A: ROUSSEAU AND THE PROBLEM OF REASON

I. INTRODUCTION

In order to provide a guide to political action we need to do the following:

a) Give some account of the nature of the "Good Society", of a society which satisfies human needs and embodies only legitimate power relationships rather than relationships based on force;

b) Give an account of the nature of the obstacles in the way of the realisation of such a society; and

c) Suggest political strategies based on the above.

These three steps involve a number of interrelated philosophical problems. Here I wish to attempt to formulate these problems. I shall adopt a quasi-historical method in doing so. I shall take two thinkers who have made what seem to me to be significant contributions to clearly articulating the problems involved, and I shall try to illustrate the nature of the problems by showing the way in which each has attempted in a different way to specify these basic problems. I wish to show how each took one or more major themes and synthesised such themes into a new way of formulating the problematic.

I have stressed here the question of formulating a problem, rather than the question of solving the problem because it seems to me that by and large the critics of this mode of thought have based their criticism on a misunderstanding of the basic problem. It is not simply a question of disagreement over the most desirable set of political institutions. It is a disagreement over the nature of human reality, the nature of society, and the nature of philosophical and scientific method. I hope that approaching the question in this manner will help to show the way in which this disagreement is embedded within the historically produced philosophical assumptions of different schools.

It seems to me that the philosophy of Rousseau constitutes a significant break with previous ways of thinking about human reality but I do not intend to try to demonstrate this. In any historical approach to philosophical problems one risks being dragged further and further back in time, and finally losing oneself in pre-socratic

fog. I shall therefore not attempt to show in what way Rousseau's ideas differ from earlier philosophers, and in any event even if there were shown to be no difference at all this would not be particularly important, since what I am interested in is not the extent to which he differed from his predecessors, but the extent to which his formulation of problems can throw light on the problems themselves.

The problem of interpreting Rousseau's work immediately introduces us to two very different ways of interpreting the nature of human reality. For there are two very different interpretations of what Rousseau said. One school interprets him as a "totalitarian democrat" who wishes to subordinate the individual utterly to the state. Now it seems to me that there is scarcely an iota of evidence in Rousseau's writing for such an interpretation, yet it has been widespread in both the nineteenth and the twentieth centuries. It seems to me that those who read Rousseau in this way have, in an important sense, not *read* Rousseau at all. That is, they have not followed carefully this articulation of the way in which he sees the problem of human reality. Rather they have assumed that there is only one way of seeing the problem, and that, therefore, Rousseau must be seeing it that way too. Instead of grasping the fact that Rousseau is trying to formulate *the problem* in a radically new way, they have assumed that he is only presenting *a new solution* to an agreed problem. It is only by forcing what he says into the strait-jacket of this perspective that they have managed to misinterpret him so thoroughly. I shall not here attempt to refute point by point the "totalitarian democrat" thesis, by reference to these critics. What is necessary, rather, is simply to read Rousseau, to go carefully through the two main texts, the *Discourse on Inequality*, and the *Social Contract*, and to watch the argument grow in those texts.

II. REASON AND DIVISION: "A DISCOURSE ON THE ORIGIN OF INEQUALITY"

To understand social inequality, and also to understand the nature of Law, Rousseau argues, we need to understand man "as nature made him" (1 154). But this raises serious methodological problems.

For we are presented with men not as they are 'naturally', but as they are in society, and as they are in society at the end of a long process of historical change: "For it is by no means a light undertaking to distinguish properly between what is original and what is artificial in the actual nature of man, or to form a true idea of a state which no longer exists, perhaps never did exist, and probably never will exist" (1 155). The question is whether the inequality observed in any particular society is inevitable or not, and to answer this question we have to discover whether it is rooted either in some necessary (biologically based) human drives, or whether it is some kind of necessary prerequisite for ordered society. This is a question about which elements in any given society are alterable and which are not. Although Rousseau formulates the question in apparently historical terms, as though it were about an original state and the alteration of that original state, it is clear from the above quotation that he understands it to be a hypothetical question which cannot be answered simply by historical description. The historical formula is an expository device, not part of the structure of his argument. It is not simply a question of the unavailability of historical data about the original state of nature. If this were the case the alternative might be to attempt a speculative reconstruction of what it might have been like. Such speculation would of course be useless, but Rousseau's method is not speculative either.

What then is Rousseau's method? He strongly criticises previous writers on the topic for having attempted to answer the question empirically: "Every one of them, in short, constantly dwelling on wants, avidity, oppression, desires and price, has transferred to the state of nature ideas which were acquired in society; so that, in speaking of the savage they described the social man" (1 161). That is, these writers have described people as they are at present, and have constructed a model of the state of nature by first removing from these people the existing state and social restraints, and then speculating how they would behave in such a situation. Rousseau's alternative is to "let us begin by laying the facts aside, as they do not affect the question. The investigations we may enter into, in treating this subject, must not be considered as historical truths, but only as mere conditional

and historical reasonings, rather calculated to explain the nature of things than to ascertain their actual origins" (1 161). He does not make explicit how this is to be done, but his actual analyses do make clear what method he is using. It involves an analysis of the ways in which various concepts are related to one another, in an attempt to establish whether some concepts are more fundamental than others. What counts as a more fundamental concept is established by showing the ways in which the definition of a particular concept is only comprehensible in terms of some other concept, which is, therefore, logically prior. Once this logical series has been made explicit, it then becomes possible not only to show which are the most basic elements, but also to give an account of the actual stages in historical development which is based on something other than pure speculation. Thus the *Discourse of Inequality* turns into an historical account, but it does not begin as such: the historical account is neither purely empirical nor purely speculative. I shall discuss the methodological legitimacy of using an analysis of the logic of concepts as a basis for an account of the structure of history later. For the moment I wish simply to establish that this is what Rousseau is attempting to do.

He begins by distinguishing between acts based on physical instincts and capacities, and acts based on ideas. It is only in the second case that the question of the logical relation between ideas arises. He postulates two basic instincts prior to reason and two basic metaphysical and moral capacities. The first instinct is the urge to self-preservation, as embodied in hunger and thirst, and the basic physical skills required to satisfy these needs. The second instinct is compassion, "an innate repugnance at seeing a fellow-creature suffer" (1 182). He suggests that this instinct also is shared with animals. Now the postulating of an instinctive compassion is obviously much more dubious than the postulating of an instinct of self-preservation. Although Rousseau quite often refers to this instinct later in his argument, it does not seem to me to be logically required for the argument, so I shall not consider it any further. Much more important are what he describes as the capacities connected with the metaphysical and moral, as opposed to the physical, side of man.

These are:

a) That "man has some share in his own operations in his character as a free agent" (1 169); and

b) Man has a "faculty for self-improvement, which, by the help of circumstances, gradually develops all the rest of our faculties, and is inherent in the species as in the individual." (1 170).

It is important to realise that both of these can only be postulated as capacities requiring development, rather than as fully-developed faculties. This applies by definition to the capacity for self-improvement, but it also applies to the will, because freedom rests on choice, which rests on the awareness of alternatives, which rests on knowledge and imagination. Thus we can see that these two capacities are interrelated. Self-improvement, by widening our knowledge, also widens our freedom of choice: "It is by the activity of the passions that our reason is improved, for we desire knowledge only because we wish to enjoy; and it is impossible to conceive why a person who has neither fears nor desires should give himself the trouble of reasoning. The passions, again, originate in our wants, and their progress depends on that of our knowledge; for we cannot desire or fear anything, except from the idea we have of it, or from the simple impulse of nature. Now, savage man, being destitute of every species of intelligence, can have no passions save those of the latter kind: his desires never go beyond his physical wants. The only goods he recognises in the universe are food, a female, and sleep: the only evils he fears are pain and hunger". (1 171). The crucial point here is that we have to distinguish between the "simple impulses of nature", which are few and of limited consequence, and those passions which are based on ideas, and may increase in number and in scope as our knowledge develops. In so far as social conflict depends on the multitude of passions we are faced with the prospect of the process of self-perfection going hand in hand with increased conflict, and increased socially induced inequality and misery.

Placed in a benign environment the individuals in our model have no motive for lengthy conflict: "They maintained no kind of intercourse with one another, and were consequently strangers to vanity, deference, esteem and contempt; they had not the least idea of meum and

tuum, and no true conception of justice; they looked upon every violence to which they were subjected, rather as an injury that might easily be repaired than as a crime that ought to be punished." All these sources of conflict presuppose a society with concepts of rank, property, and rights. They cannot be presupposed outside of society. Even such an apparently basic urge as sexual jealousy has a strong social component. It depends on notions of the uniqueness and beauty of a particular individual, and it is necessary to give an account of the status of these concepts before assuming that sexual jealousy is natural. Similarly, exploitation of the strong by the weak presupposes relations of social dependence: "As the bonds of servitude are formed merely by the mutual dependence of men on one another and the reciprocal needs that unite them, it is impossible to make any man a slave, unless he be reduced to a situation in which he cannot do without the help of others: and, since such a situation does not exist in a state of nature, every one is there his own master, and the law of the strongest is of no effect." (1 190). In a situation of limited and easily satisfied needs, the trouble required to keep a slave under physical control is far greater than the trouble required to do the limited amount of work oneself.

In the condition described in this model, moral categories simply do not apply. We cannot describe these people as either good or bad. They are not bad because they do not know what it is to be good. They are immediately involved in the world with the satisfaction of their impulses. Although they may be faced with the choice between food and sleep, they are not faced with that range of choice which would be susceptible to analysis in terms of moral categories. Moral choice, as opposed to instinct, assumes a reflective independence from the situation. The possibility of evil and of selfishness of a destructive kind presupposes the development of reason: "It is reason that engenders self-love (amour-propre), and reflection that confirms at: it is reason that turns man's mind back on itself, and divides him from everything that could disturb or afflict him. It is philosophy that isolates him, and bids him say, at sight of the misfortune of others: 'Perish if you will, I am secure.'" (1 184). The point here is not that the premoral individual gives up his kill out of compassion

to any one who comes along. It is rather that any self-regarding action beyond the immediate satisfaction of impulse requires a concept of self which is something more than mere physical awareness. It requires the concept of an ego articulated in time and bearing specific long-term interests. Such an ego-concept cannot be simply assumed. It must be accounted for as the product of reflection and of reason. Reflection and reasoning require the use of various concepts, and these concepts cannot be assumed to have existed fully-formed in the mind of each natural individual. Instead, Rousseau, argues, we must understand the origin of these categories within the relation between man and his environment. Man in the state of nature, satisfying his simple desires with ease has no need to develop more elaborate thought categories. Or, to put it in non-historical terms, the categories of reflective thought cannot be understood independently of human action in the world.

Rousseau therefore places the original impulse for the development of reason in "the obstacles of nature": "But difficulties soon presented themselves, and it became necessary to learn how to surmount them" (I 193, 192). Increasing population means the necessity for learning to live in an increasing variety of environments. This developed both mental skills and productive skills, and as tools were developed in the context of immediate problems, so did the range of other goods which could be made with these tools extend. More tools, more products, a more sedentary way of life, and more lasting contacts with other people went hand-in-hand. But, as the capacity to produce new conveniences increased, so did dependence on these conveniences: "For,besides continuing thus to enervate body and mind, these conveniences lost with use almost all their power to please, and even degenerated into real needs, till the want of them became far more disagreeable than the possession of them had been pleasant." (I 196). Furthermore, as social relations, and family relations began to stabilise, the developing reflective capacity lead to comparison. "A tender and pleasant feeling insinuated itself into their souls, and the least opposition turned it into an impetuous fury: with love arose jealousy; discord triumphed, and human blood was sacrificed to the gentlest of all passions." (I 197). Social intercourse developed an interest in public esteem, and hence developed on one side vanity and

contempt, and on the other side shame and envy. Because of the growing importance of esteem, civility was necessary, and an injury which previously would have been forgotten as soon as the physical pain had departed would now remain as a symbol of another's contempt, demanding revenge.

Nevertheless, Rousseau argues, "this period of expansion of the human faculties, keeping a just mean between the indolence of the primitive state and the petulant activity of our egoism, must have been the happiest and most stable of epochs" (1 198). Human needs were still relatively limited, and the causes for conflict which existed were not such as permanently to threaten the stability of the society. While there was conflict, there was also no possibility of permanent dominance of one individual or group over another. The reason for this was that the family units remained self-sufficient. Dominance assumes both a motive and a relation of dependence which can be manipulated. If another does not depend on me in any way I cannot control him. Thus neither in this state nor in the pure state of nature was dominance and exploitation possible. "But from the moment one man began to stand in need of the help of another; from the moment it appeared advantageous to any one man to have enough provisions for two, equality disappeared, property was introduced, work became indispensable, and vast forests became smiling fields, which man had to water with the sweat of his brow, and where slavery and misery were soon seen to germinate and grow up with the crops" (1 199).

The basis for this lay in the division of labour, which produced social interdependence, and the parallel development of agriculture, which required stable property rights over particular pieces of ground. In these conditions natural inequalities of strength and skill were translated into social inequalities of property ownership. Enslaved to new needs, rich and poor were equally deprived on their freedom. "Each became in some degree a slave even in becoming the master of other men: if rich, they stood in need of the service of others; if poor, of their assistance; and even a middle condition did not enable them to do without one another. Man must now therefore, have been perpetually employed in getting others to interest themselves in his lot, and in making them,

apparently at least, if not really, find their advantage in promoting his own. Thus he must have been sly and artful in his behaviour to some, and imperious and cruel to others; being under a kind of necessity to ill-use all the persons of whom he stood in need, when he could not frighten them into compliance, and did not judge it his interest to be useful to them" (1 202-3). The exercise of power made possible for the wealthy by the possession of property gave rise to an experience of the pleasure of exercising power for its own sake and the pursuit of wealth became subordinated to the pursuit of power, an insatiable need, which therefore guaranteed permanent conflict even though it inevitably left both master and slave enslaved. It is in this situation that there is the war of all against all that Hobbe's attributed to the state of nature. Conflict as a necessary relation between individuals depends on private property. The desire to own private property is itself not "natural". It is the product of a particular historical evolution.

But because private property is not natural in any way, its possession is intrinsically insecure. Even the claim that private property is based on labour is invalid, in so far as it is used to defend the right to own more than is needed for personal subsistence. By what right did the individual take possession of what can be seen as the common heritage of all mankind in order to labour on it for his own ends? "Who gave you your standing, it might be answered, and what right have you to demand payment of us for doing what we never asked you to do?.... You ought to have had the express and universal consent of mankind, before appropriating more of the common subsistence than you needed for your own maintenance" (1 204). In fact, therefore property rights are based on force. It is for this reason that there is a war of all against all. What is originally taken by force can only be held by force. However, if this force can be transmuted into authority and law, then the situation can be stabilised in such a way as to make actual violence only the rare last resort required to protect property. "The rich man, thus urged by necessity, conceived at length the profoundest plan that ever entered the mind of man: this was to employ in his favour the force of those who attacked him, to make allies of his adversaries, to inspire them with different maxims, and to give them other institutions as favourable to himself as the law of nature

was unfavourable" (1 204-5). Thus came into existence government, founded to protect the property of the rich against the majority, but defended as of mutual advantage for the peace of all. The institution of civil society institutionalised inequality and sanctioned it by law.

Yet Rousseau has still not finished his account of the origin of inequality. For the first political society, although set up in the interest of the rich, was also set up with the manipulated consent of the poor. It was contractually based, not a right of unlimited authority. Otherwise, he argues, the poor would not have agreed to it, the rich would have had to achieve control by conquest, and conquest never becomes law. It is always purely force. Tyrannies sanctified by law, as common then as now, cannot therefore be understood purely in terms of conquest. There is a process of inevitable decline in political institutions.

Rousseau begins his account of this process of decline as though he has not already demonstrated that government only comes into existence to preserve existing inequalities. Political distinctions which come into existence with the founding of political society of themselves produce civil distinctions, and these can be manipulated by unscrupulous magistrates wishing to usurp illegitimate power. This once more introduces conflict into society, a conflict which is naturally related to wealth, since wealth can be used to buy all other forms of distinction. Oppression, misery, inequality and mistrust lay the groundwork for absolute despotism, which is the destruction of any law, and a return to the law of the strongest.

Rousseau summarises the development of inequality as follows: "We shall find that the establishment of laws and of the rights of property was its first term, the institution of magistracy the second, and the conversion of legitimate into arbitrary power the third and last; so that the condition of rich and poor was authorised by the first period; that of powerful and weak by the second; and only by the third that of master and slave, which is the last degree of inequality" (1 214-218).

Let us now look back over the course of the argument to see

what has been established. There seems to me to be four major propositions:

1) Humans are social beings and historical beings. The capacities that define their nature develop in social interaction, and their "nature", as defined by both needs and skills, varies historically. This historical alteration is a function both of interaction with the natural world and of interaction with other humans. (Rousseau is aware that language plays a vital role in this process, but after wrestling with the problem of the origin of language over several pages gives up in despair, since language seems to be necessary for the development of society, but society seems to be necessary for the development of language.)

2) In this historical process humans develop their capacity to reason, but reason, by enabling individuals to think themselves separate from the natural world and from other people, leads to division, conflict and misery.

3) At the same time there is an increasing interdependence between individuals. This interdependence is both the result and the vehicle of the development of divisive reason. As a result of this, "the savage lives within himself, while social man lives constantly outside himself, and only knows how to live in the opinion of others, so that he seems to receive the consciousness of his own existence merely from the judgement of others concerning him" (1 220). Because of this interdependence, each has to try to manipulate the other, and all are trapped in a set of relationships over which they have no control.

4) This social division is embodied in the institution of private property. It is only through the institution of private property that it becomes possible for conflict to be regularised in the form of exploitation and inequality. Private property seems to be linked with the division of labour, that is with interdependence, in establishing the possibility of exploitation, although Rousseau does not make this connection explicit in the *Discourse*.

These four propositions are all presented as descriptive. They describe the nature of human reality in "advanced" societies such as that Rousseau lived in, and, *a fortiori*, such as we live in. Rousseau's reaction to this is neither one of total despair, nor the advocacy of an attempted return to the state of nature. Such a return, he recognises, is impossible, precisely because of the historical nature of human beings. Only a 'natural man' could live in the state of nature. We are no longer natural, since our reality is inseparable from our culture. His problem is, rather, given the fact of a developed reason, and the fact of social interdependence in advanced society, to determine whether and how it would be possible to use that reason as a synthetic rather than as a divisive force, and to make that interdependence the basis for freedom rather than for exploitation and oppression. It is this that he attempts to do in the *Social Contract*.

Some writers have suggested that the pessimistic descriptions of contemporary society in the *Discourse* are incompatible with the more optimistic *Social Contract*. This is quite incorrect. It rests partly on a failure to grasp the extent to which the *Social Contract* assumes the four main points made in the *Discourse*. In particular, it results from a failure to notice the apparently ambiguous status of reason in the *Discourse*. Although Rousseau shows reason as a divisive force, he also speaks of the human faculty for self-improvement and perfectibility, and links this to the development of the understanding. Before making his most savage attack on contemporary society, he writes: "Behold then all human faculties developed, memory and imagination in full play, egoism interested, reason active, and the mind almost at the highest point of its perfection" (1 202). This is not purely ironic. At the end of the appendix he attacks those who are "not apprehensive of degrading your species, by renouncing its advances in order to renounce its vices" (1 228). The development of reason is a very real advance. It is clear that the unpleasant state of contemporary society can only be transcended by continuing this advance, and by attempting to develop reason's synthesising capacity in counterpoint to its apparently destructive tendency to analyse, divide and separate.

In an early draft of the *Social Contract*, in a chapter which was not included in the final version Rousseau makes this point very clearly. As it is not easily available, it is worth quoting from it at some length. Referring to the ideal period in which faculties had been developed to some extent but in which inequality had not yet been institutionalised he writes: "That perfect independence and that liberty without rule, even had it remained joined to the ancient innocence, would still have had one essential vice harmful to the progress of our most excellent faculties, namely the want of that union of the parts which constitutes the whole. The earth would be covered with men amongst whom there would be scarcely any communication; we would touch one another in a few points, without being united by any; each would remain isolated among the others, each would think only of himself; our understanding would not develop, we would live without feeling anything, we would die without having lived; our whole happiness would consist in not knowing our misery; there would be neither goodness in our hearts nor morality in our actions, and we would never have tasted the most delicious sentiment of the soul, which is the love of virtue" (Oc3 283 : see also E 181). Concluding the chapter, and referring to the state of humankind as described in the *Discourse*, and summarised in this chapter, he writes: "Let us try to draw from the very evil the remedy that must cure it.... Let us show ... in the perfected art the compensation for the ills which art in its beginnings did to nature" (OC3 288).

III. REASON AND COOPERATION: THE SOCIAL CONTRACT

The *Social Contract* is a much more controversial text than the *Discourse on Inequality*, but I hope that much of the controversy will have been avoided by the above account of the *Discourse*, which helps to make clear exactly what questions Rousseau is asking in the *Social Contract*.

Rousseau commences his argument in Chapter 1 with a much quoted passage which yet does not seem to me to explain unambiguously what his intentions are: "Man is born free; and everywhere he is in chains. One thinks himself the master of others, and still remains a greater slave than they. How did this change come about? I do not know. What can make it legitimate? That question I think I can answer" (SC 3). Does he, then, intend to legitimate the chains? That is certainly what he says in this passage, but, although, as we shall see later, his practical attitude to governments is quietistic, it is not what he does in the theoretical analyses at the beginning of the *Social Contract*. A better guide to his intentions is the first sentence of the book, in the short introductory paragraph: "I mean to inquire if, in the civil order, there can be any sure and legitimate rule of administration, men being taken as they are and laws as they might be" (SC 3). Clearly here he is interested in giving an account of the kinds of laws which would be legitimate, rather than in legitimising all laws. The question is, then, under what circumstances does an individual have a moral duty to obey the law.

We have already seen, in the *Discourse on Inequality*, how men are. They are no longer simple and limited; they have developed minds and developed needs. Moreover, they are factually dependent on one another in their economic and social activities. They are social individuals, not isolated individuals. This social nature is not extrinsic and accidental. It permeates them in the form of a culture in terms of which they think and act. It is in this context that we have to analyse legitimacy, and to interpret what is, for Rousseau, the concept from which all moral argument derives: the concept of freedom.

The right of governments to expect obedience to the law cannot be based on force, in the form of the "right of the strongest", or of the "right of conquest", or of any "right" derived from the fact of slavery. Force, conquest and enslavement only establish situations of fact, not of right. The war of all against all subsists, and the dominated may attempt to alter the situation by themselves using force. Rousseau uses two criteria in his criticisms of the arguments from force to right. Firstly, freedom is an absolute value: "To renounce liberty is to renounce being a man, to surrender the rights of humanity and even its duties ...Such a renunciation is incompatible with man's nature; to remove all liberty from his will is to remove all morality from his acts" (SC 8). Secondly, there must be positive reasons for entering the state of society: "To say that a man gives himself gratuitously, is to say what is absurd and inconceivable; such an act is null and void, from the mere fact that he who does it is out of his mind" (SC 7). These points are fairly obvious, but it is worth making them explicit, because they indicate that Rousseau's whole theory is designed to preserve the interests of the individual. The problem is, then, to determine what those interests are and how they can best be preserved.

In the second chapter of the early draft of the *Social Contract*, referred to earlier, Rousseau criticises what he sees as the facile view that in the state of nature man's reason leads him to recognise the necessity of obeying the laws of nature.

The theory is inadequate because it fails to give an account of how reason could develop to the point at which the individual could abstract himself from his particular concerns and adopt the point of view of the general good in order to understand and submit to the law of nature: "But where is the man who can thus separate himself from himself, and if care for his own preservation is the first precept of nature, can one force him to look to the species in general in order to impose on himself duties the relation of which to his particular constitution he does not see at all? ... Does not it still remain to be seen how his personal interest requires that he submit to the General Will?" (OC3 286). In the situation in which it is postulated that

natural law will prevail and produce society, humans still remain divided by reason, and the individual's interest remains defined in terms which are from the beginning incompatible with the interests of others (cf. OC3 283).

Law can therefore only derive its legitimacy from the agreement of those to whom it applies, which fact Rousseau expresses in terms of the idea of an original contract. This contract must be such as to be compatible with the real interests of the individuals involved, if we are not to have to admit that law is never legitimate but always based on force, even if it is only the force of a majority against a minority. Am I forced to obey the law, or do I obey it freely? This is Rousseau's question.

