CHAPTER ONE

WHAT IS HUMAN AGENCY?

I

1

I would like to explore in this paper what is involved in the notion of a self, of a responsible human agent. What is it that we attribute to ourselves as human agents which we would not attribute to animals?

This question takes us very far indeed, and into several issues of capital importance in philosophy. I am not even going to try to sound them all. But I'd like to make a preliminary exploration of the terrain, using as my guide a key notion which has been introduced recently by Harry Frankfurt, in order to see how well the territory of the self may be mapped with its aid.

The key notion is the distinction between first- and second-order desires which Frankfurt makes in his 'Freedom of the will and the concept of a person'.[1] I can be said to have a second-order desire when I have a desire whose object is my having a certain (first-order) desire. The intuition underlying Frankfurt's introduction of this notion is that it is essential to the characterization of a human agent or person, that is to the demarcation of human agents from other kinds of agent. As he puts it,

> Human beings are not alone in having desires and motives, or in making choices. They share these things with members of certain other species, some of which even appear to engage in deliberation and to make decisions based on prior thought. It seems to be peculiarly characteristic of humans, however, that they are able to form ... second order desires ...[2]

Put in other terms, we think of (at least higher) animals as having desires, even as having to choose between desires in some cases, or at least as inhibiting some desires for the sake of others. But what is distinctively

[1] H. Frankfurt, 'Freedom of the will and the concept of a person', *Journal of Philosophy*, 67:1 (Jan. 1971), pp. 5–20. [2] *Ibid.*, p. 6.

human is the power to *evaluate* our desires, to regard some as desirable and others are undesirable. This is why 'no animal other than man ... appears to have the capacity for reflective self-evaluation that is manifested in the formation of second-order desires'.[3]

I agree with Frankfurt that this capacity to evaluate desires is bound up with our power of self-evaluation, which in turn is an essential feature of the mode of agency we recognize as human. But I believe we can come closer to defining what is involved in this mode of agency if we make a further distinction, between two broad kinds of evaluation of desire.

Thus someone might be weighing two desired actions to determine the more convenient, or how to make different desires compossible (for instance, he might resolve to put off eating although hungry, because later he could both eat and swim), or how to get the most overall satisfaction. Or he might be pondering to see which of two desired objects attracts him most, as one ponders a pastry tray to see if one will take an éclair or a mille feuilles.

But what is missing in the above cases is a qualitative evaluation of my desires; the kind of thing we have, for instance, when I refrain from acting on a given motive – say, spite, or envy – because I consider it base or unworthy. In this kind of case our desires are classified in such categories as higher and lower, virtuous and vicious, more and less fulfilling, more and less refined, profound and superficial, noble and base. They are judged as belonging to qualitatively different modes of life: fragmented or integrated, alienated or free, saintly, or merely human, courageous or pusillanimous and so on.

Intuitively, the difference might be put in this way. In the first case, which we may call weak evaluation, we are concerned with outcomes; in the second, strong evaluation, with the quality of our motivation. But just put this way, it is a little too quick. For what is important is that strong evaluation is concerned with the qualitative *worth* of different desires. This is what is missing in the typical cases where, for example, I choose a holiday in the south rather than the north, or choose to go to lunch at the beach rather than eat now in town. For in these cases, the favoured alternative is not selected because of the worth of the underlying motivation. There is 'nothing to choose' between the motivations here.

But this does not mean (a) that in weak evaluation the motivations are homogeneous. We may not be weighing two objects of the same desire, or put somewhat differently, two outcomes with the same desirability

[3] *Ibid.*, p. 7.

characterization. Take the example of someone who is hesitating between taking a holiday in the south or in the north. What the holiday in the north has going for it is the tremendous beauty of the wild, the untracked wastes, etc.; what the south has going for it is the lush tropical land, the sense of well-being, the joy of swimming in the sea, etc. Or I might put it to myself that one holiday is more exhilarating, the other is more relaxing.

The alternatives have different desirability characterizations; in this sense they are qualitatively distinct. But what is missing in this case is a distinction between the desires as to worth, and that is why it is not a strong evaluation. I ultimately opt for the south over the north not because there is something more worthy about relaxing than being exhilarated, but just because 'I feel like it'.

It follows *a fortiori* (b) that weak evaluations are not simply quantitative either. That is, the alternatives cannot necessarily be expressed in some common units of calculation and in this sense rendered commensurable. This has often been obscured by the recurring ambition of our rationalist civilization to turn practical reflection as much as possible into calculation, an ambition whose major expression has been the doctrine of utilitarianism.

The bent of utilitarianism has been to do away with qualitative distinctions of worth on the grounds that they represent confused perceptions of the real bases of our preferences which are quantitative. The hope has been that once we have done away with strong evaluation we will be able to calculate. Utilitarianism has, I believe, been wrong on both counts. For decisions between alternatives which are not distinguished as to worth are not necessarily amenable to calculation – for instance, the choice between the two holidays above is clearly not so amenable, or only in part (or *some* of the considerations relevant to my choice of holiday might be quantifiable in a strict sense, for example, cost). Nor is there any calculation when I stare at the pastry tray and try to decide whether to have an éclair or a mille feuilles.

All these weak evaluations are only 'quantitative' in the weak sense that they do not involve qualitative distinctions of worth. We sometimes explain our choices of this kind by saying that one alternative was 'more fun', or 'better value'; but there is no genuine quantification behind these expressions; they are just cover terms for 'preferred'. Utilitarians are certainly right from their own standpoint in rejecting strong evaluation, for doing away with this is a necessary condition of reducing practical reason to calculation. But it is far from being a sufficient condition.

Nor can we say (c) that weak evaluation is only concerned with outcomes, and never with desires; that all cases of second-order desires are strong evaluations. For I can have what Frankfurt calls 'second-order volitions' on the basis of weak evaluations. I have a second-order volition when I want certain first-order desires to be the ones which move me to action. So I can want the desire to lunch-and-swim-later to be prepotent, because I know that I will have a better time all things considered, though I fear that I will break down since you are offering me lunch now. And I can have second-order desires on the same kind of basis: I might want my addiction to rich desserts to abate so that I can control my weight. But in both of these cases by hypothesis the alternatives would not be distinguished in that one of the desires was unworthy or base, or alienating, or trivial, or dishonourable, or something of the sort; in short there would be no qualitative distinction of the worth of the motivations.

And just as one can desire not to have a desire one has on the basis of weak evaluation, so one can desire a desire one has not got. Roman banqueters had and acted on this kind of second-order desire when they went to the vomitorium, so as to restore appetite and be able to go on eating with pleasure. This contrasts sharply with the case where I aspire to a desire out of a strong evaluation, where I see it as admirable, for instance, as when I want to be capable of a great and single-minded love or loyalty.[4]

The distinction between the two kinds of evaluation, then, doesn't simply turn on that between quantitative and qualitative evaluation, or on the presence or absence of second-order desires. It concerns rather whether desires are distinguished as to worth. And for this we can perhaps set out two interlocking criteria.

(1) In weak evaluation, for something to be judged good it is sufficient that it be desired, whereas in strong evaluation there is also a use of 'good' or some other evaluative term for which being desired is not sufficient; indeed some desires or desired consummations can be judged as bad, base, ignoble, trivial, superficial, unworthy, and so on.

It follows from this that (2) when in weak evaluation one desired

[4] We might add a fourth reservation and protest that strong evaluation is generally not of desires or motivations, but of qualities of action. I eschew some action because that is a cowardly way to *behave*, or a base *action*. The point is well taken if we mean that we are not speaking of desires alone, but we are seriously mistaken if we think that what is evaluated here are actions *as distinct from* motivations. Cowardly or other kinds of base behaviour are such partly in virtue of their motivation. So that strong evaluation necessarily involves a qualitative distinction of desires.

alternative is set aside, it is only on grounds of its contingent incompatibility with a more desired alternative. I go to lunch later, although hungry now, because then I shall be able to lunch and swim. But I should be happy to have the best of both worlds: if the pool were open now, I could assuage my immediate hunger as well as enjoying a swim at lunch-time.

But with strong evaluation this is not necessarily the case. Some desired consummation may be eschewed not because it is incompatible with another, or if because of incompatibility this will not be contingent. Thus I refrain from committing some cowardly act, although very tempted to do so, but this is not because this act at this moment would make any other desired act impossible, as lunching now would make swimming impossible, but rather because it is base.

But of course there is also a way in which we could characterize this alternative which would bring out incompatibility. If we examine my evaluative vision more closely, we shall see that I value courageous action as part of a mode of life; I aspire to be a certain kind of person. This would be compromised by my giving in to this craven impulse. Here there is incompatibility. But this incompatibility is no longer contingent. It is not just a matter of circumstances which makes it impossible to give in to the impulse to flee and still cleave to a courageous, upright mode of life. Such a mode of life *consists* among other things in withstanding such craven impulses.

That there should be incompatibility of a non-contingent kind here is not adventitious. For strong evaluation deploys a language of evaluative distinctions, in which different desires are described as noble or base, integrating or fragmenting, courageous or cowardly, clairvoyant or blind, and so on. But this means that they are characterized contrastively. Each concept of one of the above pairs can only be understood in relation to the other. No one can have an idea what courage is unless he knows what cowardice is, just as no one can have a notion of 'red', say, without some other colour terms with which it contrasts. It is essential to both 'red' and 'courage' that we understand with what they are contrasted. And of course with evaluative terms, as with colour terms, the contrast may not just be with one other, but with several. And indeed, refining an evaluative vocabulary by introducing new terms would alter the sense of the existing terms, even as it would with our colour vocabulary.

