrumentalism regards political action as being confined to purely means-end calculations about how best to achieve objectives determined by external considerations. As in some accounts of bureaucracy's functioning, in devising optimal strategies to secure given policy objectives, the assumption of instrumentalism is that political argument is concerned only with finding the best means of achieving these goals. The Humean account limits practical reasoning to the cognitive servicing of given desires; similarly the instrumentalist view of political argument, based on this theory, limits the scope for debate to

purely informational questions.

This limitation is only sustainable if there is a tenable distinction between the acceptable partisanship of policy, and a neutral body of facts which can be mobilized by the bureaucracy to inform politicians how policy may best be implemented. This is clear from, for example, Regan's cooperative utilitarian theory,²⁰ where all that agents are thought to require is information about other agents' intentions, in deciding how to maximize given utility measures. Again, in Rawls' picture of the well-ordered society, the values embodied in the two principles of justice determine what can be pursued politically. More generally, it is the supposition of neutralist theory that it is possible to limit political activities to neutral arbitration between competing conceptions of value, where this neutrality consists either in the implementation of policies embodying shared values, such as equal respect, or in the total avoidance of value-based commitments in politics. This position leaves political argument, seen as a confrontation between controversial values, with little work to do. Value-based conflict is effectively excluded from the political arena.

Cognitivism offers little assistance in producing a corrective to this picture, since its modelling of morality on scientific truths commits it to a view of morality as being at least potentially the object of rational agreement. A non-Humean alternative to cognitivism would, by contrast, endorse a pluralistic theory of value, but would resist the inference from this that there can be no rational value-based disputes, in the same way as there can be no rational disputes over, say, matters of taste. Assuming that sets of facts must have the property of consistency between members of the sets, the Humean is apt to conclude that pluralism merely gives further evidence of the non-factual status of values, and the cognitivist is left having either to jettison pluralism or else withdraw the claim that values have objective factual status.

What by contrast appears to be a promising suggestion is that there can, indeed, contrary to Hume's arguments, be reasons advanced for evaluations, but to deny that this means that value must have some ultimate foundation, whether in subjective preferences or in facts about the world. There can be rational arguments about values even though there is no agreed rational

procedure for resolving value-based disputes. This anti-foundationalist position can be applied to the argument advanced by Hume in the passage quoted earlier from the Enquiry, where he asserts that a chain of practical reasoning must terminate in a reason for action which is not in turn based on any other such reason. There are, according to the counter-Humean claim now under consideration, indefinitely long chains of reasons which can be mobilized in support of values we may hold. But it is a logical truth, and not one issuing from differences of kind between reason and passion, that these chains must end somewhere. Hume's argument in the Enquiry passage shows not that reasons are necessarily conditional upon prior motivations, but that chains of reasons for action are finite. That a piece of practical reasoning starts from some original evaluative judgment does not show that this judgment must be a preference of the agent's, only that some such judgment must be present in order to avoid an infinite regress in explaining action.

It is important to grasp this logical point because it is easily confused with another claim Hume makes,²¹ and which might be thought to be implied by it, that only 'passions' can play any causal role in producing action. With this new causal claim, the argument from the Enquiry ascribes to motivation an originating role in practical reasoning because it is only motivations which have the causal power to make an agent act. In itself, of course, this is a fairly odd claim for Hume to make, in view of his reductive attitude towards causation: if statements of cause are simply (temporally) ordered pairs of classes of natural phenomena,²² there is no obvious reason why rational judgments should not feature in the first of the classes. Leaving this aside, there is reason to doubt, even without a Humean analysis of causation, that we have to refer the causes of action to a distinct psychological faculty of motivation. At this point I adapt an argument used by Davidson against J.S. Mill's theory of causation.²³

Davidson's point is that a specification of causes need not require a full account of the background conditions prevailing before a given event occurred. Against Mill, we cannot infer that a background condition was the cause of an event merely because the event would not have occurred if the condition had not held. Davidson makes this point against Mill's claim²⁴ that in order to

account causally for an event, such as a man's falling off a ladder, we must produce an indefinitely long list of conditions relating to the man's weight, his footwear, laws of gravitation, and so forth, on the grounds that these conditions were all necessary for the event's occurrence.

It seems, against Mill's views, that our notion of cause is relativized to a set of background conditions which may also vary in its composition from one case to another. For example, a machine can be in one of two (equiprobable) states, 1 and 2, and I can pull one of two levers, A and B. If I pull A and the machine is in state 1, it delivers a reward, but if I pull B while it is in this state it gives me an electric shock; but if it is state 2, pulling A produces the shock and B the reward. As a result, if I pull A and get the reward, the machine being in state 1, my pulling is the cause of my getting the reward, even though with different background conditions (the machine's being in state 2), this action would have caused me to receive the shock. My action causes the event which follows it even though, with different background conditions, the same action (under the description pulling A) would have been followed by a different event.

This suggests that our understanding of the notion of an event E's cause is something like the following: an event sufficient to bring about E, given the prevailing conditions (the causal background). In the case under examination, of agent-motivation, this means that we can give a causal explanation of an agent's action if we can specify an event which, relative to a description of the causal background, was sufficient to produce the action. On this account there need be no attempt at providing a complete description of all the conditions which had to obtain in order to produce the consequence. We can identify the cause of an agent's action in a statement about his reasons for acting even though this presupposes that some set of background conditions holds.

The fact that in certain cases, as in the machine example, the cause may only be operative relative to a contingent or variable background, does not refute the causal statement. Suppose a person runs away from an object which he believes to be an unexploded bomb. If someone inquires about the cause of this person's action, it is enough to be told that there was this immediate danger; or at least that the person believed, mistakenly as it happened, that there was. In this case we rest content with a statement providing

a sufficient explanation - a causal explanation - of the action. The fact that the knowledge that there was a bomb would not have provoked the same reaction in someone else, does not mean that we have to include in every causal account of the first person's action his state of fear or timorous disposition or whatever. Similarly, we can say, if there was a bomb, that this caused the person to run away, even though the statement that there was a bomb could not provide a causal explanation of the person's action if it turned out that what he took to be a bomb was some other, harmless, object. The statement that the bomb caused the person to run away even when the same action would have occurred in different circumstances (there being no bomb); and, conversely, the bomb still counts as causing the action of running away, even though the presence of a bomb might not have caused a person unaware of its presence, or less fearful, to run away.

This argument, if it is right, provides some leverage against the Humean claim that there must be a distinct psychological faculty, that of motivation, which is causally required for an agent to act. The Humean claim seems less compelling if there is not, in general, any single description of a given event which we can isolate as the cause of a given event or action. There can be a full statement of cause in accounting for a person's action even though such a statement tacitly presupposes that certain necessary background conditions hold. And there seems no grounds, other than a dogmatic adherence to the psychology of bipartition, not to claim that the causes of actions may be the reasons which also explain them - reasons which can be cited in the causal statement as the agent's (relevant) beliefs.

If this is right, then 'reason' in the form of beliefs, such as a belief about the presence of a bomb, can cause a person's action - we receive a full answer to the question what caused a person to run away if we are told that he believed there was a bomb (whether or not this belief was true). We do not need to invoke any distinct psychological category in explaining the causes of action, supposedly distinguished from that of 'reason' on the grounds of its causal efficacy. Reason, in the form of beliefs, for example, can figure in statements about actions' causes, and in no more elliptical a sense (appropriate desires or motivations here forming part of the background) than that in which causal statements referring to agents' motivations are elliptical in that

they presuppose relevant beliefs.

Much more could of course be said on this matter, particularly on matters such as causal overdetermination and deviant causal chains, a more thorough analysis of the notion of a background condition, and so on.²⁵ For the present purposes the important point is that if it is acceptable in general to maintain that a satisfactory singular statement of cause in any given case need not include a specification of all those conditions are required for a given effect to result, then we need not identify a purpose-built psychological category, that of motivation, as the locus of causal explanation for action.

A full statement of an action's cause need include no reference to any motivation - though a motivation in such cases will form part of the background. In the circumstances (which here includes affective states of the agent) the man's belief that there was a bomb caused him to run away, even though he would not have done so if he had had no desire to avoid injury; equally, this desire caused him to run off, even though, with this desire, he would not have done so if he had not believed there was a bomb. This is not a case of causal overdetermination, but merely different descriptions of the action's cause relative to different specifications of the background. It shows the logical truth that a cause which is sufficient to bring about an event given an accompanying set of necessary conditions, can itself become embedded in such a set with respect to a different cause of the same event.

If this argument is right, there cannot be any rigid distinction between classes of reasons for action based on a criterion of causal efficacy. Beliefs can figure among a person's reasons for acting - in particular, beliefs embodying evaluations. The Humean objection to cognitivism, that it cannot account for the motivational capacity of moral values, is answerable if we reject the cognitivist assumption that there is this rigid dichotomy between reason, as a cognitive or representative faculty, and motivation, as the faculty which causes agents to act.

Thus the Humean objection to cognitivism raised by Williams in the quotation above, that we need some explanation of how moral facts can make any claims on our agency, does not apply to the present account of evaluation. The cognitivist conceded that there

was a realm of facts, and sought to assimilate morality to this realm - thereby assuming the burden both of justifying this assimilation, and explaining how morality, so construed, can ever have any relevance to agency. The rival account, rejecting as it does any blank distinction of 'fact' from 'value', makes no ontological commitments about the status of moral obligations or judgments referring to them, and holds that the notion of evaluation is applicable across the Humean psychological categories. On this view, to make an evaluation is not to report on some external state of affairs, but neither is it merely to report on, or express, some internal subjective state.²⁶

Williams' objection seems to be that we need to add something to a statement that an agent entertains a certain (moral) belief in order to impel him into acting - that something being a motivation. But if this is just a causal claim to the effect that a belief by itself is insufficient to cause any action, the same is equally true of motivation. Nor is it the case, given (true) statements about an agent's motivation and his beliefs about how best to satisfy this motivation, that an appropriate action must ensue as an infallible causal consequence. This is not necessarily because of akrasia, though that is one obvious possibility, and difficult to account for in the Humean scheme,²⁷ but also because agents can often have reason to do that which they have other and better reason not to do.

On the present account, then, the notion of an evaluation cuts across the Humean psychological categories, and cannot be allocated, as both the Humean and the cognitivist suppose, to either one of them exclusively. Reason, in the form of beliefs, can play a causal role in the explanation of agents' actions, and one of the ways (not the only one) in which it can do this is by making evaluative judgments which figure in a causal account of why an agent acted as he did. To this claim it may be objected that we need some explanation of how an agent reaches a belief of this sort. A Humean might further draw attention to the problem, already mentioned, which the cognitivist faces, concerning the apparent difference between the methodology of the natural sciences and the means by which we are meant, according to cognitivism, to discover moral truths: does not this alternative theory run into the same problem, that we have little idea of how to go about establishing moral truths? Unless this problem can be

solved, it may be said, we are left with a sharp difference between scientific and moral epistemology, a difference plausibly explained by a distinction of the form drawn by Humean theory between facts and values as separate ontic categories.

There is little point in representing natural science and morality as sharing a common method of inquiry: but this does not show, even on a realist view of science, that the Humean psychological categories are necessarily beyond dispute. The claim that there is an independent physical world which scientific theory attempts to represent in its theorizing, even when coupled with the belief that the objects of moral evaluations are not, in this sense, independent, does not mean that the Humean categories must apply. For nothing has yet been said in this argument to suggest that there must correspond to the moral evaluations a specifically motivational psychological category. It is possible to accept that there is no independently existing realm of moral facts as conceived by cognitivism, while denying that we must see moral judgments as being produced by, or as expressing, motivations.

The Humean needs this further claim in order to explain how, in his general scheme, morality can motivate agents to act. But if we reject this scheme, there need be no perplexity about how agents can be impelled to act morally: this perplexity need only arise if we conclude from the lack of any external moral realm which is the object of our cognitive representations that there must instead be subjective, and specifically motivational states which explain our moral judgments and their part in action. On the anti-foundationalist position outlined above, this conclusion does not follow. We can deny that there is any external moral realm while allowing moral beliefs - that is, propositional attitudes embodying moral judgments - to play an essential role in the causal explanation of action.

The argument of this section so far has been that the Humean reason/motivation dichotomy needs to be replaced by an account of evaluation which cuts across these categories. One area in which the need for an alternative to Humean theory is particularly noticeable is its analysis of emotional states. As the passage quoted above from the Treatise makes plain, Hume saw the emotions or 'passions' as 'original existences', which as such do not have 'any representative quality, which renders [them] a copy of any

other existence or modification'. It follows that there can be no question of emotions' being rational or irrational, since in Hume's view to experience an emotion is to be in a state analogous or identical to a physiological condition such as thirst or sickness, and because the latter have no 'reference to any other object', the emotions, likewise, cannot be said to conform or fail to conform to reason.

It is fair to say that this picture of the emotions has been an important factor in the dismissive attitude of instrumentalism towards political argument. It is this which, as much as anything, explains the neglect of rhetoric as a persuasive device in political dispute, and its academic neglect: any argument which makes use of rhetoric must be non-rational, because the rhetorician's appeal is directed at the emotions, and the emotions are by their nature non-rational. For this reason rhetoric is subjected to much the same opprobrium as is suffered by emotivism as an ethical theory.²⁸

The latter is often attacked on the grounds that it is irrationalist, maintaining as it does that the meanings of our moral vocabulary, or the rules governing our use of moral utterances, are essentially statements or expressions of emotion - that a moral term actually expresses the emotions of the speaker, or else that the criterion by which we decide whether a given moral utterance has been made is whether it has mentioned or expressed an emotion of the speaker. But if it turns out that the emotions are rationally founded, contrary to the Humean claim, it is no longer possible to pin the charge of irrationalism on the emotivist: it will then be possible to accept one version or other of the emotivist thesis, while denying that this commits us to thinking of moral judgments as non-rational or irrational. Similarly, a non-Humean view of the emotions can deflect the charge made against both rhetoric and other forms of emotional appeal embodied in political argument, that they must be fundamentally non-rational methods of persuasion. If reason does play a part in the formation of emotional states, it becomes possible to claim both that political argument is rational, and that it includes rhetorical appeals in its armoury of persuasive techniques.

In what follows I will sketch an alternative, non-Humean view of the emotions as affective states including an irreducible

cognitive component. The purpose of this will be heuristic, in the sense that the analysis is not intended to provide anything approaching a full theory of the emotions. Instead the aim is to provide a basis from which to develop a reformulated version of emotivism free of the Humean assumptions informing traditional versions of emotivist theory. Once we have a clearer view of this alternative theory, it should be possible to give some account of the rationally persuasive character of political argument.

It is undeniable that we do not, in ordinary language, predicate truth and falsity of emotions, as we do of beliefs. The Humean position obviously goes well beyond this, however, to deny that we can identify any rational component in the overall structure of an emotional state. Indeed, the Humean position as quoted above seems to deny even that emotions are intentional states - that they are directed towards some object - let alone the claim that emotions require the formation of some judgment in respect of the object. It is, despite this, a fact of language that we do speak of emotions as being object-directed in this way,²⁹ that we talk of people as being angry at someone's betrayal, proud of their achievements, fearful of the consequences, and so on, in expressions for which the occurrence of the emotion-words sans phrase can be taken as convenient shorthand. Even where the object of the emotion is not given explicitly, it would be odd to ascribe an emotion to someone while also admitting that there is no object whatsoever towards which the emotion is directed. We can, ambiguously, say of a person that he is frightened of, angry about, nothing. But on one reading, this statement ascribes a non-rational emotion to the person, while on the other, it ascribes no emotion at all - on one possible reading, to say of a person that she is frightened of nothing is to say that she is not frightened.

It may be said that there are cases where an emotion is ascribed to a person in the absence, even implicitly, of any object towards which the emotion is directed. In response to the question why he is sad, a person may say that he is just sad, without being able to cite any reason - as Antonio remarks, for example, at the start of The Merchant of Venice³⁰. And this seems to be a quite intelligible statement. In this form, however, the statement looks more inviting than an amplified version of it in which the person acknowledges that there is no object of his

sadness - that there is nothing he is sad about, or (still less invitingly) that he is sad about nothing, a claim which displays the same ambiguity as that considered in the previous paragraph. But perhaps sadness is peculiar in this respect, that it may in some cases resist explication of this form. It is, however, less plausible to think that this is true of a wide range of other emotions - fear, love, hatred, shame, envy, pride, contempt being some of the examples - where we would doubt the sincerity of rationality of a person who claimed to be under their influence while acknowledging that there was no object whatsoever towards which the emotion in question was directed.

Both possibilities identified above may of course be (separately) realized: a person may be afraid of nothing either in the sense that he feels no fear, or that he fears without (good) cause. It is significant that our language of emotions admits of this ambiguity, and that both possibilities can be seen as cases of inappropriate or even, in extreme cases, irrational emotional responses. The division of possibilities suggests a composite structure for the emotions, in contrast with Hume's monadic theory. Stated in its crudest form, this composite theory would claim that emotions comprise a cognitive, object-directed component, together with an evaluation or attitude. It is a further question whether the object-directed component in all emotional states must take the form of full-blown beliefs; it might be said that we ascribe certain emotions, such as fear, to animals, without our thinking that they must therefore be capable of holding propositional attitudes about the object of their fear. In these cases, however, there is certainly an object of fear towards which the emotion is directed.

In at least occurrent, as opposed to dispositional, examples of human emotion, indeed, there is reason to go further, and claim that these emotional states do include beliefs about their object. In the bomb case, for example, it is very odd to say that the person entertains no belief whatsoever about the object, and in effect impossible to rationalize his action of running away as the act of a frightened man without imputing quite specific beliefs to him in respect of this object. Examples like this suggest that one of our reasons for ascribing beliefs to persons experiencing emotions such as fear is the explanatory value of such ascriptions - assuming that the person holds a given belief is sometimes the

only means of rationalizing their behaviour.

This is particularly clear in cases where the object of the emotion does not in fact exist. If there is no bomb, it is hard to make sense of the man's running away without imputing to him some such belief. Of course, it may be said that in such cases we need some independent means of deciding whether the action in question really is caused by fear or has some entirely different explanation (in which case we may well be obliged to redescribe the action itself). There is an evident danger of interpretative circularity in problematic cases like these: the action of running away may only be identifiable as such given the emotional explanation, while on the other hand we are only able to identify the emotion of fear through knowing that the person ran away. But whatever the success of efforts to break this circle, it is clearly intelligible that a person should act on a belief, true or false, which features as an essential component of the person's emotional state. And, as was argued above, in such cases the belief can be cited as the cause of the person's action.