The traditional liberal answer to the question of legitimacy is to divide matters up into two domains, the private and the public, and to argue that it is legitimate that people should be constrained by law in the public domain, but not in the private domain. Thus we begin with a sphere of inalienable individual rights counterposed to a sphere of limited sovereignty. The problem is, then, to draw the dividing line in the right place. From Rousseau's position, however, this is the wrong way of proceeding. He has already shown that the human is a social being, dependent externally in all acts on others, dependent internally on a socially produced culture. Instead of distinguishing between private acts and public acts, we need to distinguish between a private component and a public component in each act. Rousseau does not formulate his theory in these terms. Rather, he distinguishes between the General Will and the particular will, and he argues that the principle of the necessary social contract is: "Each of us puts his person and all his power in common under the supreme direction of the General Will, and, in our corporate capacity, we receive each member as an indivisible part of the whole" (SC 13). The body thereby created is called the State, and the General Will, when legislating, is called the Sovereign. Although he often tends to personalise the Sovereign and the General Will, it is clear that what he is referring to in either case is not a particular individual or body, but a principle. Thus we have to distinguish between the problem of

establishing exactly how the principle should be embodied in actual government. Rousseau's point is that what is involved in the problem of sovereignty is not a question about the relation between one group of people (the king, or the parliament, or whatever) and another group of people, but a relation between "The public and the individuals" (SC 13). That is, each individual is present on both sides of the equation, but present in a different way. But this immediately produces a problem. Because the individual is present in different ways on each side - as having a particular will on the one hand, and as having a general will on the other - there is no guarantee that the equation will balance. This fact produces problems, but it also makes it clear that the general will is a principle and not an existent. Rousseau writes: "For if the clashing of particular interests made the establishment of societies necessary, the agreement of these very interests made it possible. The common element in these different interests is what forms the social tie; and, were there no point of agreement between them all, no society could exist" (SC 20). Every particular act which I perform presupposes a social context, and in willing the act, I must will the social context. In willing to get to town for a date at 2 o'clock, I necessarily will the whole system which makes the buses run on time. This is the public component of my act. The General Will as a principle is the sum of these public components of each individual's will. It can be described as a will because it is, in fact, implicitly willed by each individual in each act. But it remains in its purest form a principle because its implications are not necessarily realised by each individual when willing. The General Will is that principle, in terms of which society must be organised if each individual is to be able to satisfy his or her individual interests. It is for this reason that laws are only legitimate if they are the embodiment of the general will. Because the General Will is general, it must apply equally to all. This does not mean that each individual in the state must be treated alike. It means that the principles in terms of which they are treated differently must be laid down independly of the particular interests of specific individuals, and must be justified rationally. Runciman has defined a principle of social justice which seems to me to express very precisely what Rousseau understands by the General Will: "We should ask, in

assessing the merits of rival claims, what criteria of assessment would have been established by rational persons if we could suppose them to have been required to agree, from a state of primordial equality, on principles by which they would be prepared to make claims and to concede them" (Runciman, *Relative Deprivation and Social Justice*, 297). That is, it is not that the General Will does not take individual needs into account. It is that the General Will is embodied in a decision in which each individual, instead of fighting for his or her own particular interest, has abstracted his or herself from the particular position in society, and has considered what would be the most equitable way in which treatment should be meted out to the whole set of roles that make up the society. We may imagine that if a group of people were drawing up the rules for a game, if each individual knew what position he or she was going to occupy in the game, each would argue for a set of rules that favoured a player in that particular position. But if they did not know which positions they were to be in, it would be in the interest of each to draw up a set of rules which favoured no particular position, and gave each an equal set of advantages and handicaps. Rousseau describes this position as follows: "Why is it that the General Will is always in the right, and that all continually will the happiness of each one, unless it is because there is not a man who does not think of 'each' as meaning him, and consider himself in voting for all? This proves that equality of rights and the idea of justice which such equality creates originate in the preference each man gives to himself, and accordingly in the very nature of man. It proves that the General Will, to be really such, must be general in its object as well as its essence; that it must both come from all and apply to all; and that it loses its natural rectitude when it is directed to some particular and determinate object, because in such a case we are judging of something foreign to us, and have no true principle of equity to guide us"(SC 24-25). This point is illustrated in the case of crime: not knowing whether I am to be criminal or victim, I agree that there should be punishment for crime. It is in the sense that, if I am the criminal, I may be said to will the punishment: it is logically implied by the general component of my will. (cf. SC 28.)

Since the General Will (and hence the sovereign) always lays

down principles, but never regulates particular matters, it is necessary to distinguish between the General Will and the government, and between laws, which are general in scope, and are laid down by the General Will, and decrees, which are particular, and which are acts of the government (or magistracy). For example "the law may indeed decree that there shall be privileges, but cannot confer them on anybody by name. It may set up several classes of citizens, and even lay down the qualifications for membership of these classes, but it cannot nominate such and such persons as belonging to them" (SC 30). The actual application of the laws, therefore, is the task of a special body of magistrates. But it is clear that these may not make law, and that they are bound by the law in all respects. The General Will cannot err - not because it is infallible in the way in which the Pope is infallible, but because it is not an actual will at all. We may describe it as a principle in terms of which whatever body lays down laws should will if those laws are to be recognised as legitimate. The magistrates, on the other hand, are fallible. They do not embody the General Will, but express their own particular wills. The body of magistrates is described as the Prince: "If finally the Prince should come to have a particular will more active than the will of the Sovereign, and should employ the public force in his hands in obedience to this particular will ... The social union would evaporate instantly, and the body politic would be dissolved" (SC 49-50).

The General Will should be obeyed by the individual because it is his or her own will. But the recognition of this duty assumes the individual's understanding of the relation between the private and public components of the will, and this, for reasons which Rousseau has himself given in his criticism of pure natural law theory, cannot be assumed. This implies that there must exist the right to use force against those who transgress: "In order then that the social compact may not be an empty formula, it tacitly includes the undertaking, which alone can give force to the rest, that whoever refuses to obey the General Will shall be compelled to do so by the whole body" (SC 15). This is noncontroversial, but Rousseau then goes on to say: "This means nothing less than that he will be forced to be free" (SC 15). We can only make sense of this in terms of the account of the social nature of

human reality. What does freedom mean? If we assume a fixed human essence, and a fixed set of needs, then we might define freedom, as does Hobbes, as the absence of restraint on the impulsive satisfaction of needs. We might then agree that each individual would gain by giving up some of his freedom in order to ensure a minimum of social stability in which each could satisfy the remaining needs, but we would describe this as giving up some freedoms, not as gaining freedom, and we would not say that the individual being forced to obey the law was being forced to be free. But this approach assumes an essential independence of the individual from society, and this Rousseau has rejected. Once we assume that human beings are intrinsically both internally and externally involved in society, and further, that human reality is historically constituted, then the problem of freedom becomes more complex.

In the original state of nature man is pre-moral, in that he acts in terms of immediate physical needs, and "ought" questions do not arise. If he wants to eat a particular bunch of bananas he does not stop to ask "Ought I to want to eat these bananas?" But as a result of the development of reason, and hence of reflection, these questions do begin to get asked. We then have to answer them in terms of some model of what human beings should be like. In particular, when it becomes clear that there is a wider range of choices than at first appeared, it also becomes evident that there was a wider range of restraints operating than had previously been recognised. What made me eat those bananas, when I needn't have done so? My appetite did. It is meaningful to say that my appetite "made me" do something, because we can conceive of a situation in which "I" resisted "my" appetite. Now there are obviously some very thorny theoretical problems involved here, which Rousseau does not go into. What he does do is to state an important position: "The mere impulse of appetite is slavery, while obedience to a law which we prescribe to ourselves is liberty" (SC 16). He distinguishes between natural liberty, which is absence of physical restraint, and civil liberty, which is something more. In defending civil liberty he is judging in terms of an ideal model of rational action. Reason involves the capacity to abstract oneself from the particular present set of stimuli and to formulate general principles. The implicit social

principle which derives from reason is the principle of communication, of sharing experience. Thus on the one hand the rational individual is not subject to a particular impulse, and so is free, and on the other hand the rational individual is interested in communicating with fellows, and thereby establishing both a general understanding and a General Will. Reason is therefore both the principle of the General Will and the principle of freedom.

The relation between the General Will and freedom can be explicated in terms of the idea of contradiction. If the particular component of my will is in some way incompatible with the general component of my will, then my will is logically self-contradictory, although it is not necessarily practically self-contradictory, since there may well be enough other people acting in such a way as to preserve the social fabric on which my particular will depends for it not to make any difference that I am tending to destroy it. Nevertheless even the logical contradiction suffices to make my will in an important sense irrational. In so far as reason is the only principle I have for organising my behaviour an irrational act is outside my own control, and hence also unfree.

This interpretation also helps to indicate that there is a wide range of personal acts which are not in contradiction to the General Will. The General Will does not prescribe specific actions to me, since it is only the public component of my will which is subject to the General Will, and there are usually a wide range of particular acts available which do not contradict the public component. If I want to catch a bus it is irrational for me at the same time to act in a way which is incompatible with the continuation of the bus service (say by trying to avoid paying the fare). But this still leaves me a wide range of buses and destinations to choose from. Thus Rousseau does not mean that the ideal is for the individual to become totally absorbed without remainder in the General Will. Nor does his account of freedom necessarily imply that physical impulses should be suppressed. A physical impulse is a reality, and it would be irrational for a corporal being to attempt to deny the existence of such a reality.

Rather it must be integrated into the larger whole of the individual's life, and not allowed to exist as a separate and dominating entity.

Admittedly, Rousseau is slightly ambiguous in his analysis of the relation between individual and general will. He writes initially of "the total alienation of each associate, together with all his rights, to the whole community" (SC 12). Elsewhere he clarifies this somewhat: "Each man alienates, I admit, by the social compact, only such part of his powers, goods, and liberty as it is important for the community to control; but it must also be granted that the Sovereign is the sole judge of what is important" (SC 24). Clearly the distinction here is between what in each act is relevant to the General Will and what is not. This distinction is made explicit in the treatment of specific topics such as property and religion. In his *Lettres Ecrites de la Montagne*, in which he defends himself against attacks made on him by the Geneva government, he clarifies his views on religion in the last chapter of the *Social Contract*. He distinguishes between three elements in religion, the cult practices, the theology, and the moral teaching, and he writes: "As to the part of religion which has to do with morals, that is to say, justice, the public good, and obedience to natural and positive law, social virtues and all the duties of man and the Citizen, the government has the right to be informed of them: it is in this sole respect that Religion enters directly under its jurisdiction, and that it may banish, not error, of which it is not the judge, but any harmful sentiment which tends to cut the social ties" (OC 3 694-5). Of course, the distinction between what is and what is not relevant to the General Will immediately raises the problem of criteria, to which I shall return at a later point.

One final point should be made in discussing the question of the freedom of the individual in Rousseau's theory. That is that Rousseau made it quite clear, at least on the theoretical level, that, if the laws do not conform to the General Will, if, therefore, the compact has been broken, "each regains his original rights and resumes his natural liberty" (SC 12). He builds into his definition of law that it should be the expression of the General Will. This does not

mean that anything normally described as a law is the expression of the General Will. It means that if it does not embody the principles of the General Will, it is not a proper law, it is not legitimate, and there is no moral reason for obeying it. Similarly, he builds this concept of law into his definition of "Republic", and when he refers to either republic, or law, or sovereign he is usually referring to these as principles, not as existents.

Thus the principle of Rousseau's theory of legitimate law is that law should be rationally designed to further the needs of the community as a whole in such a way as to allow each individual to seek his or her own happiness in a way which does not damage the communal substructure on which the individual's actions depend. Action of this kind is at the same time free because it is rational. In this theory there are two dialectical tensions: the tension between the individual and the society, and the tension within the individual between reason and impulse. The whole point of Rousseau's argument is that these tensions are a necessary part of human reality. We may not counterpose the individual and the state, or the reason and impulse, as independent opposites. They are dialectically interdependent opposites. The problem, however, is to spell out in detail the practical implications of this both for our understanding of human reality, and for moral behaviour and political institutions. Rousseau is aware of some, although by no means all, of the difficulties in the way of doing this.

The most obvious problem is that of how to discover what the General Will is in a concrete situation. There is, Rousseau points out, often a great deal of difference between the General Will and the sum of the particular components of the wills of each individual. If each individual seeks only his or her private interest, and leaves up to somebody else the problem of safeguarding the general social interest on which they all depend, then the final result is not going to be what anybody intended. A contemporary example of this is environmental pollution. No individual pollutes the environment, since the cast-offs of any particular individual could easily be absorbed into the environment. It is the "Will of All", the sum of each of these particular components, which produces the pollution. Thus although the "General Will is always right and tends to the public

advantage ... it does not follow that the deliberations of the people are always equally correct ... the people is never corrupted, but it is often deceived, and on such occasions only does it seem to will what is bad" (SC 22-3). Rousseau's solution to this problem is not clearly stated. Sometimes he says that a vote under suitable conditions will always give the General Will (with or without the guidance of an imported "Legislator"), whereas at other times he says that a majority vote is never a guarantee, since a majority can be wrong just as easily as a minority. He never seems to be clear as to when he is discussing the principle, and when he is discussing a practice. In his initial discussion of the difference between the General Will and the Will of All he makes one important proposal for guaranteeing the emergence of the General Will: "There should be no partial society within the state, and ... each citizen should think only his own thoughts" (SC 23). His point here is that any group has its own "General Will", that is, its own set of conditions which will further the interest of all the members of that group. But this will, general for the members of the group is particular in relation to the Sovereign. It is also much more powerful than any individual particular will. Rousseau believed that in a situation in which no particular will is powerful, the end result is likely to be a fair compromise which represents the general interest, whereas if there exist particular organised factions one of these is likely to be able to impose its particular will in a winner-takes-all result which totally leaves out of account the interests of the losers. Also, it is important to remember that Rousseau was writing at a time when "the people" were disorganised, and faction was the prerogative of the rich, (which is not that far from being the case today). It was under these circumstances that he thought that faction would play into the hands of the powerful and reinforce privilege. (See Judith Shklar, *Men and Citizens*, 185.)

Although he does make some practical suggestions, Rousseau finally moves into the realm of fantasy. Because he has shown so convincingly that decadence in the state is a function of historically produced decadence in the people, he has to accept that those people who need laws cannot give them to themselves; and those peoples who are

sufficiently rational to give laws to themselves do not need them. To break out of this circle he introduces from outside the figure whom he calls the Legislator.

This Legislator must come from outside the society and has no role in the society other than that of formulating the law. He neither acts as Prince, in having command over men, nor as Sovereign in decreeing the law. Rather he reveals the law, which the sovereign people may then legislate. To be able to do this he must be "a superior intelligence beholding all the passions of men without experiencing any of them... This intelligence would have to be wholly unrelated to our nature, while knowing it through and through; its happiness would have to be independent of us, and yet ready to occupy itself with ours; and lastly, it would have, in the march of time, to look forward to a distant glory, and, working in one century, to be able to enjoy in the next. It would take gods to give men laws" (SC 32). This is more helpful as an account of the rational nature of the principle embodied in the General Will than as a contribution to the practical problem of legislation. The Legislator must not only be able to discover the laws, he must also, by force of his own personal authority (charisma) be able to remove the divisive effects of the particularistic culture and to transform each individual into being a conscious part of the whole. (This portrait of the Legislator is not intended to be entirely theoretical, as it is partly based on his conception of the role played by Lycurgus in Sparta and Numa in Rome. But essentially the Legislator can only occur accidentally, and in contradiction to his theory of history. Paraphrasing Marx's third thesis on Feuerbach, we must ask: "Who will educate the legislator?" Unless this question can be answered a radical critique of society along Rousseauan lines must end up in the quietistic position which Rousseau himself actually adopted.)

Rousseau's critique in the *Discourse on Inequality* and in the *Social Contract* undoubtedly is radical. It is an attack on wealth and on privilege; he lays down criteria for just laws, and says that if these very strict criteria are not met then the individual is no longer obliged to obey the law. Yet he concludes his violent attack on social and political inequality in the *Discourse* by writing that moral individuals

in this situation "will respect the sacred bonds of their respective communities ... they will scrupulously obey the laws, and all those who make or administer them" (1 229). In part, of course, we may interpret this kind of conclusion as being a prudent rider necessary in a world of despotic rulers not over-fond of advocates of revolution. In his *Letters from the Mountain*, he does not seem quite able to make up his mind whether he is an activist or not. On one page he writes, "Everything considered, I gave preference to the government of my country. That was natural and reasonable, and I would have been criticised had I not done it. But I did not exclude other governments; on the contrary: I have shown that each has its reasons which may make it preferable to any other government according to the men, the time and the place. Thus, far from destroying all governments I have established them all" (OC3 811, see also Rousseau, Juge de Jean-Jacques, 3rd Dialogue OC3 935). But two pages further on, at the beginning of the next letter, he writes of Geneva: "Nothing is freer than your legitimate state; nothing is more servile than your present state" (813). Thus although he may be establishing the existing laws, he is certainly not justifying the existing government. In part his protestations of respectability involve an illegitimate play on the ambiguity of the relation between his special definition of law and the ordinary usage. But more is involved than a mere attempt at camouflage. His argument leads logically to a pessimistic position. Laws develop historically, and reflect the state of development of the people. (See *Polysynodic* OC3 638.) Therefore the people would neither be able to recognise nor to obey good laws. This is expressed most strikingly in his discussion of serfdom in his *Considerations on the Government of Poland and on its Projected Reformation* . After an attack on serfdom he continues: "I feel the difficulty in the project from emancipating your people. What I fear is not just the misunderstood interest, the pride and the prejudice of the masters. That obstacle vanquished, I fear the vices and the cowardice of the serfs. Freedom is a rich food (De Bon Suc) which requires a strong digestion; very healthy stomachs are required to stand it. Laugh at those degraded peoples who, permitting themselves to be aroused by agitators, dare speak of freedom without having the least idea of what it means and who, their hearts full of the vices of slaves, imagine that in order to be free one only needs to be mutinous.

Proud and holy liberty! If these poor people only knew you, if they knew at what price one acquires and conserves you, if they felt how much more austere than the hard yoke of the tyrant are your laws; their weak souls, slaves to the passions that would have to be stifled would fear you a hundred times more than servitude; they would flee from you in terror as from a burden about to crush them" (OC3 975). He therefore advises against immediate emancipation and suggests a scheme whereby individual serfs could gradually be emancipated by demonstrating their moral worth.

In drawing up his plans for Corsica and for Poland he was being consulted, or thought he was being consulted, by heads of state, and so felt some hope of success in bringing about change. But in *Emile* he makes it clear that all one can really hope for is individual salvation. Emile is not educated to play a part in a free society. He is brought up rather to remain aloof from any society in which he may have to live, to resist being "carried away by the passions and prejudices of men" (E 217).

Any theory which argues from the historical nature of man and of society is likely to lead to this position, unless it can historicise the future as well as the past, by elucidating the possibility of unifying reason arising within the context of a divided society.

To the four major propositions derived from the *Discourse on Inequality* the *Social Contract* adds a fifth: that although reason is initially divisive it contains through its universality, the possibility of overcoming the divisions in society. Through reason human beings may discover a new kind of individuality which complements rather than contradicts the individuality of each other. Interdependence is inevitable within contemporary society, but through reason that interdependence may be changed from a basis for domination and inequality to the basis for a new freedom.

IV. THE PROBLEM OF REASON

So far I have simply interpreted Rousseau's theory, without attempting to explore and either criticise or defend his method of argument. I suggested at the beginning that his method involved the analysis of the logic of various concepts, but it must be clear from the exposition that he does not do this in any systematic way. Although the five main propositions which I have isolated constitute a coherent system, each conceals a host of difficulties. These propositions may be summarised as follows:

1. Humans are social beings and historical beings.

2. Reason arises in the historical development of human reality, and initially plays a divisive role in society.

3. Parallel to the development of divisive reason is the development of social interdependence: "Our needs bring us together in proportion as our passions divide us, and the more we become the enemies of our fellows, the less we can do without them" (Ch 2 Geneva Manuscript of SC, OC3 282).

4. Social division and dependence is embodied in the institution of private property.

5. Reason contains the possibility of utilising the factual interdependence to develop a sense of unity and a universal General Will, which involves real individual freedom and social harmony.

To which may be added as a sixth proposition the pessimistic conclusion that:

6. But this remains only a theoretical possibility since the existing social institutions reinforce the divisive aspects of reason, rather than its unifying potential.

I think it is obvious merely from reading these propositions

that they contain the central ideas of Kantian ethics, and of Hegel's dialectic, while Marx offers a way out of the dilemma posed by the sixth proposition. These writers, and Sartre after them, merely refine and elaborate Rousseau's essential insights, and the best way, perhaps, to test their validity is to see how successfully these writers do manage to do this. This will involve a study of the philosophical methods underlying the theory, since this method is also made more explicit in the later philosophers. Thus in showing the development of the theory I shall at the same time be discussing its claim to truth, and at the end of the study I shall attempt to analyse the problems left outstanding. Here I wish merely to indicate in a little more detail some of the central difficulties in the theory as we have formulated it so far.

1) The first problem is the problem of human rationality. Both reason as a concept and rationality as a part of human reality remain ill-defined; indeed they remain so ill-defined that some critics have interpreted Rousseau as an apostle of feeling, rather than of reason. There are at least four different concepts which Rousseau uses in a rather unclear way in relation to one another: these are reason, feeling, nature, and conscience. Partly this arises from the fact that he is at times speaking of men as they are in society, and at other times speaking of possibilities. It is in *Emile* that he discusses reason in the most consequent fashion, and from this it would seem that he sees no fundamental conflict between reason and "nature" in the special sense in which he uses that term, or between reason and "feeling", as opposed to "passion". He writes: "But remember, in the first place, that when I want to train a natural man, I do not want to make him a savage and to send him back to the woods, but that living in the whirl of social life it is enough that he should not let himself be carried away by the passions and prejudices of men; let him see with his eyes and feel with his heart, let him own no sway but that of reason" (E 217). It is clear from this that when the passions are under control nature, feeling and reason are in harmony. (See also *Nouvelle Heloise* 357). In understanding some apparent contradictions in Rousseau, it is necessary to distinguish between reason as (or "healthy reason", see *Nouvelle Heloise* 362) a principle, and the reasoning powers of a particular individual,

which may well be corrupt. Nevertheless, even when this distinction has been made, there is still a problem about the relation between reason and conscience. At the beginning of *Emile* he writes that it is reason alone that "teaches us to know good and evil. Therefore conscience, which makes us love the one and hate the other, though it is independent of reason, cannot develop without it" (E 34). But later, in the *Creed of a Savoyard Priest*, the priest says "Too often does reason deceive us; we have only too good a right to doubt her; but conscience never deceives us; she is the true guide of man; it is to the soul what instinct is to the body; he who obeys his conscience is following nature and he need not fear that he will go astray" (E 249-250). And yet, a few pages later, at the end of the first part of his creed he instructs his listener to "give to my words only the authority of reason" (E 258), and does not consider inviting him to judge of the truth of his statements by reference to conscience or feeling. Similarly, in his later discussion of revealed religion his appeal is continually to reason: "In vain do they exclaim 'Submit your reason'; a deceiver might say as much; I must have reasons for submitting my reason" (E 261). We may infer from the above, that Rousseau believed that understanding could come only through the use of reason, but that he was not entirely clear on the relation between reason and action. Partly this is related to the fear that reason of itself cannot provide a sufficient motive for virtue in a world in which there is no reason to believe that virtue will also lead to personal well-being (See 278 footnote). In terms of the General Will, we may say that reason may enable the individual to discover what the General Will is, but that he or she would only lose out by following the General Will if there were no sanction to ensure that others, less rational, also did so.

Thus a major problem is to explicate precisely what is meant by the term reason, and to make clear the relation between reason as an epistemological category, and behaviour.

2) This problem of conceptual clarification is related to the problem of justifying the claims that human beings are, at least potentially, free and rational. At the beginning of the *Discourse on Inequality* Rousseau affirms these two propositions without much in the way of argument.

Freedom and reason are linked in that both involve an escape from the particular and the immediate. The awareness of alternatives depends on the capacity to generalise, as Rousseau hints, because to generalise is to relate the immediate experience in some way to other past and future experiences. It is to experience the given in the context of what is not given, and this is at least analagous to experiencing the given as non-necessary, and hence as subject to choice. Rousseau begins to explore this problem more explicitly in *The Creed of a Savoyard Priest*. There are theories which deny both the capacity to have objective knowledge and the capacity to be free. Such theories argue from physical causality, and suggest that, as part of a causal network, all human ideas have to be understood as "impressions" resulting from the interaction between the body and the world. If this is the case, then it is not possible to escape from particularity: "But who am I? What right have I to decide what is it that determines my judgements? If they are inevitable, if they are the results of the impressions that I receive, I am wasting my strength in such enquiries; they would be made or not without any interference of mine" (E 232). I am continually aware of receiving sensations, either from the outside world or from my memory. Under these circumstances, how can I be sure that the sense of a self which I have is not one more sensation, the mere by-product of the material world? The reason is, Rousseau argues, that "I find that I have the power of comparing them, so I perceive that I am endowed with an active force of which I was not previously aware" (E 232). To compare is to judge, to postulate a relation. The relation between two sensations, even if it is the simple relation of difference or similarity, is never *in* either of the sensations: "To my mind, the distinctive faculty of an active or intelligent being is the power of understanding this word "is". I seek in vain in the merely sensitive entity that intelligent force which compares and judges; I can find no trace of it in its nature. This passive entity will be aware of each object separately, it will even be aware of the whole formed by the two together, but having no power to place them side by side it can never compare them, it can never form a judgement with regard to them" (E 232-233). This is a very important argument which Kant modifies and develps much more fully and precisely in the *Critique of Pure Reason*. Through an argument of this type it is possible to establish

against a purely deterministic materialism, a defence of freedom which is not purely deterministic materialism, a defence of freedom which is not purely speculative, but which derives its force from the presupposition made by such a materalism of the possibility of knowledge. But the problem which results, which Rousseau does not touch upon, and which Kant wrestles with unsuccessfully, is to give a coherent account of the relation between this freedom and the fact of human materiality. This leads us into the next problem.