This means that in strong evaluation, we can characterize the alternatives contrastively; and indeed, it can be the case that we must do so if we are to express what is really desirable in the favoured alternative. But this

is not so with weak evaluation.[5] Of course, in each case, we are free to express the alternatives in a number of ways, some of which are and some of which are not contrastive. Thus I can describe my first issue above as between going to lunch *now* and going to lunch *later*; and this is a contrastive description in that it is essential to the identity of one of these alternatives that it not be the other. This is because the term 'now' only has sense through contrast with other terms like 'later', 'earlier', 'tomorrow', and so on. Indeed given the context (e.g., that one cannot decide to lunch in the past), and the contrastive background necessary to 'now', it would be enough to pose my issue to ask myself, 'shall I lunch now?' (or perhaps, 'had I better lunch later?').

But if I want to identify the alternatives in terms of their desirability, the characterization ceases to be contrastive. What lunching now has going for it is that I am hungry, and it is unpleasant to wait while one is hungry and a great pleasure to eat. What eating later has going for it is that I can swim. But I can identify the pleasures of eating quite independently from those of swimming; indeed, I may have enjoyed eating long before swimming entered my life. Not being contrastively described, these two desired consummations are incompatible, where they are, only contingently and circumstantially.

Reciprocally, I can describe the issue of my strong evaluations non-contrastively. I can say that the choice is between saving my life, or perhaps avoiding pain and embarrassment, on one hand, and upholding my honour on the other. Now certainly I can understand preserving my life, and what is desirable about it, without any acquaintance with honour, and the same goes for avoiding pain and embarrassment. And even if the reverse is not quite the case, no one could understand 'honour' without some reference to our desire to avoid death, pain, or embarrassment; for while one preserves honour, among other things, by a certain stance towards such things, even so saving one's honour is not simply contrastively defined with saving one's own life, avoiding pain and so on; there are many cases where one can save one's life without any taint to honour, indeed without the question even arising.

And this non-contrastive description may even be the most apposite for

[5] It might be objected that utilitarians too make use of a qualitative contrast, i.e., that between pleasure and pain. But this is precisely not a qualitative contrast of *desires* of desired consummations, which is what we are considering here. Only pleasure is what is desired, according to utilitarian theory; pain we are averse to. Of course, we might want to contrast the *avoidance of pain*, which in one sense of the term we desire, and pleasure. It is exactly this contrast which utilitarians have notoriously failed to make.

certain purposes. Since there are certainly contingent conditions underlying my being faced with this dread choice of death or dishonour – if only the colonel had not sent me to the front line at just that moment when the enemy were attacking – it is indeed in virtue of a contingent set of circumstances that I must now risk my life to avoid dishonour. But if I focus again on what makes the alternative to be rejected undesirable, that is that running away is in this case incompatible with honour, the incompatibility is no longer a contingent one: honourable conduct just consists in standing in face of such threat to life when this kind of issue is at stake. Or to put it in one word, running is to be eschewed because it is 'cowardly', a word which carries the sense of a non-contingent conflict with honourable conduct.

Thus while another pair of alternatives can be described either contrastively or non-contrastively, when we come to the desirability (or undesirability) characterizations in virtue of which one alternative is rejected, the alternatives in strong evaluation must be contrastively described. For in strong evaluation, where we deploy a language of evaluative distinctions, the rejected desire is not so rejected because of some mere contingent or circumstantial conflict with another goal. Being cowardly does not compete with other goods by taking up the time or energy I need to pursue them, and it may not alter my circumstances in such a way as to prevent my pursuing them. The conflict is deeper; it is not contingent.[6]

2

The utilitarian strand in our civilization would induce us to abandon the language of qualitative contrast, and this means, of course, abandoning our strong evaluative languages, for their terms are only defined in contrast. And we can be tempted to redefine issues we are reflecting on in this non-qualitative fashion.

For instance, let us say that I am addicted to over-eating. I find it hard to resist treating myself to rich desserts. As I struggle with this issue, in the reflection in which I determine that moderation is better, I can be looking at the alternatives in a language of qualitative contrast. I can be reflecting that someone who has so little control over his appetites that he would let his health go to pot over cream-cake is not an admirable person. I yearn to

[6] I am indebted for the present formulation of this point to the vigorous objections of Anne Wilbur Mackenzie against the whole enterprise of distinguishing strong from weak evaluation.

be free of this addiction, to be the kind of person whose mere bodily appetites respond to his higher aspirations, and don't carry on remorselessly and irresistibly dragging him to incapacity and degradation.

But then I might be induced to see my problem in a quite different light. I might be induced to see it as a question of quantity of satisfaction. Eating too much cake increases the cholesterol in my blood, makes me fat, ruins my health, prevents me from enjoying all sorts of other desired consummations; so it isn't worth it. Here I have stepped away from the contrastive language of strong evaluation. Avoiding high cholesterol content, obesity, ill health, or being able to climb stairs, and so on, can all be defined quite independently from my eating habits. Someone might even invent some drug which would allow me to go on eating rich desserts and also enjoy all those other goods, whereas no drug would allow me to eat my cake and attain the dignity of an autonomous, self-desciplined agent which I pined after on my first reading of the issue.

It may be that being talked around to see things in this non-qualitative light will help me solve my problem, that somehow it was too deeply disturbing when I put it in terms of dignity versus degradation, and now I can come to grips with it. But this is a separate question from deciding which way of putting it is more illuminating and true to reality. This is a question about what our motivation really is, how we should truly characterize the meaning things have for us.

This is a conflict of self-interpretations. Which one we adopt will partly shape the meanings things have for us. But the question can arise which is more valid, more faithful to reality. To be in error here is thus not just to make a misdescription, as when I describe a motor-vehicle as a car when it is really a truck. We think of misidentification here as in some sense distorting the reality concerned. For the man who is trying to talk me out of seeing my problem as one of dignity versus degradation, I have made a crucial misidentification. But it is not just that I have called a fear of too high cholesterol content by the name 'degradation'; it is rather that infantile fears of punishment or loss of parental love have been irrationally transferred to obesity, or the pleasures of eating, or something of the sort (to follow a rather vulgar-Freudian line). My experience of obesity, eating, etc. is shaped by this. But if I can get over this 'hang-up' and see the real nature of the underlying anxiety, I will see that it is largely groundless, that is I do not really incur the risk of punishment or loss of love; in fact there is quite another list of things at stake here: ill health, inability to enjoy the outdoor life, early death by heart-attack, and so on.

So might go a modern variant of the utilitarian thrust, trying to reduce our qualitative contrasts to some homogeneous medium. In this it would be much more plausible and sophisticated than earlier variants which talked as though it were just a matter of simple misidentification, that what people sought who pined after honour, dignity, integrity, and so on were simply other pleasurable states to which they gave these high-sounding names.

There are of course ripostes to these attempts to reduce our evaluations to a non-qualitative form. We can entertain the counter-surmise that the rejection of qualitative distinctions is itself an illusion, powered perhaps by an inability to look at one's life in the light of some of the distinctions, a failure of moral nerve, as it were; or else by the draw of a certain objectifying stance towards the world. We might hold that the most hard-bitten utilitarians are themselves moved by qualitative distinctions which remain unadmitted, that they admire the mode of life in which one calculates consciously and clairvoyantly as something higher than the life of self-indulgent illusion, and do not simply elect it as more satisfying.

We cannot resolve this issue here. The point of introducing the distinction between strong and weak evaluation is to contrast the different kinds of self that each involves. In examining this it will, I think, become overwhelmingly plausible that we are not beings whose only authentic evaluations are non-qualitative as the utilitarian tradition suggests.

A subject who only evaluates weakly – that is, makes decisions like that of eating now or later, taking a holiday in the north or in the south – such a subject we might call a simple weigher of alternatives. And the other, who deploys a language of evaluative contrasts ranging over desires, we might call a strong evaluator.

Now we can concur that a simple weigher is already reflective in a minimal sense, in that he evaluates courses of action, and sometimes is capable of acting out of that evaluation as against under the impress of immediate desire. And this is a necessary feature of what we call a self or a person. He has reflection, evaluation and will. But in contrast to the strong evaluator he lacks something else which we often speak of with the metaphor of 'depth'.

The strong evaluator envisages his alternatives through a richer language. The desirable is not only defined for him by what he desires, or what he desires plus a calculation of consequences; it is also defined by a qualitative characterization of desires as higher and lower, noble and base, and so on. Reflection is not just a matter, where it is not calculation of consequences, of registering the conclusion that alternative A is more attractive to me, or draws me more than B. Rather the higher desirability

of A over B is something I can articulate if I am reflecting as a strong evaluator. I have a vocabulary of worth.

In other words, the reflection of the simple weigher terminates in the inarticulate experience that A is more attractive than B. I am presented with the pastry tray, I concentrate on it, hesitate between an éclair and a mille feuilles. It becomes clear to me that I feel more like an éclair now, so I take it. Of course, one can say a lot more about the attractiveness of the alternatives in other cases of simple weighing. For instance, when I was choosing between a holiday in the north and one in the south, I talked about the tremendous beauty of the north, of the wild, the sense of untracked wastes, etc., or the lush tropical land, the sense of well-being, the joys of swimming in the sea, etc. All this can be expressed. What cannot be expressed is what makes the south, my ultimate choice, superior.

Thus faced with incommensurables, which is our usual predicament, the simple weigher's experiences of the superiority of A over B are inarticulable. The role of reflection is not to make these articulate, but rather to step back from the immediate situation, to calculate consequences, to compensate for the immediate force of one desire which might not be the most advantageous to follow (as when I put off lunch to swim-with-lunch later), to get over hesitation by concentrating on the inarticulate 'feel' of the alternatives (do I really feel like an éclair or a mille feuilles?).

But the strong evaluator is not similarly inarticulate. There is the beginning of a language in which to express the superiority of one alternative, the language of higher and lower, noble and base, courageous and cowardly, integrated and fragmented, and so on. The strong evaluator can articulate superiority just because he has a language of contrastive characterization.[7]

So within an experience of reflective choice between incommensurables, strong evaluation is a condition of articulacy, and to acquire a

[7] It is because the alternatives are characterized in a language of qualitative contrast that strong evaluative choices show the feature we mentioned above, that the rejected alternative is not rejected because of some merely contingent or circumstantial conflict with the goal chosen. To have a language of qualitative contrast is to characterize the noble essentially as in contrast to the base, the courageous to the cowardly, and so on.