This does not of course pretend to be a complete account of the emotions, only to provide some idea of how reason, in the form of beliefs, can bring about emotional states. More would need to be said, for example, about the different ways in which emotions can be irrational, through false belief, through the inappropriateness of a (true) belief to the emotion, or the disproportion between the belief or its object, and the emotion.³¹ These cases provide further grounds for not identifying emotions with beliefs, since not all of them are instances where the irrationality of the emotion is to be attributed to false belief. A man may falsely believe that there is a bomb; but he may also be too frightened, or not enough, by some object.

These are questions concerning the relationship between the agent's belief and his emotional response. What is required is some general account of appropriateness: of how well, in other words, an agent's emotional responses accord with his beliefs. But, again, we should not assume that these possibilities call for some fuller causal account of the relationship between beliefs and emotional states. For it may be in the nature of the inappropriate responses that no rational account can be given: the criterion of a rational response may, reductively, be whether it can be causally explained by beliefs known to be, or plausibly thought to

be, held by the agent. With regard to the emotions, agents' actions are rational to whatever degree we are able to explain them as the causal product of (rational) beliefs.

So far, then, this section has argued for the following claims: first, that beliefs, contrary to Hume's arguments, can operate as the causes of action, to the extent that the question what caused an agent to act in a certain way can intelligibly be answered by reference to a belief of the agent's; second, again contrary to Hume, that beliefs, qua representative cognitive states, are an essential component of emotional states, and our criteria in assessing the rationality or appropriateness of emotional responses include a consideration of the agent's beliefs.

If these arguments are right, there is no force in the Humean claim that cognitive states necessarily lack motivational force, and reason can after all cause persons to act. This is not to point back towards a full-scale cognitivist theory of evaluation. But it does suggest how a set of shared evaluations can make interpretative sense of agents' behaviour. It need not, of course, be the case that evaluations are themselves the objects of the beliefs which form part of the emotional states: as in the bomb example, evaluation comes in at a different point, in our understanding of the relevance of a given belief to explaining the person's action. It is the existence of this shared context of evaluation which, as with other forms of social explanations, makes the action intelligible.³²

Once it is recognized that evaluation comes in at this point, in determining criteria of relevance, the problem faced by cognitivism, that value has to be ontologized as part of the world in order to be knowable, no longer looks so threatening. We do not have to suppose that value is in any sense part of the ultimate fabric of the universe in order to account for evaluations of the sort present in emotional responses. At the same time, beliefs are not, as in Humean theory, entirely excluded from the analysis of such responses, since they are necessarily implicated in explaining persons' responses to objects (events, situations) in the external world.

The significance of this analysis of the emotions for our concerns in this chapter is twofold. First, and more immediately, it provides an alternative to the Humean account of the passions by suggesting how emotional states can be rational or irrational.

As a result, rhetoric, as a mode of persuasion based partly on an appeal to the audience's emotions, can to this extent be a rational argumentative process, contrary to the assumptions underlying instrumentalist writings in political theory. Second, this partially cognitive analysis of the emotions offers a more general challenge to the metaphysical assumptions behind instrumentalism. If it turns out that evaluations are not merely the projections of internal subjective states, but belong to a public context of evaluation, it is less easy to depict political argument as merely being a darkling plain where ignorant armies clash by night. The existence of shared standards of evaluation opens at least the possibility that argument can proceed rationally, rather than merely involving the exchange of incommensurable subjective values.

This does not of course mean that the norms of rational interchange envisaged by, for example, Rawls, Ackerman or Habermas must therefore adequately model, even in an idealized form, the value-based disputes characteristic of political argument.³³ For one thing, nothing has been said so far to support the assumption made by these theorists that argument can only be rational if it is conducted within a set of procedures capable in principle of eliciting agreement. A counter-example to this model of rational discourse need not invoke the possibility that two wholly alien cultures may confront one another in some situation of conflict while each remains true to its own canons of rationality. It may simply be, within a unitary culture, that there is more than one rationally arguable position in some cases of political or moral disagreement, and no method agreed by the disputants for achieving consensus.

This response to the discursive model of rationality only holds up, of course, if there is no rational requirement to reach consensus in (practical, political, moral, etc.) judgments. It might be thought, however, that the foregoing argument was based precisely on this possibility, that consensus in evaluations is a condition of rational intelligibility in understanding, for example, others' emotional responses. If so, it may be said, something like the Ackerman/ Habermas model of discursive rationality must underlie this account, since otherwise there would be nothing to guarantee the rational acceptability of some patterns of emotional response rather than others.

But it does not follow from the fact that there are consensual patterns of evaluation which render emotional responses rationally intelligible, that as a condition of rational discourse there must be standards of argument which necessarily produce agreement among the participants. For, to adopt a Wittgensteinian form of argument, it may be that there are certain general conditions on intelligibility - Wittgenstein's agreement in judgments - within which discussion (either within a specific culture or in human discourse generally) must proceed. Given these general conditions, however, there may be disagreement within the consensual parameters. In the case of moral judgments, there may be certain general features which any utterance must display in order to count as such a judgment, even though this leaves wide scope for disagreement over its content. Similarly, with emotional responses there may be some general characteristics which we can identify as comprising such responses, without there being any means of adjudicating between two conflicting assessments of a person's response to a given situation. In such cases, it is familiar that there can be this divergence, with each side able to offer some reason by way of accounting for their judgment, but unable to invalidate the other's judgment.

As a means of clarifying the thesis for which this chapter is arguing, it is helpful to compare Aristotle's observations on the role of 'opinion' (endoxa) in rhetorical practical syllogisms, to which the present argument is indebted.³⁴ Aristotle focuses on the incomplete syllogism or enthymeme, as a typical argumentative structure deployed in political rhetoric: the enthymeme differs from the full syllogism in that one of its premisses is not explicitly stated, but remains implicit, as an opinion which, though not apodeictically certain, is presumptively true. For example, Aristotle says³⁵ that to show that someone was victor in a contest in which the prize was a crown, it is enough to say that he won a victory at the Olympic games, as everyone knows that the prize at the games is a crown, the remaining information being filled in by the audience.

Political arguments, analogously, can make appeals to the emotions while presupposing evaluations which remain implicit. Here the assumption is that there is some community of evaluation which makes implicit assumptions about value in interpreting political arguments. This may, of course, go well beyond arousing

an audience to feel a particular emotion, though it covers that possibility: for example, to make an argument against a proposed policy, it is often enough to assert that it has a consequence which can immediately be recognized as undesirable, without the need to say so explicitly. To claim that a policy of raising interest rates will discourage investment presupposes, and need not state directly, that discouraging investment is undesirable. In the same way a rhetorician may be able to induce a certain emotional response in an audience merely by getting them to accept a particular belief, leaving the underlying evaluation implicit. For example, it may be enough, in order to move an audience to feel anger or contempt towards a person, to give them evidence that he has betrayed them. The speaker's task is to persuade them that this is so, not to make them accept the underlying evaluation that traitors should be the object of such emotions.

We can, then, regard political argument as aiming in part to induce an audience to form emotional attitudes towards policies or persons, by persuading them to form judgments or beliefs. The elliptical nature of much political argument presupposes that there are to a considerable degree shared evaluations to which a speaker tries to appeal in making an argument. This clearly goes beyond the Humean-inspired instrumentalist claim that there is nothing to be said about rational persuasion in political argument: the possibilities mentioned above concern a speaker's efforts to persuade an audience to accept some judgment or belief. The faculty-division of Humean psychology seems to preclude such possibilities.

At the same time, there is little prospect that we will be able to provide some form of ultimate rational foundation for political (or any other) argument, as a means of persuasion. To argue as this chapter has that we can identify rational techniques of persuasion is not to claim that it is possible to produce a set of neutral procedures, which could tell us what judgments to induce in an audience in any given circumstances, and how to achieve this. The Humean is at least right to this extent, that there can be no ultimate rational foundations of this kind for argument. But this does not mean that there is nothing at all to say about rational persuasion in politics. As Skinner and others have argued,³⁶ there are forms of argument which it is rational for a speaker to use given, for example, pre-existing linguistic

conventions governing what can or must be said to elicit a certain response from an audience. While the judgments which, on Skinner's neo-Wittgensteinian view, are implicit in the possibility of such communication cannot be given any ultimate foundation, there can still be rational argument within the conventions, a point neglected by the reductive views Skinner attacks, which regard principled arguments as irrelevant to the real business of politics.

To accept the existence of this shared evaluative background is not, however, to abandon the pluralistic view of value which has already been advanced. The claim that there are these implicit evaluations does not require a unitary theory of value in general. For it does not follow from the fact that there are shared evaluations embodied in a normative vocabulary or, more broadly, underlying emotional responses, that these evaluations must be reducible to some single dimension of value, nor that certain values enjoy absolute priority over others. It is quite compatible with the Skinnerian position to claim that values conflict with each other in practical action, or that their translation into action may be practically ambiguous. Political decision-making is not merely a matter of deciding which values are of practical relevance in relation to a particular issue of policy, and then proceeding to implement the favoured policy. This is the supposition of instrumentalism, but there is no reason to assume, even in the presence of unitary values, that the translation from value to policy must be straightforward.

It is in this area, if anywhere, the gap between policy and value, that political argument is located. Treatments of political conflict have perhaps concentrated excessively on the grander cases of confrontation between opposed values, neglecting the more humdrum possibility that this conflict may take place not between absolutely opposed values, but between opposed ideas about how values translate into policy, or about how different values have to be weighed against one another in a given question of policy. This is not to deny that there are, in democratic political cultures, deep and value-based differences between for example the ideologies of socialist egalitarianism and free-market individualism, though even here it is important to remember that there are certain institutional features of these cultures which accentuate conflict. But accepting that dispute at this level may

be irreducible to any common values does not explain the phenomenology of political conflict. As Skinner has persistently argued, we still need some explanation, accepting these aspects of conflict, of how politicians can, in the face of these differences, engage in argument and commend policies to their audience as most successfully promoting accepted values.

Instrumentalism, by contrast, is inclined to regard political disputes as characterized by differences of tactics in the presence of an agreed strategy. On this view, the existence of endemic political conflict merely shows that there are thought to be different routes to an agreed destination, for example maximizing welfare. But the fact that politicians of different parties may claim to pursue similar (macro-economic, etc.) objectives does not show that all political disagreement is about what means to adopt in pursuance of agreed ends. There is the point, due to Wiggins,³⁷ that in deliberation generally practical reasoning can deal not only with the best causal means to secure a fixed end, but also, given some desired objective, what best counts as fulfilling the objective.

While there may be some procedure for rationally resolving disputes of the first kind, at least where there are agreed methods of balancing out the costs of implementation, there is little reason to think that this carries over to disputes of the second kind. Within a shared commitment to the value of liberty, for example, a conservative may argue that a proposed measure regulating the workings of trade unions promotes freedom, while a socialist argues no less vigorously that the measure diminishes freedom. In the senses in which the protagonists intend their respective claims, of course, it may do both, but the significant point is that the policy cannot, in the light of this, merely be seen as offering one causal route to an end which is practically agreed. Consensus at the level of value can mask radical and often irreconcilable disputes about the practical implications of value. The reductivism of the means-end view of political dispute becomes all the more plain when the perspective is widened to include disputes over how different and mutually irreducible values may bear upon an issue of policy. Here, a fortiori, a nominally shared commitment to a set of values may remain even though different political groupings have implacably opposed views on the practical policies sanctioned by those values.

The claim that political argument requires shared evaluations does not then preclude a pluralistic theory of value. It is possible to maintain that the forms of argument deployed by politicians presuppose, if they are to be intelligible, values which are in some sense shared with their audience, while also acknowledging that there need be no single course of policy which is clearly dictated by considerations of value. Whether or not policy disagreements take the form of a conflict between opposed values, or the differing practical conclusions drawn from a single value by those involved in the dispute, there is no reason to think that, at the level of practical agency, there must be some agreed procedure for resolving the dispute.

Instrumentalism falsely claims that all such disputes must be rationally terminable by reference to ends which are themselves non-rationally given. On the present view, by contrast, political dispute is characteristically not open to rational resolution in the light of agreed ends: there is often no single demonstrably correct answer to questions of policy. An understanding of the complex relationship between political argument and the values to which its protagonists appeal means accepting that pluralism is an inevitable feature of political debate and that one of its most characteristic manifestations is in political conflict - a truth which is most clearly attested by the absence of any agreed solution to problems of policy, and the invocation by each side of values which they claim to support or be furthered by their favoured policy.

It is, however, important to remember that value-pluralism in political conflict is not necessarily to be taken as an indication that there is no room for rational argument in politics. The fact that there are conflicting policies in some matter of public concern, with no means of reducing their supporters to rational agreement over which should be adopted, does not mean that no rational claims whatsoever can be advanced by the parties to the debate in arguing their case. It would, indeed, be strange if this were so, since it would leave political argument inexplicable: the gap cannot be wide, in general, between the forms of (ostensible) rationalizations which the parties make of their policy proposals and, on the other hand, the forms of explanation which an observer could offer to reconstruct the rational grounds on which these arguments were made. Only if there is this possibility of giving a

rational explanation of why a person made the argument he did can we begin to make sense of political argument - which is not to say that we have to think that the real rationale for such an argument is the same as its purported rationale.

This point about understanding holds even though there is no final rational method for assessing the arguments presented in the course of a political debate. But this does not mean that political decision-making is wholly arbitrary. One reason for this is implicit in what has already been said about the constraints within which politicians operate in making their arguments, of rational comprehensibility. If the appeals made by political arguments are directed towards interests or evaluations which the audience is intended to recognize, it is unlikely that someone deciding between the arguments for competing policies will regard it as merely arbitrary which to accept. This is due to the point, made by Raz,³⁸ that even where an agent is confronted with a choice between alternatives, such as two different careers, neither of which is clearly preferable to the other, it does not therefore become a matter of indifference which career is chosen, as it would with a choice between two identical quantities of the same good, or goods with uniform relative weightings. Indecision between a pair of practical alternatives may be a sign not of indifference but precisely of the importance attached to them.

Utilitarianism notably fails to account for this distinction, regarding such cases as merely involving similar quantities of a single, homogeneous good. But it does not follow from the fact that a person prefers neither of pair of goods to the other that it must be a matter of complete indifference to him which of the two he has. The lack of a decisive preference may be evidence of indifference, in the utilitarian's sense; but it may indicate instead that there is no single scale on which to measure the claims to be made on behalf of each of the goods. They may be so disparate as to preclude this sort of comparison. They are each, nonetheless, goods, and an agent may be rationally concerned about which should be chosen.

Rational concern may as often emerge in an inability to decide between alternatives. There may be no solution through the application of some uniform calculative method, such as utilitarianism purportedly offers, to decision-making. The way from indecision to action lies not through the discovery of such a

rational explanation of why a person made the argument he did can we begin to make sense of political argument - which is not to say that we have to think that the real rationale for such an argument is the same as its purported rationale.

This point about understanding holds even though there is no final rational method for assessing the arguments presented in the course of a political debate. But this does not mean that political decision-making is wholly arbitrary. One reason for this is implicit in what has already been said about the constraints within which politicians operate in making their arguments, of rational comprehensibility. If the appeals made by political arguments are directed towards interests or evaluations which the audience is intended to recognize, it is unlikely that someone deciding between the arguments for competing policies will regard it as merely arbitrary which to accept. This is due to the point, made by Raz,³⁸ that even where an agent is confronted with a choice between alternatives, such as two different careers, neither of which is clearly preferable to the other, it does not therefore become a matter of indifference which career is chosen, as it would with a choice between two identical quantities of the same good, or goods with uniform relative weightings. Indecision between a pair of practical alternatives may be a sign not of indifference but precisely of the importance attached to them.

Utilitarianism notably fails to account for this distinction, regarding such cases as merely involving similar quantities of a single, homogeneous good. But it does not follow from the fact that a person prefers neither of pair of goods to the other that it must be a matter of complete indifference to him which of the two he has. The lack of a decisive preference may be evidence of indifference, in the utilitarian's sense; but it may indicate instead that there is no single scale on which to measure the claims to be made on behalf of each of the goods. They may be so disparate as to preclude this sort of comparison. They are each, nonetheless, goods, and an agent may be rationally concerned about which should be chosen.

Rational concern may as often emerge in an inability to decide between alternatives. There may be no solution through the application of some uniform calculative method, such as utilitarianism purportedly offers, to decision-making. The way from indecision to action lies not through the discovery of such a

method, but through recognizing that one is faced with a decision between values, or between interpretations of value. But there is little reason to think that any definitive rational procedure can be laid down for this decision. An agent's conception of rational choice may itself depend on which of a set of competing values he chooses, as Kierkegaard, for example, makes clear in Either/ Or.³⁹ To this extent, any choice involving fundamental values, political decisions among them, is a matter of deciding which values to follow and how to apply them in practice.

IV

As commentators have pointed out, the absence of any rigid rule-governed procedure for practical deliberation does not show that there can be no distinction at all between acceptable and unacceptable forms of argument. We may, even in cases of radical indecision, be able to dismiss certain practical proposals as being clearly out of the question, though this may not come about as the result of deliberation - an obvious way in which a possibility can be rejected is by its never being entertained.⁴⁰ Here we should certainly regard anyone who did seriously consider such a possibility as having gone wrong in some way, or as failing to understand the limits within which deliberation takes place.

But these limits are not to be seen, as in utilitarianism, as set ends to which all deliberation must be understood as devising means, nor, as in the Kantian brand of instrumentalism, as having been dictated by a class of reasons which have categorical weight in practical deliberation. A more adequate understanding of practical deliberation, whether public or private, than that offered by instrumentalism accepts that there is more room for deliberative manoeuvre than any hierarchical or means-end view of practical reasoning acknowledges and that such constraints as do bear upon reasoning may apply without either remaining static or being precise in their scope.