3) When Rousseau states that the mere impulse of appetite is slavery, while obedience to a law which we prescribe to ourselves is liberty, he is stating a principle, but not providing any criteria for the application of that principle. He is distinguishing between real and false needs, which seems intuitively to be a reasonable distinction to make. It would seem to be reasonable to say that the need for water is a real need, but that the need for, say, electric toothbrushes is a false need. But although it is easy to think of examples of one or the other, it is much more difficult to lay down precise criteria which can be used to determine whether or not a particular need or appetite is false and enslaving, or real. Even to be able to approach this problem we need a much more precise idea of the status of material appetites in the total human person, and this in turn requires us to explicate the relation between materiality and freedom. This is important for both political and moral theory. We may say that Rousseau establishes that there is a dialectical relation between the person as subject and the person as object, but that he does not elucidate the structure of that dialectic.

4) Similarly, (and analogously) he establishes that there is a dialectical relation between individual and society, but he does not go very far towards elucidating that dialectic. Once more this can be formulated in terms of a problem of criteria. Admitted, there is a private and a public component in each act, but exactly how does one tell which is which? Similarly, we may agree that individuals define and understand themselves in terms of socially produced cultural concepts, but those concepts were produced by individuals in their social interaction, and so we must also learn exactly to what extent it is

possible for individuals to go beyond a given culture. These two questions require:

a) a much more detailed account of the actual ways in which people are interdependent in society than Rousseau gives; and
b) a more precise account of the structure of the process of learning and communication through which the culture is transmitted, and in which it is created.

This refers us back to point three, which may be elaborated to include a reference to the relation of the person as subject to the person as both social and physical object. An investigation of these two points will enable us to achieve some clarity on the question of the possibility of a good society. If we know,

a) how the development of culture occurs in the dialectic between the individual and the social and material environment, and
b) the nature of the independent logic of change which the set of relationships between people develops, as Rousseau has pointed out, when these relationships are not consciously coordinated; then we should be able to say under what conditions, if any, unifying reason might develop unintendedly and make possible the good society.

5) Lastly, there is the problem of institutions. Rousseau's prayer for the legislator is no solution. Even if we reach the point at which reason becomes a factor in social organisation, it will still be necessary to design actual social institutions. Rousseau provides two important pointers, but nothing more.

The first pointer is to the crucial role of property institutions. In the *Discourse on Inequality* he traces the origin of all inequality, and the origin of government as an agent of oppression, to the institution of private property. In the *Social Contract* he does not pay much attention to property relations, except to make the vital point, against Locke and the theorists of laissez-faire capitalism, that "the right which each individual has to his own estate is always subordinate to the right which the community has over all" (SC 18). However, in his writings on Poland and Corsica he attacks not only private

property, but also money, which, as the means of trade, reinforces the division of labour and the dependence of individuals and thereby encourages tyranny. "It must be established as a sure maxim that wherever money is of prime necessity, the nation detaches itself from agriculture to fling itself into more lucrative professions. Those who enrich themselves by commerce and industry when they have enough money invest it in land which others cultivate for them; the whole nation is divided into rich loafers who own the land and unhappy peasants who do not have enough to live on while cultivating it ... Soon the countryside becomes depopulated and the towns overflow with vagabonds, little by little bread shortages occur, public misery increases together with the opulence of particular individuals, and the two together lead to all the vices which finally cause the ruin of the nation" (Corse OC3 920). He wishes therefore to encourage self-sufficient peasant farmers and, if exchange is necessary, to replace markets by government supervised barter. He believes, clearly regretfully (see OC3 931) that it is impossible to actually abolish private property, but wishes to enclose it within the strictest possible limits, "togive it a measure, a rule and a brake which contains it, directs it, subjugates it, and keeps it always subordinate to the public good" (OC3 931). This was to be done by creating a public domain to be worked in common on a corvée basis, and by setting strict limits to the maximum size of landholdings.

Although suggestive, both the analysis and the proferred solution are crude. His attack on money, although it presages later analyses of the role of capital accumulation in bringing about social evils, is still couched in terms of an economic theory which sees land as the only true source of wealth. He does not have the categories for dealing with industrial capitalist society (although it is worth noting that his description of competition in civil society in the *Discourse on Inequality* evidently had some influence on Adam Smith, who gave the first coherent account of the economics of capitalism, although Smith draws the opposite conclusions see Colletti, *From Rousseau to Lenin* 155 and 196-7). His solution to the problem of property is therefore also couched in terms of peasant society, but this does not detract from the principle of the attack on private property. Our

problem is to apply this attack to an analysis of the role of property in capitalist society and to draw conclusions from this about the sort of property relations required for a good society.

His second pointer is to the dynamic of corruption which he believes is inherent in institutions of government, but which nevertheless must be guarded against in all possible ways. The government, like any group of people, has its own "general will", its own corporate interests, which are particular in relation to the general will of the society as a whole. Moreover, the more active the government, the more able is it at the same time to impose its own particular will. The more passive the government the less can it impose its own will, but also the less can it carry out its duties, so the state is threatened on the one hand with dissolution from weak government, and on the other hand with tyranny from strong government. There is always a conflict of interest between government and people: "The citizen wills only the laws and the observation of the laws. Each individual amongst the people knows well that if there are exceptions they will not be in his favour. Thus all fear exceptions, and he who fears exceptions loves the law. Among leaders, it is quite a different matter: their very state is a state of preference, and they seek preference everywhere" (Montagne OC 3 893).

Thus for Rousseau there are two distinct problems of political institutions. The first is the problem of sovereignty, that is to know how the General Will may best find expression. To this he really has no answer but to stress the importance of an education in civic awareness, and to point out that inequality immediately destroys the possibility of the General Will emerging. It is clear that institutions which are to facilitate the expression of the General Will must be such as to favour free and rational debate, but what institutions will do that most effectively is another question.

The second is the problem of government. Sovereignty can never be delegated, but the executive authority can, and in any but the smallest state must be delegated. Rousseau believes that the best form of executive authority is what he calls elective aristocracy,

which we would call representative democracy. But even here he does not believe that we should trust governments. He believes, probably correctly, that there is no way in which we can adequately institutionalise mistrust of governments. But it is clear that once more equality in the society must play an important role: "Wherever wealth predominates ... apparent power is in the hands of the magistrates and real power is in the hands of the rich" (Corse OC3 939). A second, minor, point is the idea expressed in the essay on Poland that all administrative offices should be elective (OC3 1019). Finally he stresses the importance of continual public control over government. The problem is to work out how best this can be done in a more complex society than those envisaged by Rousseau.

B: KANT: TIME AND THE KINGDOM OF ENDS

I. INTRODUCTION

If Rousseau is sometimes unclear in his doctrine of the General Will, Kant is downright confused. But if his only achievement in political theory were to confuse the notion of the general will he would not be worth discussing. On the contrary he is important both for what his more developed confusions reveal about Rousseau's theory, and for the extent to which he at the same time manages to develop certain aspects of Rousseau's theory to a much higher level of philosophical clarity, and thereby to reveal more clearly both the advantages of the theory and some of the problems in it.

Kant begins his discussion of the state in the *Metaphysics of Morals* with what is clearly a conceptual definition: "A state (*civitas*) is a union of an aggregate of men under rightful laws. In so far as these laws are necessary *a priori* and follow automatically from concepts of external right in general (and are not just set up by statute), the form of the state will be that of a state in the absolute sense, i.e. as the idea of what a state ought to be according to pure principles of right. This idea can serve as an internal guide (*norma*) for every actual case where men unite to form a commonwealth" (PW 138). From this definition it is clear that he is using the term "state" to refer to an ideal model, and hence that whatever is predicated of the state in this sense cannot necessarily be applied to actual "states" as they exist in the world.

Next, he develops a model which involves a slight elaboration on Rousseau's theory. He distinguishes between the *ruling power* (or sovereignty) in the person of the legislator, the *executive power* in the person of the individual who governs in accordance with the law, and the *judicial power* (which allots to everyone what is his by law) in the person of the judge" (PW 138). He then concludes that "The legislative power can belong only to the united will of the people ... only the unanimous and combined will of everyone whereby each decides the same for all and all decide the same for each ... in other words, the central united will of the people ... can legislate" (PW 139). Thus he would appear to be following Rousseau in giving legislative power to the general

will as sovereign, and ascribing executive power to the prince, "the ruler of the state" (PW 141), who must remain subject to the laws made by the sovereign general will. But he takes two steps which make his theory much more conservative than that of Rousseau. The first is to redefine the general will in such a way as to effectively exlude all but property-owners. The second is suddenly to transmute his theoretical analysis into a positive description of existing states.

In defining the attributes of the members of the state who are to unite for the purpose of legislating he includes "the attribute of civil *independence* which allows him to owe his existence and sustenance not to the arbitrary will of anyone else among the people, but purely to his own rights and powers as a member of the commonwealth" (PW 139). He thus introduces a distinction between active voting citizens and passive voteless citizens. In his essay *On the common saying: 'This may be true in theory, but it does not apply in practice'*, he makes the same point in a little more detail. He specifies that a citizen "must be his *own master* ... and must have some *property* (which can include any skill, trade, fine art or science) to support himself" (PW 78). But he is referring here only to independent artisans who own their own tools, rather than wage workers, for he specifically excludes the wage-labourers "who allows someone else to make use of him" (PW 78 footnote). Interestingly, he presents virtually no argument to justify this, other than to say that the citizen "must in the true sense of the word *serve* no one but the commonwealth" (PW 78). This is a mere pun on the word "serve". Kant here utterly fails to transcend the perspective of his class and society.

This however is a relatively minor point compared with his second step. For, after setting out the principles of right, as above, he goes on to write: "Whether in fact an actual contract originally preceded their submission to the state's authority, whether the power came first and the law only appeared after it, or whether they ought to have followed this order ... these are completely futile arguments for a people which is already subject to civil law; and they constitute a menace to the state. For if the subject, having delved out the ultimate origin, were then to offer resistance to the authority currently

in power, he might by the laws of this authority (i.e. with complete justice) be punished, eliminated, or banished as an outlaw" (PW 143). Here, then, he is identifying the "laws of the authority currently in power" with "justice". And he can do this only because he suddenly and with no attempted justification whatever, begins to identify the legislative authority with the ruler, and speaks of the ruler as being the sovereign of the state, and, as the source of all laws, as being incapable of any unjust action (PW 145 note).

The characteristic of infallibility which Rousseau ascribed to the general will as a principle is here ascribed to a particular government. The result is effectively to place the ruler above the law. Consequently, Kant has a much more specifically anti-revolutionary doctrine: "All resistance against the supreme legislative power, all incitement of the subjects to violent expressions of discontent, all defiance which breaks out into rebellion, is the greatest and most punishable crime in a commonwealth, for it destroys its very foundations. This prohibition is *absolute*. And even if the power of the state or its agent, the head of state, has violated the original contract by authorising the government to act tyrannically, and has thereby, in the eyes of the subject, forfeited the right to legislate, the subject is still not entitled to offer counter-resistance. The reason for this is that the people, under an existing civil constitution, has no longer any right to judge how the constitution should be administered." (*Theory and Practice* PW 81.) The argument put forward to justify this last point is simply that there would not be any third party to judge in a dispute between government and citizens. Kant does not explore the possibility that this might imply that the final right of judgement should be left with the people, rather than with the government.

Although revolution is in no circumstances legitimate, Kant is quite consistent in arguing that "if a revolution has succeeded and a new constitution has been established, the unlawfulness of its origin and success cannot free the subjects from the obligation to accommodate themselves as good citizens to the new order of things, and they cannot refuse to obey in an honest way the authority now in power" (PW 147). This is a consistent conclusion because Kant has deprived himself of

any criterion with which to distinguish between legitimate and illegitimate government. The only way in which he can mitigate the harshness of his theory is by addressing moral injunctions to the monarch. He argues that the spirit of the original contract implies an obligation to make the mode of government conform to the original idea, and that any monarch therefore has the duty to alter the mode of government to coincide with this original intention. He specifies that although the citizen may not resist any act of government, nevertheless he or she must "with the approval of the ruler, be entitled to make public his opinion on whatever of the ruler's measures seem to him to constitute an injustice against the commonwealth" (PW 84). The argument underlying this is that we must always assume that the monarch acts in good faith, and hence that any apparent injustice can only be the result of error or ignorance of the possible consequences on the part of the ruler. Thus the ruler cannot both claim to be acting in the interests of the citizens and at the same time disallow independent and public thought. This is an important principle, and indicates Kant's democratic intent, but from his position he still has to stipulate that the citizen must act "with the approval of the ruler", thereby effectively destroying the principle.

That Kant was by no means simply a reactionary is also indicated by his attitude to the French Revolution. In *The Contest of Faculties*, published in 1798, after the terror, he writes of the revolution: "It may be so filled with misery and atrocities that no right-thinking man would ever decide to make the same experiment again at such a price, even if he could hope to carry it out successfully at the second attempt. But I maintain that this revolution has aroused in the hearts and desires of all spectators who are not themselves caught up in it a *sympathy* which borders almost on enthusiasm, although the very utterance of this sympathy was fraught with danger. It cannot therefore have been caused by anything other than a moral disposition within the human race" (PW 182). This is not merely a casual aside. It is the core of his argument for the improvability of the human race, which is based on the recognition of the revolution as a moral idealism designed to give a people the constitution which it wants. It is the existence of this moral impulse, and of spontaneous sympathy for it

amongst people unaffected by the advantages accruing from the new constitution, which demonstrates a drive towards improvement within the human race.

The ambivalence towards revolution can perhaps partly be accounted for, in the same way as his attitude towards property, as a manifestation of an inability to transcend his class perspective. But such an "explanation" is not in itself adequate. It is necessary also to trace the difficulties to their possible roots in the underlying theory. In fact it seems to me that Kant's problem of sovereignty and revolution is related to a more fundamental problem about the nature of human reality, a problem which manifests itself not only in his political theory, but also in his ethics, his theory of history, his epistemology, and his ontology.

II. FREEDOM, HAPPINESS AND VIRTUE

We have seen that for Rousseau the right of the state to coerce recalcitrant citizens to make them obey the general will is based on the principle that to be free is to act rationally, and that the general will is a component of the rational will of each individual. Kant develops this principle with great consistency, but also, as we shall see, on the basis of an anthropology which leads him to make a sharp distinction between virtue and happiness. It is this distinction which he uses as a final argument to justify the absolute right of the ruler. In the *Metaphysics of Morals* he writes that the welfare of the state "must not be understood as synonymous with the *well-being* and *happiness* of the citizens, for it may well be possible to attain these in a more convenient and desirable way within a state of nature (as Rousseau declares), or even under a despotic regime. On the contrary the welfare of the state should be seen as that condition in which the constitution most closely approximates to the principles of right; and reason, by *a categorical imperative*, obliges us to strive for its realisation" (PW 142-43). Here a clear distinction is made between the goal of human existence as dictated by reason, and the idea of welfare and happiness. In *Theory and Practice* Kant spells out the implications of this for the right to revolt, when he says that the error

of believing that there is such a right arises "in part from the usual fallacy of allowing the principle of happiness to influence the judgement wherever the principle of right is involved" (PW 82-83). That is, the right to revolt is usually based on the principle that despotic acts by the ruler cause the citizens to suffer in various ways, and that the avoidance of suffering is a basic reason for entering into the state, so that when the state is no longer performing that function, it is no longer legitimate. Kant can only reject this argument by suggesting that there is some principle other than the avoidance of suffering (or the pursuit of happiness) underlying the principle of right. To argue thus requires a particular anthropology in terms of which happiness can be judged as being only a secondary goal of human existence. Of course, once this is established, it does not immediately legitimise despotism, because it still remains to be shown that the principle itself is compatible with despotism, and in fact Kant's theory becomes contradictory, as we have seen, when he attempts to do this. But the important point is that the establishment of a principle other than happiness at least makes it possible to confuse the issue by removing the most obvious argument against the legitimacy of despotism.

Thus to understand the problems in Kant's political theory we need to look at his theory of ethics. In the *Groundwork of the Metaphysics of Morals* he begins by attempting to isolate the principle underlying our ordinary rational knowledge of morality. His main point here is that we usually are only willing to judge that an action is good if it is also disinterested. If the consequences of an action are desirable, we are naturally pleased that the action has occurred; when a large corporation makes a well-publicised donation to charity we welcome the donation, but we are also aware that the corporation's action is part of its regular image-building campaign, so we do not judge it good in moral terms. This does not mean that we necessarily judge it bad. Rather we may simply recognise it as a prudential action to which the categories of good and bad do not apply. Of course if the wider aim of the company is, say, to encourage the consumption of alcoholic drugs by its publicity campaign, we might judge the action to be bad, even though at another level it had certain immediate

consequences which were desirable. That is, the act itself cannot be judged to be good unless we relate it to its motive. Thus Kant asserts "It is impossible to conceive anything at all in the world, or even out of it, which can be taken as good without qualification, except a *good will*" (ML 61). We need to distinguish between actions done from inclination and actions done from duty, that is purely from a sense that the action *ought* to be done, irrespective of its consequences for the actor. To understand the nature and weight of this "ought" we need to explicate further what is implied by the concept "duty". We can derive from every action an "ought" statement from which it follows. This ought statement is what Kant calls the subjective principle of the volition, or the "maxim" in terms of which one is acting. He therefore reformulates the principle of the good will as follows: "An action done from duty has its moral worth, *not in the purpose* to be attained by it, but in the maxim in accordance with which it is decided upon; it depends therefore, not on the realisation of the object of the action, but solely on the *principle of volition* in accordance with which, irrespective of all objects of the faculty of desire, the action has been performed" (ML 67-68).

The distinction between the maxim and the purpose is the distinction that we have already made; it is the distinction between willing a desirable end, and willing it for moral reasons. However, the statement that the morality of the action does not depend on the realisation of its object has given rise to some misunderstanding. Kant is not saying that it does not matter what you do, as long as your will is good. For him the will is practical, i.e. active, and must give rise to acts in the world. These acts, further, must be consistent with the will. That is, if you will something, you must take every possible step to ensure that the result is achieved. If you do not do this, then you have not really willed that result. But nevertheless if, after having taken all necessary steps you still do not achieve the goal, this does not mean that we refuse to describe your maxim as good. The *genuineness* of your maxim can only be assessed by the extent to which you attempt to carry it out, but its *goodness* does not depend on your success in carrying it out.

On what, then, does its goodness depend? Since we cannot judge the morality of the maxim in terms of its specific end, in terms of what are we to judge it? Kant's answer is this: "Since I have robbed the will of every inducement that might arise for it as a consequence of obeying any particular law, nothing is left but the conformity of actions to universal law as such, and this alone must serve the will as its principle. That is to say, I ought never to act except in such a way *that I can also will that my maxim should become a universal law* (ML 70). The maxim is not legitimised by the particular purpose which it postulates. Rather, the purpose is legitimised by the form of the maxim. This follows logically from the argument that there can be no particular event which we can declare without qualification to be good. Thus if we are to use the moral concepts "good" and "dutiful" we cannot derive them from the absolute value inherent in some event. At this point in the argument Kant is not arguing that we can legitimately use such concepts. He is only trying to explicate what is implied when we do use them. To say that the maxim must be universalisable is to say that it must have no relation to my own particular interests, hence, that it must be disinterested.

But this principle needs considerable further elaboration before we can fully understand what is meant by the idea that a maxim should be a possible universal law. The above argument is first reformulated by Kant in terms of the concept of the Categorical Imperative. The characteristic of a rational being is the ability to act in accordance with the idea of a law. This is what is meant by the idea that there is a maxim underlying every action. The action is not merely a physical reaction to a situation. It involves a comprehension of the situation, and a judgement of the situation which, in practical terms becomes an action. For a rational being there are always different possible ways of acting in a particular situation. The actual act therefore involves a particular judgement of the situation. This judgement is usually implicit, but, because it is a judgement, it can be embodied in a formula, and this formula does not only refer to the situation in question. It is of such a nature that it could be applied to other similar situations. It is in this sense that it is a maxim.

Such a maxim is an Imperative when it embodies a principle which is necessitating for a will (cf. ML 81). Kant distinguishes between two kinds of imperative:

1) A hypothetical Imperative spells out the maxim of action which is to be adopted if a particular goal is to be reached. In this case the goal itself, and hence the act prescribed by the categorical imperative, remains problematic. Within the category of hypothetical imperatives Kant makes a further distinction which slightly confuses the issue. This is the distinction between a problematic practical principle, which relates to a possible purpose, and an assertoric practical principle, which relates to an actual purpose. Ths distinction is in itself trivial and unexceptionable, but Kant uses it to introduce "one end that can be presupposed as actual in all rational beings" ... the purpose of happiness, "which we can assume with certainty that they all *do* have by a natural necessity" (ML 83). However, this should not be interpreted to mean that all rational beings (including humans in so far as they are rational) act by a natural necessity in such a way as to achieve their own happiness. If this were the case, then we would have to abandon the concept of duty immediately. The Imperative "So act as always to maximise your own happiness" would be such as to rule out the possibility of any other categorical imperative. At this point in the argument it is not at all clear what Kant does mean when he speaks of a natural necessity, however. He complicates the issue even further by writing "unfortunately, however, the concept of happiness is so indeterminate a concept that although every man wants to attain happiness, he can never say definitely and in unison with himself what it really is that he wants and wills" (ML 85). Such a comment immediately makes one suspect that the concept is purely analytical referring only to the satisfactory achievement of goals, but not specifying any gaols as such. However, Kant then goes on to say that all the elements which belong to the concept happiness are empirical, but that the problem of indetermination lies in the fact that it is not possible to put these elements together in such a way as to specify the desired totality of happiness. This admission is, I think, sufficient to render problematic the whole concept of happiness as used by Kant, and I shall return to this point later.

2) The second kind of imperative is the Categorical Imperative. If a hypothetical imperative specifies the means without commanding the goal, then the Categorical Imperative may be defined as one which commands simultaneously the end and the means. But, as we have seen, there is no particular purpose which can be considered to be good without qualification, and hence which can be categorically commanded. This raises two questions. What could be the content of such a categorical imperative; and what would make it valid? The answer to the first question is this: "If I conceive a *categorical* imperative, I know at once what it contains. For since besides the law this imperative contains only the necessity that our maxim should conform to this law, while the law, as we have seen, contains no condition to limit it, there remains nothing over to which the maxim has to conform except the universality of a law as such; and it is this conformity alone that the imperative properly asserts to be necessary" (ML 88). This is a reformulation of the principle of universality. Previously it was derived from the principle of disinterested duty. Here it is derived from an imperative which commands simultaneously both an end and a means. In this case the end and the means must, formally at least, coincide. This can only be the case if the end is conformity to law as specified in the imperative itself.

Now this argument may appear to be tantologous, and it is certainly on such a high level of abstraction that it appears to be very difficult to grasp. For the moment let us remember that Kant is investigating only the necessary structure of an imperative which would be categorical. He is not arguing that there is such an imperative that we ought to obey. An imperative may be said always to contain both an instruction to act in a particular way, and a reason for acting in that way. In a hypothetical imperative the reason is given by the purpose which is postulated. Thus the hypothetical imperative: "You must catch this train if you want to get to work on time" contains the instruction "Catch this train", and the explanation "So as to get to work on time". A Categorical Imperative could not have an explanation of this type, because the goal of getting to work on time remains itself problematic. That is, the reason given implicitly or explicitly within the categorical imperative would have to be such that it did not refer

to some other end beyond itself. It would not be dependent on some further purpose in the way in which getting to work on time as a goal depends on the further purpose of keeping my job. Thus the relation between instruction and explanation in a Categorical Imperative would simply be: "Act thus, because you must (ought to?)". This is an unconditional injunction to obey the law because it is the law. A law which is obeyed because it is the law is necessarily universal; it contains no reference to extrinsic purposes, and hence no reference to situations in which it would be legitimate not to obey it. In so far as duty is similarly independent of extrinsic purpose, if it is to be duty rather than inclination, it follows that the principle of duty corresponds to the categorical imperative: "Act only on that maxim through which you can at the same time will that it should become a universal law" (ML 88).