With this in mind we can see right away how the holiday preference could become articulate, too. We might decide to go south rather than north because we will have a more humanly meaningful or uplifting experience in visiting some ancient civilization than in being away from any traces of man. With this example we can also see that languages of strong evaluation do not have to be exclusively ethical, as one might have surmised from the examples above; they can also be aesthetic and of other kinds as well.

strongly evaluative language is to become (more) articulate about one's preferences. I cannot tell you perhaps very volubly why Bach is greater than Liszt, say, but I am not totally inarticulate: I can speak of the 'depth' of Bach, for instance, a word one only understands against a corresponding use of 'shallow', which, unfortunately, applies to Liszt. In this regard I am way ahead of where I am in articulating why I now prefer that éclair to the mille feuilles; about this I can say nothing (not even that it tastes better, which I could say, for instance, in explaining my preference for éclairs over brussels-sprouts; but even this is on the verge of inarticulacy – compare our replying above that Bach 'sounds better'). And I am also way ahead of where I might be if I had never aquired any language to talk about music, if it were a quite inarticulable experience for me (of course, it would then be a very different experience).

To be a strong evaluator is thus to be capable of a reflection which is more articulate. But it is also in an important sense deeper.

A strong evaluator, by which we mean a subject who strongly evaluates desires, goes deeper, because he characterizes his motivation at greater depth. To characterize one desire or inclination as worthier, or nobler, or more integrated, etc. than others is to speak of it in terms of the kind of quality of life which it expresses and sustains. I eschew the cowardly act above because I want to be a courageous and honourable human being. Whereas for the simple weigher what is at stake is the desirability of different consummations, those defined by his *de facto* desires, for the strong evaluator reflection also examines the different possible modes of being of the agent. Motivations or desires do not only count in virtue of the attraction of the consummations but also in virtue of the kind of life and kind of subject that these desires properly belong to.[8]

[8] To be a strong evaluator is thus to see desires in an additional dimension. And this is in fact essential to our important evaluative distinctions. It has been remarked for instance that the criteria of a courageous act cannot be given simply in terms of external achievement in a given context. Someone may rush the machine-guns out of stupidity, or drunk with frenzy, or because he has had too much of life. It is not sufficient just that he see the danger, a condition which is met in the last two cases. Or suppose a man is driven with some uncontrollable lust, or hatred, or desire for revenge, so that he runs out into danger. This is not courage either, so long as we see him as *driven*.

Courage requires that we face danger, feel the fear which is appropriate, and nevertheless over-rule the impulse to flee because we in some sense dominate it, because we are moved by something higher than mere impulse or the mere desire to live. It may be glory, or the love of country, or the love of some individuals we are saving, or a sense of our own integrity. Implicit in all of these is that the courageous man is moved by what we can at least think of as seen by him to be higher. If someone for instance thought that there was nothing higher than life and the avoidance of pain, and believed that no one could sanely and responsibly think otherwise, he would have no place in his vocabulary for physical

But this additional dimension can be said to add depth, because now we are reflecting about our desires in terms of the kind of being we are in having them or carrying them out. Whereas a reflection about what we feel like more, which is all a simple weigher can do in assessing motivations, keeps us as it were at the periphery; a reflection on the kind of beings we are takes us to the centre of our existence as agents. Strong evaluation is not just a condition of articulacy about preferences, but also about the quality of life, the kind of beings we are or want to be. It is in this sense deeper.

And this is what lies behind our ordinary use of the metaphor of depth applied to people. Someone is shallow in our view when we feel that he is insensitive, unaware or unconcerned about issues touching the quality of his life which seem to us basic or important. He lives on the surface because he seeks to fulfil desires without being touched by the 'deeper' issues, what these desires express and sustain in the way of modes of life; or his concern with such issues seems to us to touch on trivial or unimportant questions, for example, he is concerned about the glamour of his life, or how it will appear, rather than the (to us) real issues of the quality of life.

The compleat Utilitarian would be an impossibly shallow character, and we can gauge how much self-declared Utilitarians really live their ideology by what importance they attribute to depth.

3

Thus the strong evaluator has articulacy and depth which the simple weigher lacks. He has, one might say, articulacy about depth. But where there is articulacy there is the possibility of a plurality of visions which there was not before. The simple weigher may hesitate, as before the éclair and mille feuilles, and his momentary preference may go back and forth. But we would not say that he envisages his situation of choice now one way, now another. With strong evaluation, however, there can be and often is a plurality of ways of envisaging my predicament, and the choice may be not just between what is clearly the higher and the lower, but between two incommensurable ways of looking at this choice.

Let us say that at the age of 44 I am tempted to pack up, abandon my job and go to some other quite different job in Nepal. One needs to renew the sources of creativity, I tell myself, one can fall into a deadening routine,

courage. Any act that might appear to qualify for this title would have to be classified by him as foolhardy, mad, or moronically insensitive to reality, or something of the kind. If we can think of gangsters as being heroic it is because in this post-Romantic age we see something admirable in people living some grand design to the ultimate end, whatever it be.

go stale, simply go through the motions of teaching the same old courses; this way is premature death. Rather rejuvenation is something one can win by courage and decisive action; one must be ready to make a break, try something totally new; and so on, and so on. All this I tell myself when the mood is on me. But then at other moments, this seems like a lot of adolescent nonsense. In fact nothing in life is won without discipline, hanging in, being able to last through periods of mere slogging until something greater grows out of them. One has to have a long breath, and standing loyalties to a certain job, a certain community; and the only meaningful life is that which is deepened by carrying through these commitments, living through the dead periods in order to lay the foundations for the creative ones; and so on.

We see that, unlike the choice between éclairs and mille feuilles, or the vacations north and south, where we have two incommensurable objects attracting us, we have here 'objects' – courses of action – which can only be characterized through the qualities of life they represent, and characterized contrastively. It is part of the desirability characterization of each that it has an undesirability story to tell about the other. But the struggle here is between two such characterizations, and this introduces a new incommensurability. When I am feeling that the break to Nepal is the thing, my desire to stay is a kind of pusillanimity, a weary enmiring in routine, the growing action of a sclerosis that I can only cure by 'splitting'. It is far from being the quietly courageous loyalty to an original line of life, hanging in during the dry period to permit a fuller flowering. And when I am for staying, my trip to Nepal looks like a lot of adolescent nonsense, an attempt to be young again by just refusing to act my age, hardly great liberation, renewal and all that.

We have here a reflection about what to do which is carried on in a struggle of self-interpretations, like our example above of the man struggling against his addiction to rich desserts. The question at issue concerns which is the truer, more authentic, more illusion-free interpretation, and which on the other hand involves a distortion of the meanings things have for me. Resolving this issue is restoring commensurability.

II

1

Starting off from the intuition that the capacity for second-order desires, or evaluating desires, is essential to human agency, I have tried to distinguish two kinds of such evaluation. I hope the discussion has also served to make

this basic intuition more plausible, if indeed it lacked any plausibility at the outset. It must be clear that an agent who could not evaluate desires at all would lack the minimum degree of reflectiveness which we associate with a human agent, and would also lack a crucial part of the background for what we describe as the exercise of will.

I should also like to add, but with perhaps less certainty of universal agreement, that the capacity for strong evaluation in particular is essential to our notion of the human subject; that without it an agent would lack a kind of depth we consider essential to humanity, without which we would find human communication impossible (the capacity for which is another essential feature of human agency). But I will not try to argue this case here. The question would revolve around whether one could draw a convincing portrait of a human subject to whom strong evaluation was quite foreign (is Camus' Meursault such a case?), since in fact the human beings we are and live with are all strong evaluators.

But for the remainder of this paper, I should like to examine another avenue of the self, that of responsibility, with the aid of the key notion of second-order desire. For we think of persons as responsible as well, in a way that animals are not, and this too seems bound up with the capacity to evaluate desires.

There is one sense of responsibility which is already implicit in the notion of will. A being capable of evaluating desires may find that the upshot of such evaluation is in conflict with the most urgent desire. Indeed, we might think of it as a necessary feature of the capacity to evaluate desires that one be able to distinguish the better one from the one that presses most strongly.

But in at least our modern notion of the self, responsibility has a stronger sense. We think of the agent not only as partly responsible for what he does, for the degree to which he acts in line with his evaluations, but also as responsible in some sense for these evaluations.

This sense is even suggested by the word 'evaluation', which belongs to the modern, one might almost say post-Nietzschean, vocabulary of moral life. For it relates to the verb 'evaluate', and the verb here implies that this is something we do, that our evaluations emerge from our activity of evaluation, and in this sense are our responsibility.

This active sense is conveyed in Frankfurt's formulation where he speaks of persons as exhibiting 'reflective self-evaluation that is manifested in the formation of second-order desires'.

Or we might put the suggestion another way. We have certain *de facto*, first-order desires. These are given, as it were. But then we form evaluations, or second-order desires. But these are not just given, they are also endorsed, and in this sense they engage our responsibility.

How are we to understand this responsibility? An influential strand of thought in the modern world has wanted to understand it in terms of choice. The Nietzschean term 'value', suggested by our 'evaluation', carries this idea that our 'values' are our creations, that they ultimately repose on our espousing them. But to say that they ultimately repose on our espousing them is to say that they issue ultimately from a radical choice, that is, a choice which is not grounded in any reasons. For to the extent that a choice is grounded in reasons, these are simply taken as valid and are not themselves chosen. If our 'values' are to be thought of as chosen, then they must repose finally on a radical choice in the above sense.