It should be noted that none of the above analysis relies on the assumption that there is any structural difference between individual and collective practical reasoning. All the arguments so far could still be true even if it were the case, as instrumentalism assumes, that the structures of practical reasoning which apply at the level of individual deliberation are

applicable also to collective deliberation. Kantianism most clearly takes this course, by regarding the circumstances of political action as not merely analogous to, but identical with, the situation of the individual agent: the conditions in which politicians take their decisions, for Kant, must conform to a universal rational pattern. While the utilitarian is not as explicit as this about the reduction of all practical agency to the level of the individual, the same aggregative procedures, of assessing the relative intensity of preferences, apply in each case.

While it is true that utilitarianism in, for example, the theories of Parfit and Regan has addressed itself to problems of coordinating individuals' actions so as to achieve the best overall outcome, both these treatments take individual preferences themselves as fixed (albeit idealized in certain respects), and see the role of public policy as limited to deciding, given these preferences, how best to satisfy them. Though this is not the same as modelling collective decision procedures on individual deliberation (the individual agent need not go through the process of weighing relative intensities of preferences against one another), the persuasive aspect of political debate disappears, and with it the possibility that a new understanding of public action can emerge through debate.

This chapter has tried to indicate some ways in which practical reasoning, even at the individual level, goes beyond the limits imposed on it by instrumentalism, and to draw some conclusions for our understanding of politics. The starting-point for an alternative to instrumentalism lies in questioning the Enlightenment philosophical psychology underlying much of its thinking. This means abandoning any rigid distinction of 'reason' and 'passion' as presupposed, in their different ways, by both Humean and Kantian theory.

In its place, we should adopt a notion of evaluation which cuts across the Enlightenment psychological categories. While this alternative notion would not pretend to provide a general theory of rational action from which the content of morality could be inferred, it would make possible a better understanding of moral and other concerns as they feature in political debate. If we abandon the Kantian notion of a deliberative hierarchy, it becomes clearer that political rationality may consist not merely in

conformity to categorical norms of action, but in a recognition that moral considerations do not always enjoy absolute priority in political decision-making. It is the mark of rationality to recognize that there is no decision-theoretic formula which can dissolve political conflict in the way Kant and his followers have supposed.

Whether or not there are formal differences marking off moral from non-moral reasons for action, pluralism (in the broader sense distinguished above) holds that there can be no absolute priority among classes of reasons for action. If Kantianism is abandoned, and we take the view that there are no categorically binding moral reasons for action, we remove one justification for political neutrality as proposed, for example, by Larmore - that the limits of the political are marked by a principle of equal respect, which applies generally, and demarcates the legitimate area of public intrusion on individuals' pursuit of their conception of value. Without this reliance on the categorical status of certain moral obligations, it is hard to see what compelling reason there is for the separation of public from private fields of action, particularly in a theory such as Larmore's which, unlike Kant's, operates with a wide-ranging form of pluralism in civil society.

A full commitment to pluralism requires the rejection of neutralist theories such as Rawls' and Larmore's, which confine the role of public policy to implementing or working within agreed conceptions of value. The two-level pluralism distinguished earlier precludes any such rigid division of public from private realms: pluralism at the level of value itself, where there may be no single scale on which to compare different values, and at the level of action, where even within a single value it may be indeterminate what practical action the value sanctions or requires. Looked at like this, pluralism can manifest itself in uncertainty over what forms of practical policy are favoured by neutrality, unless the Larmorean liberal can show that the commitment to neutrality is itself value-neutral. As should be clear by now, the problem in showing this is that different value-based commitments give rise to differing perceptions of what areas are neutral, and how the terms of competition between value-free and value-based concerns are to be set. A consequence of the diversity of value is that there can be conflicting attitudes towards politics and the scope of its operations. Neutrality is

breached if the political order institutionalizes one conception of this scope over others.

In rejecting neutrality, we are not forced to stipulate some other, presumably wider, scope for legitimate political intervention in citizens' personal lives. It is Larmore's complaint that previous liberal theory has adopted an 'expressivist' theory of the state,⁴¹ according to which, as in Rawls' theory, the political arrangements of a society embody substantial moral ideals: such ideals, in Larmore's view, ought to be confined to civil society. But Larmore ends by producing an expressivist theory of his own, as the basis for neutrality itself proves to be a moral ideal. The significant point, however, is that we need not react to the problems in Larmore's account by substituting an alternative, avowedly anti-neutral theory of the relationship between state and civil society, endowing the state with a rich array of moral attributes. Instead, we should recognize that an acceptance of the full pluralism outlined above leaves it indeterminate how the relationship of state to civil society is to be conceived, and what part, if any, moral ideals can play in mediating this relationship. The same considerations of value-based indeterminacy apply equally to the wider question - which must always itself be political - of the role of politics in wider human concerns.

On this view there can be no final philosophical account, as recent works of neutralist theory have aimed to provide, of the legitimate bounds of political action. This is not to claim that we cannot draw any effective distinction between the categories of public and private, nor that it is impossible to provide any philosophical theory of what counts, at any given time, as the political - though it is implied by the arguments of this chapter that we could not identify certain concerns as intrinsically political, except at an unhelpfully general level. Our notion of the political is closely involved at any time with what goods we value and how far we think political processes capable of securing them. To make more sense of this notion than instrumentalism does, we need to recognize that politics cannot merely occupy the negative space left by (suitably mediated) private conceptions of value. Our understanding of value and its practical ramifications in part results from involvement in political argument. And that means that argument in politics is a rational process -

enlightening us about our commitments and about what, given these commitments, we have good reason to value.

Notes

1. See A Theory of Justice (Oxford 1972); 'Kantian Constructivism in Moral Theory', Journal of Philosophy 1980; 'The Basic Liberties and Their Priority', in S. McMurrin (ed.), The Tanner Lectures on Human Values 1982.

2. Derek Parfit, Reasons and Persons (Oxford 1984); Peter Singer, Animal Liberation (New York 1975); John Harris, The Value of Life (Oxford 1985) and 'The Survival Lottery' in Singer (ed.), Applied Ethics (Oxford 1986); also Jonathan Glover, Causing Death and Saving Lives (London 1977).

3. Charles Larmore, Patterns of Moral Complexity (Cambridge 1987).

4. For the introduction and explication of this term see Rawls, A Theory of Justice, p42f.

5. For an incisive discussion of these and related issues see Joseph Raz, The Morality of Freedom (Oxford 1986), esp. the chapter on 'Incommensurability'.

6. Rawls, 'The Independence of Moral Theory', Proceedings and Addresses of the American Philosophical Association 1975.

7. Alan Gewirth, Reason and Morality (Chicago 1978); see also his Human Rights (Chicago 1980).

8. Bruce Ackerman, Social Justice and the Liberal State (Yale 1980), passim; Rawls, A Theory of Justice Part I; Larmore, op. cit. Chapter 3.

9. David Hume, A Treatise of Human Nature, ed. L. A. Selby-Bigge, (Oxford 1888), p415.

10. Donald Regan, Utilitarianism and Cooperation (Oxford 1980).

11. Hume, op. cit. p415; p458.
12. David Hume, Enquiries Concerning the Human Understanding and Concerning the Principles of Morals, ed. L. A. Selby-Bigge, (Oxford 1962), p293.
13. See above, note 9.
14. On this point, concerning the separation of the grounds of moral obligation from motivation, see Larmore op. cit. pp80-86.
15. Robert Nozick, Anarchy, State and Utopia (Oxford 1974).
16. See Immanuel Kant, The Critique of Judgment, ed. J. C. Meredith (Oxford 1952); Ronald Beiner, Political Judgment (London 1983), Chapter 3.
17. Bernard Williams, 'Ought and Moral Obligation', reprinted in his Moral Luck (Cambridge 1981), p122.
18. It is of course a keenly debated issue in the philosophy of science how directly research programmes are related to understanding the structure of the natural world: for a highly sceptical view see Paul Feyerabend, Against Method (London 1978); also Imre Lakatos, 'Falsification and the Methodology of Scientific Research Programmes' in I.Lakatos & A.Musgrave (eds.), Criticism and the Growth of Knowledge (Cambridge 1970); Thomas Kuhn, The Structure of Scientific Revolutions (Chicago 1962); and, in some ways less sceptically, Karl Popper, The Logic of Scientific Discovery (New York 1959).
19. See e.g. Thomas Nagel, The View From Nowhere (Oxford 1986); Gilbert Harman, The Nature of Morality (Oxford 1977); Charles Stevenson, Ethics and Language (Yale 1944).
20. Regan, op. cit.
21. Hume, Treatise p414; p492; p521; p526.
22. Ibid., p164; p166.

23. Donald Davidson, Essays on Action and Events (Oxford 1980), p150f.
24. John Stuart Mill, A System of Logic Book III Chapter V iii.
25. The intuitive idea is that the background comprises a set of individually or collectively necessary (but not collectively nor, a fortiori individually sufficient) conditions for the occurrence of a given event. Given this background, a separate condition may be sufficient for the event to occur, even though (as in the machine example) this latter condition might itself belong to a set of necessary conditions of the form just mentioned, relative to a different sufficient condition. Obviously fuller analysis is required here.
26. On these issues see Sabina Lovibond, Realism and Imagination in Ethics (Oxford 1983), passim.
27. For Hume, akrasia is impossible if it is explicated as acting contrary to reason; he celebratedly remarks that it is not unreasonable to prefer any good to any other if one entertains 'a more ardent affection' for it (Treatise p416). This leaves open the possibility that someone can act against their greater deliberated desire. But Hume cannot accept that an agent can knowingly (and so akratically) do this without undermining his general thesis that action cannot be contrary to reason. For further detail on this, and contrast with Aristotle's views on akrasia, see my 1987 MA Dissertation Aristotle on Politics and the Human Good, Appendix (copy deposited in York University Morrell Library).
28. See Bernard Williams, 'Morality and the Emotions' reprinted in his Problems of the Self (Cambridge 1973); Alasdair MacIntyre, After Virtue (London 1982), Chapters 2 and 3.
29. For this point see George Pitcher, 'Emotion', Mind 1965 pp226-46.
30. 'In sooth I know not why I am so sad', Antonio, in The Merchant of Venice I i line 1. It should be said that the thesis

argued for in the text only fails if there is in fact no object of the agent's emotion - whether or not the agent knows what it is as for example in cases where the true object of fear is either suppressed or represented as something else by the agent.

31. Pitcher, ibid., p228f.

32. See e.g. C.Hookway & P.Pettit (eds.), Action and Interpretation (Cambridge 1980).

33. Rawls and Ackerman, op. cit.; Jurgen Habermas, Legitimation Crisis (London 1976), and The Theory of Communicative Action (Boston 1987).

34. In the Rhetoric, Book I.

35. Ibid., 1357a16-22.

36. See e.g. Quentin Skinner, 'Conventions and the Understanding of Speech Acts', Philosophical Quarterly 1970, pp118-32; and 'Some Problems in the Analysis of Political Thought and Action', Political Theory 1974 pp277-303.

37. David Wiggins, 'Deliberation and Practical Reason', reprinted in his Needs, Values, Truth (Oxford 1987), p225.

38. Raz, op. cit.

39. See MacIntyre, op. cit. pp39-45.

40. As Bernard Williams points out in his Ethics and the Limits of Philosophy (London 1985). Of course, this does not mean that all undeliberated options would necessarily be rejected by a fully-informed agent if she were presented with them.

41. Larmore, op. cit. p73.

Chapter Five

Rationality and Politics

Political Rationality

I

In the last chapter I examined the Enlightenment doctrines forming the basis of instrumentalism and questioned the rigid distinction between cognition and motivation. Enlightenment philosophical psychology, as it has featured in modern instrumentalist writings, has invoked this distinction in depicting political life as being minimally subject to rational correction: the concerns of politics are, on this view, exhausted by means-end calculations, of how best to promote objectives which are themselves taken as given. The link between such apparently diverse theories as utilitarianism and Kantianism lies in their common assumption that there can be no rational debate over ends in politics. Whatever ends political agents may promote are subject in Kant's view to the veto of morality, and conditioned, both in Kantianism and utilitarianism, by motivations which are definitionally non-rational.

This Enlightenment style of thinking has left a profound mark on modern liberal philosophy. But it has also left liberalism with a large problem. If we accept that the final ends of political action are determined by extra-political considerations - by, for example, the 'conceptions of the good' held by individuals or groups in civil society - we have to face the fact that irreducibly different and incompatible conceptions of value are competing for statutory advantage or public funding. Given this fact, how is the post-Enlightenment liberal to retain the view of politics as concerned solely with instrumental calculation, as there are a number of mutually inconsistent final ends which it is politically possible to promote? It is not merely that public resources are finite, that some must prosper at others' expense. Rather, in some cases, there could be no rapprochement between the values held by different groups within civil society, as they logically exclude one another. In questions of public censorship, for example, there is no means of reconciling the wishes of those favouring license with those favouring restriction.

This liberal quandary springs directly from the attempt to

combine instrumentalism with a belief in the plurality of value. The work of Rawls and Dworkin¹ can be seen as liberal efforts to resolve this problem. On the face of it, the gap can be narrowed from either direction - either by denying that pluralism leads to unresolvable conflicts of value, or by limiting the scope within which politics can operate, leaving a pluralistic civil society in the space which it vacates. These varieties of response are versions of neutrality, the claim that political institutions and policy should, as far as possible, remain impartial between competing conceptions of value, with more or less optimism about the prospects of reconciliation between conflicting values.

As has already been argued, we should be sceptical about both of these forms of neutralist claim. In Rawls' theory, it is doubtful whether the 'construction'² which he now prefers as a method of modelling moral consensus will produce a set of political institutions embodying shared values or policies commanding universal assent. Similarly, the neutralist attempt to resolve the problem, by portraying the public realm as a minimal residuum left over when private conceptions of value have been accounted for, fails to recognize that the clash between values may emerge through rival views of where political action is appropriate or justifiable. The means of arbitration between values sought by the neutralist appears to be unavailable.

Towards the end of the last chapter I sketched an alternative method for understanding value and its relationship to politics to that offered by Enlightenment-influenced liberalism. This alternative would dispense with any sharp psychological or ontological division between reason and motivation of the form presupposed by both Humean and Kantian theory. If we reject this dichotomy several consequences follow. First, we should not think in terms of any hard ontic distinction of 'fact' and 'value', the former representing external states of affairs in the world and the latter internal affective states of agents. Second, we should abandon the idea that philosophical reflection on morality must divide itself into theories construing morality as representing states of affairs, and on the other hand theories reducing morality to purported objectifications of agents' internal states. Third, there is no warrant, once the distinction is questioned, for supposing that there is a basis for the psychological faculty-division of reason and motivation in the latter's alleged causal

role in action.

It is worth pausing to consider some of the consequences of these remarks. Most obviously, it becomes doubtful whether practical reasoning in general, and its political aspect in particular, can be dealt with satisfactorily by the purely executive account of rational agency offered both by Hume and Kant.³ We cannot regard practical rationality as involving only the calculation, from given motivations, of the best means to satisfy them. As has already been observed, one of the problems faced by neutralist liberalism is that it gives little guidance as to how we should deal with political decisions where there are different and inconsistent objectives which can be promoted through policy. Without any neutral procedure for deciding between them, there is not much to show how the instrumental view of practical reasoning can account for the fact that political action often involves choices between competing ends.

These comments apply quite generally to the Enlightenment theories of rationality under discussion, not only to political action - the theories have the same problem in coping with the idea that an individual agent may be faced with choices between final ends. A further assumption made by instrumentalism is that collective action can be seen as a vector product of the desires of individual members of the society. Given this reducibility, political action becomes a collective analogue of the processes followed by individuals in arriving at a practical decision. The same procedures, of selecting those motivations which present themselves as strongest, and devising means of satisfying them, are applicable, on this view, to both individual and collective decision-making. This reducibility produces a view of political action as only rational to the extent that it devises the best means of achieving ends given in advance, and determined wholly by individuals' motivations or preferences.

This account of political deliberation is supported by the fact/value distinction. Individuals' preferences are taken as fixed data which have to be serviced by those who take political decisions: the politician's role is reduced, on this basis, to providing relevant information about how to achieve determinate ends. The evaluative side of political argument can only be explained in these terms as the expression by politicians of individuals' fundamentally non-rational preferences. This leaves

it unclear how liberal theorists can produce any conclusive argument, by their own standards of rationality, for belief in neutrality as the supreme political value; or how to support the pluralistic doctrines which liberals have traditionally professed when pluralism is presumably to be seen as merely another non-rational preference.

A further consequence of the Enlightenment schematism, which again unites the seemingly disparate theories of Kantianism and utilitarianism, is their common denial that there is any insoluble conflict between morality and the values promoted by political action. Kant's remarks on the difference between the 'political moralist' and the 'moral politician'⁴ make it clear that politics plays only a subservient or residual role in action. Similarly, utilitarian theory regards actions as either morally required or proscribed - either a given course of action maximizes utility (however explicated, and relative to context) or not, and there is no sense in the idea that political action may run afoul of moral requirements. This again tells against one version of pluralism, that value is sufficiently diverse to leave us uncertain in some situations as to how best we should go about acting in accordance with our ideas of value.

This chapter aims to set out an alternative basis for understanding politics and its relationship with value, beginning from the critique of the Enlightenment theories offered in the last chapter. The next section will prepare some of the ground by providing a theoretical account of evaluative judgments which tries to retain standards of judgmental correctness while stopping short of full-blown cognitivism. This will not mean abandoning pluralism entirely, for reasons which will be given in the next section. But it will signal the end of the neutralist hope for a state which can arbitrate impartially between competing ideals of value.

This will not however be a restatement, as might be thought from this brief description, of classical Millian liberalism, favouring a form of pluralism within a quite definite conception of the desirable society - one with the maximum possible individuality, creative living, free expression and so forth. The point is not that these are not goods in themselves, but that there may be other goods with which the pursuit of such a society conflicts. A full commitment to pluralism acknowledges that there

is no single dominant set of values which society should promote, and no way of structuring political institutions from considerations of value, as Rawls aims to do: that is, the institutions are designed to conform to a conception of justice, and this is held to be the supreme social and political value. Pluralism subverts the aim, which in political philosophy dates back at least as far as Plato, of institutionalizing morality through political structures.⁵ In place of this, there can only be the acceptance that these structures themselves play an important role in deciding what conceptions of value are current within a society and how far it is possible to live in accordance with them.