The main problem in interpreting the Categorical Imperative is that of establishing criteria for what would count as a universal law. Kant explicates this partly by presenting a number of alternative formulations of the law, and partly by offering examples of the application of the law.

In examining these formulations and examples we shall discover that the apparently arbitrary and harsh form of the Categorical Imperative conceals a rich concept of human purpose. The first formulation is "Act as if the maxim of your action were to become through your will a universal law of nature" (ML 89). By "nature" here Kant means a complete system of laws. That is, he is saying that an act conforms to the Categorical Imperative if I could will the maxim which it embodies as one element in a complete system of law governing my behaviour. Why have we moved from the concept of a single universal law to the concept of a system of laws? The answer would seem to be that it is always a question of whether a law could be universally willed as a law *for me* as a rational being. That is, it must always be considered as one of a set of laws governing the modes of behaviour of a complex rational being. If the law were, *per impossibile*, to be considered in isolation, there would be no criteria with which to judge whether it could be universal or not. The second reformulation of the

law brings out more clearly the implication of the fact that it is the individual who must be able to will the maxim as a universal law. Any act is performed for some purpose. We have seen that in terms of the original categorical imperative the purpose of obeying the law is simply that it is the law. But this needs further explication. It implies that a certain kind of activity, law-governed activity by a rational being, is an end in itself. This means that the rational being, in so far as it is defined by its capacity for law-governed activity, is an end in itself. Thus the second formulation of the Categorical Imperative is: "Act 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" (ML 96).

From the principle that the subject of all ends is to be found in every rational being as an end in himself it follows that a rational being, when acting in conformity to law, is making the law. That is, what specifies a rational being is the capacity to discover and act in accordance with universal rational principles. The relation of the rational being to the law is not one of subordination and superordination. It is a free relation of discovery. Hence we may say that the will of every rational being is a will which makes universal law, in the sense that by its will it turns a rational principle into a maxim for conduct. The law as a maxim has no hold over the will other than the will's own affirmation of it as a law, and this affirmation comes from the will's awareness of its own status as a rational being which is an end in itself. To recognise oneself qua rational being as an end-in-itself is to will universal law. Thus such a rational being is autonomous, as law-maker, and act free from particular interest as a maker of universal law. This specifies once more that a good will, acting out of duty, acts free from personal interest. A will which acts out of interest is determined to its act by the interest, and this destroys its autonomy.

The idea of a law of nature describes the form of the Categorical Imperative. The idea of the end-in-itself specifies the matter or purpose. Both are made explicit in their relation to one another in Kant's final formulation: "All maxims as proceeding from

our own making of law ought to harmonise with a possible kingdom of ends as a kingdom of nature" (ML 104). By a kingdom here is meant "a systematic union of different rational beings under common laws" (ML 100). So any act, if it is to be done out of duty, must be an act such that one could accept the maxim underlying it as a natural law in a system of laws governing a community of which one is oneself a member. It is evident from this formulation that Kant has here explicated what is implied in Rousseau's concept of the General Will as a principle. But so far we have a clearly stated principle, but no criteria for its application. In the *Grundlegung* Kant considers a number of examples of actions which might or might not be compatible with the Categorical Imperative, and in so doing he spells out the main criteria.

Here he distinguishes between two types of incompatible principle. The first are principles which cannot even be conceived as universal laws. For example, the maxim "Whenever I believe myself short of money, I will borrow money and promise to pay it back, though I know that this will never be done" (ML 90) cannot become a universal law since the notion of borrowing requires trust, and the principle, by destroying the possibility of trust, would make it impossible for anybody to borrow money at all. Now this is not a purely formal logical contradiction. It is an incompatibility between ends and means, in the sense that the maxim is one which I could not will while maintaining the goal of a kingdom of ends. (See *Murphy* 70). To put it another way, the reason that I cannot will it is not that it is in some way a logical contradiction, but that it is a non-viable practical law, and the criterion for a viable law is that provided by the notion of the kingdom of ends.

The second form of maxim brings out more clearly the notion of purpose as a criterion. This is the maxim which can be conceived, but cannot be willed. The clearest example of what is involved here is contained in the question: "Does my maxim of neglecting my natural gifts, besides agreeing in itself with my tendency to indulgence, agree also with what is called duty?" (ML 90). His answer is that a system of nature could subsist under these conditions, but that a rational

being could not possibly will such a system of nature, "For as a rational being he necessarily wills that all his powers should be developed, since they serve him, and are given him, for all sorts of ends" (ML 90). Thus Kant's general criterion for a moral maxim is that it should promote the full development of all the powers and capacities of rational beings. This is important from two points of view. Firstly, it indicates that the Categorical Imperative is not purely formal, but requires us to take into account the specific nature of a specific community of rational beings. Secondly, it introduces the idea of historical development and even hints at the possibility of the historical development of reason itself, in so far as reason is the most important of the powers of a rational being. Kant returns to this theme in his writings on history, which I shall discuss shortly.

Up till now Kant has been explicating what is implied in the concept of a Categorical Imperative. He has not yet tried to answer the question as to whether such an imperative is possible. In his discussion of the different kinds of imperative he points out that as far as hypothetical imperatives are concerned, once the end is given, the necessity of willing the means follows analytically. That is, once the particular end has been chosen, it immediately becomes imperative to will the means. But this is not the case with the Categorical Imperative, since this is not based on an externally given end. Also, since the Categorical Imperative presents itself as necessary, it cannot be derived in any way from experience. To say that my will is subject to the Categorical Imperative, and therefore that I ought to adopt the Categorical Imperative, is to assert something which cannot be derived purely analytically from the notion of will. The Categorical Imperative is therefore on the one hand *a priori*, in so far as it is not learnt from experience, and on the other hand synthetic, in that it cannot be deduced merely from the definition of the will. This means that to validate the categorical imperative we have to discover some third idea which provides a necessary connection between it and the will. Only if we can do this can we say that my will ought necessarily to be subject to the Categorical Imperative.

Kant's answer to this problem is that what provides the link is the idea of freedom. He defines this as follows: "Will is a kind of causality belonging to living beings so far as they are rational. *Freedom* would then be the property this causality has of being able to work independently of *determination* by alien causes" (ML 114). That is, a free will is a will which gives itself its own laws, or its own maxims of action. A free will is therefore autonomous, and, as we have seen, to be autonomous means to be subject to the Categorical Imperative, in so far as it involves treating oneself and every rational being as a maker of universal law. An autonomous will is the opposite of a heteronomous will, which is subject to determination from inclinations imposing themselves on it from outside. A heteronomous will is a will which is unaware of the possibility of choice, and is subject to external ends. However, once the possibility of choice arises, that is, once the will distances itself from particular given ends and places them in question, we are faced with the problems of what ends ought to be chosen. This is solved by the discovery of something which is an end in itself, that is, the rational being itself, and through it the community of rational beings organised in a kingdom of ends. Thus all free beings are subject to the Categorical Imperative.

Our last problem is to establish whether a rational being is also a free being. This question Kant believes cannot be answered completely on a theoretical level, but it can be answered in a way which is adequate for our purposes: "Now I assert that every being who cannot act except *under the Idea of freedom* is by this alone ... from a practical point of view ... really free" (ML 115). The reason for this is that to act under the Idea of freedom, or under the assumption that one is free, is to be faced with ought questions. To believe that I am free is to say: I am faced with alternatives- which ought I to choose? This question immediately raises the problem of ends, and can therefore only be finally answered by the discovery of an end in itself as embodied in the Categorical Imperative. Thus our question is simplified. In order to conclude that rational beings are subject to the Categorical Imperative we do not have to establish whether they are free, but only whether they act under the idea of freedom. The answer is that "we cannot possibly conceive of a reason as being consciously directed

from outside in regard to its judgements; for in that case the subject would attribute the determination of his power of judgement, not to his reason, but to an impulsion. Reason must look upon itself as the author of its own principles independently of alien influences" (ML 116). That is, underlying the adherence of any rational being to its own judgements is the assumption that these judgements are determined by the rational being itself, and not influenced by external impulsions. A rational being is a being who will only accept judgements which have been arrived at by a process of reasoning which is autonomous in the sense that each step in the argument is intrinsic to the argument, and not validated by some unanalysed prejudice imposed on the argument in some way. So a rational being necessarily acts under the idea of its own freedom.

This argument is sufficient to validate the Categorical Imperative, but it is not sufficient to fully comprehend it. To do this we would have to be able to give a positive account of freedom, going beyond the mere assertion that we are unable to act other than under the presupposition of freedom. This will lead us into Kant's epistemology and ontology, but before following this path I would like to examine one of the central problems in Kant's ethical theory, a problem already referred to in discussing the right of rebellion, but only touched on in passing in the exposition of the ethics. This is the problem of the relation between duty and happiness.

Kant discusses this relationship more fully in *The Critique of Practical Reason*, where he offers the following set of definitions.

"*Life* is the faculty of a being by which it acts according to the laws of the faculty of desire. The *faculty of desire* is the faculty such a being has of causing, through its ideas, the reality of the objects of these ideas. *Pleasure* is the idea of the agreement of an object or an action with the *subjective* conditions of life, i.e. with the faculty through which an idea causes the reality of its object" (CPrR 9 note 7). "Pleasure from the conception of the existence of a thing, in so far as it is a determining ground of the desire for this thing, is based upon the susceptibility of the subject because it depends upon the actual

presence of an object. Thus it belongs to sense (feeling) and not to the understanding, which expresses a relation of a conception to an object by concepts, and not the relation of a conception to the subject by feelings. It is only practical in so far as the faculty of desire is determined by the sensation of agreeableness which the subject expects from the actual existence of the object. Now a rational being's consciousness of the agreeableness of life which without interruption accompanies his whole existence is happiness, and to make this the supreme ground for the determination of choice constitutes the principle of self-love" (CPrR 20).

"The principle of one's own happiness, however much reason and understanding may be used in it, contains no other determinants for the will than those that belong to the lower faculty of desire. Either, then, no higher faculty of desire exists, or else pure reason alone must of itself be practical, i.e. it must be able to determine the will by the mere form of the practical rule without presupposing any feeling or consequently any idea of the pleasant or the unpleasant as the matter of the faculty of desire and as the empirical condition of its principles" (CPpR 23).

"To be happy is necessarily the desire of every rational but finite being, and thus it is an unavoidable determinant of its faculty of desire" (CPrR 24).

Thus Kant describes happiness as the consciousness of the sensation of agreeableness which accompanies the satisfaction of one's desires. Happiness is enduring pleasure. But this pleasure must not be thought of solely in physical terms. Kant includes among his example of pleasures an instructive book, a fine speech, an intellectual conversation, and an act of charity (although it is perhaps unintentionally revealing that in setting out hypothetical cases of comparing pleasures with one another, his examples contrast each of the above pleasures with more "physical" pleasures such as hunting, eating, gambling and going to the theatre). The point that he is making is that the satisfaction of a desire causes a feeling of agreeableness, and that we choose between various possible objects of desire in terms of an evaluation of the amount or intensity of the accompanying agreeable feeling. It is in this sense that all these possible pleasures

are pleasures of the "lower faculty of desire". Since happiness depends on an internal feeling, there can be no guarantee that what produces such a feeling in me will at the same time produce this feeling in other people. If my happiness does coincide with the happiness of the others this can only be accidental. It is for this reason that duty, as the principle of my relation to others, cannot be derived from happiness, as the principle of my relation to myself. This relation to myself is determined by "my finite nature as a being of needs" (CPrR 24), and it is these needs, apparently, which come between my happiness and the other's happiness.

Now, if we assume that my happiness is related to the satisfaction of my needs and desires, this has one of three possible consequences for the happiness of other people. Either my needs are necessarily incompatible with the needs of other people, or the relation between them is undetermined, and potentially either compatible or incompatible, depending on the circumstances, or else my needs are essentially compatible with those of others. Kant assumes the second possibility. But is this a legitimate assumption? In one sense it is, if we are dealing with the felt needs of people as exhibited in their acts in a particular society. But Kant is not only doing this. He is also implicitly attempting to investigate the underlying essence of human behaviour. And, precisely, he shows that this underlying essence can only be fully realised in a kingdom of ends. It would seem to be meaningful, in terms of Kant's theory, to say that human beings have a need for autonomy, and that this need can only be satisfied in a kingdom of ends, and thus not only is this need compatible with the needs of others, but it actually requires the satisfaction of their needs for its satisfaction. Kant seems to be at least partly aware of this in his various discussions of the motivation for acting dutifully. He distinguishes here between acting out of interest, and taking an interest in something. Although no interest impels me to obey the categorical imperative, nevertheless, if I did not take an interest in it, I would not obey it at all. Kant argues that the effect on me of the moral law is to make me doubt the validity of my sensual (pathological) motives, and to question my self-conceit. I experience this humiliation of my self-conceit as respect for the law. This respect

is a feeling, but it is not one of those feelings of the lower faculty of desire related to happiness. It is a purely intellectual moral feeling (cf. CPrR 77-79).

To obey the moral law is to affirm oneself as an end in itself, and to do this is to assert one's dignity: "The law-making which determines all value must for this reason have a dignity ... that is, an unconditioned and incomparable worth ... for the appreciation of which as necessarily given by a rational being, the word *'reverence'* is the only becoming expression" (ML 103). This is an awareness of oneself as free, and hence as connected with an order of things beyond the world of sense and subject to one's own personality, not to the mechanism of nature (cf. CPrR 89). In being conscious of my freedom in obeying the moral law I experience an intellectual contentment, a contentment beyond that which I can experience in that sensuous contentment which rests in the satisfaction of the inclinations. Although this feeling is not complete bliss, it is at least analagous to the blissful self-suffiency which can be ascribed only to the supreme being (cf. CPrR 123). Nevertheless, Kant insists that this cannot be called happiness. However, it is clear from the above that this feeling of self-contentment is in some ways at least envisaged as being superior to mere sensual happiness. Why, then, does Kant insist so strongly on the dichotomy between happiness and virtue? If there is some feeling in some ways superior to happiness this is merely a definitional accident. It results from the fact that he has defined happiness in essential relation to the satisfaction of inclinations. By using the word in this way he has implied a dualism in human motivation which is perhaps not there at all. But the fact that he has done this is not itself the result of an accident of definition. It is a necessary result of his own epistemology. For within his theory it is not possible to make sense of the relation between the moral law as a rational principle and the particular feelings of a material human being: "It is, however, wholly impossible to comprehend ... that is, to make intelligible *a priori* ... how a mere thought containing nothing sensible in itself can bring about a sensation of pleasure or displeasure" (ML 128). Because he cannot theoretically link the two, he is obliged to maintain an independent theory of happiness based only on an account of humans as

sensible beings.

This leads into further problems. Kant postulates the concept of the highest good, the combination of virtue and happiness, as the necessary highest end of a morally determined will (determined will) and a true object thereof. This follows from the idea of the kingdom of ends, which implies the idea of the full satisfaction of needs, as well as the idea of virtue. This coincidence would only be certain either if virtue produced happiness, or if happiness were identical with virtue. We have seen that neither of these is the case. There is no guarantee that the virtuous person will also be happy, since happiness as the satisfaction of desires requires the concordance of the natural world, and virtue itself cannot ensure this.

Nevertheless this concordance seems to be necessary to preserve the status of the moral law. If that which is commanded by the moral law, the search for the highest good in the kingdom of ends, is in fact impossible, then the moral law loses its force. Since there is no direct relation between virtue and happiness, the only other possibility is that there might be an indirect relation "mediated by an intelligible author of nature" (CPrR 119). Are we entitled to postulate the existence of such a mediation?

We certainly have no empirical grounds for doing so. But, Kant argues, if we examine the nature of ethical theory, of reason in so far as it is "practical", in the sense of providing principles of practice, we should be able to provide an alternative source of knowledge. The idea of freedom cannot be empirically established either. We can never show that a particular act was undetermined. Instead, we establish the idea of freedom on the basis of our moral experience, by showing that we cannot act excep- under the idea of freedom. The affirmation of freedom then, is a "postulate of pure practical reason", that is, "a theoretical proposition which is not as such demonstrable, but which is an inseparable corollary of an *a priori* unconditionally valid practical law" (CPrR 127). Kant argues that there are two further postulates of pure practical reason which have to be accepted as implied by the Categorical Imperativel Firstly, virtue, as the complete rational

control of self in obedience to the moral law, is in fact not attainable by a rational being in the world of sense. Its possibility can therefore only be conceived as the result of an endless progression in moral development. That is, the possibility of fully rational behaviour is not simply given. It has to be achieved in an historical process. But this is only possible for the individual on the supposition of immortality. Secondly, the idea of the coincidence of virtue and happiness requires the postulate of a will which both wills the moral law and is at the same time the supreme cause of nature. This is God(cf. CPrR 129-30). The introduction of God transmutes the Kingdom of Ends into the Kingdom of God, "a world wherein reasonable beings singlemindedly devote themselves to the moral law ... (and) in which nature and morality come into a harmony" (CPrR 133). It is this that legitimises the distinction between the wellbeing of the citizens and the duty of the citizens to obey the law in actual societies.

Rousseau's society in which the general will is actually embodied is explicated by Kant in his account of the kingdom of ends, in which the freedom of each individual is expressed in the legislative acts of each will, and in which treating each other as an end is seen to be the prerequisite for one's own freedom. But, like Rousseau, Kant can offer no convincing account of how such a society might be possible. Rousseau's legislator becomes Kant's God. The possibility of actual historical development becomes the promise of immortality and an ideal historical development. Yet Kant does also have a theory of real historical development, and before examining the reasons for his failure, it is worth considering this theory, which is set out in the essay *"Idea for a Universal History with a Cosmopolitan Purpose"*.

His aim in the essay is to establish whether it might be possible to discover a principle which could be used to uncover any direction in human history. He first distinguishes between

a) behaviour which is purely instinctive and completely governed by natural laws, and
b) purely rational action in accordance with a pre-arranged plan.

Actual human history conforms to neither of these models, "for we find that, despite the apparent wisdom of individual actions here

and there, everything as a whole is made up of folly and childish vanity, and often of childish malice and destructiveness" (PW 42). Thus, if there is some purpose in human history, it is certainly not a conscious human purpose. But this leaves open the further possibility that there might be some hidden purpose in nature itself which manifests itself in obscure ways in this apparent confusion, and it is this possibility that Kant wishes to investigate.

He does so by making the assumption that nature can only be understood teleologically. The problem, therefore, is to discover what is the purpose for man, and what is the mechanism whereby that purpose is achieved. If we are to understand nature teleologically, then we must accept that no being has a capacity which it does not need, and that therefore any capacity which it possesses in an undeveloped form is destined to be fully developed in the course of time. The main such human capacity is reason. Reason does not work instinctively, neither is it developed instinctively: "(It) requires trial, practice and instruction to enable it to progress gradually from one stage of insight to the next" (PW 42). For this reason, reason can only come to full development in the species, not in the individual; no individual would have the time or the range of experience to develop on his own the full panoply of reason. Since humans are guided by reason, rather than by instinct, it is clear that nature's purpose for man is not merely happiness, which could most easily be achieved by the appropriate instincts. The purpose must be a sense of autonomy, responsibility, or, as Kant puts it, self-esteem.

The method chosen by nature for the development of human reason, and for the establishment of the necessary material substrate for happiness and well being, is antagonism between individuals in society: "the *unsocial-sociability* of men, that is, their tendency to come together in society, coupled, however, with a continual resistance, which constantly threatens to break this society up" (PW 44). This tendency is rooted in human nature, in the fact that humans need to live in society, but that at the same time they all wish to continue to live as individuals, to impose their own wills on others, and to acquire the status in the society which is required to do this. It is

in this stru-gle that human capacities are developed, and that culture replaces barbarism. Through this development " a *pathologically* enforced social union is transformed into a *moral* whole" (PW 45). Thus humans who would naturally prefe to remain as close to the comfortable instinctual level as possible, instead find themselves forced to develop their talents.

The end-point of this process requires the establishment of a society embodying justice and freedom. This can only be the last and most difficult task of the human race. Man, through his animal inclinations, remains an animal, but, as also a potentially rational being, an animal who needs, and knows he needs, a master. But the master, the highest authority in civil society, is himself a man, and so in turn also an animal who needs a master. A further, and perhaps equally difficult prerequisite for this establishment of a perfect civil constitution, is the bringing about of peace between states. Once more, the general threat of war, arising out of the unsociableness of men, is such as to encourage moves towards peace.

At present we are still far from this end-point: "We are *cultivated* to a high degree by art and science. We are *civilised* to the point of excess in all kinds of social courtesies and proprieties. But we are still a long way from the point at which we could consider ourselves *morally* mature" (PW 49). Nevertheless there is a certain amount of evidence that we are at least moving in the direction of a genuinely cosmopolitan society. Trade between states, and industry within states, upon which each state depends for its power, have reached a stage where states are interdependent, and where despotism within a state is likely to be disadvantageous to the economy: "If the citizen is deterred from seeking his personal welfare in any way he chooses which is consistent with the freedom of others, the vitality of business in general and hence also the strength of the whole are held in check" (PW 50). If, therefore, we consider the passage of history up to the present we shall both be able to recognise a progression, and also be able better to understand what is required for further development. Examining each state, "we should observe how their inherent defects led to their overthrow, but in such a way that a germ of enlightenment

always survived, developing further with each revolution, and prepared the way for a subsequent higher level of improvement" (PW 52).

Kant here clearly adopts and develops the Rousseauean idea of the historical development of reason. He combines in this analysis teleological assumptions with at least some empirical remarks, particularly about the role of economic relations as expressed in unsocial-sociability and in trade relations between nations, and also suggests, in the last quotation, a method of investigating particular historical events in such a way as to illuminate the development of reason within historical conflict and change. The central idea is of a pattern which works itself out in human history through human agency but without human design. Kant interprets this in terms of a teleology of nature. Later theorists will accept the central idea, but attempt to account for it in other terms.

III. SENSIBLE VERSUS INTELLIGIBLE

We need to discover why Kant relied on the concept of natural teleology. This question is bound up with his epistemology. The problems associated with it can be illustrated firstly by examining the relation between the treatment of reason in the ethics and in the theory of history. In the theory of history, rational behaviour is the end-point of human history. Fully-developed humanrreason does not occur before this end-point has been reached. From this it follows that the Categorical Imperative, as a rational affirmation of the rational actor as an end in itself, can only be understood by human beings at the end of the historical process. To be more precise, the Categorical Imperative can only be grasped at the point at which the kingdom of ends has become possible. In this case, the argument for the existence of God and for the immortality of the soul fall away. Both relied on the impirical impossibility of the kingdom of ends; they were introduced in order to provide another kind of possibility for it, and therefore to make it legitimate to will the supreme good. But in terms of the new theory the task of bringing about the kingdom of ends can only be formulated at a point in time when the hidden purpose of nature has already developed human culture to the point where it is possible "by our own

rational projects (to) accelerate the coming of this period which will be so welcome to our descendants" (PW 50).

This contradiction in Kant's work is related to the inability, discussed earlier, to give a coherent account of the relation between sensible happiness and moral contentment. In both cases Kant shows himself incapable of adequately synthesising human rationality (whether as a potentiality or as a developed capacity) and human materiality. But what is significant is that he recognises that there is a problem, and by the ways in which he wrestles with it both clarifies the nature of the problem, and also indicates the direction in which a solution is to be found.

Let us begin by considering the way in which he formulates this problem in the *Grundlegung*. We have seen that he there argues that in order to validate the Categorical Imperative we only have to be able to show that a rational being necessarily acts under the idea of freedom, and that we cannot in fact give a positive account of the nature of freedom. We cannot say how it is that human beings, as material beings subject to the laws of nature, can be free. But he nevertheless argues that we must at least be able to show that it is legitimate to conceive of ourselves as free. And, he suggests, we can only do so by means of the following distinction: "One observation is possible without any need for subtle reflection and, we may assume, can be made by the most ordinary intelligence ... The observation is this .. that all ideas coming to us apart from our own volition (as do those of the senses) enable us to know objects only as they affect ourselves: what they may be in themselves remains unknown. Consequently, ideas of this kind, even with the greatest effort of attention and clarification brought to bear by understanding, serve only for knowledge of *appearances*, never of *things in themselves*. Once this distinction has been made (it may be merely by noting the difference between ideas given to us from without, we ourselves being passive, and those which we produce entirely from ourselves, and so manifest our own activity), it follows of itself that behind appearances we must admit and assume something else which is not appearance - namely, things in themselves - although, since we can never be acquainted with

these, but only with the way in which they affect us, we must resign ourselves to the fact that we can never get any nearer to them, and can never know what they are in themselves. This must yield us a distinction, however rough, between the *sensible world* and the *intelligible world*, the first of which can vary a great deal according to differences of sensibility in sundry observers, while the second, which is its ground, always remains the same. Even as regards himself - so far as man is acquainted with himself by inner sesation - he cannot claim to know what he is in himself ... beyond this character of himself as a subject made up, as it is, of mere appearances he must suppose there to be something else which is its ground - namely, his ego as this may be constituted in itself; and thus as regards mere perception and the capacity for recieving sensations he must count himself as belonging to the *sensible world*, but as regards whatever there may be in him of pure activity (whatever comes into consciousness, not through affection of the senses but immediately) he must count himself as belonging to the *intellectual world*, of which, however, he knows nothing further" (ML 118-19). In this passage, Kant sums up, in a rough way, the main conclusion of the *Critique of Pure Reason*. He establishes a basic dichotomy between the sensible world and the intellectual world, and hence also between the "sensible" I and the "intellectual" I. It is this dichotomy, as we have seen, which is at the root of his problems about the nature of happiness and of the good society.