This is, of course, the line taken by Sartre in *L'Être et le Néant*, where he argues that the fundamental project which defines us reposes on a radical choice. The choice, Sartre puts it with his characteristic flair for striking formulae, is 'absurde, en ce sens, qu'il est ce par quoi ... toutes les raisons viennent à l'être'.[9] This idea of radical choice is also defended by an influential Anglo-Saxon school of moral philosophers.

But in fact we cannot understand our responsibility for our evaluations through the notion of radical choice – not if we are to go on seeing ourselves as strong evaluators, as agents with depth. For a radical choice *between* strong evaluations is quite conceivable, but not a radical choice *of* such evaluations.

To see this we might examine a famous Sartrian example, which turns out, I believe, to illustrate the exact opposite of Sartre's thesis, the example in *L'Existentialisme est un Humanisme* of the young man who is torn between remaining with his ailing mother and going off to join the Resistance. Sartre's point is that there is no way of adjudicating between these two strong claims on his moral allegiance through reason or the reliance on some kind of over-reaching considerations. He has to settle the question, whichever way he goes, by radical choice.

Sartre's portrayal of the dilemma is very powerful. But what makes it plausible is precisely what undermines his position. We see a grievous moral dilemma because the young man is faced here with two powerful moral *claims*. On the one hand his ailing mother may well die if he leaves her, and die in the most terrible sorrow, not even sure that her son still lives; on the other hand is the call of his country, conquered and laid waste by the enemy, and not only his country, for this enemy is destroying

[9] J. P. Sartre, *L'Être et le Néant* (Paris, 1943), p. 559.

the very foundation of civilized and ethical relations between men. A cruel dilemma, indeed. But it is a dilemma only because the claims themselves are not created by radical choice. If they were the grievous nature of the predicament would dissolve, for that would mean that the young man could do away with the dilemma at any moment by simply declaring one of the rival claims as dead and inoperative. Indeed, if serious moral claims were created by radical choice, the young man could have a grievous dilemma about whether to go and get an ice cream cone, and then again he could decide not to.

It is no argument against the view that evaluations do not repose on radical choice that there are moral dilemmas. Why should it even be surprising that the evaluations we feel called upon to assent to may conflict, even grievously, in some situations? I would argue that the reverse is the case, that moral dilemmas become inconceivable on the theory of radical choice.

Now in this hypothetical case the young man has to resolve the matter by radical choice. He simply has to plump for the Resistance or for staying at home with his mother. He has no language in which the superiority of one alternative over the other can be articulated; indeed, he has not even an inchoate sense of the superiority of one over the other; they seem quite incommensurable to him. He just throws himself one way.

This is a perfectly understandable sense of radical choice. But then imagine extending this to all cases of moral action. Let us apply it to the case where I have an ailing mother and no rival obligation. Do I stay, or do I go for a holiday on the Riviera? There is no doubt I should stay. Of course, I *may* not stay. In this sense, there is also a 'radical choice' open: whether to do what we ought or not (although here I might put forward all sorts of rationalizations for going to the Côte d'Azur: I owe it to myself, after all I have faithfully taken care of her all these years while my brothers and sisters have gone off, and so on). But the question is whether we can construe the determination of what we ought to do here as issuing from a radical choice.

What would this look like? Presumably, we would be faced with the two choices, to stay with my mother or go to the south. On the level of radical choice these alternatives have as yet no contrastive characterization, that is, one is not the path of duty, while the other is that of selfish indulgence, or whatever.

This contrastive description will be created by radical choice. So what does this choice consist in? Well, I might ponder the two possibilities, and then I might just find myself doing one rather than another. But this brings

us to the limit where choice fades into non-choice. Do I really choose if I just start doing one of the alternatives? And above all, this kind of resolution has no place for the judgement 'I owe it to my mother to stay', which is supposed to issue from the choice.

What is it to have this judgement issue from radical choice? Not that on pondering the alternatives, the sense grows more and more strongly that this judgement is *right*, for this would not be an account of radical choice, but rather of our coming to see that our obligation lay here. This account would present obligations as issuing not from radical choice but from some kind of vision of our moral predicament. This choice would be grounded. What is it then for radical choice to issue in this judgement? Is it just that I find myself assenting to the judgement, as in the previous paragraph I found myself doing one of the two actions? But then what force has 'assenting to the judgement'? I can certainly just find myself saying 'I owe it to my mother', but this is surely not what it is to assent. I can, I suppose, find myself feeling suddenly, 'I owe this to my mother'; but then what grounds are there for thinking of this as a choice?

In order for us to speak of choice, we cannot just find ourselves in one of the alternatives. We have in some sense to experience the pull of each and give our assent to one. But what kind of pull do the alternatives have here? What draws me to the Côte d'Azur is perhaps unproblematic enough, but what draws me to stay with my mother cannot be the sense that I owe it to her, for that *ex hypothesi* has to issue from the choice. It can only be a *de facto* desire, like my desire for the sun and sea of the Côte d'Azur. But then the choice here is like the choice of the two holidays in the previous section. I feel the attraction of these two incommensurable alternatives, and after I ponder them I find that one begins to become prepotent, it draws me more. Or perhaps, the matter obstinately refuses to resolve itself, and I say at one moment, 'what the hell, I'll stay'.

The agent of radical choice has to choose, if he chooses at all, like a simple weigher. And this means that he cannot properly speaking be a strong evaluator. For all his putative strong evaluations issue from simple weighings. The application of a contrastive language which makes a preference articulate reposes on fiat, a choice made between incommensurables. But then the application of the contrastive language would be in an important sense bogus. For by hypothesis the experience on which the application reposed would be more properly characterized by a preference between incommensurables; the fundamental experience which was supposed to justify this language would in fact be that of the simple weigher, not of the strong evaluator. For again by hypothesis,

what leads him to call one alternative higher or more worthy is not that in his experience it appears to be so, for then his evaluations would be judgements, not choices; but rather that he is led to plump for one rather than the other after considering the attractiveness of both alternatives.

But of course, even this account of choice would not be acceptable to the theorist of radical choice. He would refuse the assimilation of these choices to such decisions as whether to go south or north for my holiday. For these choices are not supposed to be simply the registration of my preferences, but radical choices. But what is a radical choice which is not even a registration of preference? Well, it may be that I just decide, just throw myself one way rather than another. I just say, 'what the hell, I'll stay'. But this, of course, I can do in the holiday choice case, where, for instance, I cannot seem to make up my mind which is preferable. This does not distinguish the two cases.

Perhaps then it is that in radical choice I do not consult preferences at all. It is not that I try to see which I prefer, and then failing to get a result, I throw myself one way or the other; but rather, this kind of choice is made quite without regard to preferences. But then with regard to what is it made? Here we border on incoherence. A choice made without regard to anything, without the agent feeling any solicitation to one alternative or the other, or in complete disregard of such solicitation: is this still choice? What could it be? Well, suddenly he just goes and takes one of the alternatives. Yet, but this he could do in a fit of abstraction. What makes it a choice? It must be something to do with what he is thinking out of which this act comes. But what could that be? Can it just be that he is thinking something like 'I must take one of them, I must take one of them', repeating it to himself in a fever? Surely not. Rather he must be pondering the alternatives, be in some way considering their desirability, and the choice must be in some way related to that. Perhaps he judges that A is by all criteria more desirable, and then he chooses B. But if this is a choice and not just an inexplicable movement, it must have been accompanied by something like: 'damn it, why should I always choose by the book, I'll take B'; or maybe he just suddenly felt that he really wanted B. In either case, his choice clearly relates to his preference, however suddenly arising and from whatever reversal of criteria. But a choice utterly unrelated to the desirability of the alternatives would not be intelligible as a choice.

The theory of radical choice in fact is deeply incoherent, for it wants to maintain both strong evaluation and radical choice. It wants to have strong evaluations and yet deny their status as judgements. And the result is that on close examination, it crumbles; in order to maintain its coherence the theory of radical choice in fact mutates into something quite different. Either we take seriously the kinds of consideration which weigh in our moral decisions, and then we are forced to recognize that these are for the most part evaluations which do not issue from radical choice; or else we try at all costs to keep our radical choice independent of any such evaluations, but then it ceases to be a choice of strong evaluations, and becomes a simple expression of preference, and if we go farther and try to make it independent even of our *de facto* preferences, then we fall ultimately into a criteria-less leap which can not properly be described as choice at all.

In fact the theory maintains a semblance of plausibility by surreptitiously assuming strong evaluation beyond the reach of radical choice, and that in two ways. First, the real answer to our attempted assimilation of radical moral choice to the mere preference of a simple weigher is that the choices talked about in the theory are about basic and fundamental issues, like the choice of our young man above between his mother and the Resistance. But these issues are basic and fundamental not in virtue of radical choice; their importance is given, or revealed in an evaluation which is constated, not chosen. The real force of the theory of radical choice comes from the sense that there are different moral perspectives, that there is a plurality of moral visions, as we said in the previous section, between which it seems very hard to adjudicate. We can conclude that the only way of deciding between these is by the kind of radical choice that our young man had to take.

And this in turn leads to a second strong evaluation beyond the reach of choice. If this is the predicament of man, then it is plainly a more honest, more clairvoyant, less confused and self-deluding stance to be aware of this and take the full responsibility for the radical choice. The stance of 'good faith' is higher, and this not in virtue of radical choice, but in virtue of our characterization of the human predicament in which radical choice has such an important place. Granted this is the moral predicament of man, it is more honest, courageous, self-clairvoyant, hence a higher mode of life, to choose in lucidity than it is to hide one's choices behind the supposed structure of things, to flee from one's responsibility at the expense of lying to oneself, of a deep self-duplicity.

When we see what makes the theory of radical choice plausible we see how strong evaluation is something inescapable in our conception of the agent and his experience; and this because it is bound up with our notion of the self. So that it creeps back in even where it is supposed to have been excluded.

2

We can see this from a different angle if we consider another way of showing the theory of radical choice to be wrong. I mentioned in the last section that strong evaluators can be called deep because what weighs with them are not only the consummations desired but also what kind of life, what quality of agent they are to be. This is closely connected with the notion of identity.