This account obviously requires a realignment of the state-civil society relationship offered by neutrality. Some remarks towards this end will be offered later in the chapter. But first we need to establish a more solid foundation for pluralism while retaining standards of relevance and appropriateness in judgments. The next section will attempt to achieve this by developing a projectivist account of value.

II

As the last chapter argued, we should reject the idea that philosophical reflection on value divides into two jointly exhaustive positions, cognitivism and noncognitivism. If we accept the fact/value gap, the noncognitivist criticism goes through, that the cognitivist has no explanation of how moral beliefs can motivate agents to act. But this criticism has less force if we reject the distinction in favour of a more holistic pattern of explanation, according to which we should not assign the causal antecedents of action to a distinct psychological faculty, such as that (in Humean theory) of motivation, but rather think of causation in Davidsonian terms,⁶ as making sense only in relation to a specification of background conditions. Thus we should see action as consisting not in the conjunction of a causal statement about an agent's belief coupled with a further statement about his motivation, but in the applicability to the agent of an overall explanatory schema which includes essential reference to his beliefs and desires. In saying this, we do not reject the Humean

view that action is to be regarded as the outcome of beliefs and desires of the agent; but claim that the differences between beliefs and desires as factors in action does not consist in causal properties supposedly peculiar to desires.

This altered view of agency breaks down the instrumentalist assumption that there is a radical difference of type between cognition and motivation. On the instrumentalist view of agency, motivation is a brute fact in the world, beyond rational correction, while beliefs can only signify states of affairs in the world.⁷ If motivation had any cognitive component, it would no longer be capable of producing states of affairs in the world, because it would merely record them. There is then in Humean theory no possibility that desires can undergo rational correction. All that can happen is that an agent's motivational set alters for some non-rational (eg psychological) reason. One of the consequences of this is that political persuasion, understood as a method of rationally inducing agents to alter their desires (as opposed to the means of satisfying them), is impossible because it is in the nature of desires to be unalterable. Thus instrumentalism rejects rhetorical persuasion as either ineffectual, since aimed at changing the unalterable, or else non-rational. Either way, rhetoric is dismissed as an essentially disreputable form of argument.

But if we reject the Humean bipartition between reason and passion there is less warrant for this dismissal, or for the thought that rational persuasion is alien to political (or other forms of) argument. In developing an alternative to instrumentalism on non-Humean foundations, it is important to appreciate that what has been said does nothing to overturn an emotivist account of value along the lines Hume is often thought to have proposed, though it does require a different picture of the structure of emotions to Hume's own.

This different picture does indeed demand an account of the relationship between value and the emotions which differs from that offered by standard emotivist theory, since these have usually rested more or less explicitly on a Humean view of the emotions. On the most basic version of the theory, for example, a statement of value was straightforwardly reducible to a first-personal statement expressive or descriptive of some emotion of the speaker, construed as a qualitative internal state. Thus in

Stevenson's theory,⁸ 'This is good' means something like 'I approve of this', coupled with a prescription that others also feel approval.

This gloss on moral judgments, combining as it does an attempt to explain both the speaker's attitude and his attempt to prescribe to others what attitudes they should feel towards the object, faces insuperable problems when coupled with a Humean view of such attitudes as incorrigible psychological states. If, as Hume states, 'passions' are 'original existences',⁹ without any reference to external objects, as beliefs have, it is hard to see how the Humean account of valuation can be sustained with the alleged prescriptivity of moral judgments. There is no reason to think that one person's subjective valuation of an object can be altered by another's prescriptions concerning that object. Humean philosophical psychology reduces a Stevensonian view of the meaning of moral statements to the conclusion that these statements are, most of the time, entirely ineffectual.

Something like the Stevensonian account becomes more plausible, however, if we abandon the Humean assumption that psychological states such as those involved in this form of valuation must exclude a cognitive component. If it is possible that these states embody cognitive judgments about objects, a Stevensonian account of the nature of evaluative judgments looks more inviting. It then becomes possible to explain the prescriptive element in a moral or other evaluative judgment as not merely stating some subjective personal preference, but as embodying rational judgments. There need be no answer to the question whether these rational judgments are made concerning some external realm of facts as traditionally interpreted by moral realism. Instead, this revised form of emotivist account of value would assign to emotional states a non-contingent role in the formation of moral judgments, but would link the rationality of such judgments to beliefs involved in such states.

On the revised version of emotivism, then, there would be no suggestion that moral judgments were merely statements of subjective preference, as on standard Humean emotivist theory. Indeed, the subjectivity of the emotions themselves would to some extent be compromised on the revised account. It would still of course be true that emotions were ascribed to specific persons. But the ascription of an emotion to an agent would occur within a

broader interpretative context which serves to rationalize and explain the emotion. To rationalize an emotion in this sense is to provide a reason or set of reasons for the emotion's having occurred, and these reasons depend on the agent's beliefs - that is, the explanatory value of these reasons is conditional upon their referring at least tacitly to the agent's beliefs. In this sense there is only reason to impute an emotional state to an agent if there is an appropriate public context for doing so, which rationalizes the emotion by interpreting actions in terms of beliefs the agent has or may for independent reasons be assumed to have.

This form of interpretation, ascribing beliefs to agents as a means of understanding their emotional states, is clearly only possible if there are agreed norms of behaviour on the basis of which we can interpret actions arising from emotions. The rationality of emotional states consists in our being able to ascribe to persons emotions which are explicable given beliefs they are known to hold or could reasonably be expected to hold. This need not result in behaviourism, since it is not always the case that the grounds on which the relevant belief is imputed to the agent is the fact of his being in the emotional state itself. Our best reasons for holding agents to be in emotional states is the adequacy of the explanatory schemata which claim that such states exist and explain certain forms of behaviour.

Traditional forms of emotivist theory sought to establish a relationship between moral utterances and the rules for expressing emotions in language. These attempts attracted a good deal of criticism on the grounds that no general rules exist to compel a speaker to express some emotion whenever he is making a moral judgment.¹⁰ As far as this goes, the criticism is certainly justified: there are no general rules which require a speaker to express emotion in making such a judgment (which is not to deny that in some cases it may be impossible to grasp the sense of a moral utterance without recognizing in it certain locutions expressive of emotion). If we abandon the project of finding any such general linguistic relationship between morality and the emotions, however, it still seems plausible to suppose that emotional states play an important role both in persons' dispositions to moral judgment and in their moral actions.

Thus the claim is not that it is a condition of making any

intelligible moral utterance that it uses terms expressive of emotion. It is only too obvious that there are moral judgments whose expression need make no use whatever of terms directly indicating the speaker's emotions. At the same time, however, it is plain that terms which do play an essential part in the making of moral judgments may themselves require for their correct use that the speaker feels a certain emotional attitude towards the subject of moral appraisal. In this case, it would be claimed that at least some moral expressions demand in order to be understood that the speaker is taken to feel and express a certain emotion. This would not make some ultimate reductive claim, as was made by old-style emotivism, that there is nothing more to the making of moral statements than expressing emotions, or that the criterion by which we decide whether a moral utterance has been made is whether an expression of emotion has been used. But this revised emotivist position would hold that there are cases where the expression of emotion does enter, and not contingently, into the making of a moral utterance.

There are some moral expressions which can be understood only as expressions of emotion. The semantics of the moral terms themselves are such that a speaker who used them without feeling the appropriate emotion would have made an incorrect or insincere statement. This possibility of insincerity is important because it shows that the conventional meaning of the terms in question is such that an utterance of them can only be made by getting the hearer to recognize an intention to express the relevant emotion. This possibility of insincere usage does not, then, undermine the general emotivist thesis, but on the contrary reinforces it, as uttering an evaluative expression while not feeling the appropriate emotion would show that a moral judgment had not, contrary to appearances, been made.

As an example of this general thesis, we may consider a term like 'coward', as it occurs in moral utterances. It is clear that the term is standardly used only when the speaker feels certain emotions, such as contempt, about the person so described. This emotional response appears, moreover, to be an integral part of the (negative) moral evaluation: it is not clear that this evaluation is equally well expressed by some other formulation which makes no mention of cowardice, and passes the same moral verdict in terms lacking the emotional force of 'coward'.¹¹ There

may be non-standard cases, where the speaker is prepared to identify someone's behaviour in these terms, without disapproving of the conduct in question (and perhaps feeling a certain admiration for it). But this does not break the link between emotional reaction and moral assessment in the standard case. The fact that a speaker may admire a coward's behaviour (as such) does not show that the term, where it does pass a negative moral verdict, can sincerely be used without expressing emotion. In such cases, then, we are not to regard the making of moral judgments as involving two separable components, the first an evaluative verdict and the second an emotional addition to it; the best account of the person's behaviour may be that it is contemptible, and no reformulation not expressive of emotion can capture this moral assessment.

One way in which the emotivist thesis may be justified, then, is that evaluative terms may express emotion as an essential feature of certain moral judgments. As a consequence of this, it is reasonable to suppose that there are at least some moral notions which are fully intelligible only given a grasp of their emotive content. According to this view, the sense of the relevant moral notions would depend upon the expression of emotion, and it would usually be a condition of learning the use of such moral expressions that the language-user grasped paradigmatic cases as involving a certain emotion. On this account, the circumstances for learning and using the evaluative expression would be ones where a certain (specifiable) emotion was expressed. Recognizing and intending to express the emotion would then enter into the semantic rules for using the term.

From this position it is easier to see how a modified form of emotivism could explain how the emotions can play a part in moral thinking, without becoming impaled on either horn of the dilemma discussed earlier. We can understand moral thinking as rational without falling into cognitivism. On this account, rationality consists not in the supposed correspondence of moral statements with some independent realm of facts about the world, but in the conformity of moral judgments to accepted patterns of evaluation: it is the intelligibility of moral judgments founded on emotions which sanctions their rationality.

This is not to say that there is some algorithm which could in principle determine the rationality or otherwise of a given moral

judgment. Dispute can enter at several points. To take the cowardice example, while the conceptual link between sentiments of disapproval and acts of cowardice is well-established, there is room for disagreement over whether a particular action counts as cowardly; a fictional example is Garcin's agonizing in Sartre's play Huis Clos over whether he has upheld his pacifist principles or, less admirably, run away in the face of the enemy. Clearly the moral evaluations and their associated emotional responses will differ depending on which is true. At a criteriological level, it might be disputed whether a certain kind of action was cowardly. It might be debated, for example, whether conscientious objection to conscription could be described in these terms. It is possible to imagine arguments being offered in these disputes, particularly in the first case. At the same time, there is no need to think that this discussion must be rationally terminable, or that there are quasi-forensic rules which can mediate disputes even of the first kind. Beyond this, disagreement may arise over the level of emotional response deemed appropriate. Here again, it is clear that there is some link between the level of this response and its associated moral evaluation - in the cowardice example, it is plausible to think that the degree of contempt aroused in an observer by the action in question will be reflected in his moral assessment of it.

Given that the possibility of such disagreements about emotional responses, it might be asked whether there is anything which can set standards of correctness or appropriateness in such responses, in the absence of a systematic procedure for determining their nature and level. It might be thought, in view of the forms of indeterminacy distinguished above, that the supposed rationality of emotionally-based evaluations was illusory. And this could be taken as confirming the fundamentally non-rational nature of evaluations; since these are founded on non-representational 'passions', there is no reason to expect that they will concur.

A partial reply to this is to repeat the point made in the last chapter that, whatever their evaluative role, emotional states undeniably incorporate cognitive judgments about their objects.¹² To this extent they cannot merely be seen as non-rational and hence non-representational states, at least in their occurrent forms.¹³ As was argued before, it is reasonable to suppose that

these object-directed attitudes are beliefs: the warrant for this step being, again, the explanatory value of imputing object-directed beliefs to agents in emotional states: for example, it was sufficient to explain why someone was in a state of fear to point out the agent's belief that some danger was present. Often, of course, the relevant belief is not mentioned directly (though it has to be if the object is not in fact present), but understood elliptically. It is plain that this elliptical form of explanation would be impossible if there were no uniformity in agents' evaluations in such situations. It is this evaluative convergence which makes it possible to understand their emotional reactions.

The question is whether there are any criteria of rational adequacy in the cases where this sort of convergence does not exist. These, notoriously, are cases where observers may arrive at different or even wholly opposed evaluations of some situation. Part of the answer is to be found by considering what claims can intelligibly be made regarding a proposed reaction to some moral or political decision. It is a condition of rational adequacy in some cases of evaluative conflict that the disputants are capable of making out comprehensible claims for the reactions they favour. It is important to appreciate how much this already excludes from argument, and to recognize here an instance of the point already made, that deliberative priorities are often most obvious in the omission of certain concerns from practical consideration. Obviously there can be debates about such priorities, but tacit evaluations are also evident in what is excluded from discussion. One prominent example in political debate is in the means considered appropriate to promote (commonly agreed) ends. While there may be disagreements over how best to realize certain ends, there may also be other possible courses of action which are dismissed by never being considered, even within a political culture prizing executive values of effectiveness.

Divergent political evaluations may be the result of factual disagreements, but it is also clear that they may well not be. There is also the possibility that different criteria of factual relevance will be mobilized in support of the competing evaluations - a consideration which is enough by itself to cast doubt on any rigid distinction of neutral 'facts' from non-neutral 'values'. Thus the disputants may fail to agree on what counts as a relevant consideration, or how a certain item of information

bears upon the issue under discussion. These possibilities need not raise the spectre of some grand conflict between ultimate values. They may involve instead different conceptions of the relationship between values and practical action: for example, disagreement may arise over which policy or course of action promotes some value such as equality, or liberty (the chances of disagreement appreciate, of course, where there is the possibility of conflict between values).

There seems little hope of reviving the Enlightenment project of providing rational foundations for morality - if, at least, this is taken to mean a procedure for reaching definitive moral conclusions from agreed initial premises. In so far as liberal neutrality is premised on the success of such a project, it miscarries. Similarly, Kantian-influenced theories of discursive rationality such as Habermas's seem to hold out no prospect of final agreement on values. Even when the inequalities of wealth and power which Habermas holds responsible for the distorted communication between groups in capitalist societies are notionally absent, it is hard to see why disagreement over values must disappear. Those enjoying unequal political or social advantages may subscribe to different values. But conflicts of values may persist when these inequalities are removed. In Rawls' Original Position thought-experiment, for example, the absence of racial, sexual and class distinctions between the parties need not preclude, as Rawls himself makes clear, a confrontation between values (at the same time, Rawls adheres in the construction of his theory to a discursive model of rationality in which fundamental disagreement about certain values, such as that of justice, seems to be regarded as pathological).

This restates the point already made, that there is a fairly poor fit between pluralism and neutrality. In summary, the problem is that the neutrality of procedures for resolving conflict is likely to be disrupted once pluralism is allowed into the reckoning. For pluralism, on one natural understanding of the notion, just means that there is no procedure capable of resolving all conflicts of values. If the values are discrete and embodied in different groups or sub-cultures, pluralism may surface in radically different conceptions of procedural neutrality - or, indeed, of its desirability. Even without this fragmentation, there may be no means of settling conflicts which occur within

sets of values shared by the community as a whole: for example, if the community values effectiveness in public life, it may face conflicts with the moral demands it imposes on politicians as professional operators. It is at best doubtful whether these considerations can merely be offset against one another on some ultimate balance-sheet. If there is a genuine plurality of values in society, there will be no means of projecting them all onto some single scale as a means of reconciling any conflicts which may arise between them. Unless neutrality is construed as permitting a free-for-all in which civil society disputes are allowed to take their course without political intervention, this possibility of conflict is likely to produce a number of competing political solutions.

Procedural neutrality seems to be ruled out once we abandon the Enlightenment foundationalist project. It is natural to suppose that if the rational foundations exist, then there should be some mechanism for resolving (apparent) moral conflicts. But if pluralism is true then either these foundations do not exist or else their existence is not inconsistent with rationally interminable moral conflict. It may be asked at this point what is left of the project of rationalizing moral evaluation when the stronger, conflict-resolving form of foundationalism is abandoned. One of the unsatisfactory aspects of noncognitivist accounts of value, as Wiggins has noted,¹⁴ is that they seem not to do justice to the phenomenology of moral judgments - that they appear, to the persons making them, to be something more than mere subjective expressions of preference. Moral judgments reach out, as it were, to a realm of value beyond the individual who makes them. Even if moral realism is rejected, we still need some explanation of how the moral life could appear to have this character. One possibility is that foundationalism fails, but we can still provide partial rationalizations of moral judgments - though not the full dispute-resolving form of moral justification sought by foundationalism. Moral realism, on this account, remains at best an unproven dogma: our moral rationalizations cannot be reduced to consistency, so it must remain a moot point whether or not there is some ultimate standard to which moral judgments appeal. As Wittgenstein wrote,

When I have exhausted the justifications I have reached bedrock, and my spade is turned. Then I am inclined to say: 'This is simply what we do'.¹⁵

The shift from the first person singular to first person plural in this quote is presumably meant to indicate a standard of judgment beyond the purely subjective. Although collectively we can produce some reasons for our intersubjectively-agreed moral judgments, these fall short of the ultimate justification sought by foundationalism. How cohesive an answer this gives, and how resistant it is to challenge, depends heavily on the extension of the 'we' in the final sentence. Clearly, and trivially, its limits can be set so as to embrace only those who happen to agree on a set of moral beliefs (with, as Wittgenstein would have insisted, their associated practices or 'forms of life'). If, however, its limits are extended beyond this, the possibility arises that moral conflict may be endemic in certain cultures, such as our own. The price exacted for this greater authenticity is a corresponding diminution in our powers of moral justification, even in internal disputes within the culture. There may be ample ground still to be cut away between the disputants at the point when the spade turns.

The intelligibility of rival moral justifications depends, as already noted, on shared patterns of judgment, even in the face of substantive disagreement: dispute can only occur within a form of life if these patterns make allowance for it. (Wittgenstein would perhaps have claimed that it was constitutive of mathematics, for example, as a human practice that it makes no such allowance.) Thus in the case of certain so-called 'thick'¹⁶ ethical notions, such as that of cowardice, or lying, it is true that there are certain agreed criteria for the application of the terms which denote them; at the same time, their correct use involves the making of certain evaluations, and their being accompanied by certain emotional responses. Dispute, when it occurs, is not (except in the context of philosophical discussion about the meaning of the terms) merely criteriological. What is open to dispute is whether the terms' criteria of application are satisfied or not. If they are not, it is clearly inappropriate to feel the accompanying emotional response; but if they are, the terms cannot correctly be used without feeling the emotion. It is in this sense that their use is embedded in a 'form of life'.