The passage brings out very clearly the unquestioned assumption on which Kant bases his theory. This assumption is a theory of perception. The object "affects" the perceiving subject, and this affection is experienced by the subject as the passive reception of a sensation. The sensation is caused by the object, but it is not 'of' the object, in the sense of in some way representing the object as it is in itself. Now this theory of perception, which we may call the "common sense-scientific" point of view, treats the perceiving subject as an object among other objects in a causal network. The 'subject-object' reacts to the other objects in nature in terms of causal laws reflecting its own nature as a receiver and the other objects nature's as emitters of signals. There is no room for the concept of 'freedom' in this model, and determinist philosophers have used such a model to

deny human freedom. In the passage here Kant merely asserts that the model ignores our own experience of ourselves as active, and argues that this aspect of ourselves is sufficient to justify the claim that in so far as we are also 'things in themselves' we are pure activity. This argument obviously begs a lot of questions, and it is these questions which Kant had earlier attempted to resolve in the *Critique of Pure Reason*.

In the *Critique* he shows that no part of human experience can be accounted for purely in terms of the model of perception which I have elaborated. But instead of rejecting the model outright, he believes that he can give an account of human experience by adding certain additional features to the model. The model itself, which I shall describe as the sense-datum theory, remains unquestioned. Indeed it is so taken for granted that Kant does not consider it necessary to argue it or even to state it in detail. He begins his discussion of perception with the following definitions: Intuition "is only possible, to man at least, in so far as the mind is affected in a certain way. The capacity (receptivity) for receiving representations through the mode in which we are affected by objects is entitled *sensibility*" (KS 65). The nature of this sensibility is not spelled out here, but scattered references throughout the text do give an idea of how Kant conceived it to be. Sensation is "the sequence of one impression upon another", each being a unity "contained in a single moment" (KS 131). That is, "different perceptions ... occur in the mind separately and singly" (KS 144). Each sense-datum is discrete. It contains in itself no reference to any other such datum. Sensation involves a sequence of such discrete data. Any actual "manifold of intuition", however, is a collection of such sense data, rather than only a single datum. In this collection the sense data are conjoined or related to one another in some way. But the conjunction itself is not an additional datum of the senses. In themselves the sense-data are a mere "rhapsody of perceptions" (KW 193). Further "In experience, however, perceptions come together only in an accidental order, so that no necessity determining their connection is or can be revealed in the perceptions themselves. For apprehension is only a placing together of the manifold of empirical intuition; and we can find in it no representation of

any necessity which determines the appearances thus combined to have connected existence in space and time" (KS 209). That is, there is nothing in any of the discrete sense-data which requires us to order them in any specific way. Lacking any relation to any other sense-datum the manner in which the sense-data are to be put together to form a picture is undetermined. If we wish to relate these sense-data to one another we are apparently not subject to any necessity in the way in which we do so; we can put them together any way in which we choose.

At first glance this theory looks very odd, but in fact it is logically implied by the "commonsense-scientific" point of view. It was also one of the dominant theories of perception, in the empiricist school, at the time Kant was writing. The empiricists tried in various ways to show how the connecting relationships which we in fact use can be derived from these sense-data. For example, Berkeley, in his *New Theory of Vision*, recognises that there is no particular sense-datum of 'distance', and attempts to show how we derive the concept of space from the fusion and correlation of various sense-data. Similarly Hume points out that although we may perceive event A following event B, we never have a sense-datum of 'causality'. He therefore attempts to show how the concept of causality may be derived from certain patterns of succession in the sense-data.

Although Kant does not question the underlying model of a stream of discrete sense-data, what he does do is to work it out to its logical conclusion and to show that at that point it becomes quite unable to account for any of our experience at all. The problem is this. When I see, say, a dog walking down the road I believe I am seeing something happening independent of me. If I can analyse what I believe I am seeing into sense data I will discover that it is made up of a large number of separate data. Exactly what is to count as a single datum is not clear, but for the sake of the model we can imagine that as I scan the dog and the road I take in first the head, then the body, then focus on the tail in a number of successive wagging positions, then, as it barks, swing my attention suddenly back to its head, listen to the bark, focus on the road in front of the dog to see what it is

barking at, and so on. In describing this as 'seeing a dog walking down the road' I am implying that the way in which I put all this together in one 'picture' is necessarily determined by the dog on the road. Although I have a sense-datum sequence head-body-tail-head, I am not allowed to put this together in a picture of a two-headed dog. There is only one way in which I necessarily have to organise the picture (which is a picture in several sensory dimensions, including the bark of the dog, the heat of the sun, and so on). I believe that it is necessary to describe this as a barking dog on the road which is emitting a barking sound. But what entitles me to believe this? One might answer, as, in essence, does Hume, that I have already learnt to associate heads, dogs and barks in this particular way through my prior experience of these sense-data, which are habitually associated with one another in a certain way. That is, I come to this particular dog-on-road already equipped with dog-categories which I can use to order my sense-data in what seems to me to be the right way. I justify my claim that this is the necessary way in which they must be organised by referring to these categories and to their mode of acquisition. But this is really to beg the question, for we will discover that in order to build up my dog-categories I needed at some point to use more fundamental categories. For instance, I say that the dog barked, rather than that the opening of the dog's mouth was followed by a noise. If we have the concept of cause, we might well be entitled to say that the dog's open mouth (and whatever was going on inside), caused the noise. But the problem is, from where do we get the concept itself? As a relation, it cannot be itself given as a sense-datum, but the dog-categories which we use to order the sense-data are only possible if we already possess this and other concepts. The empiricists attempted to derive these basic categories in various ways from the sense-data. Kant shows that none of them can in fact be so derived.

Before examining his argument, I should first briefly consider Heidegger's argument that Kant does not in fact use the sense-datum theory. Heidegger points out that Kant distinguishes between two kinds of intuition, sensible intuition, in which the subject is passive in regard to the object, and what Heidegger calls an infinite intuition: "An understanding which through its self-consciousness could supply to

itself the manifold of intuition - an understanding, that is to say, through whose representation the objects of the representation should at the same time exist" (KS 157). This is the mode of knowledge of a being which creates that which it has knowledge of. Heidegger suggests that Kant's intention is essentially to contrast human knowledge, as finite, with such an infinite intuition, and thus to ask how it is possible for a limited being to have any knowledge at all. He argues that for Kant the essence of sensibility lies in the finitude of intuition. The organs which serve affection are sense organs, therefore, because they belong to finite intuition, i.e. to sensibility. Thus Kant was the first to arrive at an ontological, non-sensuous concept of sensibility. (cf. Heidegger 31-32.) On this analysis, Kant's theory is not dependent in any way on the sense-datum theory. Such a reading makes it possible to interpret Kant as a very coherent thinker, but only, I fear, at the expense of ignoring large parts of his text, including the brief passages I have quoted, and those passages in which he is clearly dichotomising between the sensible world and the intellectual world in mind-body terms, and not in terms of the finite and the infinite. Nevertheless, Heidegger's reading does indicate the underlying problem with which Kant is dealing. This is the problem of the relation between my experience, as necessarily finite and particular, and the totality which I am experiencing. How can I transcent my particular finite perspective in order to make statements which are true for other people and at other times, and in this sense are universal and infinite? The sense-datum model is one way of posing this problem, and it was, unfortunately, the way which Kant chose. Heidegger's more general formulation also opens up a wider range of possible solutions, but for the moment we need to follow Kant through his model.

Sitting, then, on my momentary discrete sense-datum, what can I derive from it - and what can I not derive from it? Kant formulates the question as follows: "Whether there is any knowledge that is thus independent of experience and even of all impressions of the senses. Such knowledge is entitled *a priori*, and distinguished from the *empirical*, which has its sources *a posteriori*, that is, in experience" (KS 42-43). The problem is, firstly, to establish which concepts are *a priori*, in the sense of not being derived from sense-data,

and then, secondly, to justify our right to use these concepts. The two criteria for *a priority* can be established negatively, as "necessity" and "universality". That is, since in the stream of sense-data the sequence is always contingent and personal, any concept which claims the necessity of a certain relationship, and therefore at the same time claims that it is universally valid, cannot have been acquired purely from sensation. That is, when I move from the affirmation "this seems to be a dog's head" to "this is a dog" I am claiming that the sense-data are organised in a way which is necessary, and which is valid for any other subject experiencing the dog.

In discussions of Kant's writing there seems to be some confusion about what Kant means by "necessity". The question is whether the necessity to which he refers is the alleged necessity of general statements of the type "All dogs bark", or the necessity which I claim when I say "This dog is barking now", meaning that anybody else present would necessarily also perceive this dog as barking, and that the way in which I have conjoined the perception of the dog and the perception of the bark is the only way in which they could be conjoined. Kant himself is slightly ambiguous in his usage. He neither distinguishes clearly between the two types of necessity nor explains the relation between the two. He is at his clearest in the dicussion of the question in the *Prolegomenon*, but even there a certain amount of ambiguity persists. He distinguishes between "Judgements of Perception" and "Judgements of Experience" : "Empirical judgements, so far as they have objective validity are JUDGEMENTS OF EXPERIENCE; those which are only subjectively valid I call mere JUDGEMENTS OF PERCEPTION ... All our judgements are at first mere judgements of perception, they are valid only for us, i.e. for our subject, and only afterwards do we give them a new reference, namely to an object, and want the judgement to be valid for us at all times and equally for everybody; for if a judgement agrees with an object, all judgements about the same object must agree with one another, and thus the objective validity of the judgements of experience means nothing other than its necessary universal validity. ... We know the object ... through the universally valid and necessary connection of the given perceptions ... as for example when I say air is elastic, this judgement is at first only a judgement of

perception, I only refer two sensations in my senses to one another. But if I would have it called a judgement of experience, I demand that this connection shall stand under a condition which makes it universally valid. I require that I and everybody must always necessarily conjoin the same perceptions under the same circumstances ... Hence judgements are either merely subjective, when representations are referred to a consciousness in one subject alone and are unified in it, or are objective, when they are unified in a consciousness in general, i.e. necessarily unified in it." (Prolegomenon 56-64). Here it is clear in the theoretical discussion that what is at issue is the relation between particular perceptions and a particular object. But in the example Kant uses an instance which could be either particular or general: This air is elastic, for me and for anybody else experiencing it; or All air is elastic. The other examples which he uses in the passage vary in clarity. The first group refer to the subjectivity of feelings: the warmth of a room, the sweetness of sugar, and the bitterness of wormwood, all of which he says are subjectively valid only, and "only express a reference of two sensations to the same subject, namely myself, and this only in my present state of perception" (58).

Although these examples would seem to refer to the problem of a particular object, they are not good examples from Kant's point of view, because part of their apparent strength comes from an aspect which is in fact quite irrelevant to the argument: the fact that our perceptions of tastes and feelings is very relative to the momentary state of our body. My judgement as to whether the room is warm or cool will depend largely on the temperature from which I have just come. But this aspect of the problem can easily be solved by including the relevant conditions in the judgement: "A room at this temperature feels warm to anybody coming in from the snow". This is now an objective statement linking a certain temperature to human bodies in a particular way. But this is not really Kant's problem at all. What he wants to know is the extent to which I can legitimately claim that the taste of sweetness and the sight of the sugar are necessarily linked in a particular way even for me now. His other example is even worse. He distinguishes between the habitual conjunction of the perception of

the sun shining and of the warmth of the stone, on the one hand, and the judgement of experience that the sun warms the stone, on the other hand. The problem with this example is that it is highly complex and at this point Kant does not analyse the complexity. It contains at least three objective judgements: The sun is shining, the stone is warm, and the shining sun is causing the warmth of the stone. The causal statement, on which Kant focusses here, is more complex than the other two, both because it depends on them, and because its mode of verification is likely to be more complicated. This means that one can read the example either as referring to a general statement "the sun shining warms stones", or as referring to a particular but necessary (objective) statement, "the shining sun is warming this stone".

A clearer example would go like this. "This apparent stone seems to me to be warm" is a judgement of perception which describes the possibly accidental conjunction of two sense-data (assuming for the sake of the argument that we can describe the 'look' of the stone as a single sense-datum). "This stone is warm" is a judgement of experience, which claims that this conjunction of sense-data is in fact necessary.

In trying to clarify the various types of necessary judgement which Kant envisages, we need to distinguish between at least the following:

1) "This stone is warm". This is a particular statement about a particular object and claims that the two sense-data are necessarily conjoined in this particular way.
2) "The shining sun causes the stone to become warm". This states a more complex, but still particular relationship, and it evidently involves reference to a wide range of sense-data, in that I only say that it is *the sun* which is causing the stone to become warm in the light of my previous experience of the sun and of stones. This statement also claims a necessary relation between the various sense-data involved in my perception of the shining sun and the warm stone.
3) "The shining sun always warms stones". This is based upon a succession of judgements of type 2, and is necessary in so far as they are necessary.

4) "Every event has a cause". This is a very different type of statement, but it seems to be assumed in some way by the earlier statements. It is necessary in a much stronger sense than is statement 3, which says that the objective situation is such that I can only order this particular set of sense-data in this particular way, but does not say that the facual situation has to be this way. Statement 4, however, does make the claim that this particular event not only factually does have a cause, but has to have a cause (although not necessarily this particular cause).

The first two statements deal with ways in which we necessarily have to organise particular sets of perceptions. The third statement deals with what we, on the basis of these organisations of our perceptions, believe to be the factual relation between certain classes of things. The fourth statement, however, says something about the way in which things necessar-ly have to be related to one another. Kant believes that we can affirm it with an absolute certainty which we can never achieve for mere empirical generalisations like type 3, in which we always have to admit that we might be mistaken, and that we might discover new evidence which will make us modify our judgement. He also believes that there is a relation between the two types of necessity, involved respectively in propositions 1 and 2, and propostion 4. The relationship seems to be this. The *a priori* element in the particular judgement regarding a particular object must be such that it can be formulated into a principle. Embodied in the judgement of experience this principle is mediated by the empirical element in the judgement in such a way as to achieve only factual necessity. Freed from the particular empirical content of the judgement however, it is in some sense absolutely necessary as an *a priori* principle. Thus one can establish the same *a priori* principle in one of two ways. Either one can uncover an element in a judgement of experience which cannot be given in sense-datum form, or one discovers a principle which is held with a certainty which cannot be arrived at by empirical generalisation. For example, speaking of geometry Kant writes: "as empirical concepts, and also as that upon which these are grounded, as empirical intuition, neither the concepts nor the intuitions can yield any synthetic propositions except

such as is itself also merely empirical (that is, a proposition of experience), and which for that reason can never possess the necessity and absolute universality which are characteristic of all geometrical propositions". That is, the principles of geometry are such, Kant believed, that they could not be refuted by experience, and therefore also cannot be derived from experience.

However, there is a danger in this second mode of procedure in that the degree of certainty with which a principle is held is sometimes at least a function of nothing other than the state of scientific knowledge at the time. For this reason if too much stress is laid upon the arguments from the certainty of principles, and not enough on the more fundamental argument from the inadequacy of sense-data, then Kant's theory may appear to be more dependent that it in fact is on the assumptions of Newtonian physics and Euclidean geometry. What I have been trying to do in the above rather obscure paragraph is to make sense of an apparent ambiguity in Kant's use of the term "necessity", in order to show that the fundamental necessity with which he is concerned is that involved in the judgement of experience which asserts a necessary connection between sense-data. The question of the necessary validity of certain principles describing the nature of the objects of our experience can be shown to derive from this. From this point of view, it will be best to see in Kant's use of Newtonian physics and Euclidean geometry merely an attempt to illustrate the confidence with which we do make objective judgements about moving bodies and things in space, i.e. unity of object precedes unity of nature.

Necessity and universality, then, cannot be established by perception, from which it follows that any judgement of necessity and universality must contain an *a priori* element which does not come from the sense-data. However, some statements are *a priori* true because they are analytic. They are logically deduced from the definition of a particular concept, and so are themselves true by definition. But the kind of *a priori* statements for which we are looking could not be true by definition, since they involve affirmations about the world, not just about the meaning of words. Kant therefore distinguishes between analytic *a priori* judgements, which are true by definition, and synthetic

a priori judgements, judgements which are *a priori* but in which the predicate tells us more about the world than merely what is contained by definition in the subject. Strawson argues that Kant never gives a satisfactory account of the distinction between analytic and synthetic *a priori* judgements, nor can such an account be derived from the remarks that he does make on the subject. (cf. Strawson, p.43.) In one sense this is true, because the whole concept is constructed in relation to the essentially incoherent sense-datum theory. But nevertheless we can isolate the class of statements to which Kant is referring, since they are just those statements which are impossible in terms of the simple model of the theory with which Kant begins. That there are such statements, and that they are of fundamental importance, tells us something about the simple model. What it tells us is another question. It may well be neither what Kant thinks it is, nor what Strawson thinks it is. But for the moment we can accept the negative definition of synthetic *a priori* propositions as propositions which cannot be formulated within the sense-datum model, and we can see the *Critique* as, initially at least, an attempt to discover whether there are any such propositions. To a certain extent Kant's own method of exposition is misleading, since in the Introduction he gives a number of examples of what he believes to be synthetic *a priori* propositions, and then formulates the central problem: "Now the proper problem of pure reason is contained in the question: How are *a priori* synthetic judgements possible?" (KS 55). But in the actual text he begins all over again, as is to be expected, by first showing of particular propositions or classes of propositions that they are synthetic and *a priori*, and only then trying to explain how they might be possible. Since synthetic *a priori* principles are fundamental to our claims to knowledge, Kant terms his procedure a transcendental inquiry: "I entitle *transcendental* all knowledge which is occupied not so much with objects as with the mode of our knowledge of objects in so far as this mode of knowledge is to be possible *a priori*" (KS 59). A transcendental inquiry is an inquiry into the possibility of transcendence, that is into the possibility of transcending my here-now sense-datum, my immediate particularity, in such a way as to have objective knowledge.

We experience ourselves as living in a spatially and temporally

ordered world. Kant begins his analysis of *a priori* knowledge by looking at these two fundamental features of our world. Of each, we need to ask whether our idea of it is given in sense-data. And of each, according to Kant, the answer is "no". His argument is terse but clear: "In order that certain sensations be referred to something outside me (that is, to something in another region of space from that in which I find myself), and similarly in order that I may be able to represent them as outside and alongside one another, and accordingly as not only different but as in different places, the representation of space must be presupposed. The representation of space cannot, therefore, be empirically obtained from the relations of outer appearance. On the contrary, this outer experience is itself possible at all only through that representation". (KS 68). If we know that various peculiarly shaped objects are part of a jigsaw puzzle, we can begin to put them together. But, if we do not already have the idea of a jigsaw puzzle, we cannot derive the idea of it from some scattered pieces. This is a rather loose analogy, but it can be used to illustrate Kant's point that any theory which attempts to show how sense-data can be put together to form a spatial picture necessarily begins by assuming that we have some idea or concept of space, and that the problem is merely to situate things within that space. However, the theoretical problem is actually to account for the initial idea of space. It is this that cannot be accounted for by the sense-data themselves. If each sense-datum is a discrete mental entity, as it must be on the model, it can carry no intimation of an "outer", or of spatial relations between objects in that outer.

Space has further peculiarities. Any object in space is experienced by us as being spatially related to all other objects in space. That is, we experience space as one. If we compare the notion of "space" with a concept like chair, we can see that the concept "chair" has a number of exemplars, each of which is separate from the others. But different spaces are related to one another in a different way from the way in which different chairs are related to one another. Different spaces coexist as part of the same single Space. Furthermore, Kant argues, space is given as stretching off in all directions into infinity. We do not develop a concept of space by abstracting from given spaces.

We begin immediately with one infinite space. For this reason, Kant calls space an *a priori* intuition, rather than an *a priori* concept.

Strawson describes the above arguments as "unilluminating", and "weak in the extreme" (Strawson 58). He concludes that the only extended argument which Kant offers is the argument from geometry, to the effect that geometry is a body of synthetic *a priori* propositions which describes the structure of space with a certainty which could not be derived from mere experience. This argument falls down because it appears that Euclidean geometry does not in fact describe our space accurately. However, it seems to me that Strawson is wrong to place the weight which he does on the argument from geometry. The above argument is the basic one and it is far from unilluminating once one recognises that it is directed against the simple sense-datum model. If, as the argument shows, that model cannot account for the fact that our experienced world is spatial, then we have to change the model. I do not think that Strawson distinguishes sufficiently clearly here between the argument against the old model, and the argument for the new model which Kant proposes to take its place.

It is an important part of Kant's method to show the necessity of employing a particular concept by showing that only if we use that concept can certain important parts of our knowledge be possible. This involves what he calls a "transcendental exposition": "I understand by a transcendental exposition the explanation of a concept, as a principle from which the possibility of other synthetic *a priori* knowledge can be understood." In the first edition of the *Critique* Kant included remarks on geometry in what he later referred to as his "metaphysical exposition" of the concept 'space'. In the second edition he distinguishes clearly between the metaphysical exposition, which "contains that which exhibits the concept as given *a priori*, and the transcendental exposition which shows the concept as a principle from which further knowledge can be derived. Thus geometry is not part of what shows the concept of space as given *a priori*, but only knowledge which depends upon the concept of space. However, Kant could replace geometry here by the principle that all objects are spatially related to all other objects, which is equally a synthetic *a priori* principle

the possibility of which can be understood from the concept of space as an *a priori* intuition. Thus Euclidean geometry does not play a necessary role in his argument.

Kant's central argument for the *a priority* of time is similar to that for space: "Neither co-existence nor succession would ever come within our perception, if the representation of time were not presupposed as underlying them *a priori*. Only on the presupposition of time can we represent to ourselves a number of things as existing at one and the same time (simultaneously) or at different times (successively)" (KS 74). We cannot "think away" time, even in imagination. Like space, it is given as one, as a single time in which all events are temporally related to all other events, and it is given as unlimited. It is therefore a pure form of sensible intuition, as the grasping of a single object.

Once more, the basic argument, quoted, goes right to the crux of the matter: we can only organise our discrete sense-data into some temporal order once we have the concept of time. Each sense-datum in itself is atemporal.

Thus Kant effectively shows here that the sense-datum model as already outlined cannot account for the most basic aspect of our experience, the fact that it is spatially and temporally organised. But we now need to begin to look at the model with which he wishes to replace the simple sense-datum model. His argument is simple. If my experience is made up of a stream of sense-data, and if nevertheless it is spatially and temporally ordered by means of *a priori* intuitions of space and time, this can only be because my mind is in some way necessarily ordering them in this way. This point is established right at the beginning of the *Aesthetic*, when the sense-datum theory is spelled out: "That in which alone the sensations can be posited and ordered in a certain form, cannot itself be sensation; and therefore, while the matter of all appearance is given to us *a posteriori* only, its form must lie ready for the sensations *a priori* in the mind, and so must allow of being considered apart from all sensation" (KS 66). Kant distinguishes space and time as the forms, respectively of "outer sense" and "inner

sense". Time is the more fundamental, in that all outer experience is in one sense inner also, insofar as it is my experience. Therefore, all outer objects are subject both to time and to space, whereas inner experiences are subject to time only.

Kant believes that his argument has established the following, of time as well as of space: "Our exposition therefore establishes the *reality*, that is, the objective validity, of space in respect of whatever can be presented to us outwardly as object, but also at the same time the *ideality* of space in respect of things when they are considered in themselves through reason, that is, without regard to the constitution of our sensibility" (KS 72). Space and time, therefore, are empirically real, but transcendentally ideal. All objects of possible experience are subject to the forms of time and space, since these are the forms of our own sensibility. But we cannot say of things in themselves, which by definition we cannot experience, that they are subject to the conditions of time and space.

This is particularly significant with regard to time. If time is the form of the inner sense, that is, the form in which my experience of myself is given to me, it follows "that the subject, which is the object of the sense, can be represented through it only as appearance, not as that subject would judge of itself, if its intuition were self-activity only, that is, were intellectual" (KS 88). In terms of this model, the stream of sense-data is made up both of impressions coming from outside, and also of particular impressions of my internal life, such as images, memories and so on.