By 'identity' I mean that use of the term where we talk about 'finding one's identity', or going through an 'identity crisis'. Now our identity is defined by our fundamental evaluations. The answer to the question 'What is my identity?' cannot be given by any list of properties of other ranges, about my physical description, provenance, background, capacities, and so on. All these can figure in my identity, but only as assumed in a certain way. If my being of a certain lineage is to me of central importance, if I am proud of it, and see it as conferring on me membership in a certain class of people whom I see as marked off by certain qualities which I value in myself as an agent and which come to me from this background, then it will be part of my identity. This will be strengthened if I believe that men's moral qualities are to a great extent nourished by their background, so that to turn against one's background is to reject oneself in an important way.

So my lineage is part of my identity because it is bound up with certain qualities I value, or because I believe that I must value these qualities since they are so integrally part of me that to disvalue them would be to reject myself. In either case, the concept of identity is bound up with that of certain strong evaluations which are inseparable from myself. This either because I identify myself by my strong evaluations, as someone who essentially has these convictions; or else because I see certain of my other properties as admitting of only one kind of strong evaluation by myself, because these properties so centrally touch what I am as an agent, that is, as a strong evaluator, that I cannot really repudiate them in the full sense. For I would be thereby repudiating myself, inwardly riven, and hence incapable of fully authentic evaluation.

Our identity is therefore defined by certain evaluations which are inseparable from ourselves as agents. Shorn of these we would cease to be ourselves, by which we do not mean trivially that we would be different in the sense of having some properties other than those we now have – which would indeed be the case after any change, however minor – but that shorn of these we would lose the very possibility of being an agent who evaluates; that our existence as persons, and hence our ability to adhere as

persons to certain evaluations, would be impossible outside the horizon of these essential evaluations, that we would break down as persons, be incapable of being persons in the full sense.

Thus, if I were forced by torture or brainwashing to abandon these convictions by which I define my identity, I would be shattered, I would no longer be a subject capable of knowing where I stood and what the meanings of things were for me, I would suffer a terrifying breakdown of precisely those capacities which define a human agent. Or if, to take the other example, I were somehow induced to repudiate my lineage, I would be crippled as a person, because I would be repudiating an essential part of that out of which I evaluate and determine the meanings of things for me. Such repudiation would both be itself inauthentic and would make me incapable of other authentic evaluations.

The notion of identity refers us to certain evaluations which are essential because they are the indispensable horizon or foundation out of which we reflect and evaluate as persons. To lose this horizon, or not to have found it, is indeed a terrifying experience of disaggregation and loss. This is why we can speak of an 'identity-crisis' when we have lost our grip on who we are. A self decides and acts out of certain fundamental evaluations.

This is what is impossible in the theory of radical choice. The agent of radical choice would at the moment of choice have *ex hypothesi* no horizon of evaluation. He would be utterly without identity. He would be a kind of extensionless point, a pure leap into the void. But such a thing is an impossibility, or rather could only be the description of the most terrible mental alienation. The subject of radical choice is another avatar of that recurrent figure which our civilization aspires to realize, the disembodied ego, the subject who can objectify all being, including his own, and choose in radical freedom. But this promised total self-possession would in fact be the most total self-loss.

3

What then is the sense we can give to the responsibility of the agent, if we are not to understand it in terms of radical choice? Do we have to conclude that we are not in any sense responsible for our evaluations?

I think not. For there is another sense in which we are responsible. Our evaluations are not chosen. On the contrary they are articulations of our sense of what is worthy, or higher, or more integrated, or more fulfilling, and so on. But as *articulations*, they offer another purchase for the concept of responsibility. Let us examine this.

Much of our motivation – our desires, aspirations, evaluation – is not simply given. We give it a formulation in words or images. Indeed, by the fact that we are linguistic animals our desires and aspirations cannot but be articulated in one way or another. Thus we are not simply moved by psychic forces comparable to such forces as gravity or electro-magnetism, which we can see as given in a straightforward way, but rather by psychic 'forces'[10] which are articulated or interpreted in a certain way.

Now these articulations are not simply descriptions, if we mean by this characterizations of a fully independent object, that is, an object which is altered neither in what it is, nor in the degree or manner of its evidence to us by the description. In this way my characterization of this table as brown, or this line of mountains as jagged, is a simple description.

On the contrary, articulations are attempts to formulate what is initially inchoate, or confused, or badly formulated. But this kind of formation or reformulation does not leave its object unchanged. To give a certain articulation is to shape our sense of what we desire or what we hold important in a certain way.

Let us take the case above of the man who is fighting obesity and who is talked into seeing it as a merely quantitative question of more satisfaction, rather than as a matter of dignity and degradation. As a result of this change, his inner struggle itself becomes transformed, and is now quite a different experience.

The opposed motivations – the craving for cream cake and his dissatisfaction with himself at such indulgence – which are the 'objects' undergoing rediscription here, are not independent in the sense outlined above. When he comes to accept the new interpretation of his desire to control himself, the desire itself has altered. True, it may be said on one level to have the same goal, that he stop eating cream cake, but since it is no longer understood as a seeking for dignity and self-respect it has become quite a different kind of motivation.

Of course, even here we often try to preserve the identity of the objects undergoing redescription – so deeply rooted is the ordinary descriptive model. We might think of the change, say, in terms of some immature sense of shame and degradation being detached from our desire to resist over-indulgence, which has now simply the rational goal of increasing

[10] I put the expression in quotes here because the underlying motivation which we want to speak of in terms of psychic 'forces' or 'drives' is only accessible through interpretation of behaviour or feeling. The line here between metaphor and basic theory is very hard to draw. Cf. Paul Ricoeur, *De L'Interprétation* (Paris, 1965) and my 'Force et sens' in G. Madison (ed.), *Sens et Existence* (Paris, 1975).

over-all satisfaction. In this way we might maintain the impression that the elements are just rearranged while remaining the same. But on a closer look we see that on this reading, too, the sense of shame does not remain self-identical through the change. It dissipates altogether, or becomes something quite different.

We can say therefore that our self-interpretations are partly constitutive of our experience. For an altered description of our motivation can be inseparable from a change in this motivation. But to assert this connection is not to put forward a causal hypothesis: it is not to say that we alter our descriptions and then *as a result* our experience of our predicament alters. Rather it is that certain modes of experience are not possible without certain self-descriptions. The particular quality of experience in the obesity case where I approach the alternative purely as a balance of utility, where I am free from the menace of degradation and self-contempt, cannot be without my characterizing the two rival desires in this 'deflated' way, as two different kinds of advantage. This deflated description is part of the objectifying, calculative way I now experience the choice. We can say that it is 'constitutive' of this experience, and this is the term I shall use for this relation.

But the fact that self-interpretations are constitutive of experience says nothing about how changes in both descriptions and experience are brought about. It would appear in fact that change can be brought about in two different ways. In some circumstances we are led to reflect, on our own or in exchange with others, and can sometimes win through to a new way of seeing our predicament, and hence a change in our experiences. But more fundamentally, it would apear that certain descriptions of experience are unacceptable or incomprehensible to some people because of the nature of their experience. To someone who strongly experiences the fight against obesity in terms of degradation, the 'deflated' descriptions appear a wicked travesty, a shameless avoidance of moral reality – rather as we react to the hiding of political crime through Orwellian language, for example renaming mass murder a final 'solution'.

That description and experience are bound together in this constitutive relation admits of causal influences in both directions: it can sometimes allow us to alter experience by coming to fresh insight; but more fundamentally it circumscribes insight through the deeply embedded shape of experience for us.

Because of this constitutive relation, our descriptions of our motivations, and our attempts to formulate what we hold important, are not simple descriptions in that their objects are not fully independent. And

yet they are not simply arbitrary either, such that anything goes. There are more or less adequate, more or less truthful, more self-clairvoyant or self-deluding interpretations. Because of this double fact, because an articulation can be *wrong*, and yet it shapes what it is wrong about, we sometimes see erroneous articulations as involving a distortion of the reality concerned. We do not just speak of error but frequently also of illusion or delusion.

We could put the point this way. Our attempts to formulate what we hold important must, like descriptions, strive to be faithful to something. But what they strive to be faithful to is not an independent object with a fixed degree and manner of evidence, but rather a largely inarticulate sense of what is of decisive importance. An articulation of this 'object' tends to make it something different from what it was before.

And by the same token a new articulation does not leave its 'object' evident or obscure to us in the same manner or degree as before. In the fact of shaping it, it makes it accessible and/or inaccessible in new ways. This is in fact well illustrated by our example of the man fighting obesity.

Now our articulations, just because they partly shape their objects, engage our responsibility in a way that simple descriptions do not. This happens in two related ways which correspond to the two directions of causal influence mentioned above.

First, because our insights into our own motivations and into what is important and of value are often limited by the shape of our experience, failure to understand a certain insight, or see the point of some moral advice proffered, is often taken as a judgement on the character of the person concerned. An insensitive person, or a fanatic, cannot see what he is doing to others, the kind of suffering he is inflicting on them. He cannot see, for instance, that this act is a deep affront to someone's sense of honour, or perhaps deeply undermines his sense of worth. He is proof to all remonstrating on our part.

He cannot listen to us because he has closed off all sensitivity to questions of honour, or perhaps to the sense of personal worth, in himself, say; and this might in turn be related to earlier experiences which he has undergone. These earlier experiences account for a shape of his current experience in which these issues figure as specious and of no account, and this current shape makes it impossible for him to allow the insights we are pressing on him. He cannot admit them without his whole stance towards these matters crumbling; and this stance may be motivationally of deep importance to him.