This is important in contributing towards the sense which Wiggins identifies, that moral thinking is something more than a roundabout statement of subjective internal states. Moral responses can be seen as examples of projection: the ascription to

objects (including states of affairs) of certain evaluative properties. It is of course true that projection of this kind demands that we come to view the object or action concerned in a certain light, and there is no notion that the object 'really' possesses the relevant properties in any sense of 'really' which does not refer to the projection itself. This would perhaps not satisfy Wiggins, who might reply that once we have acknowledged that this projection is going on, we cannot retain any belief in the objectivity of moral judgments which morality requires to secure any hold on us as agents. But there is no reason to think that the motivational grip of concepts which occur in the agent's evaluations is in any way lessened by the acknowledgment that the attribution of these concepts is only a matter of externalization by the agent. Wiggins leaves it unexplained how moral judgments secure their motivational hold.

The arguments of the present section have been designed to suggest that we need not conclude that moral thinking is either motiveless or meaningless. The belief that such a dilemma exists is partly the legacy of Enlightenment styles of philosophical thinking which partition the world between fact and value or between cognition and motivation. Once we abandon these habits of thinking, it becomes easier to see how we can reach an understanding of moral (and other) evaluation which is both rational and, contrary to assumptions about rational requirements of coherence in morality, which allows for pluralism.

On the account of value presented in this section, pluralism is compatible with the claim that there are rational criteria by which evaluations can be assessed - though there need not be a single method of rational assessment, nor need there be an answer to all evaluative questions. Intelligibility places constraints on the reasons which may be produced in support of a given moral assessment, even though there may be no means, in many cases, of arbitration when these reasons conflict. But this need not reduce moral argument to rationally unstructured conversational manoeuvring. Nor need the claim be that the discussion creates wholly internal standards of rational acceptability, as Rorty sometimes seems to maintain.¹⁷ Even without the prospect of rational terminability envisaged by some models of neutral dialogue, moral argument can still be guided by rational considerations other than the residual constraints of Rorty's

'great conversation of mankind'.¹⁸

These thoughts are admittedly very partial and inadequate. This section has however tried to make it clear that there is a tension between neutrality and pluralism as political ideals, which recent liberal theorists have largely failed to reckon with. To accept pluralism is to question the neutralist aim of arbitrating between different conceptions of value. This is not to say that pluralism entails any necessary conflict between the values in question. But where different groups or the ideals they support are in competition for public benefits such as funding, there is always the possibility of conflict - not necessarily because the values themselves are at odds, as might be the case, for example, with ideals of equality and natural aristocracy, but because public support of this form cannot be administered even-handedly.

Pluralism, and in particular the possibility of the more radical forms of value-based conflict, seems to pose the Rortian threat of purely self-rationalizing unstructured discourse. It may seem, moreover, that the only defence against this threat is a procedural conception of rationality of the form favoured by liberal neutralists like Ackerman.¹⁹ But, as the present section has argued, we can dissolve this apparent dilemma if it is coherent to think of evaluative discourse not as proceeding according to fixed rules, but as rationally structured by considerations of intelligibility. It is because we can situate evaluations in an overall setting which makes persons' actions understandable that we can describe them as rational; the converse also holds, that irrationality in this connection just consists in the contextual unintelligibility of persons' evaluative utterances. This leaves, and is intended to leave, considerable scope for disagreement within the bounds of intelligibility. But it is not by reference to some (internal or external) semantic model that we determine whether a given evaluative utterance is rationally acceptable. The mistake is to look beyond the shared language for some ultimate guarantor or arbiter of acceptability in evaluative judgments.

In On Certainty Wittgenstein addressed these issues in the following terms.

Giving grounds, justifying the evidence, comes to an end; - but the end is not certain propositions' striking us immediately as true, i.e. it is not a kind of seeing

on our part; it is our acting, which lies at the bottom of the language-game.²⁰

The foundations of the language-game cannot themselves be described as being true or false, since they decide what is to count as true or false. The point is not, presumably, that action merely replaces propositional foundations for moral or other forms of discourse. In the next section the focus will turn to political activity as an illustration of the Wittgensteinian arguments presented so far: the aim will be to discover how cohesive a notion of political rationality remains once the foundationalist project is abandoned, and to what extent a praxis-based epistemology can replace more orthodox categories of understanding political action. This should help towards recovering a non-instrumentalist view of politics in which there is a more dynamic relationship between agents and reasons for action than in much modern liberal writing.

III

In some ways the account of practical reasoning and action to be developed in this section resembles Habermas's theory of 'communicative action'.²¹ Both share an emphasis on the importance of action in rational communication. Habermas's theory, however, centrally claims that the process of reaching ideally undistorted communication is essentially a matter of generating agreement between the participants. It should be clear from the foregoing why the present account rejects Habermas's approach, with its strongly Kantian assumptions about rationality. By contrast, this section will argue that there is no effective (i.e. practically operable) notion of undistorted communication, with its assumption that all imbalances of power have been eliminated. What is left is a model of rationality in which these imbalances are regarded as being endemic to political action, and are regarded as expressing the pervasiveness of disagreement. A world in which the use of political power is as entrenched as it is already contains discrepant practical judgments.

The previous section argued that rational discourse can proceed in spite of evaluative disagreements. The existence of intelligible disagreement between political agents suggests that

there may be norms of discursive rationality even in the absence of 'agreement in judgments'. As one prominent strand in liberal thought has always stressed, it is possible to see the prevalence of such disagreements not as the product of some error, but as being a positive good. Indeed, as Mill argued,²² the prevalence of contrary opinions or judgments may sustain rather than damage their integrity as discrete bodies of doctrine. It is the experience of conflict which in some cases enables the participants to understand, from within such a doctrine, what their own rational purposes are.

If certain forms of evaluative disagreement are rationally interminable, as argued in the previous section, there is no reason to think that a single paradigm of rational interchange, such as is offered by liberal neutralists or by Habermas, will enjoy the universal applicability these writers expect. Pluralism, in the sense in which it has been understood in this work, just is the claim that there is no means by which certain fundamental moral disputes can be rationally resolved. This already casts serious doubt on the project of deriving political institutions and processes from supposedly consensual moral notions: if the notions are modelled accurately, they will have to allow for the possibility of practical conflict.

There are, it should be said, other obstacles in the path of any very general theories of moral rationalization. One notable example is the role of notions of character in explaining action. While these need not be mirrored directly in agents' practical reasoning, it seems undeniable that they are frequently of use in explaining actions. The most direct way in which such notions could feature in deliberation is through an agent's acting out of a reflective conception of himself as a certain kind of person, i.e. as having a certain character. But, though this does happen, an agent's reasoning may express his character while not mirroring it directly in his subjective deliberation - for example, in the respective weights given to different concerns (which is not to imply that there need be any single scale on which all such concerns can be weighed). The notion of character may be indispensable if we are to bring subjective and intersubjective accounts of action into equilibrium.

Once the notion of character is introduced into analysis of rational agency, then, there is no longer any reason to suppose

that there is any completely general structure of rationalization in practical reasoning. If not, some degree of agent-relativization seems to be called for in analyzing rational action. It is common in utilitarian writing to assume that the belief/desire model can handle changes in deliberated wants by allowing for the possibility of improved, or at least altered, information over time. But the motivational discontinuities under consideration here are more radical than this purely cognitive possibility. Nor are they adequately handled by the thought that a person's desires may change in the manner envisaged by Humean philosophical psychology. The latter gives no countenance to the idea that such changes may be rational, rather than brute (e.g. physiological) facts of nature. By contrast, the present account construes these changes as the result of a rational process which is, however, not in principle open to the kind of synchronic deliberation allowed for by the Enlightenment models.

In part the more dynamic conception of practical reasoning is manifested in the relationship between character and moral reasoning. One way in which a person's character may significantly change is in the kinds of deliberative concerns which have importance for him (a change which may equally affect his attitude to non-moral practical concerns). This is a case in which the synchronic model of practical reasoning proves inadequate - the situation is not one where it would be possible in principle, with perfect information, to complete the reasoning in one go, so to speak, because the example involves changes in the agent's perception of what has deliberative value. The agent could not, therefore, execute the reasoning at the start, unless he foresaw that he would undergo such changes; and this realization would only give him reason to act on his future values if he saw them as being in some sense an improvement on his present ones - which is as much as to say that the latter had ceased to have any deliberative weight for him.

At the most fundamental level this comes out in the way practical concerns are distributed over a person's life. It is integral to our notion of a life that the person living it can develop an understanding of his practical commitments through practical experience, and that those commitments sometimes undergo radical transformation as a result of this. The forms of rationalization which a person can apply to his past actions in

such circumstances are in their nature different from those available to him beforehand. These transformations are such that the agent's understanding of what counts as an effective rationalization will itself depend on what he has lived through, not only in the obvious sense that his deeds determine what has to be explained in the first place, but also what forms of rationalization count for him as providing an explanation of what he has done. It is not a question of applying some pre-determined theory to the vagaries of subjective experience, but of working out an understanding of action from experience itself.

This emphasis on the experiential character of subjective rationalization does not have to assume that the process of accounting for one's past actions is therefore trivially self-justifying. It may be said that the situation is all too familiar in which an agent is drawn to a particular form of justification because it conveniently squares with what he actually did. But even the mode of justification which consists solely of a redescription in commendatory terms of an agent's actions has to make itself intelligible to its intended audience if it is to achieve its object. The audience must not only recognize the redescription as an intended justification, but as being applicable to the action in question. The project of justification has to work within the public language constraints already noted. To claim that such forms of rationalization may be circular and therefore self-defeating ignores the anti-foundationalist thrust of Wittgenstein's arguments. If meaning is, as these arguments claim, to be explicated not through an internalized semantic model but through public practices or language-games, the justification for our using evaluative terms must in some way consist in these practices and not in their relationship to some ultimate model of their meaning.

A clear example of the limiting of options in rationalization is the character of Garcin in Sartre's play Huis Clos, mentioned earlier. Garcin is tormented by the thought that his past life has failed to substantiate his own self-image as a man of action and courage. His fumbling attempts at self-justification waver between two accounts of his actions, which depict him either as adhering to his pacifist principles or as running away from danger. Part of Garcin's agony, as Sartre represents it, is his own realization that these alternatives exist, and that he may have acted out of

cowardice while believing himself at the time to be motivated by principle. It is essential to the situation that either of these descriptions, while mutually incompatible, could have been applied in principle to Garcin's projected action at the time he was deliberating it. But it is made clear, and Garcin is aware, that subsequent events in his life have pushed out of reach the account of his actions which he wants - a fact which is brought out remorselessly by those to whom he is trying to justify himself. The circumstances of Garcin's death make it impossible for him to redescribe his actions as he wishes - not because death makes anything impossible (Sartre's dramatization removes that source of difficulty) but because they simply remove his desired form of self-justification.

So far it has been argued that there is an irreducible experiential element in rationalization, not only in the obvious sense that what has to be rationalized (retrospectively) depends on what an agent has done, but also that the repertoire of justifications open to an agent is limited by his past actions. Hegel's historicist teleology is one example of this experiential claim.²³ But the present account differs from Hegel's in maintaining that there is a plurality of standpoints from which an action can be rationalized: there is no definitive position - in Hegelian terms, that of absolute Spirit having achieved full self-consciousness - from which the course of history, or an agent's life, must be understood (a claim which is not, of course, inconsistent with the argument above about the experiential limits of rationalization). At the same time, there is no assumption that an agent must be able to rationalize his actions in some way or other, merely by dint of having consciously acted (a further implication of Hegel's views).

As in the case of Garcin, the only possible rationalization may be one which, given the agent's other dispositions and commitments, was not deliberable ex ante. This is particularly clear in circumstances involving akrasia. For if akratic action is possible at all, it must be seen as non-deliberable. An agent can only go wrong in this way if his actions fail in some way to correspond to his deliberated reasons - which is not, however, to claim that the akratic action cannot be captured by some intentional description. Since akrasia has this surd character,²⁴ a correct appraisal of akratic actions has to take into account

their non-rationalizable status. Garcin's difficulty, in trying to come to terms with his past, is that the only form of explanation which he regards as rationally acceptable is inapplicable to his actions. His quest for rationalization cannot deal with the brute fact of his cowardice.

This example amplifies the point made earlier in this section about character as a limiting condition on rationalization. Garcin's situation is not one of deficient information. It is not that he could, had he known at the time, acted better - that there is to hand, ex post, some account of his action which could have featured among his reasons for acting. Rather the best account of his action is one which itself acknowledges the limits of rationalization, and uses instead a character-based notion (that of cowardice) in describing what he did. But this characteristic does not stand on an equal footing with other forms of rational explanation, as it could not rationally have figured in Garcin's own deliberations - both because of the nature of Garcin, and because of his cowardice. The account of his action which can be given afterwards is not one which would have been available to him at the time of acting. There are attributes of character which are manifested in the agent's inability to accept certain ways of rationalizing his actions.

These remarks to a significant extent apply to political action also, in its individual and, more ambiguously, its collective forms. At the individual level it is plain that character-based notions are frequently employed in political explanation, to an extent belied by much theoretical writing: for example, the whole question of a politician's style of professional operation is closely involved with notions of character. Such questions are not easily handled by theories which take policy-making to be impersonal matters of executive or administrative decision. It may be said, of course, that these considerations, while undeniably a feature of real-life politics, would ideally be absent from a rationally ordered political culture. In particular, it may be said that vagaries of character are only of significance in imperfectly democratic societies, where political decisions depend to an inordinate degree on personal fiat.

It is not possible here to discuss in any detail the relative merits of representative and delegatory or participatory theories of democracy. No doubt it is true that issues of personal style

their non-rationalizable status. Garcin's difficulty, in trying to come to terms with his past, is that the only form of explanation which he regards as rationally acceptable is inapplicable to his actions. His quest for rationalization cannot deal with the brute fact of his cowardice.

This example amplifies the point made earlier in this section about character as a limiting condition on rationalization. Garcin's situation is not one of deficient information. It is not that he could, had he known at the time, acted better - that there is to hand, ex post, some account of his action which could have featured among his reasons for acting. Rather the best account of his action is one which itself acknowledges the limits on rationalization, and uses instead a character-based notion (that of cowardice) in describing what he did. But this characteristic does not stand on an equal footing with other forms of rational explanation, as it could not rationally have figured in Garcin's own deliberations - both because of the nature of Garcin, and because of his cowardice. The account of his action which can be given afterwards is not one which would have been available to him at the time of acting. There are attributes of character which are manifested in the agent's inability to accept certain ways of rationalizing his actions.

These remarks to a significant extent apply to political action also, in its individual and, more ambiguously, its collective forms. At the individual level it is plain that character-based notions are frequently employed in political explanation, to an extent belied by much theoretical writing: for example, the whole question of a politician's style of professional operation is closely involved with notions of character. Such questions are not easily handled by theories which take policy-making to be impersonal matters of executive or administrative decision. It may be said, of course, that these considerations, while undeniably a feature of real-life politics, would ideally be absent from a rationally ordered political culture. In particular, it may be said that vagaries of character are only of significance in imperfectly democratic societies, where political decisions depend to an inordinate degree on personal fiat.

It is not possible here to discuss in any detail the relative merits of representative and delegatory or participatory theories of democracy. No doubt it is true that issues of personal style

will be more to the fore in regimes where power is concentrated in relatively few hands. But it is mistaken to conclude from this that these dimensions of political action are eliminable - or that ideally they should be. To some extent, in any case, the question of forms of political organization is irrelevant, since if the general claim about character as an explanatory notion is correct, it applies to political action however decentralized or participatory the form of government. Nor is it part of the present thesis that these considerations are always foremost or of fundamental explanatory significance. There are, however, forms of political behaviour which can only be satisfactorily accounted for using character-based notions.

Most obviously these occur when a conflict of policy is embodied in two (or more) personalities proposing them. These may take place even where the policy issues, viewed in isolation, seem fairly clear-cut. The power struggle in the Soviet Communist Party between Stalin and Trotsky after Lenin's death in 1924 is a case in point.²⁵ While the ostensible political issues concerned the direction of economic policy and whether to pursue the path of 'socialism in one country', it is clear that (though these issues in themselves were of course substantive) the episode is fully intelligible only as a clash between the dominant personalities in the party not merely over policies, but power - a point vividly confirmed by Stalin's conversion in the thirties, after the Trotskyite faction had been defeated and purged, to the very policy of industrialization he had condemned as anti-Leninist during the power struggle. In this case, it would be naive to view Stalin's original adherence to a proletariat-peasant alliance and his subsequent volte-face as deriving purely from considerations of policy (though Trotsky later represented the power struggle as being wholly abstracted from questions of personality²⁶).

These events seem to leave out of reach the Marxian objective of a neutral politics consisting solely in the 'administration of things', just as, in the context of liberal democratic politics, the neutralist aim of finding some purely procedural means of resolving political dispute is illusory. In part this is because there are certain institutional factors in democratic political societies which serve to accentuate rather than eliminate conflict. But it is doubtful whether political conflict would disappear in the absence of these factors. This is apparent in the

fluctuations in the level of ideological conflict which a given system may manifest over time.

In Britain after the Second World War, for example, the nationalization programme undertaken by the Attlee government was broadly accepted by subsequent Conservative administrations, and a new 'Butskellist' consensus formed around the welfare state. This social democratic consensus continued until the Conservative election victory of 1979. It is hard to avoid the conclusion that the end of consensus politics was largely the effect of Margaret Thatcher's personality as Conservative party leader and her adversarial style.²⁷ The separation of style and content is to some extent over-schematic, since (as in this example) the more confrontational style of Thatcherite politics accompanied a radical revision of policy. This however could be taken as a confirmation rather than a refutation of the argument of this section, depending on the direction in which the (partial) reduction is taken: the example suggests not that personality is reducible to policy so much as that the formulation of policy is inseparable from questions of personal style. Part of the reason why a person may find certain forms of policy and styles of argument compelling is that they exhibit a dynamism and clarity (or, unfavourably viewed, excessive simplicity) which suits his personality.