I can reflect upon this internal life, but in so doing I receive only further impressions, like them subject to temporal conditions. I therefore can know my "real" self only as it appears to be in reflection, and not as it might be in itself if perceived in some sort of direct intuition not subject to the conditions of my own form of sensibility. This is, from Kant's premises, a valid argument, and it illustrates clearly a major weakness in the empiricist analysis of the self. That is that the simple sense-datum model begins with the assumption of a consciousness which is receiving the sense-data,but then dissolves that

consciousness into itself being only a series of sense-data. It then becomes impossible to comprehend the status of the sense-data themselves, since it is no longer clear to what these appearances appear. Kant makes this explicit by formally removing the machinery of perception, the observer who organises perceptions through the forms of space and time, from the sphere of possible experience.

Kant distinguishes carefully between "sensibility" and "understanding": "Objects are *given* to us by means of sensibility, and it alone yields us *intuitions;* they are *thought* through the understanding, and from the understanding arise *concepts*" (KS 65). In intuition we are in some sort of direct relation to the object, and the object is given to us in its singularity. It is because space and time are given to us as singular that they must be considered intuitions, rather than concepts. Then we think, however, we classify what we are thinking about in terms of concepts. Each concept refers to a number of different possible examples of the concept, and when we classify a particular object by applying a particular concept to it we are putting it together with the other objects to which the concept applies. Any form of knowledge, from the most minimal form of recognition upwards involves some such classification, and hence involves concepts. Kant further distinguishes between pure concepts and empirical concepts. An empirical concept contains in it some element which refers to sensation. The concept "dog", for example, refers to our actual sensible experience of dogs. Pure concepts, however, contain "only the form of the thought of an object in general" (KS 92). Kant states this as an assumption, but in fact its status is that of a hypothesis to be proved. Pure concepts, in so far as they contain no reference to sensation, must be such as can be known *a priori*. Our problem is first to discover whether they are any such concepts, and then to justify their use.

Kant answers the first half of the question by investigating the function of concepts in thinking. A concept is used to bring together a number of different representations. This can only be done by means of a judgement: "In every judgement there is a concept which holds of many representations, and among them of a given representation

that is immediately related to an object" (KS 105). A judgement, then, is a way of unifying aspects of our experience by means of concepts. Thus if we can exhaustively classify all the different forms of judgement, we have a list of all the different ways in which it is possible to bring together aspects of experience. Each form of judgement represents a particular way in which different bits of our experience, different representations, can be brought together. Kant believed that such an exhaustive classification of the forms of judgement lay ready to hand in the table of judgements established by normal logicians. I shall not go into the details of this table, partly because it is generally recognised to be wrong, and partly because we are here interested in the implications of the possibility of *a priori* judgements for our understanding of human reality, rather than in the details of the various types of *a priori* judgement. From our perspective it does not matter whether Kant establishes only one synthetic *a priori* proposition or a hundred. In each case it makes it necessary to rethink the sense-datum model. The table of judgements shows the various forms of synthesis in judgements. To show the relevance of this table for our possession of possible *a priori* concepts in experience, it is necessary to show the relation between synthesis in judgement and synthesis in experience. This can be done by showing that, in terms of our model, any togetherness of the sense-data must be contributed by the understanding itself, and so that the forms of synthesis in the sense-data must be similar to the forms of synthesis in judgement. Kant expresses this in the following dense paragraph: "What must first be given - with a view to the *a priori* knowledge of all objects - is the *manifold* of pure intuition; the second factor involved is the *synthesis* of this manifold by means of the imagination. But even this does not yet yield knowledge. The concepts which give *unity* to this pure synthesis, and which consist solely in the representation of this necessary synthetic unity, furnish the third requisite for the knowledge of an object; and they rest on the understanding" (KS 112).

Kant himself wrote two very different versions of the transcendental deduction of the legitimacy of the categories. The essence of the argument is the same in each case, but different points are stressed, and the sequence is different in the two cases. I shall construct the

argument in such a way as to show it as a criticism of the sense-datum theory. From this point of view the argument is developed in three main stages. In the first stage Kant shows that the use of the *a priori* categories is necessary if we are to have experience of an objective world. In the second stage he shows that unless we do have such an experience we cannot have any idea of an I as the subject to which appearances appear. In the third stage he attempts to show how the sense-datum model can be adapted in the direction of transcendental idealism to allow for this.

1) Ojbectivity

When Kant speaks of "experience" he is referring to "empirical knowledge" (KS 173), in so far as empirical knowledge is "the determinate relation of given representations to an object" (KS 156). That is, the notion of experience implies that the experiencing subject interpret the sequence of representations as having a particular necessary relationship to one another which is a function of their reference to an object which is being experienced. To postulate an object of experience is to claim that there is a rule-governed connection between certain sense-data: "Now we find that our thought of the relation of all knowledge to its object carries with it an element of necessity; the object is viewed as that which prevents our modes of knowledge from being haphazard or arbitrary, and which determines them *a priori* in some definite fashion. For in so far as they are to relate to an object, they must necessarily agree with one another, that is, must possess that unity which constitutes the concept of an object" (KS 134-5).

If we compare this with the simple sense-datum model, it is evident that on that model there can be no experience of objects. Sense-data are discrete. Each is "completely foreign to every other, standing apart in isolation" (KS 130). There can be no unity between sense-data unless each datum in some way refers to the other data with which it is united in the concept of an object. Unless there is such reference, then any way in which they are subsequently put together must be purely arbitrary.

This means that the combination of sense-data cannot itself be given through sensation: "But the combination of a manifold in general can never come to us through the senses, and cannot, therefore, be already contained in the pure form of sensible intuition" (KS 151). The combination is a particular pattern in which the basic units are arranged, and this pattern is necessarily something other than the basic units, and cannot be given as one of the basic units, that is, as another sense-datum. Thus if there is to be experience of objects, it must involve combinatory concepts other than those concepts which can be given purely empirically. That is, it must involve *a priori* concepts of relationships, and, as we have already seen, the only such concepts are the Categories derived from the table of judgements. So experience of objects is impossible without the application of the Categories.

However, this argument still leaves open the possibility that we ought to abandon our claim to have experience of objects, rather than accept the claim to the legitimacy of the Categories. How can we justify our claim to have experience of a world of objects? Would it not be possible to construct an account of human experience which made no reference to objects at all? Kant's answer to this question is that this could only be done at the expense of destroying also the subject to which sense-data appear, and that there is thus no possibility of a coherent account based solely on the sense-datum theory.

2) Subjectivity

The argument is simple, though Kant's formulation of it is not. He writes: "Only in so far as I can grasp the manifold of the representations in one consciousness, do I call them one and all *mine*. For otherwise I should have as many-coloured and diverse a self as I have representations of which I am conscious to myself. Synthetic unity of the manifold of intuitions, as generated *a priori*, is thus the ground of the identity of apperception itself, which precedes *a priori* all *my* determinate thought" (KS 154). In order to talk about me having a particular experience (of a sense-datum), it must be possible to distinguish between myself and the particular experience. That is, we must have some

concept of an I to whom all these appearances are appearing. If there is no such identical I, then we are left only with a "many-coloured and diverse self" which is nothing other than the sequence of appearances, and so is not *one* self at all, but rather a sequence of discrete selves as a correlate of the sequence of appearances. What, then, is necessary in addition to account for the possibility of an identical I? Kant's answer is that there can be no I, no "identity of apperception itself", unless there is a "synthetic unity of the manifold as generated *a priori*", that is, unless there are in fact synthetic connections between the various sense-data. Only if there is a certain necessity in the relation of bits of my experience to one another is it possible for me to distinguish between myself and my experience, and hence to have a concept of myself as a subject identical through time.

There are two factors involved here. Firstly, it must be possible to have some experience of the I as going beyond this particular appearance, and as being the possible or actual subject of other appearances in the past or the future. Secondly, in order to distinguish the subject from the experience it must be possible to distinguish, in that temporal sequence of possible appearances, between the order of appearances for the subject and some other order experienced by the subject as being subject to a rule outside its control. That is, for there to be a concept of a self which has experience, a part at least of our experience must be experience of objects, where 'object' is defined as the unity of a rule for combining a particular series of appearances.

3) Transcendental Idealism

Kant's argument to establish that the notion of an appearance depends on the notion of an identical subject, which in turn depends upon the notion of an objective order underlying appearances, is interwoven at every point with his account of what might be described as the mechanism underlying this. In terms of the model we have a discrete series of appearances, and also a set of *a priori* categories which, when applied to the appearances, provide us with a unified world of objects which impose rules on the order of the appearances. The question is: how are the categories applied? We have seen, on the one hand, that we have

no reason to believe that the categories in some way reflect relations between objects as they exist independent of our possible experience, since we have no way of knowing anything about such things - in-themselves. But, on the other hand, the categories are not applied by me at my whim. It is not as though I have certain building blocks. In that situation what I could make would be limited both by the nature of my tools and by the nature of my building blocks, but within those limits I would have complete freedom. Kant, however, insists that the categories are not to be applied with any such gay abandon. Indeed, the categories are not tools which I can use as I please. They are, rather, elements in the structure of my mind, and as such they bind my judging and experiencing in particular ways.

All knowledge involves synthesis; that is, it involves bringing bits of experience together. In order to understand the nature of the categories we need to begin by investigating the nature of this synthesis. Synthesis involves, according to Kent, three elements:

a) a number of separate elements;
b) the bringing together of those separate elements; and
c) the assembly of them into a particular pattern.

In the simple sense-datum model, the perceiving subject never has more than one datum before it at a time. If each datum is to appear, and then to slip away into an unfathomable past, then there can be no synthesis. That is, each datum must be first apprehended, and then brought together with other data, or, in Kant's terminology, reproduced. Kant describes the first part of the process as the synthesis of apprehension in intuition: "Every intuition contains in itself a manifold which can be represented as a manifold only in so far as the mind distinguishes the time in the sequence of one impression upon another; for each representation, *in so far as it is contained in a single moment*, can never be anything but absolute unity. In order that unity of intuition may arise out of this manifold ... it must first be run through and held together. This act I name the *synthesis of apprehension*" (KS 131). The synthesis of apprehension is that which preserves each 'now' as one of a sequence of 'nows'.

But for a particular synthesis, it is not enough that each

representation should be preserved. It must be reproduced in the present as one of the elements out of which the synthesis is to be constructed: "Experience as such necessarily presupposes the reproduceibility of appearances ... if I were always to drop out of thought the preceding representations ... and did not reproduce them while advancing to those that follow, a complete representation would never be obtained" (KS 133). This Kant calls the synthesis of reproduction in imagination. Fina-ly, the diverse representations which have been apprehended and then reproduced together must be organised into a whole. This involves combining them in a certain pattern, or according to a certain rule. This rule is what Kant calls a concept: "If, in counting, I forget that the units, which I now hover before me, have been added to one another in successuion, I should never know that a total is being produced through this successive addition of unit to unit, and so would remain ignorant of the number. For the concept of the number is nothing but the consciousness of this unity of synthesis" (KS 134). To be conscious that I now have before me a number is to be aware that I am proceeding according to a rule in synthesising in this particular way. To apply a particular concept to a particular set of reproduced representations is to classify that set as being governed by a particular rule, and hence as being a particular kind of object. Kant therefore calls this aspect of synthesis the "synthesis of recognition in a concept".

Now in so far as all my experience is *mine* it must be subject to this triple synthesis, and hence all the various aspects of my experience must be related to one another in terms of my conceptual equipment, and hence in terms of those Categories which are the basic elements of my conceptual equipment. But this still leaves open the question as to *which* categories must be applied to which synthesis of the manifold of perception. Kant solves this problem by pointing out that before reaching the stage of the conceptual synthesis of the understanding, the sense-data have already been ordered in a preliminary way by the two basic forms of sensibility, space and time. In particular, since time is the form of the inner sense, all our representations are already temporally related to one another. Thus if we can distinguish between different patterns of temporal relations this will provide us with

a basis for determining the applicability of the different categories. That is, if a particular category is to be applied to the manifold of intuition it must be possible to show that that category can be interpreted as expressing a particular rule-governed order of possible representations in time. For example, what distinguishes an object from a process is that the order in which I look at the different aspects of the object is quite indifferent, whereas the order in which I look at different aspects of a process is irreversible. I can look either way long the road, as I feel like it, but I cannot choose the order of positions in which I perceive the moving car. In each case there can be formulated a particular rule which determines the possible order of perceptions. This rule Kant calls the "schema" of the category in question.

Kant introduces his discussion of the schemata by pointing out that, whereas an empirical concept has been abstracted from experience, and therefore is "homogeneous" with experience, in so far as it merely reflects the structure of a particular aspect of experience, the same cannot be said for *a priori* concepts. Since they are derived quite independently of experience they do not initially contain reference to some aspect of experience, and so it is not clear how they are to be applied to experience. It is therefore problematic "how pure concepts can be applicable to appearances" (KS 180). Kant concludes: "obviously there must be some third thing, which is homogeneous on the one hand with the category, and on the other hand with the appearance, and which thus makes the application of the former to the latter possible. This mediating representation must be pure, that is, void of all empirical content and yet, at the same time, while it must in one respect be *intellectual*, it must in another be *sensible*. Such a representation is the *transcendental schema*" (KS 181). A rule of temporal order is sensible, in that time is the form of sensibility, but it is purely intellectual in that it can be constructed entirely *a priori* by using only the idea of the unity of apperception, the idea that all experience, as *my* experience, has to be synthesised in one consciousness, and so must be linked by rules with all other aspects of my experience.

Kant does not seem to me to be very clear in his account of

the exact relation between a concept and a schema. He defines a schema as the "representation of a universal procedure of imagination in providing an image for a concept" (KS 182). But he has previously defined "concept" formally as "something universal which serves as a rule", and in the following paragraph he writes: "The concept 'dog' signifies a rule according to which my imagination can delineate the figure of a four-footed animal in a general manner, without limitation to any single determinate figure such as experience, or any possible image that I can represent *in concreto*, actually presents. Thus schematism of our understanding ... is an art concealed in the depths of the human soul ... " (KS 182-183). Here he defines the concept itself as a rule for generating an image, and seems, in the second sentence, to be specifically equating this with schematism. It would seem that we might say that empirical concepts such as 'dog' are essentially schematised. The concept of schematisation defines the relation between a concept and an image, and Kant introduces it at this point to show that this relation is not immediately given in the case of *a priori* concepts, but that it must be sought by investigating the structure of the way in which empirical concepts are related to images. The empirical concept is a rule for synthesising empirical data, so the *a priori* concept must be a rule for synthesising the form in which empirical data are given; that is, the pure form of temporal succession.

Kant's model can therefore be summarised as follows: The sense-data, or representations received by the subject are first organised into spatial and temporal order by the pure forms of sensibility, and then this order functions as the basis for a further ordering into rule-bound sequences of possible and actual experiences by the understanding, using certain *a priori* categories, without which coherent experience would be impossible, but which are themselves not derived from experience. Knowledge is objective in the sense that the way in which we combine sense-data in experience is necessarily determined by an objective rule, but the particular rule used is always a necessary structure of the unity of apperception, rather than a rule imposed by the object in so far as it is a thing-in-itself existing independent of our possible experience.

I would now like to look more closely at certain aspects of this model. What does it say about the relation between objects, what does it say about the subject and about the problem of knowledge of the subject, and finally to what extent do the more baroque aspects of the theory of transcendental idealisms indicate real problems?

1. Relations between objects

When I add up the sense-data 'dog's head','dog's body', 'dog's tail', 'dog's head, and 'dog's bark', and end up with a dob, I am applying a rule for the assemblage of particular kinds of sense-data presented in a particular kind of order. The rule could be seen as a blue-print for the manufacture of the experience of a dog. When provided by imagination with toy bricks I can use the blue-print to construct an imaginary dog. When provided by sensibility with real bricks of a certain kind I use the blue-print to construct an experienced dog. Now the important point which Kant is making here is this: A dog is a very complex animal. It exists in three spatial dimensions, in time, and also in what one might call the kinetic and auditory dimensions. It has a particular sort of inside, as well as a particular sort of outside. Yet I glance briefly at an object on the road, and almost immediately identify it as a dog. In so doing I am, in an important way, making predictions about the object. I am predicting that it would have a certain appearance if I looked at it from another direction, that it would bleed if cut, that it would have a brain of a certain size if operated on, and so on. Now I could in specific instances be wrong in these predictions. In this particular case it might be a clever clockwork imitation, or a picture on a screen, or a four-footed dog-like martian asking to be taken to my leader. But the possibility of error on some occasions is not important. What is important is to discover why I can ever do this with any confidence. Kant's answer is that I can do this because an object is always defined in terms of a rule for the assembly of possible experiences. I can therefore always use the rule to go beyond the actual limited experience of it that I have had, the few out of the many possible sense-data that I have received, and to predict other possible experiences of it. I can use these predictions to verify that it actually *is* a dog. But

if the rule were not there to make the predictions with, the possibility of verification would never arise, and there could be no ordered experience whatsoever. Thus objectivity and predictability are bound up together. Predictability, furthermore, is bound up in an important way with the concept of causality. This is obviously so in so far as one uses causal laws to predict the outcome of processes. But it is also the case in predicting stability, or continuity. Both continuity and predictable change depend on the notion of rule-governedness. The statement "There is no event without a cause" is in fact a law of continuity before it is a law of change, for it says that things will remain the same unless affected by other things. This relation between identity and causality was clearly perceived by Hume: "We readily suppose an object may continue individually the same, though several times absent from and present to the senses ... But this conclusion beyond the impression of our senses can be founded only on the connection of *cause and effect;* nor can we otherwise have any security that the object is not changed upon us, however much the new object may resemble that which was formerly present to the senses. Whenever we discover such a perfect resemblance, we consider whether it be common in that species of objects; whether possibly or probably any cause could operate in producing the change and resemblance; and according as we determine concerning these causes and effects, we form our judgement concerning the identity of the object ... the only (relation) that can be traced beyond our senses, and inform us of existence and objects, which we do not see or feel, is causation" (*Treatise of Human Nature*, Vol. I, 77-78). Thus, in effectively undermining the notion of causality, Hume also undermined the notion of identity, and so the possibility of any coherent experience.

It is useful to interpret Kant's argument as following the reverse path. By establishing that the possibility of the identity of the subject depends on the possibility of the rule-governed identity of the object, he has at the same time established the principle of causality, defined as succession according to a rule. All he has to do is to show that the concept of a rule can also be used to distinguish within the field of objectivity between strictly causal and other relationships. The distinction is that in the perception of an object

the order of succession of different appearances is bound up with my own behaviour, with the way in which I direct my attention to the object, whereas in the perception of an event, that is, an objective sequence of happenings, the sequence of appearances is not determined solely by the way in which I direct my attention: "In the ... example of a house my perceptions could begin with the apprehension of the roof and end with the basement, or could begin from below and end above; and I could similarly apprehend the manifold of the empirical intuition either from right to left or from left to right. In the series of these perceptions there was thus no determinate order specifying at what point I must begin in order to connect the manifold empirically. But in the perception of an event there is always a rule that makes the order in which the perceptions (in the apprehension of this appearance) follow upon one another a *necessary* order" (KS 221).

It I look at a house, there is a certain rule-governedness in the sequence of representations; I cannot see the back and the front simultaneously, and I cannot combine the representations in such a way as to place the front-door on the second floor. But within that context the actual order in which these representations appear to me is indeterminate. If, on the other hand, I watch a boat float down the river then not only can I not combine the representations in such a way as to have the portholes in the funnel, but there is also a certain irreversibility in the order in which the representations appear to me. I must see the boat upstream before I see it downstream. Thus in each case the mode of combining the representations is rule-bound, but there is a different type of rule in each situation. Withouth the concept of a necessary order of representations we could not have the concept of cause. It is for this reason that causality is an *a priori* concept. We cannot derive the idea of necessary succession from actual sequences of representations, and the idea of necessary succession according to a rule is constitutive of the possibility of experience, since "were I to posit the antecedent and the event were not to follow necessarily thereupon, I should have to regard the succession as a merely subjective play of my fancy; and if I still represented it to myself as something objective, I should have to call it a mere dream" (KS 227).

Several critics have objected that in this argument Kant has confused the necessity of succession according to a rule in the subject, with the necessity of succession according to a rule in the object. Could we not conceive of a world of objects, that is, of rule governed manifolds of perception, in which things nevertheless happened for no reason at all? Kant's answer to this, I think, is to point out the relation between causality, predictability and identity. Such a world is not conceivable as a possible world of experience, because in it there would be no possibility of establishing the identity of the subject of experience. From the point of view of the subject's identity such a world would be strictly equivalent of the pure sense-datum model. For the unity of the subject the unity of the object, or the unity of nature, is also necessary. This means that if, within a stable world, an apparently uncaused event occurs, we must necessarily posit not that the event is uncaused, but that the cause is unknown. I do not think that this principle is vitiated by the fact that in microphysics and quantum theory any laws formulated are statistically based. This proves that an electron could not develop a sense of its own identity, but does not affect those of us who live in the macro-physical world where physical laws are not statistical.

Since we exprefience a world of objects which behave in consistent ways, we are justified in formalising our experience of this consistency in scientific laws, although of course the laws cannot claim *a priori* necessity in the way in which the principles on which they are based can. Without the concept cause, for example, we could make no predictions, but the predictions which we do make may turn out to be false. Without the concept we could not make either correct or falst predictions. Which predictions we actually do make with it is an empirical question.

Kant insists throughout his analysis that our experience is always of things as they appear to us, and never as they are in themselves, independent of human experience. He uses the term object not to refer to the thin in itself, but to refer to the rule-bound relation between particular groups of representations. The categories, including the category of causality are used to order these representations. They

are *a priori* necessary to account for the possibility that all these representations have of cohering in one unified experience. We have established their necessity only with reference to this purpose. This means that we have no right whatsoever to apply these categories to things as they are in themselves. From the perspective of Kant's model this is obvious, for the forms of time and space and the categories of the understanding are part of the mental apparatus of the subject. They are immovable spectacles through which all perceptions are refracted, and the index of refraction by definition is not part of the nature of the objects emitting the light rays.

There are obvious problems with this theory. Kant begins with the "common-sense scientific" picture of an "image" in some way being caused by the interaction between the object and the perceptual equipment of the subject. But he ends up by destroying that picture by pointing out that one cannot in fact ascribe either "objectness" or causal powers to the thing-in-itself, since these categories only apply within the field of possible experience, and the thing-in-itself is by definition outside this field. This means that Kant's new model is defective, but it has nothing to do with the effectiveness of his attack on the old model. To understand the full implications of this attack we now need to look more carefully at Kant's concept of the subject.

2. Knowledge of the Subject

The sense-datum model may be described as a model in which the subject is purely passive. The subject is simply a receiver. As Kant puts it, a purely sensible intuition is nothing but receptivity (see KS 151). It cannot go beyond any one representation to relate it to others. Its history would be rather like an unused movie, in which a series of pictures sit impotently next to one another on the strip of cellulose. By looking at each one separately it is not possible to grasp the movement of the scene. By showing that this model is incoherent, Kant has shown that it is necessary to understand the subject as being in some way spontaneous as well as receptive. The subject synthesises

experience. Synthesis occurs in making judgements, but, much more fundamentally it occurs as the basis of our very experience of objects, since this experience is always of wholes which go beyond what is immediately given in perception, both in the sense of relating this given to already past perceptions, and also in the sense that it implies continuous rule-based prediction of future associated perceptions. Experience is always both retrospective and prospective. It always goes beyond the here-now perception, and if it did not do so the here-now perception itself would be dissolved. Thus we need to ask what the knowing subject is that it can transcent the here-now. It is clear that in answering this question we will at the same time be coming to understand the nature of human freedom.

It is also clear that the question requires us to investigate the nature of the human experience of time. The criticism of the sense-datum theory is also the criticism of the use of the present as the dominant category in human experience. We have to understand the subject not as simply a present consciousness of a present representation, but as what Strawson calls "a temporally extended *point of view* on the world" (Strawson 104). Kant attempts to resolve this problem by placing the transcendental I outside time. Time is merely the form in which the transcendental I orders sensible data, and it would appear to be for this reason that the I can have an overview of the temporal order. Heidegger interprets Kant's doctrine here not to mean that the I is outside time, but rather that the I is time: "The ego cannot be conceived as temporal i.e. intratemporal precisely because the self originally and in its innermost essence is time itself" (Heidegger 200-201). By this he means that the I cannot be subject to intra-temporal descriptions not because it exists quite independent of time, but because the I is purely temporal, and as such is the foundation for all our particular experiences of temporality. But apart from any intrinsic difficulties this theory might have, it seems to me that it is much more precise than any theory which Kant manages to formulate. Kant is always tied down by his mechanical model of perception and by his rigid distinction between the transcendental and the empirical ego. I learn about myself in reflection. But reflection, as an activity of the inner sense, is subject to the form of time, and as such reveals

me to myself only *qua* appearance, and never as thing-in-itself.