But in his kind of case we take the limits of the man's insight as a

judgement on him. It is because of what he has become – perhaps indeed, in response to some terrible strain or difficulty, but nevertheless what he has become – that he cannot see certain things, cannot understand the point of certain descriptions of experience. In the sense of 'responsibility' where we only attribute it to people in relation to outcomes that they can presently encompass or avoid, we should not speak of responsibility here. And even if we take account of what the agent could have done differently in the past, the responsibility in this sense may be very attenuated, for example when people have been marked by truly harrowing early experiences.

But in another sense of 'responsibility', one older than our modern notions of moral agency, we hold them responsible in that we judge them morally on the basis of what they see or do not see. So that a man may condemn himself by giving his sincerely held view on the nature of experience that he or others are living through, or on what is of importance to himself or what he sees as important to men in general.

This is one sense in which we think of people as responsible for their evaluations, and in a way which has nothing to do with the theory of radical choice. But we also think of ourselves as responsible for them in a more straightforward 'modern' sense.

This has to do with the other direction of causal influence in which we can sometimes alter ourselves and our experience by fresh insight. In any case, our evaluations would always be open to challenge. Because of the character of depth which we saw in the self, our evaluations are articulations of insights which are frequently partial, clouded and uncertain. But they are all the more open to challenge when we reflect that these insights are often distorted by our imperfections of character. For these two reasons evaluation is such that there is always room for re-evaluation.

Responsibility falls to us in the sense that it is always possible that fresh insight might alter my evaluations and hence even myself for the better. So that within the limits of my capacity to change myself by fresh insight, within the limits of the first direction of causal influence, I am responsible in the full direct, 'modern' sense for my evaluations.

What was said about the challengeability of evaluations applies with greatest force to our most fundamental evaluations, those which provide the terms in which other less basic ones are made. These are the evaluations which touch my identity in the sense described in the previous section. There I spoke of the self as having an identity which is defined in terms of certain essential evaluations which provide the horizon or foundation for the other evaluations one makes.

Now precisely these deepest evaluations are the ones which are least clear, least articulated, most easily subject to illusion and distortion. It is those which are closest to what I am as a subject, in the sense that shorn of them I would break down as a person, which are among the hardest for me to be clear about.

Thus the question can always be posed: ought I to re-evaluate my most basic evaluations? Have I really understood what is essential to my identity? Have I truly determined what I sense to be the highest mode of life?

Now this kind of re-evaluation will be radical; not in the sense of radical choice, however, that we choose without criteria; but rather in the sense that our looking again can be so undertaken that in principle no formulations are considered unrevisable.

What is of fundamental importance for us will already have an articulation, some notion of a certain mode of life as higher than others, or the belief that some cause is the worthiest that can be served; or the sense that belonging to this community is essential to my identity. A radical re-evaluation will call these formulations into question.

But a re-evaluation of this kind, once embarked on, is of a peculiar sort. It is unlike a less than radical evaluation which is carried on within the terms of some fundamental evaluation, when I ask myself whether it would be honest to take advantage of this income-tax loophole, or smuggle something through customs. These latter can be carried on in a language which is out of dispute. In answering the questions just mentioned the term 'honest' is taken as beyond challenge. But in radical re-evaluations by definition the most basic terms, those in which other evaluations are carried on, are precisely what is in question. It is just because all formulations are potentially under suspicion of distorting their objects that we have to see them all as revisable, that we are forced back, as it were, to the inarticulate limit from which they originate.

How, then, can such re-evaluations be carried on? There is certainly no metalanguage available in which I can assess rival self-interpretations, such as my two characterizations above of my ambition to go to Nepal. If there were, this would not be radical re-evaluation. On the contrary, the re-evaluation is carried on in the formulae available but with a stance of attention, as it were, to what these formulae are meant to articulate and with a readiness to receive any gestalt shift in our view of the situation, any quite innovative set of categories in which to see our predicament, that might come our way in inspiration.

Anyone who has struggled with a philosophical problem knows what this kind of enquiry is like. In philosophy typically we start off with a question, which we know to be badly formed at the outset. We hope that in struggling with it, we shall find that its terms are transformed, so that in the end we will answer a question which we could not properly conceive at the beginning. We are striving for conceptual innovation which will allow us to illuminate some matter, say an area of human experience, which would otherwise remain dark and confused. The alternative is to stick stubbornly to certain terms and try to understand reality by classifying it in these terms (are these propositions synthetic or analytic, is this a psychological question or a philosophical question, is this view monist or dualist?).

The same contrast can exist in our evaluations. We can attempt a radical re-evaluation, in which case we may hope that our terms will be transformed in the course of it; or we may stick to certain favoured terms, insist that all evaluations can be made in their ambit, and refuse any radical questioning. To take an extreme case, someone can adopt the utilitarian criterion and then claim to settle all further issues about action by some calculation.

The point has been made again and again by non-naturalists, existentialists, and others, that those who take this kind of line are ducking a major question: should I really decide on the utilitarian principle? But this does not mean that the alternative to this stance is a radical choice. Rather it is to look again at our most fundamental formulations, and at what they were meant to articulate, in a stance of openness, where we are ready to accept any categorical change, however radical, which might emerge. Of course we will actually start thinking of particular cases, for instance where our present evaluations recommend things which worry us, and try to puzzle further. In doing this we will be like the philosopher and his initially ill-formed question. But we may get through to something deeper.

In fact this stance of openness is very difficult. It may take discipline and time. It is difficult because this form of evaluation is deep in a sense, and total in a sense, that other less than radical ones are not. If I am questioning whether smuggling a radio into the country is honest, or judging everything by the utilitarian criterion, then I have a yardstick, a definite yardstick. But if I go to the radical questioning, then it is not exactly that I have no yardstick, in the sense that anything goes, but rather that what takes the place of the yardstick is my deepest unstructured sense of what is important, which is as yet inchoate and which I am trying to bring to definition. I am trying to see reality afresh and form more adequate

categories to describe it. To do this I am trying to open myself, use all of my deepest, unstructured sense of things in order to come to a new clarity.

Now this engages me at a depth that using a fixed yardstick does not. I am in a sense questioning the inchoate sense that led me to use the yardstick. And at the same time it engages my whole self in a way that judging by a yardstick does not. This is what makes it uncommonly difficult to reflect on our fundamental evaluations. It is much easier to take up the formulations that come most readily to hand, generally those which are going the rounds of our milieu or society, and live within them without too much probing. The obstacles in the way of going deeper are legion. There is not only the difficulty of such concentration, and the pain of uncertainty, but also all the distortions and repressions which make us want to turn away from this examination: and which make us resist change even when we do re-examine ourselves. Some of our evaluations may in fact become fixed and compulsive, so that we cannot help feeling guilty about X, or despising people like Y, even though we judge with the greatest degree of openness and depth at our command that X is perfectly all right, and that Y is a very admirable person. This casts light on another aspect of the term 'deep', as applied to people. We consider people deep to the extent, inter alia, that they are capable of this kind of radical self-reflection.

This radical evaluation is a deep reflection, and a self-reflection in a special sense: it is a reflection about the self, its most fundamental issues, and a reflection which engages the self most wholly and deeply. Because it engages the whole self without a fixed yardstick it can be called a personal reflection (the parallel to Polanyi's notion of personal knowledge is intended here); and what emerges from it is a self-resolution in a strong sense, for in this reflection the self is in question; what is at stake is the definition of those inchoate evaluations which are sensed to be essential to our identity.

Because this self-resolution is something we do, when we do it, we can be called responsible for ourselves; and because it is within limits always up to us to do it, even when we do not – indeed, the nature of our deepest evaluations constantly raises the question whether we have them right – we can be called responsible in another sense for ourselves, whether we undertake this radical evaluation or not.

4

I have been exploring some aspects of a self or human agent, following the key notion that a crucial feature of human agency is the capacity for

second-order desires or evaluation of desires. In the course of the discussion, it will have become more and more plausible, I hope, that the capacity for what I have called strong evaluation is an essential feature of a person.

I think that this has helped to cast light on the sense in which we ascribe reflection, will and also responsibility to human agents. But our conception of human agency is also of crucial importance to any potential science of the human subject, in particular to psychology.

In concluding, I would like to sketch a few of the consequences for the study of psychology of this conception. First, it evidently means that a concept like 'drive', used in motivational theory as a psychic force operating in abstraction from any interpretation, cannot find a fruitful application. The idea of measuring drive like a force in natural science is in principle misguided. Instead we would have to accept that those branches of psychology which attempt to account for fully motivated behaviour must take account of the fact that the human animal is a self-interpreting subject. And this means that these branches of the discipline must be 'hermeneutical' sciences.

I have discussed some of what is involved in this elsewhere.[11] But one consequence, which has been touched on in this symposium, is for the study of personality. For if we take the view that man is a self-interpreting animal, then we will accept that a study of personality which tries to proceed in terms of general traits alone can have only limited value. For in many cases we can only give their proper significance to the subject's articulations by means of 'idiographic' studies, which can explore the particular terms of an individual's self-interpretations. Studies exclusively in terms of general traits can be empty, or else end up with baffling inconsistencies. I believe that there is some common ground here with a point made by W. and H. Mischel in a very stimulating paper[12] that such functions as self-control are carried out more discriminatively than we can account for in terms of something like 'a unitary trait entity of conscience or honesty'.

But perhaps the most valuable fruits of a more fully developed conception of the self on the above lines, which avoided the reductiveness of drive theory, would come in dialogue with those strands in psychoanalysis which are particularly concerned with the development of the

[11] Chapter 5 below.

[12] W. Mischel and H. N. Mischel, 'Self-control and the self', in T. Mischel (ed.), *The Self* (Oxford, 1977), pp. 31–64.

self, of which a paper by Ernest Wolf gives an extremely interesting account.[13] For evidently any theory of the ontogenesis of the self, and any identification of its potential breakdowns, must also both draw and draw from, implicitly or explicitly, a portrait of the fully responsible human agent. The attempt to explore our underlying notion of responsibility could therefore both help and be helped by a study of the growth and pathologies of the self.