The matter of style in political argument, it should be noted, has ramifications beyond the character-based considerations already discussed. Some of these impose further limits on the neutralist ideal of discursive rationality. For example, it is plausible to suppose that, considerations of personality aside, certain styles of argument comport more naturally with some strains of political ideology than with others. This can lead to the acknowledgement that the discursive ideal will only be persuasive to those already disposed to accept certain patterns of (liberal) political justification. In a period in Britain where the political centre of gravity has shifted from centre-left to the right, radicals on the left often resort to forms of argument which explicitly reject the possibility of consensus and of meaningful dialogue. Here, again, it is hard to view style and content as wholly separable - particularly in view of the dialectical structure of these Marxist and neo-Marxist ideologies. In the latter, a commitment to neutrality of procedure is

misleading because it veils anterior ideological commitments, primarily concerning the location of economic and political power. To accept a certain argumentative style is, then, to be committed already to certain forms of justification and therefore to substantive - and contestable - ideological claims. Perhaps these assertions are wrong. But it is hard to maintain the appearance of neutrality when they are made in the name of genuine political commitments. It is for the neutralist to show that they are wrong from entirely neutral argumentative premises.

This naturally raises the question of the relationship of the present study to these questions of commitment. If it abandons neutrality, then it seems to be limited, on the arguments above, to those of a particular ideological persuasion; but if, on the other hand, it pretends to neutrality, it would appear that the impartial standpoint is after all available, contrary to what has been argued. Surely, it may be said, neutrality demands that the foregoing arguments are ideologically non-partisan, and if they are, neutral forms of political justification do exist after all.

An objection of this kind leads fairly quickly to larger issues concerning the relationship between theorizing and practice in political philosophy. I will offer some further comments on this topic at the end of this chapter. It should be said in advance, however, that I take the tenor of the argument so far to be sceptical about the closeness of fit between philosophical theorizing and substantive political commitments. If there were this closeness of fit, the objection would have some force, since there would be less reason to think of first-level and theoretical argument about politics as distinct: in this case, it would be harder to maintain that there is a sharp line between partisan argument in politics, and ideologically neutral philosophical reflection about politics. This chapter and those preceding it have already given pluralistic reasons for doubting whether neutrality can be imported into political structures as readily as the liberal neutralist supposes. At the same time, it does not appear to be inconsistent to combine the form of pluralism identified earlier with a theoretical account of politics which aims for generality and avoids explicit ideological commitments - beyond repudiating a certain recent strain of liberal theory.

This is, admittedly, a difficult position. It can best be sustained, I think, in the presence of a firm distinction of

levels between philosophical theorizing about politics and explicit ideology. Given this distinction, it seems possible to argue against liberal neutrality at the theoretical level, and for reasons which are themselves based on a conception of neutrality or at least of even-handedness. This possibility becomes clearer in the light of the redefined pluralism explained earlier.

Pluralism, as was argued, can be interpreted not merely as claiming that there is a diversity of values in society, but also that this may be expressed through the political competition of groups representing or supporting these values. On this view, pluralism cannot be contained within any single definitive procedural or institutional expression of the values concerned; the values are so heterogeneous that the relationship between them must reflect competition as well as coexistence. (This is not to claim that conflict can only occur between groups each of which single-mindedly espouses one and only one value: a group with a more generous range of values may experience internal conflict.)²⁸ Given this version of pluralism, it is possible, I think, to describe ideological conflict as being enacted through the conflict of values, without giving way to explicit first-level partisanship. It seems unexceptionable to maintain that liberal neutrality is inconsistent with this form of pluralism, while also arguing that the present analysis makes no explicit commitments between first-order values or ideologies. For liberal theory holds that neutrality can either be embodied in specific legislation, or at least in the terms of competition between competing ideologies. But the present account differs even from the latter procedural claim, since it denies that there is a clear notion of fair ideological competition: indeed, an incommensurability between values of the form manifested by this radical pluralism is partly the grounds of this denial.

It is, again, important to keep in view the wider implications of the present discussion. The possibility that character can play a role in explaining action is disruptive of neutrality. What differentiates characters, in part, is their embodiment of diverse and in some cases incompatible values: as was argued earlier, part of the distinction between characters consists in the significance which practical concerns have for different persons. In this sense, what distinguishes persons is just the differential significance they attach to such concerns. Pluralism is in part,

then, a character-based phenomenon: given this differential weighting of concerns, it is unlikely that the neutralist aim of steering an impartial course between values can be sustained. There need be no purely procedural means of determining what persons have good reason to value. The limits on neutrality, accordingly, derive both from the different substantive values which persons endorse, and from the absence of any agreed mechanism for deciding between these values.

Parallel considerations impede the attempt in moral philosophy to devise rational procedures for reaching moral agreement, as some recent work has suggested.²⁹ Once the focus moves from algorithmic methods of Enlightenment-influenced moral theorizing to moral or ethical notions related to character, such as that of virtue, it becomes much less obvious that there can be any mechanical procedure for resolving disagreements. If different persons embody different characteristic virtues, then at least some forms of ethical consideration must apply to them in different ways, and this seems to preclude the blanket application of moral obligations to all (similarly situated) agents in the way Kantian theory claims. This means, for example, that a given kind of practical reason will have greater force for one character than for another; it quickly follows, at a general level, that the force enjoyed by a particular category of practical reasons, namely morality, will vary from person to person rather than being applicable to all rational agents as such.

Less generally, specific kinds of moral concern may depend for their practical force on considerations related to character: one way in which this can come out is in different attitudes towards moral conflict and its resolution - or even to whether there is a moral conflict to be resolved. It is of course conceptually possible that someone can fail to recognize the latter through moral ignorance (as Socrates, in Plato's account of him, thought was true of all wrongdoing). But discrepant moral perceptions of some practical decision need not arise purely from ignorance. They may result from differences of character which cannot be explained as mere cognitive failures. One of the ways of distinguishing a person's character consists precisely in what weight she gives to forms of practical reason and how she is disposed to act on them.

Perhaps it will be said that this only reveals a difference in the extent to which people are prepared to act on a schedule of

moral obligations which is, in itself and with due allowance for agents' circumstances, both static and applicable to all agents. Reasons have already been given for doubting either that such a view captures the nature of persons' actual moral thinking, or that a universalistic theory of this kind is the definitive expression of moral rationality. Certain of our moral - or perhaps better, ethical³⁰ - notions are resilient to this sort of treatment. For example, the character-related notion of integrity is hard to bring within the universalistic theory: one aspect of integrity (not the only one) concerns the relationship between a person's character and her moral actions or dispositions. It is plain that the practical importance given to integrity as a value varies from one person to the next, and that this in itself will affect the way she views her other practical (including moral) concerns.³¹ It may be debatable whether integrity is necessarily a virtue - it is fairly obvious that a wholehearted commitment to integrity may only be possible if certain other concerns are neglected. But it is undeniably a character-related notion if it is a virtue at all, and therefore difficult to draw into a completely general theory of rational agency.

It is notable that certain virtues have been almost ignored in recent Anglophone moral philosophy, while others have been misconstrued because efforts have been made to bring them within the compass of a general rationalistic theory structured by the notion of obligation. One example of the latter is the virtue of tolerance. Attempts have been made³² in some recent writing to fit tolerance into the general structure, with the result that important features of our notion of acting tolerantly have been obscured. One of these is the centrality of the agent in accounts of tolerant action: not merely in the obvious sense that any action must be performed by an agent, but in bringing out the fact that the agent's own values and moral commitments need to be understood if his acting tolerantly is to be understood.³³ Another feature is the non-obligatoriness of tolerant action. Again, this may be better understood as a disposition to certain forms of forbearance on the part of the agent than through obligation-related notions like rights (to, for example, religious toleration), which reintroduce the idea that the action is either required or at most morally indifferent. There is little point in modifying obligations to give them greater flexibility, as duties

of 'imperfect'³⁴ or 'prima facie'³⁵ obligation, whose stringency conveniently relaxes to suit theoretical requirements.

Some virtues have simply been ignored because they resist explication in terms of moral obligations. The virtue of kindness, for example, falls into this category. This need not be because kindness is a trait not much valued in modern society, though (as I shall suggest below) there are political as well as theoretical reasons why it has been relatively neglected in recent work. From the theoretical angle, it should be fairly clear why the Enlightenment accounts of morality have little to say about kindness: in a framework which regards all actions as either rationally required or proscribed, no room is left for the idea that there can be actions (and an associated disposition to perform them) which, though non-obligatory, are however in some way productive of good. In both utilitarian and Kantian theory, there is some single course of action which is rationally required. As a result there is nothing in the idea that there are other good actions outside the class of those which are obligatory. Kindness in action is hard to understand within this framework, still less as a disposition of character. Indeed, the very notion of a disposition seems unaccountable in the Enlightenment theories: it falls prey to Kantianism's theory of metaphysical freedom and utilitarianism's aspiration to agent-neutrality in the assessment of outcomes.

As was mentioned, there are political reasons for this neglect as well, though the tacitly utilitarian assumptions informing much modern political argument certainly constitute one of those reasons. The political language of welfarism is more apt to talk of promoting benefits or satisfying entitlements than of kindness, with its unwelcome connotations, to modern ears, of paternalism or charitable condescension. This is an area in which utilitarianism, unsurprisingly, finds the argumentative context of modern politics congenial: unsurprisingly, because the egalitarian animus towards charity has its counterpart in the baseline equality of 'each to count for one and none for more than one' (it need not, of course, lead to equality in distributions). It should be stressed that this is not to claim that egalitarianism is in any sense regrettable or has undesirable consequences - on the contrary, this bias in political argument often conceals rather than exposes such inequalities in power and wealth as remain. But it does

prompt the question why a virtue like kindness may be recognized in personal life and yet be distrusted, even resented, in a political context. Not far behind that lie larger questions about the nature of political virtue - whether the notion of personal virtue has any application in a political context, or (as Machiavelli has been taken to have claimed) political virtue is entirely sui generis. It also provokes reflection on the relationship of virtue to moral notions such as that of innocence, and the latter's place in political life.

These are obviously large questions, which cannot be dealt with fully here.³⁶ The following merely attempts to relate them to the thematic concerns of this chapter. It is perhaps instructive to begin from the other end, by considering the attitude of instrumentalist moralities towards the concept of innocence. I have already argued that these make little sense of the notion of virtue in either private or, especially, political life. There is, however, a sense in which both utilitarian and Kantian morality lead to a markedly innocent view of politics. In part this derives from their application of a theory of rational action to politics which is both optimific and universally applicable, with little allowance made for the idea that political life might impose peculiar demands on those participating in it: as a result, they are unlikely to have the Machiavellian thought that politicians can be selected, in the recruitment process, for traits which preclude - and even wipe out - innocence.

There is also an emphasis, certainly present in Kantianism and at least some forms of utilitarian theory, on the public accountability of politicians. This is seen in, for example, Rawls' stress on publicity as a condition of legitimacy in government. In the utilitarian case, the worst that can happen is some version of the utilitarian form of paternalism perhaps favoured by Sidgwick, in which latter-day Platonic guardians sometimes deceive the populace in the name of greater long-term utility - but in consistently utilitarian terms, there is nothing wrong with that, as long as it achieves its aim.³⁷ No doubt deceiving the public, even for these paternalistic reasons, does require a lack of innocence. But what even the Sidgwickian guardians must be innocent about is the value of the optimific enterprise itself: there is no sense in the idea that their virtues (which can only be those of efficient administration) may

run afoul of that enterprise or that there is moral knowledge which may frustrate its implementation.

But the other variants of the theory are, so to speak, too deeply embedded in innocence to have any functional conception of it. In monistic accounts of value such as these, there is no place for the idea that an increase in knowledge may be accompanied by a loss of virtue or goodness, or, more generally, that moral knowledge may be non-additive.³⁸ This means that certain forms of moral conception are placed out of the reach of these theories - notably that of innocence itself.

These remarks should help to make it clear why innocence is not an operable notion within the Enlightenment theories, though predicable of them. Perhaps it needs to be added that this does not leave, as the sole alternative, a wearily cynical view of political activity as at best jobbery and at worst legitimized murder. Nor, indeed, should it encourage an unduly rosy view of the political activity permitted by instrumentalist theories - quite the reverse, as the lack of moral knowledge implicit in these forms of political innocence may sanction atrocities in the name of the (utilitarianly) greater good. (Admittedly this is much less likely in the Kantian Kingdom of Ends, with its strong emphasis on the value of personal integrity.)

Once we move away from the instrumentalist theories, however, a more complex relationship between virtue and innocence emerges. The first step towards this more complex view comes with the recognition that to realize even the instrumental kinds of benefit may demand non-virtuous, even vicious, dispositions of character. This may arise even while remaining within instrumentalism, if the dispositions which help politicians to maximize benefits in one case prevent maximization in another. Moving further out, we encounter the possibility that certain forms of virtue are intrinsically political, and that these may psychologically preclude innocence. For example, if effectiveness is regarded as a virtue in politicians, it may easily run counter to certain forms of innocence. While it is possible for innocence to be manifested in a political context - in, for example, the attitude of many on the British left towards Stalinist Russia in the 1930s - it is doubtful whether it has any distinctive political expression.

It is hard to see how innocence can coexist with a recognition that there is a radical mismatch between one's values, on the one

hand, and the world on the other, which may offer no means of living them through. Perhaps there just is no deliberative route onward for such a person, given his substantive commitments and his circumstances. By now it should be clear why the instrumentalist theories cannot conceive that agents may reach this kind of deliberative dead end: since they abstract from differences of character, and the subjective values which characters embody, there is no reason to think that an agent can find himself in this impasse. All that is required is that he apply the rules applicable to rational action in general. These preclude conflicts of the form depicted in tragedy. As has perhaps less frequently been observed, they equally rule out suicide.

Both tragedy and suicide can be rationally explained, however, by an account of practical reasoning which emphasizes character-based notions like that of virtue. Rather than adopting a theory of deliberation tailored to fit the world, or an abstract conception of rational agency, the present account allows for a more diversified view of action according to which character influences an agent's deliberative priorities. This can be true both of what kinds of concern have weight for the person and of the ways in which, if at all, they are translated into action. If character contours the practical landscape in these ways, it is easier to see how conflicts of the 'tragic dilemma' sort may arise. These may not be satisfactorily explicable either by realist theories of ultimate metaphysical conflict, or by a subjectivized account which locates the source of tragedy wholly in the agent. While, as has been argued, notions of character are important in understanding tragedy as they are with any other form of action, it is also necessary to retain some sense of how, in these conflicts, human purposes are negated by the world - or by how it appears to be (which itself may in part be a function of character). The same sense can lead to the belief that there is no reason to go on.

It is hard to see how such a belief can rationally arise within the Enlightenment theories. These reduce the unity of the self by understating its rational discontinuities with other selves - whether as a set of causal properties coupled with affective states, or as an intrinsically unknowable locus of rational potentialities. Curiously, these theories have tended to undermine the self while having been invoked as the rational foundation for

liberal philosophies which have enshrined the human subject as the supreme source of value in the world: it is unclear how this evaluative claim can survive the collapse of the self into (in Sandel's useful terminology³⁹) either radical disembodiment or radical situatedness. As a result, the translation of these metaphysical theories of the self into a political expression is highly ambiguous, and it is highly uncertain that modern liberalism has any effective conception of political action.

In this section I have given some indications as to how such a conception might be recovered. The salvage operation begins with a more diversified account of rationality than that permitted by the instrumentalist theories. In particular, once allowance is made for dispositional and other character-based notions in explaining action, the possibility arises that there is no completely general theory of rational action which can encompass politics and all other branches of human endeavour. One explanation for this would be that political professionalism selects for certain types of character in its recruitment processes, and that the practice of politics may only be explicable using character-related notions such as those of disposition and virtue (and of course vice). In the latter part of this section I have considered some of the implications of this revised view, such as its impact on our understanding of moral goodness. We cannot assume that all forms of goodness are mutually compatible - more specifically, there are virtues which require moral knowledge incompatible with other forms of goodness such as innocence (I pass over the question whether innocence itself is best regarded as a virtue). This being so, the Rawlsian project of projecting political institutions from (supposedly) consensual ideas of goodness seems doomed to failure: some of these ideas resist translation into political forms.

To this extent Larmore is right to reject 'expressivist'⁴⁰ theories of politics, which aim to enshrine conceptions of value in political institutions. But it is naive to think that the neutrality preferred by Larmore is a coherent response to the anomalousness of value. As this chapter has argued, any set of political procedures and institutions will only partially capture current notions of value. In place of the neutralist aspiration for a value-free or entirely consensual political order, there can only be a conflict between different groups for political power. So politics can be a channel through which values are expressed.

But a recognition of pluralism demands that this expression is at best partial, that the public commitment to certain forms of value must exclude and perhaps inhibit others. More generally, the commitment to political action itself will realize certain values at the expense of others. In the next section I examine further limits on the scope of political rationalization, looking towards a more dynamic understanding of action than that offered by instrumentalism.

IV

This chapter has so far suggested certain limits on the scope of political rationalization, and on the ability of moral theories founded on the Enlightenment understanding of rationality to guide political practice. To claim that there are these limits on the power of general theories of rational action to account for politics is not, however, to propose that there must be some alternative theory which can do this job more effectively. Although the focus has been on utilitarian and Kantian moral theory, the interpretation of pluralism offered in this chapter restricts the applicability of any general account of rational action.

In principle this may be because, as in the more drastic forms of relativism, the values concerned are associated with conflicting theories of rationality, so that any conflict which arises between them is not resolvable. But it is doubtful, first, whether there really are no criteria of rationality, such as consistency, shared even between the value-systems embodied in (historically or geographically) remote cultures. And secondly, most real-world political conflict is not of this form: it is fanciful to think of socialists and conservatives in a modern western industrial nation such as Britain, for example, as subscribing to quite different criteria of rationality. A more plausible explanation for the phenomenon is that homogeneous theories of rationality have produced internal inconsistencies - or else have failed to cope with the challenge posed by contact with previously alien cultures. A further possibility, consistent with these, is that political (and perhaps ethical) structures may have developed to the point where they can no longer be brought

within any such broad theories of rational action. On a more drastic view, they never could.