In order for experience to be mine, it has to be possible to accompany it by the "I think", and this I think expresses the identity of the transcendental subject throughout the changing series of perceptions. However, with this "I think", "I am conscious of myself, not as I appear to myself, nor as I am in myself, but only that I am. This *representation* is a *thought*, not an *intuition*. Now in order to *know* ourselves there is required in addition to the act of thought, which brings the manifold of every possible intuition to the unity of apperception, a determinate mode of intuition, whereby this manifold is given; it therefore follows that although my existence is not indeed appearance (still less mere [illegible] the determination of my existence can take place only in conformity with the form of inner sense, according to the special mode in which the manifold, which I combine, is given in inner intuition. Accordingly I have no *knowledge* of myself as I am, but merely as I appear to myself ... for knowledge of myself I require, besides the consciousness, that is, beside the thought of myself, an intuition of the manifold in me, by which I determine this thought. I exist as an intelligence which is conscious solely of its power of combination ..." (KS 168-9). Thus consciousness of my existence as a transcendent subject unifying experience, which is an idea arrived *a priori* by the use of reason, cannot be turned into knowledge of the nature of this subject without some data being provided by intuition, but since human intuition is limited by the forms of sensibility, it cannot provide any data about the subject as it is in itself, that is, as transcendental. Since the *a priori* categories can be used only to order sense-data, and obtain their legitimacy from the fact that they make it possible to convert the stream of sense-data into a coherent experience, it follows that they also cannot be applied to the transcendental I. We can therefore say nothing about *what* it is; we can only affirm *that* it is. In the second part of the *Critique* Kant undertakes a detailed rejection of the "rational doctrine of the soul" which uses the "I think" as a peg on which to hang a speculative theory based, Kant argues, on the illegitimate application of the categories. Here Kant shows that it is not even possible to apply the category of substance to the

transcendental I. The I has a purely logical meaning (KS 334).

Nevertheless Kant does give the I a much larger load to carry than he is willing to admit when theorising about it. The mechanisms of sensible intuition must be in some way part of the world of things-in-themselves, and hence in some way part of the transcendental I. He admits this at one point in his discussion of knowledge of things-in-themselves: "Through observation and analysis of appearances we penetrate to nature's inner recesses, and no one can say how far this knowledge may in time extend. But with all this knowledge, and even if the whole of nature were revealed to us, we should still never be able to answer those transcendental questions which go beyond nature. The reason of this is that it is not given to us to observe our own mind with any other intuition than that of inner sense; and that it is yet precisely in the mind that the secret of the source of our sensibility is located" (KS 287). As I pointed out earlier, the sense-datum model was constructed as a result of the scientific investigation of nature, but Kant now finds himself obliged to place these originally natural mechanisms of perception quite outside the sphere of nature. Similarly, in his discussion of the syntheses of apprehension, reproduction and recognition he is discussing aspects of the transcendental I, further mechanism for ordering sense-data, but he has to end up by implicitly declaring even these discussions illegitimate.

As we have already seen, Kant does believe that, although the speculative use of reason cannot provide any knowledge of the transcendental I, there is nevertheless another source of knowledge. Practical reason, faced with the necessity of choosing courses of action in the world, has to make certain assumptions in order for it to choose. These are the assumptions of the freedom and immortality of the subject, and the further assumption of a benevolent God. To put it another way, the fact that we can ask the question "What ought we to do?" is proof that we are, in practical terms, free. As the natural world, the world of experience, is bound by the category of causality, it follows that in so far as we are free we belong to the "intelligible" world of things-in-themselves (see above, p.28).

We are now in a position both to understand more clearly why Kant feels justified in postulating human freedom and equating it with reason, and also to understand the problems in his theory. The sense-datum theory is a causal theory par excellence. The subject is completely passive, merely reacting to stimuli from objects. Kant has shown that this theory cannot account for the nature of human experience, and that one therefore has to understand the subject as in some way active, as in some way always going beyond the immediate here-now datum towards an understanding ot it in order to grasp the world as a spatio-temporal totality. This spontaneity of the subject is related to reason, in so far as reason is the ability to join ideas together, and thereby also to synthesise partial perspectives into some concept of the whole. However, while retaining the sense-datum theory of perception, and hence the mechanistically causal account of the human being as a natural object, Kant could not really bring together these two aspects of human reality. He was thus forced to postulate within each individual the dichotomy between the transcendental I as free and rational, and the empirical I as part of the causal network of nature. This is why he cannot provide a unified theory of human fulfillment, but dichotomises between happiness as sensuous contentment, and the intellectual contentment which comes from the awareness of oneself as effectively part of the kingdom of ends. This is also why he cannot give a satisfactory account of the relation between heteronomy and autonomy, and of the relation between reason and history. The transcendental I is outside time, and hence theoretically cannot be said to change and develop (although, contradictorily, Kant introduces the idea of immortality to allow for the development of this subject). Because it cannot develop, we must assume that it is from the beginning rational and autonomous. Under these circumstances, heteronomy must come in some way from the interference of the empirical subject with the transcendental subject. Kant of course does not trivialise this into some postulated struggle between reason and the body. Rather, he skirts around the issue, and therefore leaves us with no clear understanding of how an individual might move from heteronomy to autonomy. Similarly, he can give no real account of the historical development of reason in the way that Rousseau does. The Kingdom of Ends is formally places outside history,

although at the same time he does suggest that reason develops historically. But this could only be thought coherently if human beings in so far as they are (potentially) rational, can be places inside time.

3. Transcendental Idealism and the problem of "transcendental psychology"

I suggested at the beginning of the discussion of Kant's theory of knowledge that what Kant does in the *Critique of Pure Reason* is to take the sense-datum theory of perception and to destroy it by working it through to its logical conclusion, which is transcendental idealism. Transcendental idealism is an attempt to make sense, from within the sense-datum theory, of the fact that it is necessary to see the subject as a temporally extended point of view on the world. Pure reason gives us the question: "What must the I be in order to be a temporally extended point of view on the world?" Practical reason gives us the question: "What is the relation between the I as a temporal extended point of view and the I as free?" Kant's answer to these questions is ultimately incoherent and therefore wrong, but the questions themselves are important. They are, I would argue, more important that the question as to the "bounds of sense" which recent British commentators on Kant have concentrated on; not because this latter is an unimportant question, but because it is relatively easy to answer, whereas the questions about the I are not even easy to ask.

The questions are an attempt to grasp the nature of the relationship between the I and the world. Kant's theory of this relationship seems to oscillate between two positions, depending on whether he is concentrating on the world as in-itself or the world as phenomenal. On the one hand the I is receptive; on the other hand the I is constitutive of the world, although the way in which this might be understood varies.

What exactly is the status of the noumenal world of things-in-themselves in Kant's theory? It seems to me that Kant never made up his mind. His descriptions of it vary widely. At one point in the *Critique* he writes: "In the process of warning the latter (sensibility) that it

must not presume to claim applicability to things-in-themselves but only to appearances, it (understanding) does indeed think for itself an object in itself, but only as transcendental object which is the cause of appearance and therefore not itself appearance" (KS 293). Thus here he thinks of the noumenon as the cause of the phenomena, and although he then goes on to say that we cannot apply any of the "concepts of our understanding", which would include the concept of causality, to the noumenon, nevertheless it is clear that the idea of the noumenon as cause of phenomena is intrinsic to the sense-datum theory of intuition. However, on other occasions he is much more circumspect in what he says about the noumenon. Heidegger quotes one of his posthumously published writings in which he says: "The distinction between the concept of thing-in-itself and that of appearance is not objective but merely subjective. The thing in itself is not another object but another aspect (*respectus*) of the representation with regard *to the same object*" (Heidegger 37). This is much more consistent with Heidegger's own interpretation of Kant, referred to briefly earlier. For Heidegger the distinction which Kant is making is that between finite and infinite knowledge of the same thing. Finite knowledge is the knowledge by sequence of aspects to which a finite being is necessarily limited. Infinite knowledge would be knowledge "in the round" (as well as both from inside and from outside). This would be knowledge of the thing in all its infinite aspects at one time. For Kant such knowledge is available only to the "intellectus archetypus" which creates the object of its knowledge at the same time as it knows it. If this is really the distinction which Kant is drawing here, then the noumenon is the whole, and the phenomenon is the aspect of the whole to which I as a finite being am related at any particular time. It is quite probable that Kant was moving towards such a formulation, but it seems to me that Heidegger is wrong to say that he actually got there. Kant's doctrine of the noumenon remains confused, at least in the *Critique of Pure Reason*, and so his doctrine of the nature of the I also remains confused.

In order to bring the transcendental subject as free and spontaneous back into a coherent relationship with the empirical subject and with human reality as embodied there are two alternatives left open

PHENOMENON & NOUMENON

by Kant's philosophy. The first is to abandon [illegible]tum theory entirely, and to rethink the whole relation subjec[illegible]d along the lines suggested by Heidegger, and without any presuppositions about the nature of sensibility. The second alternative is to abandon the noumenon as cause entirely, and to treat the subject as purely spontaneous, by attempting to show how it could be independently constitutive of the world. This was the path chosen by Kant's immediate successors, and Kant himself indicated that path by his notion of the *intellectus archetypus*, as sketched out in the *Critique of Judgement*.

IV. THE INTELLECTUS ARCHETYPUS

The origin of the problem discussed in the *Critique of Judgement* is twofold. On the one hand, it is an attempt to solve a problem which derives from Kant's own distinction between phenomenon and noumenon. On the other hand, it is an attempt to solve certain problems connected with Judgement.

a) Phenomenon and Noumenon

In a footnote in the Introduction Kant formulates the relation between the two in a way which brings out clearly the problem of the nature of that relationship:

> "The opposition or assistance is not between nature and freedom, but between the former as phenomenon and the effects of the latter as phenomena in the world of sense. The causality of freedom itself (of pure and practical reason) is the causality of a natural cause subordinated to nature (i.e. of the subject considered as man, and therefore as phenomenon). The intelligible, which is thought under freedom, contains the ground of the *determination* of this (natural cause) in a further inexplicable way (just as that intelligible does which constitutes the supersensible substrate of nature)." (CJ 32-3 n2)

This seems to involve a threefold relationship:

a) The causality of freedom (as noumenon) producing effects in nature through the subject as phenomenon.

b) An intelligible (noumenon) which constituts the supersensible substrate of nature.

c) The relation between nature as phenomenon and the subject as the phenomenal expression of noumenal freedom, whereby nature either opposes or assists freedom.

Kant is willing to accept that relationships (a) and (b) necessarily remain "inexplicable", but believes that it is necessary to at least give some account of how (c) is possible. The reason for this is obvious. Nature has to be conceived as entirely determined by causal laws, and the subject as phenomenon, as sensible, is also determined by such causal laws. On the other hand, the subject as free produces effects in nature. What is the relation between these effects and the laws of nature? They cannot conform to the laws of nature in the full sense, otherwise they would be completely explicable without positing the idea of freedom. The free will must, in terms of Kant's theory, add something new, and at the same time cannot add that something new in such a way as to entirely destroy the unity of nature which is connected with the unity of experience. Kant believes that a consideration of the "faculty of judgement" will provide us with a means of resolving this apparent contradiction.

b) Judgement

In the *Critique of Pure Reason*, Kant distinguishes between understanding and judgement as follows:

> "If understanding in general be defined as the faculty of laws or rules, the faculty of judgement may be termed the faculty of *subsumption* under these rules; that is, of distinguishing whether this or that does not stand under a given rule" (CPR 114).

He points out that the judgement must be conceived of as distinct from the understanding, because the application of rules cannot, without falling into an infinite regress, be considered as itself being rule-bound. Judgement is therefore "a peculiar talent, which does not, and cannot require tuition, but only exercise" (CPR 115).

In the *Critique of Judgement* he makes a further distinction between determinant judgements and reflective judgements; between subsuming a particular under a universal, and arriving at a universal from a particular:

> "The determinant judgement only subsumes under universal transcendental laws given by the understanding; the law is marked out for it, *a priori*, and it has therefore no need to seek a law for itself in order to be able to subordinate the particular in nature to the universal. But the forms of nature are so manifold, and there are so many modifications of the universal transcendental natural concepts left undetermined by the laws given, *a priori*, by the pure understanding - because these only concern the possibility of nature in general (as an object of sense) - that there must be laws for these (forms) also. These, as empirical, may be contingent from the point of view of *our* understanding; and yet, if they are to be called laws (as the concept of a nature requires), they must be regarded as necessary in virtue of a principle of the unity of the manifold, though it be unknown to us. The reflective judgement, which is obliged to ascend from the particular in nature to the universal, requires on that account a principle that it cannot borrow from experience, because its function is to establish the unity of all empirical principles under higher ones, and hence to establish the possibility of their systematic subordination" (CJ 15-16).

In treating of nature, there are certain *a priori* rules, the categories, which must necessarily be applied. But these rules do not fully account for any observed phenomena. The various laws of the natural sciences are arrived at *a posteriori*, by a process of reflective judgement in which we attempt to uncover general rules which account for the modes of behaviour of particular objects. These rules, or laws are contingent, in the sense that there is no apparent reason why they should be as they are rather than otherwise. Thus the problem arises: how do we move from the particular event to the general law, and what entitles us to do this? Consider the following three propositions:

1) "This seems to be a barking dog"

In Kant's terminology, this is a judgement of perception which describes the association of certain sense data.

2) "This dog is barking"

This is a judgement of experience. Through the application of the categories it asserts the necessary connection of the sese-data according to a particular rule.

3) "All dogs bark"

This (we shall assume, for the sake of the argument) a zoological law. It asserts a necessary relationship between being a dog and being able to bark. But that this relationship should be the case is a contingent fact as far as the understanding is concerned. It is not based on the application of further synthetic *a priori* categories of the type involved in proposition (2). How does the reflective judgement move from type (2) to type (3) propositions?

Kant's answer is the reflective judgement operates in terms, not of a rule, but of a principle, and that is the principle of the unity of nature. This principle cannot be derived from any particular experience, yet it is factually assumed by the attempt to link particular experiences together into laws, and by the attempt to derive these laws from yet more general laws. Thus far the argument, in some sense at least, seems to be acceptable. We do assume a coherence and a unity in nature, and this coherence cannot be discovered in any particular experience. Kant, however, moves from the idea of unity to the idea of purposiveness (or, in an alternative translation, finality (Zweckmässigkeit). He formulates the principle as follows:

> "This principle can be no other than the following: As universal laws of nature have their ground in our understanding, which prescribes them to nature (although only according to the universal concept of it as nature), so particular empirical laws, in respect of what is in them left undetermined by these universal laws, must be considered in accordance with such a unity as they would have if an understanding (although not our understanding) had furnished them to our cognitive faculties, so as to make possible a system of experience according to particular laws of nature ... Now the concept of an object, so far as it contains the ground of the actuality of this object, is the *purpose;* and the agreement of a thing with that constitution of things which is only possible according to purposes is called the *purposiveness* of its form. Thus the principle of judgement, in respect of the form of things of nature under empirical laws generally, is the

purposiveness of nature in its variety" (CJ 16-17).

The German word *Zweckmässigkeit* seems to be rather more ambiguous than either the word "purposiveness" or the more technical term "finality" which Meredith uses in his translation. Dictionary translations include "suitability, appropriateness, fitness, opportuneness, functionalism" (Cassells), and "expediency" (Langenscheidts). Kant introduces the term in the above passage in close association with the specific idea of purpose (*Zweck*). But at other times he appears to take advantage of its broader connotations of suitability or appropriateness, particularly when, later in the argument, he introduces the idea of *Zweckmässigkeit ohne Zweck*, purposiveness without purpose. It seems to me that at many points in the argument Kant makes use of this ambiguity to reach illicit conclusions.

To make this assumption of unity and purposiveness is to assume that nature, as phenomenon, is in harmony with our cognitive faculties in so far as these faculties necessarily attempt to unify experience by placing it under rules. We look for this unity, and, since the attainment of any goal is bound up with the feeling of pleasure, we must experience pleasure of a particular sort when we do discover unity and purposiveness in nature. An analysis of the peculiarities of this sort of pleasure will contribute to our general understanding of the reflective judgement.

This raises the possibility of discovering a subjective, aesthetic experience of unity-purposiveness, relating not to the actual objective cognition, but to the fact that a particular representation (of the imagination) is in harmony with the understanding in its search for unity. That is, the imagination brings together sense-data in a particular pattern, and the understanding unifies this representation by means of a concept. The concept itself is purely objective, but the relation between the representation and the understanding as unifying is subjective, in the sense of appertaining to the nature of the subject. As such, it is a source of possible pain or pleasure. But a pleasure arising from the relation between a particular representation of the imagination and the understanding's general need for unity

would also be such as to be valid for any subject, since all subjects share the characteristic of a unity-seeking understanding. This would be an experience which is both subjective, and at the same time necessary and universal. It is in fact the experience of the beautiful, and Kant calls judgements regarding the beautiful "judgements of taste".

> "The pleasure, therefore, in the judgement of taste is dependent on an empirical representation and cannot be bound up *a priori* with any concept (we cannot determine *a priori* what object is or is not according to taste; that we must find out by experiment). But the pleasure is the determining ground of this judgement only because we are conscious that it rests merely on reflection and on the universal, though only subjective, conditions of the harmony of that reflection with the cognition of objects in general, for which the form of the object is purposive." (CJ 28)

This means that there are two distinct principles of the reflective judgement: the principle of the unity-purposiveness of nature, which is the guiding thread in our attempt to explain natural events in terms of laws; and the principle of the possible harmony between the imagination and the understanding, which is the principle of aesthetic judgement of taste. (Kant also distinguishes a third principle, related to "the purposiveness of the subject in respect of the objects according to their form or even their formlessness, in virtue of the concept of freedom" (CJ 28-29), which is the principle underlying judgements of the sublime. This is an aesthetic principle complementary to the beautiful. I shall ignore this for the moment.)

c) <u>The Critique of Aesthetic Judgement</u>

In this section Kant first shows that the judgement of taste in the beautiful, as ordinarily understood, in fact has the characteristics sketched out in the Introduction, and then goes on to provide a deduction of the legitimacy of the principle. The first two points he makes are quite straightforward:

a) When we take delight in a beautiful object, we are interested neither in the actual existence of the object, nor in any direct agreeable

sensations which we get from it. It is not a physical pleasure, but a delight in the form of the object.

b) We accept the entire subjectivity of physical pleasures; we do not demand of other people that they enjoy the same food as we do. We accept the entire objectivity of the concepts of the understanding, in that true statements can be proved by rational argument. But, while we accept delight in the beautiful as a subjective reaction, a relation concerning my interaction with the object, rather than the object itself, we at the same time insist that the judgement of taste in the beautiful is in fact universally valid. That is, "the aesthetical universality which is ascribed to a judgement must be of a particular kind, because it does not unite the predicate of beauty with the concept of the object, considered in its whole logical sphere, and yet extends it to the whole sphere of judging persons." (CJ 49). This can only be the case if the content of the judgement can be universally communicated, which is not the case with judgements of physical pleasure. It can in fact only be the case if it refers to some experience which could be experienced by all individuals in the same situation, and this in turn can only be experience related to the process of cognition of the object; that is, to the relation between the various cognitive faculties common to all subjects. More specifically, it must be a function of a harmony between the representative imagination and the understanding brought about by the form of the object in question.

c) The next step in the argument is however, more problematic. The question is, what in the form of the object would it be that would bring about the requisite harmony between imagination and the understanding? This section of the argument is entitled "Of judgements of taste, according to the relations of the purposes which are brought into consideration in them." But, as far as I can make out, Kant does not justify his introduction of the concept "purpose" at this point. He seems to carry it over, without further explanation, from the Introduction. He begins by defining "purpose", and then goes on to distinguish between "purpose" and "purposiveness":

> "But an object, or a state of mind, or even an action is called purposive, although its possibility does not necessarily presuppose the representation of a purpose, merely because its possibility can be explained and conceived by us only so far as we assume for its ground a causality according to purposes, i.e. in accordance with a will which has regulated it according to the representation of a certain rule. There can be, then, purposiveness without purpose, so far as we do not place the causes of this form in a will, but yet can only make the explanation of its possibility intelligible to ourselves by deriving it from a will" (CJ 55).

We have already seen that the judgement of taste is not interested in the existence of the object; it is disinterested. So, in so far as it is related to purpose, it cannot be a function of the object which is purposed. It can therefore only be a function of the form of purposiveness in the representation:

> "Therefore it can be nothing else but the subjective purposiveness in the representation of an object without any purpose (either objective or subjective), and thus it is the mere form of purposiveness in the representation by which an object is *given* to us, so far as we are conscious of it, which constitutes the satisfaction that we without a concept judge to be universally communicable; and, consequently, this is the determining ground of the judgement of taste" (CJ 56).

What this seems to mean is that we judge something to be beautiful when its elements fit together so well that it appears to have been put together in that way for a purpose. And the pleasure in this purposiveness is a function of the fact that such purposiveness "contains a determining ground of the activity of the subject in respect of the excitement of its cognitive powers" (CJ 58), as has previously been established. At this point in the argument these two elements, the purposiveness of the representation, and the harmony of the cognitive powers, seem to be merely stated as separate facts involved in our judgements of the beautiful. Kant's essential point is that it is the form of an object which determines our judgement of its beauty: "It is not what gratifies in sensation but what pleases by means of its form that is fundamental for taste" (CJ 61). And it is the form in itself, not in relation to its usefulness, or in regard to the degree of perfection which it exhibits in the object, because the idea of perfection assumes an intellectual conception of the object, and is not purely aesthetic.

d) The fourth characteristic of the judgement of taste is that it refers to a necessary satisfaction. A beautiful object is not one that may or may not satisfy, or might satisfy on one occasion but not on another. Yet the basis for this experience of necessity is of an unusual kind. We cannot say that everybody will in fact agree that a particular object is beautiful, but we do say that they ought to agree. This claim implies the existence of what Kant describes as a "common sense", which is another way of saying that the judgement that a particular object is beautiful must be based upon the effect that that object has on the general relation between those cognitive powers which are shared in common by all people.

Up to this point in the argument Kant has been setting out the characteristics of judgements of taste. The idea of purposiveness seems always to refer here to "the mere form of purposiveness in the representation by which an object is given". In particular, the care which he takes to distinguish between judgements of taste, and judgement of perfection, which deal with the adequacy of the thing for its (conceptualised) purpose, seems to me only to be comprehensible if the purposiveness to which he refers is some sort of internal characteristic of the object as experienced. This internal characteristic of the "image" is such as to produce an effect in the relation between imagination and understanding, but the purposiveness is in the image. This is further borne out by a later footnote: "For in this mere judgement of taste no consideration is given to the purpose for which these natural beauties exist: whether to excite pleasure in us, or as purposes without any reference to us at all" (CJ 227 footnote). But in the deduction which establishes the legitimacy of making judgements of taste, and in which he merely spells out once more the idea of the relation between judgements of taste and the harmony of imagination and understanding, he seems to move from the idea of the purposiveness inherent in the image to the idea of the "purposiveness" of the image for bringing about the harmony between imagination and understanding. It will perhaps be best to set out the short "Deduction" in full:

"If it be admitted that, in a pure judgement of taste, the

> satisfaction in the object is combined with the mere act of judging its form, it is nothing else than its subjective purposiveness for the judgement which we feel to be mentally combined with the representation of the object. The judgement, as regards the formal rules of its action, apart from all matter (whether sensation or concept), can only be directed to the subjective conditions of its employment in general (it is applied neither to a particular mode of sense nor to a particular concept of the understanding), and consequently to that subjective (element) which we presuppose in all men (as requisite for possible cognition in general). Thus the agreement of a representation with these conditions of the judgement must be capable of being assumed as valid *a priori* for everyone. That is, we may rightly impute to everyone the pleasure or the subjective purposiveness of the representation for the relation between the cognitive faculties in the act of judging a sensible object in general" (CJ 132).

What exactly is the relation between "purposiveness" in the object and "purposiveness of the representation for the relation between the cognitive faculties"? If we take the word *Zweckmässigkeit* in its broad meaning of "suitability", then, in terms of the argument, it is reasonable to say that the pleasure in the beautiful, if it is to be both subjective and universal, must be a function of the relation between the cognitive faculties, and that, in that sense, the beautiful object is, by definition, 'suitable' for the relation between these faculties. But this is a tautology. The problem is to know *what it is* in the beautiful object that makes it thus suitable. Kant has already argued that it is the 'suitability' of the object in another sense; that is, in the sense of its own internal coherence. This at least has some content. Al l the new introduction of 'purposiveness' for our faculties does is to confuse the argument at this point. But it does (illicitly) facilitate the move from objects as purposive to objects as purposed, and thus helps to make the link with the supersensible.