Thus I believe that there are links between the rather groping remarks about identity in this paper and the much more fully developed notion of a 'cohesive self' that Kohut and Ernest Wolf have introduced. These links would greatly repay further exploration. They are made all the closer in that Kohut and Wolf are not working with a drive or psychic 'force' view of motivation. Thus sexual libido is not seen as a constant factor, but rather sexual desire and excitability have a very different impact on a cohesive self than on one which has lost its cohesion.[14]

The prospect of psychoanalytic theory which could give an adequate account of the genesis of full human responsibility, without recourse to such global and reified mechanisms as the super-ego, and with a truly plausible account of the shared subjectivity from which the mature cohesive self must emerge, is a very exciting prospect indeed.

[13] Cf. E. S. Wolf, 'Irrationality in a psychoanalytic psychology of the self', in T. Mischel (ed.), *The Self* (Oxford, 1977), pp. 203–23.

[14] *Ibid.*; see also, for example, Heinz Kohut, *The Restoration of the Self* (New York, 1977).

CHAPTER TWO

SELF-INTERPRETING ANIMALS

I

Human beings are self-interpreting animals. This is a widely echoing theme of contemporary philosophy. It is central to a thesis about the sciences of man, and what differentiates them from the sciences of nature, which passes through Dilthey and is very strong in the late twentieth century. It is one of the basic ideas of Heidegger's philosophy, early and late. Partly through his influence, it has been made the starting point for a new skein of connected conceptions of man, self-understanding and history, of which the most prominent protagonist has been Gadamer. At the same time, this conception of man as self-interpreting has been incorporated into the work of Habermas, the most important successor of the post-Marxist line of thought known somewhat strangely as critical theory.

And one could go on. Through all this cross-talk about 'hermeneutics', the question of what one means by this basic thesis, that man is a self-interpreting animal, and how one can show that it is so, may still go unanswered. These are of course tightly related questions; and I would like to try to fumble my way towards an answer to them.

It may turn out to be a mistake, but I am tempted to try to put together the full picture that this thesis means to convey by stages; to lay out, in other words, a series of claims, where the later ones build on the earlier ones, and in that sense form a connected picture. But to talk of claims implies that what is said at each stage is controversial, and that it will have to be established against opposition. So before starting it may be useful to say a word about who or what is opposing, or what the argument is all about.

The thesis that man is a self-interpreting being cannot just be stated flatly, or taken as a truism without argument, because it runs against one of the fundamental prejudices or, to sound less negative, leading ideas of modern thought and culture. It violates a paradigm of clarity and objectivity.

According to this, thinking clearly about something, with a view to

been an influential figure in this whole counter-movement. But what remains to be understood is why he has also often been ignored or rejected by major figures who have shared somewhat the same notions of action, starting with Schopenhauer but by no means ending there.

Perhaps what separates Hegel most obviously and most profoundly from those today who take the same side on the issue about action is their profoundly different reading of the same genetic view. For Heidegger, for example, the notion that action is first of all unreflected practice seems to rule out altogether as chimerical the goal of a fully explicit and self-authenticating understanding of what we are about. Disclosure is invariably accompanied by hiddenness; the explicit depends on the horizon of the implicit. The difference here is fundamental, but I believe that it too can be illuminated if we relate to it radically different readings of the qualitative view of action, which both espoused in opposition to the epistemological rationalism of the seventeenth century.

CHAPTER FOUR

THE CONCEPT OF A PERSON

I

In volume 2, chapters 3 and 4, I trace the conflict between two philosophies of social science. But the two underlying views do not just confront each other in social science. They also polarize the other sciences of man – psychology, for instance; and beyond that they inspire rival pictures of morality and human life. I want here to explore some of these deeper ramifications, by looking at two conceptions of what it is to be a person.

Where it is more than simply a synonym for 'human being', 'person' figures primarily in moral and legal discourse. A person is a being with a certain moral status, or a bearer of rights. But underlying the moral status, as its condition, are certain capacities. A person is a being who has a sense of self, has a notion of the future and the past, can hold values, make choices; in short, can adopt life-plans. At least, a person must be the kind of being who is in principle capable of all this, however damaged these capacities may be in practice.

Running through all this we can identify a necessary (but not sufficient) condition. A person must be a being with his own point of view on things. The life-plan, the choices, the sense of self must be attributable to him as in some sense their point of origin. A person is a being who can be addressed, and who can reply. Let us call a being of this kind a 're-spondent'.

Any philosophical theory of the person must address the question of what it is to be a respondent. At the same time, it is clear that persons are a sub-class of agents. We do not accord personal status to animals, to whom we do, however, attribute actions in some sense. This poses a second question which any theory must answer: what is special about agents who are also persons?

With these questions in mind, I want to present, partly in summary, partly in reconstruction, two views of what it is to be a person, which I believe underpin a host of different positions and attitudes evident in

modern culture. And clearly, our (perhaps implicit) notion of what it is to be a person will be determining for two *orders* of question: scientific ones – how are we to explain human behaviour? – and practical-moral ones – what is a good/decent/acceptable form of life?

The first view is rooted in the seventeenth-century, epistemologically grounded notion of the subject. A person is a being with consciousness, where consciousness is seen as a power to frame representations of things. Persons have consciousness, and alone possess it, or at least they have it in a manner and to a degree that animals do not. This answers the second question. But it also answers the first, the question of what makes a respondent. What makes it possible to attribute a point of view to persons is that they have a representation of things. They have the wherewithal to reply when addressed, because they respond out of their own representation of the world and their situation.

What this view takes as relatively unproblematic is the nature of agency. The important boundary is that between persons and other agents, the one marked by consciousness. The boundary between agents and mere things is not recognized as important at all, and is not seen as reflecting a qualitative distinction. This was so at the very beginning, where Descartes saw animals as complex machines; and it continues to be so today, where proponents of this first view tend to assume that some reductive account of living beings will be forthcoming. What marks out agents from other things tends to be identified by a performance criterion: animals somehow maintain and reproduce themselves through a wide variety of circumstances. They show highly complex adaptive behaviour. But understanding them in terms of performance allows for no distinction of nature between animals and machines which we have latterly designed to exhibit similarly complex adaptive behaviour.

We see this, for instance, with proponents of computer-based models of intelligence. They see no problem in offering these as explanations of animal performance. They only admit to puzzlement when it comes to relating consciousness to performance. We are conscious, but the machines which simulate our intelligent behaviour are not. But perhaps some day they might be? Speculation here is ragged and confused, the symptom of a big intellectual puzzle. By contrast, the reductive view of agency is subscribed to with serene confidence.

The second view I want to explore here does, by contrast, focus on the nature of agency. What is crucial about agents is that things matter to them. We thus cannot simply identify agents by a performance criterion, nor assimilate animals to machines.

To say things matter to agents is to say that we can attribute purposes, desires, aversions to them in a strong, original sense. There is, of course, a sense in which we can attribute purposes to a machine, and thus apply action terms to it. We say of a computing machine that it is, for example, 'calculating the payroll'. But that is because it plays this purpose in our lives. It was designed by us, and is being used by us to do this. Outside of this designer's or user's context, the attribution could not be made. What identifies the action is what I want to call here a derivative purpose. The purpose is, in other words, user-relative. If tomorrow someone else makes it run through exactly the same programme, but with the goal of calculating *pi* to the *n*th place, then *that* will be what the machine is 'doing'.

By contrast, animals and human beings are subjects of original purpose. That the cat is stalking the bird is not a derivative, or observer-relative fact about it. Nor is it a derivative fact about me that I am trying to explain two doctrines of the person.

Now one of the crucial issues dividing the first and second concepts of the person is what to make of this difference between original and derived purpose. If you take it seriously, then you can no longer accept a performance criterion for agency, because some agent's performances can be matched derivatively on machines. For the first view, the difference has to be relegated to the status of mere appearance. Some things (animals, ourselves) look to us to have purposes in a stronger, more original sense than mere machines.

But the second view does take it seriously, and hence sees the agent/thing boundary as being an important and problematic one. And it offers therefore a different answer to the question, what makes a respondent? This is no longer seen in terms of consciousness, but rather in terms of mattering itself. An agent can be a respondent, because things matter to it in an original way. What it responds out of is the original significance of things for it.

But then we have a very different conception from the first. The answers to the two questions are related in a very different way. The basic condition for being a respondent, that one have an original point of view, is something all agents fulfil. Something else needs to be said in answer to the question, what distinguishes persons from other agents?

And the answer to neither question can be given just in terms of a notion of consciousness as the power to frame representations. The answer to the respondent question clearly can not be given in these terms, because agents who have nothing like consciousness in the human sense

have original purposes. Consciousness in the characteristically human form can be seen as what we attain when we come to formulate the significance of things for us. We then have an articulate view of our self and world. But things matter to us prior to this formulation. So original purpose can not be confused with consciousness.

Nor does the notion of consciousness as representation help to understand the difference between persons and animals, for two related reasons which it is worth exploring at some length.

The first is that built into the notion of representation in this view is the idea that representations are of independent objects. I frame a representation of something which is there independently of my depicting it, and which stands as a standard for this depiction. But when we look at a certain range of formulations which are crucial to human consciousness, the articulation of our human feelings, we can see that this does not hold. Formulating how we feel, or coming to adopt a new formulation, can frequently change how we feel. When I come to see that my feeling of guilt was false, or my feeling of love self-deluded, the emotions themselves are different. The one I now experience as a compulsive malaise rather than a genuine recognition of wrong-doing, the other as a mere infatuation rather than a genuine bent of my life. It is rare that the emotion we experience can survive unchanged such a radical shift in interpretation.

We can understand this, if we examine more closely the range of human feelings like pride, shame, guilt, sense of worth, love, and so on. When we try to state what is particular to each one of these feelings, we find we can only do so if we describe the situation in which we feel them, and what we are inclined to do in it. Shame is what we feel in a situation of humiliating exposure, and we want to hide ourselves from this; fear what we feel in a situation of danger, and we want to escape it; guilt when we are aware of transgression; and so on.