These general accounts of rationality have had unfortunate consequences in other areas of theory. The pressure of deontology from Kantian and neo-Kantian theories of morality has, for example, given rise to implausibly general treatments of political obligation. Once the rationalistic structure for morality is accepted, it is natural to suppose that these can be extended to encompass politics as well: this follows from the idea that obligations are universalistic and mutually consistent, as in the Kantian model. Questions then arise about the relationship between political and (other forms of) moral obligation, whether these can conflict, and whether the former is a sub-class of the latter. As a consequence of the universalistic aspect of moral obligations, the search is launched for the grounds on which political obligations are incurred, with predictable difficulties when exceptions arise or these grounds are invalidated. Consent-based theories of political obligation are a notable example. Such theories standardly maintain⁴¹ that the grounds for incurring political obligation is the express or tacit consent of those to be governed. They have problems in sustaining universality when consent either has not been granted in the first place, or clashes with other moral obligations held to be incumbent on citizens (when for example the authorities require acquiescence or active assistance from the citizenry in carrying out racist or genocidal policies). One prominent reason for these difficulties is the attempt to assimilate a notion drawn from moral philosophy to a distinct, and in some ways quite different, category of reasons for action. One response to this recognition is just to abandon the search for any universalistic theory modelled on the rationalistic structure.

It may be more enlightening to direct attention away from this form of practical necessity, and towards a broader conception of the good of political participation. One possibility is that the latter offers opportunities for the exercise of distinctively political virtues. This does not, as might be thought, invite the same criticism as that just made of theories of political obligation, that they transfer inappropriate categories of understanding from moral to political philosophy. But on the present account the virtues lack the homogeneity and consistency

required of structures of moral obligation. There is no assumption that they must, or even can, form a self-consistent unity (the possibility of such unity importantly depends on psychological factors). As a result, the ways in which virtue can be expressed may be fragmentary and partial, and irreducible to compliance with a universally-applicable set of obligations.

In addition, the relationship between political value and the virtues may be ambiguous, even leaving aside the question of internal consistency. The value of political participation may be only partially captured by its promoting the exercise of political virtues. But even the latter may not best be represented as the trading of participatory benefits for certain forms of obligation. This trade-off model may seem appropriate to understanding, for example, the distribution of welfare benefits to citizens in return for certain fiscal exactions. But it is very unclear even in this sort of case what the ground is for incurring the specific obligation to pay certain taxes - certainly not that the person or institution obligated is the likely recipient of the public revenues so created. Nor is it very convincing to claim that eligibility for public benefits somehow grounds all political obligations: it is hard to believe, for example, that a taxpayer's eligibility for unemployment benefit in the event of redundancy creates the full panoply of obligations enforced by modern states. This is true even if the person would agree to take on these obligations when asked to do so. A person's willingness to accept an obligation in hypothetical circumstances does not in fact place the person under that obligation, the force of the latter remaining as hypothetical as the circumstances themselves.

No doubt there is good prudential reason, most of the time, to do what the law requires. But it is very doubtful whether this practical requirement is best expressed in terms of obligations. Whatever the nature of the prudential reasons for obeying the law, it is hard to see how they can be framed in the language of obligations except as (in the notoriously discredited terminology) obligations to oneself.⁴² There is good reason to doubt whether such obligations exist at all. But even if they do, they mislocate the object of citizens' political obligations as being directed to themselves rather than to the state. In this as in other areas, obligations merely fail to handle satisfactorily phenomena whose real source is to be found in subjective motivation or (as Kant

called it) inclination.

It seems more plausible instead to think of some political goods as being partly expressible in terms of certain characteristic virtues. This does not of course answer the obligation theorist's worries about how to justify state power; such a universalistic theory may simply be unavailable. An emphasis on the virtues may however help to understand the goods produced by membership of a political community - as well as indicating, in the case of certain other virtues, the goods to be had from a life outside the community. Theorists of obligation, particularly those working in the contractarian tradition, have usually seen the relationship of state and citizen as involving a trade-off between obligation and rights or claims against the state (to personal protection and welfare benefits). On the alternative view, the notion of political virtue may be better equipped to understand the values realized by political life than that of obligation, which demands, in different respects, both too little and too much of politics. Too little, because it takes a negative view of citizens' relationship to the state as exclusively concerned with imposed restraints, and confines their experience of politics to this relationship; and too much, because it draws all citizens into the net of obligation and makes little allowance for different levels of participation (or abstention). The general account of citizens' relationship with the state sought by theories of political obligation may be unsustainable in a pluralistic society.

If this is so, it appears that the principal objective of obligation theories, to refute anarchist objections to state power, is unachieved. A theory of political virtue cannot neatly fill the gap left by failed theories of political obligation. But it may be able to provide an understanding of the values promoted by a life of political action, and the value of political participation itself. Familiarly, this may mean becoming a professional politician in order to secure certain public benefits, such as improving housing or helping to eliminate poverty. But it is important that politics is not seen merely as the instrument to achieving such goods, however intrinsically worthwhile. Rather the good state of politics is one in which civic virtues are intrinsically valued, and not solely as instrumentalities. In place of the neutralist picture of a

citizenry negatively bound to certain duties of forbearance while pursuing their private interests in civil society, an emphasis on the political virtues offers an alternative view of the relationship between value and political action. According to this view, citizens' relationship to their political community is not best expressed as a network of obligations, but through opportunities for active participation in the community.

There is no maximal life of virtue. But there are perhaps certain virtues best realized through exercising these opportunities. One example, already mentioned, is the virtue of kindness. This is not to say, of course, that this virtue cannot be expressed in private life: it is a familiar sentiment, indeed a platitude, that 'charity begins at home'. But if it is not to end there as well, there must be civic contexts in which kindness can be exercised. Instrumentalism, as the dominant ethos of political life, may preclude the growth and flourishing of public virtue; in this way, some private virtues may simply lack any public manifestation. Thus a repudiation of instrumentalism rests less on the claim that there are certain exclusively political virtues, than that the pursuit of instrumental benefits excludes certain private virtues, such as kindness, from public life.

The absence of public kindness may be felt as a loss in several, and diverse, ways. To those in receipt of state welfare, for example, the impersonal and bureaucratic administration of personal benefits adversely affects both the recipients' perception of these benefits and of themselves. Relatedly, a public culture of effectiveness may lead to over-emphasis on statistically measurable forms of welfare and the neglect of the human costs incurred by its administration - both to public employees and to welfare recipients. Kindness (perhaps in contrast with related virtues such as generosity) cannot be assessed in purely volumetric terms: it depends on how a benefit is given, not just on how much. At another level, the conduct of politics itself, when devoted to the pursuit of instrumental benefits, may lose sight of these public virtues. When political action is directed exclusively towards procuring these benefits, it becomes difficult to avoid regarding politicians as omniscient and as being under a professional obligation to secure them. Appreciating kindness implicitly accepts both value lying beyond obligation and the constraints on agents' powers. Thus an ethos of kindness is

more likely to prevail where agents' powers are acknowledged as being limited and where the value of action is not judged solely in terms of its tangible outcome, as a criterion of professionalism. The atrophy of certain public virtues may betoken a depreciation in the quality of political life. This is not to say that the virtue of kindness can simply be restored to politics without otherwise disrupting the status quo. As was mentioned before, it may be unattainable - without lapsing into paternalism or charitable condescension - given existing inequalities in wealth and power.

This is not to say that political life is that even of maximal achievable virtue; again, it may simply be that there is no content in this notion, given the plurality of value. But it is notable that the conduct of politics so often proceeds in this vacuum, where virtues do not merely happen to be absent but seem to be actively excluded.

This line of thought can, of course, be traced back to Machiavelli and, through him, to Aristotle. The idea that there are goods obtainable only from political life is prominent both in the Prince and in the Discourses, in their reaffirmation of classical conceptions of virtue as (something akin to) institutionalized virility, realized through the standing engagement of the populace in public affairs. For Machiavelli, there is indeed a supreme life of virtue, which is that of the man of virtu - one who imposes his will on fortune, achieves spectacular feats of martial valour, and so forth. For Aristotle, by contrast (whose ideas are notable influences on Machiavelli's conception of civic virtue), there is no perhaps no presupposition that political life is supremely valuable. In Book Ten of the Nichomachean Ethics, for example, Aristotle seems to uphold the value of theoria or solitary contemplation as the human ideal, while his emphasis on autarkeia or self-sufficiency seems clearly inimical to a life of political engagement.⁴³ At the same time, however, Aristotle also regards polis life, in terms of his own teleological biology, as the supreme form of human development, and as offering goods not otherwise available.⁴⁴ It is tempting to conclude that this reflects an unresolved tension in Aristotle's own mind between the life of philosophical reflection and that of politics; but whether or not this is so, it remains true that some virtues, and thus some human values, can only be realized

politically.

It may be thought that this stress on the political aspect of the virtues is merely a means of rehabilitating liberal individualism, shifting the focus away from civil society to public action while retaining the liberal emphasis on personal development. The character-based notion of virtue might seem to reinforce this. But it is doubtful whether the refocussing can be achieved without disturbing the liberal conception of civil society itself - particularly of the relationship between public and private spheres of action. The liberal neutralist is at least right to this extent, that the distinction between the spheres can only be maintained if public institutions remain studiously impartial, at some level, between competing ideals of value: otherwise there is no guarantee that the integrity of civil society, as the sphere in which private concerns are transacted, will be preserved. The danger facing the Millian (non-neutralist) liberal is that this integrity will be undermined by democratic decision-making mechanisms - a danger of which Mill himself was well aware.⁴⁵

Once neutrality is abandoned, with its implicit distinction of public and private, it is much harder for liberalism to prevent the encroachment of politics on civil society. This is not, it should be said, to go so far as to claim that all concerns, whether public or ostensibly private, are political. There are reasons for doubting this beyond the purely phenomenological - that there is a fairly well defined understanding, in a society and at any given time, of which concerns count as political (which is not to deny that there is no obvious criterion which prevents any given concern from becoming the subject of political dispute). Apart from the phenomenological type of objection, the attempted conflation also fails to take account of the ethical peculiarity of politics as a field of action, which has already been mentioned in this chapter. This need not mean the invasion of civil society by politics, but it is naive to think that private concerns will remain intact and unchanged through the process of politicization - unsurprisingly, since the pressure for political action in civil society is likely to arise from dissatisfaction with the status quo. One important recent example in developed nations has been the (still continuing) revolution in attitudes towards the role of women in society.⁴⁶

There is then little reason to think that the Aristotelian-Machiavellian view of politics as a field of autonomous concerns will preserve the integrity of civil society against (what are claimed to be) abuses of power, as liberals have tried to do. As Machiavelli himself makes clear,⁴⁷ the nature of political power is such that its use cannot be circumscribed in advance - indeed he regards the versatile use of power as one of the characteristic qualities marking the man of virtu. It is not necessary to endorse Machiavelli's own preferred list of political virtues to find this a persuasive account of the use of power; in place of the martial prowess extolled by Machiavelli, we may find political virtue to consist in a readiness to engage in public debate, in a state of society which does not merely accept passively political decisions handed down from above but rather participates actively in questioning them and in helping to determine which issues are of public concern.

No doubt the relationship between citizens and political activity in (for example) western democracies is, for much of the time, a sad caricature of this picture. But this may mean only that practice usually falls short of the mark, not that the mark itself does not exist. In so far as there is any effective notion of political virtue in societies such as Britain today it presumably encompasses some form of judgmental capacity, as Beiner has argued,⁴⁸ together with (or including) some degree of professional integrity in the day-to-day conduct of public business. It has also to take account of the more stylistic sorts of consideration mentioned earlier. This need not be an entirely frivolous or aesthetic matter. To understand political behaviour through character-based notions such as that of style can be both to set it in a more comprehensible context and to see it as manifesting certain forms of virtue (or of characteristic vices).

These are forms of rational explanation which we already use in understanding political action. It is, however, undeniable that a form of instrumentalism coexists alongside this in much commentary on and in public expectation about politics. This is true of much electioneering rhetoric, for example, and in politicians' habits of justifying their past actions: in these situations politicians typically represent themselves as having acted, or proposing to act, so as to realize certain public benefits. As has already been indicated, what is important here is not whether these claims are

true (rather than, say, that the persons concerned were advancing their careers or making other personal gains). The important point is that politicians feel obliged, even insincerely, to frame their self-justifications in this way. To the extent that this rhetoric assumes that securing, or at least attempting to secure, the instrumental benefits is a condition of political justification, instrumentalism reigns supreme in the everyday discourse of politics. But while this is true, the public also expects more of politicians than mere effectiveness in promoting instrumental benefits.

It is doubtful whether this is adequately represented as the imposition of a code of conduct on professional politicians whose content derives solely from moral precepts applicable to all action. It is true that there is no very well-defined code of conduct to which politicians are answerable in their professional dealings. This however only gives further reason for doubting that the orthodox view of morality as a structure of obligations is appropriate to politics. The judgmental capacities discussed by Beiner,⁴⁹ for example, seem to defy formulation in these terms, and indeed in terms of any explicit rules (as Beiner himself persuasively argues); there is also little reason to think that such capacities can be defined as proficiency in the patterns of strategic thinking required to promote the public goods. Political judgment goes beyond the facility for implementing rules. It is a dispositional quality, and as such an attribute of character which is manifested in politicians' actions. One consequence of this is that whether a politician displays sound judgment or not in a given case is partly a contingent matter - dependent, that is, on factors some of which are outside his or her control. This is again inimical to the rule-governed instrumentalist approach, which tends to view success or failure in politics, as in other areas of life, as a matter of knowing which causal levers to operate.

This is an important aspect of the virtues in general, one which sharply distinguishes them from the Kantian 'holy will'. While the latter, in line with Kant's insistence on the totally unconditioned nature of moral agency, is held to be beyond the reach of contingency, it is characteristic of the virtues that their exercise or manifestation is not unconditioned in this way. To the extent that the virtues are embodied in agents' intentions,

they are proof against this contingency, since it is possible to identify such an intention regardless of whether the action succeeds or not (very often, of course, the means by which the intention is identified is just that the action was successfully performed). These cases look as if they can be handled easily enough by Kant's 'holy will'. But there are cases where the translation of virtue into action is not as straightforward as this. One is where the virtue cannot be directly reflected in deliberation, and so is not represented directly in the agent's conscious intentions, or more commonly, where an observer's assessment of the action in question would be different if it transpired that a reflexive conception of virtue played an essential part in the deliberation. For example, an agent who acts out of charitable concern for others is relevantly different from one who acts from a conception of himself as somebody who so acts. One example is the character of Mrs Solness in Ibsen's The Master Builder, discussed by Winch:⁵⁰ her life is so dominated by a desire to fulfil her perceived duties, with the result that she is morally dead. The inference we are invited to draw, presumably, is that virtue need not - and cannot always - be embodied in this intentional form, of conscious adherence to duty. Another obstacle takes the form of psychological conflict. This is a possibility notoriously discountenanced by Kantian morality and by Aristotle's own account of the virtues. It seems clear, however, that the latter is one form taken by the wider phenomenon of dispositional conflict. Besides the familiar possibility of conflict between virtuous and other (notably egoistic) dispositions, there can also be conflict between virtuous dispositions.

More generally, the virtues and their exercise cannot be wholly immunized against contingency, and politics provides a prime example of this. One aspect of this contingency is precisely the dispositional conflicts just mentioned, whether between virtues or between virtuous and other dispositions. Contingency lies not merely in the fact (with which even the Kantian has to contend) that being an agent of a certain kind - having certain dispositions, and at the limit even being an agent at all - is itself a contingent matter. It is also witnessed in the psychological conflicts to which virtuous dispositions can give rise. In contrast with the Kantian and, indeed, the Platonic accounts of the virtues, the Aristotelian approach allows for this

dimension of contingency. This is one reason why it is able to make more sense of the notion of political virtue.

It is not surprising that ethical theories like Kant's which have sought to eliminate contingency from the moral life should lack any distinctive conception of political virtue (the latter merely consists, for Kant, in following universal rules applicable to all rational agents as such). For it seems to be a contingent matter whether or not political circumstances, at a given time and place, will prove conducive to the exercise of virtue. Theories which regard the virtuous agent as (definitionally) immune from contingency are inclined to see the practical manifestation of virtue as the life of contemplation, an activity held to be as invulnerable as is humanly possible to the depredations of fortune; and, in the Platonic version of the theory, in so far as politics enters into such a life at all, the political task is held to be the elimination of contingency from public life. The eugenic breeding and censorship of non-celebratory art favoured by Plato is meant to secure the political stability held necessary for the guardian class to live the life of philosophical contemplation, which alone is virtuous in Plato's thought. Thus virtue is only attainable once politics is out of the way - one might say that for Plato the only purpose of politics is to secure the conditions of its own transcendence. Similar remarks, mutatis mutandis, apply to Rawls' theory: the well-ordered society is one in which value-based political dispute has in effect ceased.

This example also suggests that the notion of political virtue can survive only where politics itself is not regarded in purely instrumental terms. If politics is solely concerned with promoting external benefits, it is hard to see what specifically political virtues there could be, beyond that of effectiveness. This may or may not be a disposition of character; but, if the instrumentalist thesis is right, it cannot be part of what is ultimately valued, since the value of effectiveness consists solely in its capacity to secure other goods valued for their own sake. It is also clear from the nature of virtue itself that it cannot be understood in purely instrumental terms, whatever the context. One reason for this is that it makes little sense to ask what virtues are for, what end they serve: they exist as natural features of the moral world. This is not to relapse into a form of ethical naturalism or realism. Rather it is to acknowledge that there is an

anthropomorphic basis for value not only in projection itself, but also in the object of projection - that is, we can take the dispositions of character constitutive of virtue as natural objects. Virtues of character can be taken as natural facts in this sense. But these facts appear as instrumentalities only given a discredited functional view of nature. The account of the virtues given above is quite compatible with a wholly non-teleological view of nature.

Thus we are not required to accept Aristotle's teleological biology as a necessary feature of any neo-Aristotelian theory of the virtues. Aristotle's is a general theory of human flourishing in which the attainment of virtue is seen as integral to realizing one's species-essence. The present theory retains the Aristotelian conception while rejecting its associated biological claims. One of the advantages of these assumptions, of course, is that the theory of virtue is thereby made compatible with a naturalistic metaethics, avoiding the dangers of indeterminacy in moral judgments and, at the limit, of arbitrariness. In rejecting Aristotelian biology, however, we need not fall into this Rortian trap. Even in the absence of any naturalistically-interpreted list of the human virtues, we can retain the idea of standards of appropriateness in ascribing virtues to persons (as we can indeed with other, non-ethical dispositions).