However, that the principle of the judgement of taste should be one which is related to the harmony of the representative imagination with the rule-giving understanding is important for his general argument. We have seen that this is related to the need for unity. But the unity given by the understanding is the unity of a rule. The understanding is the faculty of rules, of laws. The imagination, on the other hand,

is the faculty which synthesises the manifold of pure intuition (see KS 112).

Kant argued in the CPR that experience always involves:

a) the 'synthesis of reproduction in the imagination', which reproduces or brings together a particular sequence of sense-data, and

b) the 'synthesis of recognition in a concept', that is, the synthesis of the imagination in accordance with a rule.

In this process, the imagination acting by itself is free: "the freedom of the imagination consists in the fact that it schematizes without any concept" (CJ 129). It is only when the imagination is operating in conjunction with the understanding that the experience becomes rule-bound. This raises the possibility of two modes of the relationship:

a) the understanding having to *impose* its rule on the imagination, or

b) the free synthesis of the imagination being such as virtually already to exhibit the rule, in which case the imagination *"in its freedom* harmonizes with the (understanding) *in its conformity to law"* (CJ 129).

Thus the experience of a beautiful object is an experience which illustrates the possibility of harmony between freedom and law.

APPENDIX

A Note on *The Critique of Judgement*

The introduction lays down the main lines of the argument. The body of the work does not really develop them logically. Instead it moves in a series of leaps, possibly under cove of the ambiguities of *Zweckmässigkeit*.

1) The exposition of the idea of the beautiful shows that the idea of the beautiful, as a universal, necessary, but singular judgement is linked to the idea of harmony between the cognitive faculties, and their harmony is related to the idea of the subjective purposiveness of the representation.

2) The deduction shows that the characters discovered in the exposition necessarily go together, and are legitimised through the universal reference to the cognitive faculties. But it also shifts from the idea of the purposiveness of the representation to the idea of purposiveness for the subject, i.e. to the idea that the beautiful object indicates (as a regulative idea) the designedness of nature for occasioning the harmony of the cognitive faculties.

3) The antinomy attempts to show that this purposiveness is rooted in the "supersensible".

The antinomy of the aesthetic judgement consists in the following two propositions:

1) Everyone has his own taste, i.e. the judgement of taste is not based on concepts, for otherwise it would admit of controversy (would be determinable by proofs).
2) There may be a quarrel about taste, i.e. "The judgement of taste is based on concepts, for otherwise, despite its diversity, we could not quarrel about it (we could not claim for our judgement the necessary assent of others) (CJ 184).

The antinomy is solved by distinguishing between determinable

and undeterminable concepts. A determinable concept is one for which an intention can be given that is, as is shown in the *Critique of Pure Reason*, any concept which refers to the phenomenal world. We also, however, have a concept of the noumenal world, because we can think it, but no intention can be given for such a concept; it is thus undeterminable. Because an undeterminable concept is contentless it cannot be used to establish a proof. Therefore it is possible to conceive of a judgement which asserts the universality of a rule - that is, is based on a concept - but which cannot admit of proof because the concept on which it is based refers to the supersensible and thus has no specific content. So we can only resolve the antinomy of taste if we assume that the concept on which the judgement of taste is based in fact refers to the supersensible. This concept is here defined by Kant as "the concept of the general ground of the subjective purposiveness of nature for the judgement" (CJ 185) and it refers to the supersensible "because (the) determining ground lies perhaps in the concept of that which may be the supersensible substrata of humanity" (CJ 185). That is, roughly, if nature is designed to 'fit' the cognitive faculties in the way contained in the experience of beauty, this cannot be a function of its phenomenal nature as a complex of objects relates by mechanical causality. It can only do so by refernce to its noumenal, supersensible, substrata. Thus:

a) the experience of beauty produces an antinomy which, like the antinomies of pure and practical reason, can only be resolved by introducing the undeterminable concept of the supersensible
b) and in this case the supersensible becomes the ground for the purposiveness of nature.

Kant is, however, careful to stress at this point that "the subjective principle, viz, the indefinite idea of the supersensible in us, can only be put forward as the sole key to the puzzle of this faculty whose sources are hidden from us; it can be made no further intelligible" (CJ 186).

We are not intended, furthermore, to assert the reality of the purposiveness of nature for our cognitive faculties. The example of the mechanical formation of beautiful rock crystals shows the

possibility of 'purposiveness' quite without purpose. In any event, the judgement of beauty is necessarily a subjective judgement and cannot be derived from a known purpose (in which case it would be a judgement of perfection). (But note here that Kant seems to have moved back from purposiveness for the cognitive faculty to purpose in the object; for he combines the two in a new way: "The property of nature that gives us occasion to perceive the inner purposiveness in the relation of our mental faculties in judging certain of its products ..." (CJ 196). Also, the argument that the concept of purposiveness is indeterminate implies that the purposiveness referred to it a relation itself directly involving the supersensible (as is the case with the purposiveness of the (phenomenal) representation to the (supersensible) understanding) - rather than involving merely phenomenal purposiven ss without purpose, referred only indirectly to the supersensible through the 'as if' whose purpose it is.

4) In the Introduction Kant gives a deduction of "The principle of the formal purposiveness of nature" as a transcendental principle of judgement" (see CJ p.17). In the Critique of Aesthetical Judgement he establishes the subjective purposiveness of nature in regard to the cognitive faculties. In the Critique of the Theological Judgement he explains the question as to whether we can ascribe objective, real purposiveness to nature:

> "Experience leads our judgement to the concept of an objective and material purposiveness, i.e. to the concept of a purpose of nature, only when we have to judge of a relation of cause to effect which we find ourselves able to apprehend as legitimate only be presupposing the idea of the effect of the causality of the cause as the fundamental condition, in the cause, of the possibility of the effect" (CJ 212-3).

That is, an objective purpose can be ascribed to a being only if a being cannot be accounted for in terms of empirical natural laws:

> "This *contingency* of its form in all empirical natural laws in reference to reason affords a ground for regarding its causality as possible only through reason. For reason, which must cognize the necessity of every form of a natural product in order to comprehend even the conditions of its genesis, cannot assume such (natural) necessity in that particular given form. The causality of its origin is then referred to the faculty of acting in accordance with purposes (a will), and the object which can only thus be

> represented as possible is represented as a purpose. ... I would say provisionally: a thing exists as a natural purpose if it is (although in a double sense) both *cause and effect of itself*. For herein lies a causality the like of which cannot be combined with the mere concept of a nature without attributing to it a purpose; it can certainly be thought without contradiction, but it cannot be comprehended" (CJ 216-217).

Kant gives the example of a tree, which reproduces itself (produces itself generically) and also produces itself in the growth process (produces itself individually) by transforming matter into new products which it uses to sustain itself. Finally, it generates itself in such a way that "the maintenance of any one part depends reciprocally on the maintenance of the rest" (CJ 218). Each part of the tree is both supported by the rest and at the same time has a function in supporting the rest.

That causality which is a category of the understanding is a law of succession; as such it precludes any reciprocal action of that which is effect on that which is cause. In the case of the causality of the will (as a category of reason) however, the object which is purposed may equally well be considered either as the effect of the steps being taken to reach it, or as the cause of those steps. (These two may be distinguished as real and ideal causes). In the case of the tree, the reciprocal relation is that between the parts and the whole; each part is both cause and effect of the whole and vice versa. In a manufactured product the parts are 'caused' by the whole, in the sense that they are designed as a function of the whole which is purposed by the maker. In the case of a natural product, however, the parts not only fit purposively together, but also actually reciprocally produce one another. Such a being is "an organised and self-organizing being". It possesses "a formative power of a self-propagating kind" and such a capacity, Kant argues, "cannot be explained by the mere mechanical faculty of motion"; for part X of whole A to carry out its function in respect of the whole it must contain within it some principle of rule referring to the whole which guides it in its functioning. It is this rule-governed behaviour of the part which cannot be explained according to "the mere mechanical faculty of nature". Further, we can have no direct intuition of the mode of

operation of such a causality; it can, therefore, only "serve as a regulative concept for the reflective judgement, to guide our investigation about objects of this kind by a distant analogy with our own causality according to purposes generally and in our meditations upon their ultimate ground" (CJ 222). But although the idea can function only regulatively, the existence of organised wholes, which can be thought only in terms of such an idea, "afford objective reality to the concept of a *purpose of nature* ... and so they give to the science of nature the basis for a teleology, i.e. a mode of judgement about natural objects according to a special principle which otherwise we should in no way be justified in introducing (because we cannot see *a priori* the possibility of this kind of causality)" (CJ 222). The next step in Kant's argument is the assertion that once we are obliged to admit of the existence of some natural purposes (organised wholes) we are then "justified, nay called upon" (CJ 226) to treat nature as a whole as a natural purpose; in other words, to assume as a principle for the reflective judgement that nothing in it is contingent, and that each being needs to be understood in its relation to the whole. But this means that we must seek the explanations in nature; nature as a whole necessarily has its ground in the supersensible, but we are not entitled to construct special explanations for particular objects or things by reference to the supersensible. The argument is a simple one:

a) The behaviour of organised natural wholes cannot be explained in terms of the mere mechanical faculty of motion.
b) These are objective natural purposes, since this mode of causality involves reciprocal interaction between whole (end) and part (means).
c) The objective existence of -such natural purposes entitles us to treat nature as a whole as purposive, as a regulative idea in the pursuit of natural science.

The dubious point in the argument relates to the question as to whether or not a particular thing - in this case, organised wholes - can or cannot be accounted for merely in terms of empirical natural laws. Kant's main reason for arguing that they cannot arises from his particular concept of natural causality. In the *Critique of Pure Reason* he defines causality as "succession according to a rule". In this sense of the term

the relation between parts and whole cannot be 'causal'. But this may only be a function of the inadequacy of Kant's concept of causality. And, in fact, he has ready to hand in his "Third Analogy of Experience" in the CPR a principle which could perhaps have been developed to give an account of the relation between whole and parts; this is the "Principle of Coexistence, according to the law of reciprocity or community": "All substances, in so far as they can be perceived in space at the same time, exist in a state of complete reciprocity of action" (CPR 161). He shows that it is a condition of the possibility of things themselves as objects of experience that the co-existence of things in space can only be cognised under the principle of their reciprocal action. (I won't go into the argument here, but it does contain the idea of reciprocal influence that is, of a 'causality' other than succession according to a rule.)

5) Dialectic of the teleological judgement.

We have so far, Kant believes, uncovered two different forms of causality; mechanical causality, which was shown in CPR to be one of the *a priori* categories of the understanding, and so constitutive of the possibility of experience; and the law of final causes, which we have discovered by considering the nature of organised wholes. These two types of causality give us two apparently contradictory maxims for the reflective judgement in its search for laws in the science of nature. These are -

> "all production of material things and their forms must be judged to be possible according to merely mechanical laws"
>
> "some products of material nature cannot be judged to be possible according to merely mechanical laws" (CJ 234).

If each of these maxims is made into a constitutive principle of the possibility of objects, we are faced with a contradiction. Although the antinomy can be resolves merely by pointing out that these are regulative maxims of the reflective judgement, the first instructing that we should always attempt to seek further mechanical causes, and the second that we may nevertheless use the idea of function in trying to understand certain objects; nevertheless, the problem remains as to whether the relation between final and mechanical causality may or may

not be resolved within nature; whether purposiveness in nature is "designed" or "undesigned". Kant's response to this is, in essence, a dogmatic assertion:

> "It is indeed quite certain that we cannot adequately cognize, much less explain, organized beings and their internal possibility according to mere mechanical principles of nature, and we can say boldy it is alike certain that it is absurd for men to make any such attempt or to hope that another Newton will arise in the future who shall make comprehensible by us the production of a blade of grass according to natural laws which no design has ordered. We must absolutely deny this insight to men. But then now do we know that in nature, if we could penetrate to the principle by which it specifies the universal laws known to us, there *cannot* lie hidden (in its mere mechanism) a sufficient ground of the possibility of organized beings without supposing any design in their production? Would it not be judged by us presumptuous to say this?" (CJ 248).

That is, no firm decision can be made either way, but we ourselves are incapable of thinking organised wholes other than in terms of 'design' as a regulative idea. Kant does not make clear his reasons for stating that it is impossible to make comprehensible by us the production of a blade of grass according to natural laws which no design has ordered. The reason for stating that it would be presumptuous to affirm the reality of design would seem to be related to the fact that, in terms of sense-datum theory, necessary connection can never be given directly in intuition. The necessary connections in nature which are determinately available to us are those which are imposed by the categories of our own understanding, which do not include the category of final cause. We can never, therefore, have an intuition of the mode of operation of a design in a final cause.

In terms of the categories of our understanding the mode of behaviour of individual objects remains undetermined, and the empirical laws of the natural sciences which we formulate on the basis of an observation of those individual objects are equally contingent. Our reason proceeds in its furthest demand to the unconditioned; it always pushes us on to seek for the unconditioned unity underlying any particular aspect of our experience. But our cognitive faculties are such that only certain categories can be asserted *a priori*, and these only in relation to their applicability to sense intuition. We can therefore never reach this unconditional unity, and it is for this

reason that particular laws and particular events remain contingent for our understanding. We proceed discursively from particular to universal; we have no direct intuition of the universal in its relation to the particular. Nevertheless, we must, in the light of the demands of our reason, posit this unconditional unity as the ground of sensible experience, and therefore posit also that sensible experience is not contingent. This means that we must think of the distinction between contingency and necessity, or between possibility and actuality, as being a function of the nature of our particular cognitive faculties; this in turn implies the idea of "a possible understanding different from the human" (CJ 253).

> "And so we are able to say: Certain natural products, from the special constitution of our understanding, *must be considered by us*, in regard to their possibility, as if produced designedly as purposes. But we do not, therefore, demand that there should be actually given a particular cause which has the representation of a purpose as its determining ground, and we do not deny that an understanding, different from (i.e. higher than) the human, might find the ground of the possibility of such products of nature in the mechanism of nature, i.e. in a causal combination for which an understanding is not explicitly assumed as cause" (CJ 254).

The characteristic of such an understanding, would be that it proceeded not from the concept to the empirical intuition which remains contingent in relation to the concept, but "from the *synthetical-universal* (the intuition of a whole as such) to the particular, i.e. from the whole to the parts" (CJ 255).

Such an understanding, the *intellectus archetypus*, would be able to represent the possibility of the parts as dependent on that of the whole. However, for our understanding "it cannot happen that the whole shall contain the ground of the possibility of the connection of the parts (which would be a contradiction in discursive cognition), but only that the *representation* of a whole may contain the ground of the possibility of its form and the connection of the parts belonging to it" (CJ 256). That is, our understanding can only conceive of the relation in terms of purpose.

This whole analysis brings out very clearly the relation between the idea of purposiveness and Kant's development of the sense-datum model. The relationship is made even more explicit when Kant

goes on to say that "if we were justified in regarding material beings as beings in themselves ... then the unity that constitutes the ground of the possibility of natural formations would be simply the unity of space. But space is no real ground of the products (in terms of Kant's model), but only their formal condition, although it has this similarity to the real ground which we seek that in it no part can be determined except in relation to the whole (the representation of which therefore lies at the ground of the possibility of the parts)" (CJ 257.

For Kant space is a subjective form of sensibility; it does not indicate any real unity in nature. What is given in intuition is the sequence of sense-data which is then ordered in space and time. However, the sense-data themselves are 'given' and retain a contingency which cannot be effectively subsumed under (natural) law because of the absence of any independent unity in the sense-data themselves.

The contingency of natural objects is a function of the form of our sensibility. The *intellectus archetypus*, therefore, would have to be equipped with a distinctive mode of intuition, "a faculty of a *complete spontaneity of intuition* would be a cognitive faculty distinct from sensibility and quite independent of it" (CJ 254).

It is not, however, at all clear what this might mean. That nature which we are investigating is the phenomenal world which is the object of an experience and hence necessarily related to our specific form of sensibility and intuition. Another form of intuition would mean another form of world. Kant seems, in fact, to be referring to an understanding which would have direct access to the noumenal world unhindered by the necessary particularity of our form of sensibility. The noumenal world has its own principle of unity (while the phenomenal world derives its unity from the transcendental ego) and the *intellectus archetypus*, with its spontaneity of intuition, would be able to grasp this unity immediately (not immediately through perception and reflection). This is further implied by the statement with which he ends the paragraph:

> "For external objects as phenomena an adequate ground related to purposes cannot be met with; this, although it lies in nature, must only be sought in the supersensible substrate of

> nature, from all possible insight into which we are cut off. Hence it is absolutely impossible for us to produce from nature itself grounds of explanation for purposive combinations, and it is necessary by the constitution of the human cognitive faculties to seek the supreme ground of these purposive combinations in an original understanding as the cause of the world" (CJ 258; see also p. 267).

The *intellectus archetypus* would have immediate access to the supersensible substrata - but as a consequence the phenomenal world would disappear for it. Because of the incoherence of the original phenomenon-noumenon distinction, it is not possible to think the *intellectus archetypus* coherently. But it does open up the possibility of dismissing the phenomenal world as partial, finite and unreal, because it is based on sensibility, and positing an ideal of knowledge which proceeds from the synthetical-universal to the parts.

For Kant, the idea of the *intellectus archetypus* is not a major theme. He is more concerned with developing the principle of teleology and showing that it can be used without at least contradicting the principle of mechanical causality, since both have to be referred to a supersensible ground in which they are reconciled. But his use of the concept of teleology leads to further 'as if' conclusions which could be synthesised, by later writers, with the *intellectus archetypus* in order to produce the system of absolute idealism.

6) "Methodology of the Teleological Judgement." - The Ultimate Purpose of Nature.

Once we have established the necessity of thinking organised wholes in terms of purpose and, therefore, the necessity also, as a regulative idea, of thinking the whole of nature teleologically, we are in a position to distinguish between 'external purposiveness' and 'internal purposiveness'. The latter is purposive in its own organisation, the former, that which is externally purposive, is something which is useful for the purposes of other beings. The principle of the teleology of nature as a whole implies that we should investigate the way in which things which are not purposive in themselves might be purposive for other things. This implies, however, the existence of something which is an 'ultimate purpose', for which all other natural objects are

mere external purposes. What is thus ultimate purpose? For Kant the answer is simple.

> "(Man) is the ultimate purpose of creation here on earth, because he is the only being upon it who can form a concept of purposes and who can, by his reason, make out of an aggregate of purposively formed things a system of purposes" (CJ 276).

In which way is man the ultimate purpose of nature? *Prima facie*, nature does not seem to be organised in such a way as to bring about the happiness of mankind. In any event, the capacity for happiness is not that which defines man as an end-in-himself. What does so is the capacity to posit his own ends. If, therefore, we are to treat nature as a whole whose ultimate purpose is man, we must investigate it from the point of view of discovering the ways in which it is related to man as a choosing being.

> "The production of the aptitude of a rational being for arbitrary purposes in general (consequently in his freedom) is *culture*. Therefore, culture alone can be the ultimate purpose which we have cause for ascribing to nature in respect to the human race" (CJ 281).

This in turn implies the development of a civil community in which "lawful authority in a whole ... is opposed to the abuse of their conflicting freedoms" (CJ 282), since this is the precodnition for the greatest development of men's natural capacities. Culture is essentially a struggle to discipline our inclinations, "the freeing of the will from the despotism of desire" (CJ 282). "But yet we cannot mistake the purpose of nature - ever aiming to win us away ... which lies hidden in us" (CJ 283-4). Here we have the idea of an historical development, in which nature particpates, to prepare mankind for "a lordship in which man alone shall have authority." That is, the principle of teleology, as a necessary regulative idea, entitles us to conceive of the Kingdom of Ends as the ultimate purpose of nature, to be brought about by an historical process in which culture develops through the establishment of civil community. Once more, as in *Critique of Practical Reason*, Kant argues that the second end of mankind, happiness, can only be conceived of as conjoined with the Kingdom of Ends through the idea of a just God as the supreme legislator who is able to make the whole of nature accord with his purposes.

SUMMARY

As I have suggested at various points in the course of the analysis, Kant has to look for a principle of unity in phenomenal nature, such as could make comprehensible the empirical laws of nature, outside phenomenal nature, because this nature, on the sense-datum theory, has no unity of its own; it is a mere rhapsody of perception. From this point of view, the argument from teleology may be summarised as follows:

1) In order to engage in natural scientific endeavour we have to assume the unity of nature;
2) Phenomenal experience, in so far as it is merely a succession of sense-data, contains no intrinsic unity. Or, to put it another way, nature as phenomenal, under the sense-datum theory, contains no ground of its own unity;
3) The unity of experience is accounted for by the application of the categories by the understanding to the sense-data. However, the categories of the understanding by themselves are inadequate to account for the multiplicity of natural laws;
4) Therefore the unity of nature must have some other ground in the supersensible;
5) But the only relation which we can conceive of between the supersensible and the sensible is that which we derive from our own will; the relationship of purpose, in terms of which an idea (as supersensible) produces a sensible result;
6) Thus the only way in which we can conceive of the unity of nature is by conceiving of nature as if it were the product of purposive understanding. This conclusion must, of course, remain an 'as if', at least until we have some other independent reason for believing that this is in fact the case.

To what extent does all this help Kant in his attempt to give an account of the relationship between phenomenon and noumenon? In so far as the problem is formulated in terms of the relation between freedom (the noumenal) and causality (the phenomenal), both the critique of aesthetic and of teleological judgement are designed to bring these

concepts into harmony. In the Introduction Kant summarises the role of the judgement as follows:

> "It makes possible the transition from the conformity of law in accordance with the former (the concept of nature) to the final purpose in accordance with the latter (the concept of freedom), and this by the concept of a purposiveness of nature" (CJ 33).

This it does in two senses. In a weak sense "purposiveness" is "law-like" behaviour. In a strong sense we can consider purposiveness as a manifestation of an actual purpose, which in turn makes explicit the necessity for comprehending the whole of nature as having its own supersensible ground, which is at the same time the ground for the compatibility of mechanical and teleological explanation, and for the compatibility of the causality of freedom with the causality of nature. The relationship of supersensible ground to nature as phenomenon is now specified as a relationship of purpose, and this at the same time guarantees the ultimate concordance of nature with the human search for the kingdom of ends.

But, for Kant all this must remain on the level of an "as-if". Because of the limitation of our own cognitive faculties we are obliged to use the idea of a natural teleology as a regulative principle for the reflective judgement. But we are not entitled to move beyond this to actually think about the supersensible reality. We cannot understand in general how or in the specific case why it operates in the way it does. To the dichotomy between transcendental I and empirical I is added the dichotomy between phenomenal nature and noumenal nature, and in neither case can we know how the one operates in the other. Even in Kant's terms, all that the *Critique of Judgement* does is to give us an additional ground for believing that the one does act in the other. Neither the empirical ego nor phenomenal nature has any real unity of its own, and that which is the ground of the unity of each is unknowable.

BIBLIOGRAPHY

A: ABBREVIATIONS

I — J.J. Rousseau, A discourse on the origin of inequality, in (tr) G.D.H. Cole, *The Social Contract and Discourses*, Dert, London, 1961.

SC — J.J. Rousseau, The Social Contract, in (tr) G.D.H. Cole, *The Social Contract and Discourses*, Dert, London, 1961.

E — J.J. Rousseau, *Emile*, (tr) B. Foxley, Dert, London, 1969.

PW — (ed.) H. Reiss, *Kant's Political Writings*, (tr) H.B. Nisbet, Cambridge University Press, 1970.

KS — *Immanual Kant's Critique of Pure Reason*, (tr) N. Kemp Smith, MacMillan, London, 1976.

CPR — I. Kant, *Critique of Pure Reason*, (tr) J. Meiklejohn, Dert, London, 1959.

CPrR — I. Kant, *Critique of Practical Reason*, (tr) L. Beck, Bobbs-Merrill, Indianapolis, 1975.

CJ — I. Kant, *The Critique of Judgement*, (tr) J. Bernard, Hafner, New York, 1951.

B: OTHER WORKS CITED

W.G. Runciman, *Relative Deprivation and Social Justice*, Penguin, Harmondsworth, 1972.

L. Colletti, From *Rousseau to Lenin: Studies in Ideology and Society*, (tr) J. Merrington and J. White, New Left Books, London, 1972.

P. Strawson, *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason*, Methven, London, 1968.

D. Hume, *Treatise on Human Nature*, Dert, London, 1956.

M. Heidegger, *Kant and the Problem of Metaphysics*, (tr) J. Churchill, Indians University Press, Indiana, 1962.

J. Murphy, *Kant: The Philosophy of Fright*, MacMillan, London, 1970.

Bibliography.

Works cited or referred to.

Anderson, (ed) Towards Socialism

Austin, J.L. Sense & Sensibilia, OUP: Oxford 1962