These dispositions are so-called 'thick'⁵¹ ethical notions, which as such may embody quasi-objective criteria of application while remaining peculiar to a certain culture. In this respect they resemble the emotional judgments discussed in the last chapter: this is not very surprising, since possessing a given virtue often depends on having - or lacking - a relevant emotional disposition. The virtue of courage, for example, seems unintelligible without an understanding of the emotion of fear, and that some persons are less disposed to feel this emotion or at least to act on it while under its influence. In such cases, as has already been argued, identifying the disposition implicitly relies on shared evaluations: we can only identify courage, for instance, given a prior understanding of what counts as reasonably provoking fear.

This is not to claim, of course, that virtues merely designate emotional dispositions. Many virtues, like that of kindness, seem to connote no particular pattern of emotional behaviour.

Nonetheless even in these latter examples the point made before in analysing the emotions holds, that shared patterns of evaluation underpin our ascriptions of virtues to agents and render them intelligible. In this sense our understanding of the virtues, in politics as in other contexts, rests on 'agreement in judgments' - a truth which naturally leaves plenty of room for disagreement about which persons manifest a given virtue or which occasions are appropriate for its exercise.

It is perhaps doubtful whether we can simply move back to a form of thinking about ethics which is based on the notion of virtue. In so far as it survives in everyday discourse about politics, it coexists uneasily with alien, instrumentalist concerns - relating, for example, to managerial efficiency or to effectiveness. But it does not require wholesale conceptual reorientation. As I have indicated, the notion remains current in some important aspects of our ethical thinking, including politics. Rejecting instrumentalism does not mean hankering for a full-scale return to the heroic values of a pre-modern age. It does, however, demand some hard thinking about what we have good reason to value. Its outcome should acknowledge that there are goods in politics beyond those computable in means-end terms - goods whose value lies in the cultivation of civic virtue.

V

It will be clear from the preceding section that I have no systematic theory to offer as a replacement for instrumentalism, still less any concrete policy proposals. In some respects I feel duly apologetic about this, in others less so. It seems fair enough, for example, to complain that this chapter has failed to give much substance to the counter-theory adumbrated in the earlier critical chapters on instrumentalism; it is not unreasonable, furthermore, to wonder whether the absence of any fully-articulated alternative leaves instrumentalism, albeit faute de mieux, in possession of the field. I have given some historical indication of how instrumentalism arose from the rationalizing philosophies of the Enlightenment: in so far as modernity remains in the thrall of these philosophies, we are bound to find some of the demands which first elicited instrumentalism compelling - notably the insistence on transparency, on the susceptibility of

public decision-making to rational justification, and the linked (though not identical) notion of public accountability. As a result, the pressure for explication cannot merely be ignored by any proposed alternative theory. What I have tried to do, in place of this, is to suggest some more drastic limits on rationalization than those entertained by instrumentalism (the sense in which I feel less apologetic about not devising a replacement for it).

It should not be thought that acknowledging these limits commits us to some quasi-mystical view of political processes as forever beyond the pale of reason, a celebration of the irrational (itself one of the banes of 'post-modern' culture). Rather the analysis offered earlier should be read as a rationalistically-minded critique of reason itself, as applied to ethical and political systems - a familiar enough Kantian form of enterprise. This work has concluded that politics is largely refractory to the totalizing and system-building bent of the Enlightenment philosophies: its relationship with other ethical concerns is partly, and irreducibly, a political question. As far as this goes, the role of theory may at best be a negative one, suggesting what courses of action are unjustified by particular ethical commitments, rather than (as the system-building tendency of recent decades has aimed to do) furnishing universal conditions of justification. It may be able to show, for instance, that a particular, economic, interpretation of liberty is too restrictive, or that it leaves other important political freedoms out of account.

In a century of appalling political crimes, this is not demanding as little as it may seem. Of course making the demand will not of itself prevent their repetition in the future. It may however serve as a reminder that a sceptical and non-instrumental view of politics is a surer theoretical basis for repudiating such actions than more ambitious forms of recent theory; and that a rationalistic theory of rationality is no more appropriate to politics than to other ethical concerns. Future theory may be better served by focussing less on the use of power to effect a well-ordered society, than on the abuses of power - and, as a concomitant of that, on empowerment, as a means to the self-betterment of the powerless.

Notes

1. John Rawls, A Theory of Justice (Oxford 1971) and 'Kantian Constructivism and Moral Theory', Journal of Philosophy 1980; Ronald Dworkin, Taking Rights Seriously (London 1977).
2. Rawls, 'Kantian Constructivism and Moral Theory', p515ff.
3. David Hume, A Treatise of Human Nature ed. L.A.Selby-Bigge (Oxford 1888) Book II; Immanuel Kant, The Groundwork of the Metaphysics of Morals tr. H.J.Paton (New York 1956).
4. Immanuel Kant, 'Perpetual Peace' in Kant's Political Writings ed. H.Reiss (Cambridge 1970), p118.
5. For example, the apparatus of artistic censorship and eugenic reproduction advocated in the Republic and the Laws.
6. Donald Davidson, Essays on Action and Events (Oxford 1980), pp156-7.
7. A distinction due to Hume: see the Treatise of Human Nature p458ff.
8. Charles Stevenson, Ethics and Language (Yale 1944), Chapter 4.
9. Hume, op. cit. p415.
10. See e.g. Alasdair MacIntyre, After Virtue (London 1981) p12ff; Bernard Williams, 'Morality and the Emotions' repr. in Problems of the Self (Cambridge 1973).
11. Bernard Williams, Ethics and the Limits of Philosophy (London 1985), p140.
12. George Pitcher, 'Emotion' in Mind 1965 p327ff.
13. For the occurrent/dispositional distinction between forms of emotion see Pitcher in ibid., p329ff.

14. David Wiggins, 'Truth, Invention, and the Meaning of Life' repr. in his Needs, Values, Truth (Oxford 1987).
15. Ludwig Wittgenstein, Philosophical Investigations (Oxford 1953), paragraph 217.
16. Williams, Ethics and the Limits of Philosophy p129.
17. Richard Rorty, Contingency, Irony, Solidarity (Cambridge 1989), e.g. p73ff.
18. Richard Rorty's 'great conversation of mankind' seems to conform to this picture. See his Philosophy and the Mirror of Nature (Princeton 1979) p392ff.
19. Bruce Ackerman, Social Justice and the Liberal State (Yale 1980).
20. Ludwig Wittgenstein, On Certainty (Oxford 1969), paragraph 204.
21. Jurgen Habermas, The Theory of Communicative Action tr. T.McCarthy (Boston 1984).
22. John Stuart Mill, On Liberty (London 1972), Chapter 2.
23. For Hegel's historicist teleology see the Preface to the Philosophy of Right and the Phenomenology; also e.g. Charles Taylor, Hegel (Cambridge 1975), pp389-427.
24. Donald Davidson, 'How is Weakness of the Will Possible?' in Essays on Action and Events.
25. See e.g. Leonard Schapiro, The Communist Party of Great Britain (London 1960), Chapter 16.
26. Ibid., p311.
27. There is a growing literature on the theory and practice of Thatcherism. See, e.g. Peter Jenkins, Mrs Thatcher's Revolution

14. David Wiggins, 'Truth, Invention, and the Meaning of Life' repr. in his Needs, Values, Truth (Oxford 1987).
15. Ludwig Wittgenstein, Philosophical Investigations (Oxford 1953), paragraph 217.
16. Williams, Ethics and the Limits of Philosophy p129.
17. Richard Rorty, Contingency, Irony, Solidarity (Cambridge 1989), e.g. p73ff.
18. Richard Rorty's 'great conversation of mankind' seems to conform to this picture. See his Philosophy and the Mirror of Nature (Princeton 1979) p392ff.
19. Bruce Ackerman, Social Justice and the Liberal State (Yale 1980).
20. Ludwig Wittgenstein, On Certainty (Oxford 1969), paragraph 204.
21. Jurgen Habermas, The Theory of Communicative Action tr. T.McCarthy (Boston 1984).
22. John Stuart Mill, On Liberty (London 1972), Chapter 2.
23. For Hegel's historicist teleology see the Preface to the Philosophy of Right and the Phenomenology; also e.g. Charles Taylor, Hegel (Cambridge 1975), pp389-427.
24. Donald Davidson, 'How is Weakness of the Will Possible?' in Essays on Action and Events.
25. See e.g. Leonard Schapiro, The Communist Party of Great Britain (London 1960), Chapter 16.
26. Ibid., p311.
27. There is a growing literature on the theory and practice of Thatcherism. See, e.g. Peter Jenkins, Mrs Thatcher's Revolution

(London 1987); Stuart Hall & Martin Jaques, The Politics of Thatcherism (London 1983); Kenneth Minogue, Thatcherism, Personality and Politics (London 1987); Peter Riddell, The Thatcher Decade (Oxford 1989); Bruce Arnold, Margaret Thatcher: A Study in Power (London 1984); and, in her own words, Margaret Thatcher's In Defence of Freedom (London 1986).

28. Bernard Williams, Moral Luck (Cambridge 1981)

29. See e.g. Alasdair MacIntyre's Whose Justice? Which Rationality? (London 1988).

30. Williams, Ethics and the Limits of Philosophy, p6.

31. See Peter Winch, 'Moral Integrity' in his

32. E.g. David D. Raphael's 'The Intolerable', in S.Mendus (ed.), Justifying Toleration (Cambridge 1988), p143ff.

33. See my 'Fatwa and Fiction: Censorship and Toleration' (Morrell Discussion Paper).

34. Raphael, op. cit.

35. David Ross, The Right and the Good (Oxford 1930), p21ff.

36. See Peter Johnson, Politics and the Limits of Innocence (

37. See Bernard Williams in his and J.J.C.Smart's Utilitarianism; For and Against (Cambridge 1973), p130.

38. Williams, Ethics and the Limits of Philosophy p148-9.

39. Michael Sandel, Liberalism and the Limits of Justice (Cambridge 1982), p21.

40. Charles Larmore, Patterns of Moral Complexity (Cambridge 1987), p73.

41. See e.g. A. John Simmons, Moral Principles and Political Obligations (Princeton 1979); Carole Pateman, The Problem of Political Obligation (Berkeley 1985); Leslie Green, 'Consent and Community' in P. Harris (ed.), Political Obligation (London 1990); Richard Flathman, Political Obligation (London 1973).
42. See Williams, Ethics and the Limits of Philosophy, p50, p182.
43. See Aristotle, Nicomachean Ethics 1177a27ff.
44. Aristotle, Politics 1252b28ff.
45. Cf. J.S. Mill, Considerations on Representative Government, e.g. Ch. 7.
46. See, e.g., V. Beechey & E. Whitelegg (eds.), Women in Britain Today (Milton Keynes 1986); D. Taylor (ed.), Women: A World Report (London 1985); L. Leghorn & K. Parker (eds.), Women's Work (Boston 1981).
46. Cf. Machiavelli, The Prince, e.g. Ch. 8.
47. Ronald Beiner, Political Judgment (London 1983), Chapter 8. Beiner's identification of political judgment as a virtue is intended to be normative as well as descriptive.
48. Beiner ibid., passim.
49. Winch, op. cit.
50. Williams, Ethics and the Limits of Philosophy p140.

Select Bibliography

1. Bruce Ackerman, Social Justice and the Liberal State (Yale 1980).
2. Hannah Arendt, (a) The Human Condition (Chicago 1958).
(b) Eichmann in Jerusalem (New York 1965).
3. Aristotle, (a) Nicomachean Ethics, ed. D.Ross (London 1925).
(b) Politics, ed. E.Barker (Oxford 1958).
(c) Rhetoric, ed. W.R.Roberts (Oxford 1924).
4. Ronald Beiner, Political Judgment (London 1983).
5. Alex Callinicos, Against Postmodernism (London 1989).
6. Stanley Cavell, Must We Mean What We Say? (Cambridge 1976).
7. Norman Daniels (ed.), Reading Rawls (Oxford 1975).
8. Donald Davidson, Essays on Action and Events (Oxford 1980).
9. John Dunn, (a) Rethinking Modern Political Theory (Cambridge 1985).
(b) Western Political Theory in the Face of the Future (Cambridge 1979).
(c) Political Obligation in its Historical Context (Cambridge 1980).
10. Ronald Dworkin, (a) Taking Rights Seriously (London 1977).
(b) Law's Empire (London 1986).
11. Paul Feyerabend, Against Method (London 1978).
12. Richard Flathman, Political Obligation (London 1973).
13. Alan Gewirth, (a) Reason and Morality (Chicago 1978).
(b) Human Rights (Chicago 1982).

14. Jonathan Glover, Causing Death and Saving Lives (London 1977).
15. Leslie Green, 'Consent and Community' in P.Harris (ed.), Political Obligation (London 1990).
16. Jurgen Habermas, (a) Legitimation Crisis (London 1976)
(b) The Theory of Communicative Action (Boston 1987).
17. John Harris, (a) 'The Survival Lottery', in P.Singer (ed.), Applied Ethics (Oxford 1986).
(b) The Value of Life (Oxford 1986).
18. Stuart Hampshire (ed.), Public and Private Morality (Cambridge 1978)
19. Georg-Wilhelm Friedrich Hegel, (a) The Philosophy of Right, ed. T.M.Knox (Oxford 1967).
(b) Phenomenology of Spirit, ed. A.V.Miller (Oxford 1977).
20. Martin Hollis, The Cunning of Reason (Cambridge 1987).
21. C.Hookway & P.Pettit (eds.), Action and Interpretation (Cambridge 1978).
22. David Hume, (a) A Treatise of Human Nature ed. L.A.Selby-Bigge (Oxford 1888).
(b) Enquiries Concerning the Human Understanding and Concerning the Principles of Morals ed. L.A.Selby-Bigge (Oxford 1962).
23. Peter Johnson, Politics, Innocence, and the Limits of Goodness (London 1988).
24. Immanuel Kant, (a) The Groundwork of the Metaphysics of Morals ed. H.J.Paton (New Jersey 1978).
(b) Political Writings ed. H.Reiss (Cambridge 1970).
(c) Critique of Pure Reason ed. N.K.Smith (New York 1963).
(d) Critique Of Practical Reason ed. T.K.Abbott (London 1909).
(e) Critique of Judgment ed. J.C.Meredith (Oxford 1952).

25. Thomas Kuhn, (a) The Structure of Scientific Revolutions (Chicago 1972).
(b) The Essential Tension (Chicago 1977).
26. Charles Larmore, Patterns of Moral Complexity (Cambridge 1987).
27. Sabina Lovibond, Realism and Imagination in Ethics (Oxford 1983).
28. Alasdair MacIntyre, (a) After Virtue (London 1981).
(b) Whose Justice? Which Rationality? (London 1988).
29. John McDowell, 'The Role of Eudaimonia in Aristotle's Ethics' in A.Rorty (ed.), Essays on Aristotle's Ethics (Berkeley 1980).
30. Susan Mendus (ed.), Justifying Toleration (Cambridge 1988)..
31. John Stuart Mill, On Liberty & Utilitarianism, ed. M.Warnock (London 1962).
32. Thomas Nagel, The View from Nowhere (Cambridge 1986).
33. Robert Nozick, Anarchy, State, and Utopia (Oxford 1974).
34. Martha Nussbaum, The Fragility of Goodness (Cambridge 1986).
35. Derek Parfit, Reasons and Persons (Oxford 1984).
36. Carole Pateman, The Problem of Political Obligation (Oxford 1979).
37. George Pitcher, 'Emotion', Mind 1965.
38. Hannah Pitkin, 'Obligation and Consent' in W.Runciman, P.Laslett & Q.Skinner (eds.), Philosophy, Politics and Society Fourth Series.
39. Plato, (a) The Republic ed. F.Cornford (Oxford 1941).
(b) Gorgias ed. T.Irwin (Oxford 1979).

40. Karl Popper, The Logic of Scientific Discovery (London 1977).
42. John Rawls, (a) A Theory of Justice (Oxford 1971).
(b) 'Kantian Constructivism and Moral Theory', Journal of Philosophy 1980.
(c) 'Justice as Fairness: Political not Metaphysical', Philosophy and Public Affairs 1985.
(d) 'Fairness to Goodness', Philosophical Review 1975.
(e) 'The Independence of Moral Theory', Proceedings and Addresses of the American Philosophical Association 1975.
43. Joseph Raz, The Morality of Freedom (Oxford 1986).
44. Donald Regan, Utilitarianism and Cooperation (Oxford 1980).
45. Richard Rorty, (a) Contingency, Irony, Solidarity (Cambridge 1989).
(b) Philosophy and the Mirror of Nature (Princeton 1980).
(c) The Consequences of Pragmatism (Minneapolis 1982).
46. David Ross, The Right and the Good (Oxford 1930).
47. Jean-Jacques Rousseau, Discourse on the Origins of Inequality ed. M. Cranston (London 1984).
48. Michael Sandel, Liberalism and the Limits of Justice (Cambridge 1982).
49. Leonard Schapiro, The Communist Party of the Soviet Union (London 1960).
50. A. John Simmons, Moral Principles and Political Obligations (Princeton 1979).
51. Peter Singer, Animal Liberation (New York 1975).
52. Quentin Skinner, (a) The Foundations of Modern Political Thought 2 Vols (Cambridge 1978).
(b) 'Conventions and the Understanding of Speech Acts', Philosophical Quarterly 1970.

(c) 'Some Problems in the Analysis of Political Thought and Action', Political Theory 1974.

53. Charles Taylor, Hegel (Cambridge 1975).

54. Michael Walzer, Spheres of Justice (New York 1983).

55. David Wiggins, Needs, Values, Truth (Oxford 1987).

56. Peter Winch, The Idea of a Social Science and Its Relation to Philosophy (London 1958).

57. Bernard Williams, (a) Problems of the Self (Cambridge 1973).

(b) (With J.J.C. Smart) Utilitarianism: For and Against (Cambridge 1973).

(c) Moral Luck (Cambridge 1981).

(d) Ethics and the Limits of Philosophy (London 1985).

58. Ludwig Wittgenstein, (a) Philosophical Investigations (Oxford 1953).

(b) On Certainty (Oxford 1969).