One could say that there is a judgement integral to each one of these emotions: 'this is shameful' for shame; 'there is danger' for fear, and so on. Not that to feel the emotion is to assent to the judgement. We can feel the emotion irrationally; and sometimes see that the judgement holds dispassionately. It is rather that feeling the emotion in question just is being struck by, or moved by, the state of affairs the judgement describes. We can sometimes make the judgement without being moved (the case of dispassionate observation); or we can feel very moved to assent to the judgement, but see rationally that it does not hold (irrational emotion). But the inner connection of feeling and judgement is

attested in the fact that we speak here of 'irrational' emotion; and that we define and distinguish the feelings by the type of situation.

It follows from this that I can describe my emotions by describing my situation, and very often must do so really to give the flavour of what I feel. But then I alter the description of my emotions in altering the description I accept of my situation. But to alter my situation-description will be to alter my feelings, if I am moved by my newly perceived predicament. And even if I am not, the old emotion will now seem to me irrational, which itself constitutes a change in what I experience. So we can understand why, in this domain, our formulations about ourselves can alter what they are about.

We could say that for these emotions, our understanding of them or the interpretations we accept are constitutive of the emotion. The understanding helps shape the emotion. And that is why the latter cannot be considered a fully independent object, and the traditional theory of consciousness as representation does not apply here.

This might be understandable on the traditional theory if our formulations were not representative at all, that is, if there were no question of right and wrong here. It might be that thinking simply made it so, that how we sincerely describe our feelings just is how we feel, and that there is no point in distinguishing between the two. We might think that there are some domains of feeling where this is so. For instance, if on sincere introspection I come up with the verdict that I like blueberries, there is no further room here for talk of error or delusion.

But this is emphatically not the case with the emotions I described above. Here we can and do delude ourselves, or imperfectly understand ourselves, and struggle for a better formulation. The peculiarity of these emotions is that it is at one and the same time the case that our formulations are constitutive of the emotion, *and* that these formulations can be right or wrong. Thus they do in a sense offer representations, but not of an independent object. This is what makes the representative theory of consciousness inapplicable in this domain.

And so consciousness in this traditional sense does not seem to be the conception we need to capture the distinction between persons and other agents. The consciousness of persons, wherein they formulate their emotions, seems to be of another sort.

The second reason why representative consciousness cannot fill the bill here also comes to mind if we consider this range of human emotions. As long as we think of agents as the subjects of strategic action, then we might be inclined to think that the superiority of persons over animals lies

in their ability to envisage a longer time scale, to understand more complex cause–effect relationships, and thus engage in calculations, and the like. These are all capacities to which the power to frame representations is essential. If we think merely in this strategic dimension, then we will tend to think that this representative power is the key to our evolution from animal to man.

But if we adopt the second view, and understand an agent essentially as a subject of significance, then what will appear evident is that there are matters of significance for human beings which are peculiarly human, and have no analogue with animals. These are just the ones I mentioned earlier, matters of pride, shame, moral goodness, evil, dignity, the sense of worth, the various human forms of love, and so on. If we look at goals like survival and reproduction, we can perhaps convince ourselves that the difference between men and animals lies in a strategic superiority of the former: we can pursue the same ends much more effectively than our dumb cousins. But when we consider these human emotions, we can see that the ends which make up a human life are *sui generis*. And then even the ends of survival and reproduction will appear in a new light. What it is to maintain and hand on a human form of life, that is, a given culture, is also a peculiarly human affair.

These human matters are also connected with consciousness in some sense. One could indeed argue that no agent could be sensitive to them who was not capable of formulating them, or at least of giving expression to them; and hence that the kind of consciousness which language brings is essential to them. We can perhaps see this if we take one example from the above list, being a moral agent. To be a moral agent is to be sensitive to certain standards. But 'sensitive' here must have a strong sense: not just that one's behaviour follow a certain standard, but also that one in some sense recognize or acknowledge the standard.

Animals can follow standards in the weaker sense. My cat will not eat fishmeal below a certain quality. With knowledge of the standard I can predict his behaviour. But there need be no recognition here that he is following a standard. This kind of thing, however, would not be sufficient to attribute moral action to an agent. We could imagine some animal who was systematically beneficent in his behaviour; what he did always redounded to the good of man and beast. We still would not think of him as a moral agent, unless there were some recognition on his part that in acting this way he was following a higher standard. Morality requires some recognition that there are higher demands on one, and hence the recognition of some distinction between kinds of goal. This has nothing

to do with the Kantian diremption between duty and inclination. Even the holy will, which gladly does the good, must have some sense that this is the good, and as such worthy to be done.

Moral agency, in other words, requires some kind of reflexive awareness of the standards one is living by (or failing to live by). And something analogous is true of the other human concerns I mentioned. And so some kind of consciousness is essential to them. I think we can say that being a linguistic animal is essential to one's having these concerns; because it is impossible to see how one could make a distinction like the one above, between, for example, things one just wants to do, and things that are worthy to be done, unless one was able to mark the distinction in some way: either by formulation in language, or at least by some expressive ceremonial which would acknowledge the higher demands.

And so when we ask what distinguishes persons from other agents, consciousness in some sense is unquestionably part of the answer. But not consciousness understood as just representation. That can help explain some of the differences; for instance, the great superiority of man as strategic agent. But when we come to the peculiarly human concerns, the consciousness they presuppose cannot be understood just as the power to frame representations of independent objects. Consciousness – perhaps we might better here say language – is as it were the medium within which they first arise as concerns for us. The medium here is in some way inseparable from the content; which is why as we saw above our self-understanding in this domain is constitutive of what we feel.

We should try to gather the threads together, and show how the two conceptions square off against each other. They both start off with our ordinary notion of a person, defined by certain capacities: a person is an agent who has a sense of self, of his/her own life, who can evaluate it, and make choices about it. This is the basis of the respect we owe persons. Even those who through some accident or misfortune are deprived of the ability to exercise these capacities are still understood as belonging to the species defined by this potentiality. The central importance of all this for our moral thinking is reflected in the fact that these capacities form an important part of what we should respect and nourish in human beings. To make someone less capable of understanding himself, evaluating and choosing is to deny totally the injunction that we should respect him as a person.

What we have in effect are two readings of what these capacities consist of. The first takes agency as unproblematic. An agent is a being who acts, hence who has certain goals and endeavours to fulfil them. But this range

of features is identified by a performance criterion, so that no difference of principle is admitted between animals and, say, complex machines, which also adaptively react to their surroundings so as to attain certain ends (albeit in a derivative way).

Along with agency, its ends too are seen as unproblematic. What is striking about persons, therefore, is their ability to conceive different possibilities, to calculate how to get them, to choose between them, and thus to plan their lives. The striking superiority of man is in strategic power. The various capacities definitive of a person are understood in terms of this power to plan. Central to this is the power to represent things clearly. We can plan well when we can lay out the possibilities clearly, when we can calculate their value to us in terms of our goals, as well as the probabilities and cost of their attainment. Our choices can then be clear and conscious.

On this view, what is essential to the peculiarly human powers of evaluating and choosing is the clarity and complexity of the computation. Evaluation is assessment in the light of our goals, which are seen ultimately as given, or perhaps as given for one part, and for the rest as arbitrarily chosen. But in either case the evaluation process takes the goal as fixed. 'Reason is and ought to be, the slave of the passions.' Choice is properly choice in the light of clear evaluation. To the human capacities thus conceived, the power of clear and distinct representation is obviously central.

So on one view, what makes an agent a person, a fully human respondent, is this power to plan. My interlocutor replies to me out of his power to make a life-plan and act on it. This is what I have to respect.

By contrast, the other view starts off quite differently. It raises the question of agency, and understands agents as in principle distinct from other things. Agents are beings for whom things matter, who are subjects of significance. This is what gives them a point of view on the world. But what distinguishes persons from other agents is not strategic power, that is, the capacity to deal with the same matter of concern more effectively. Once one focusses on the significance of things for agents, then what springs to view is that persons have qualitatively different concerns.

In other terms, once one raises the question of agency, then that of the ends of agents comes into view. And what is clear is that there are some peculiarly human ends. Hence the important difference between men and animals cannot simply consist in strategic power; it is also a matter of our recognizing certain goals. Consciousness is indeed essential to us. But this cannot be understood simply as the power to frame representations, but

also as what enables us to be open to these human concerns. Our consciousness is somehow constitutive of these matters of significance, and does not just enable us to depict them.

This supports a quite different reading of the essentially personal capacities. The essence of evaluation no longer consists in assessment in the light of fixed goals, but also and even more in the sensitivity to certain standards, those involved in the peculiarly human goals. The sense of self is the sense of where one stands in relation to these standards, and properly personal choice is one informed by these standards. The centre of gravity thus shifts in our interpretation of the personal capacities. The centre is no longer the power to plan, but rather the openness to certain matters of significance. This is now what is essential to personal agency.

II

Naturally these conceptions ramify into very different views in both the sciences of man and the practical deliberations of how we ought to live. These are the two orders of questions I mentioned at the outset: how are we to explain human behaviour? and, what is a good life? We can for the sake of simplicity consider different doctrines in science and morals as consequences of these two underlying conceptions. But of course the motivation for our holding one or the other is more complex. We may be led to adopt one, because it relates to a certain approach to science, or goes with a certain style of moral deliberation, rather than adopting the approach or the style, because they are consequences of an already established core conception of the person.

In fact the order of motivation is mixed and varies from person to person. One can adopt a given core conception because its scientific ramifications strike him as valid, but only reluctantly accept what it entails about moral deliberation. Here it may rub against the grain of his intuitions, but because it seems true for what appear unanswerable reasons, he has no option but to endorse it. For another, it may be the moral consequences which make it plausible, and the scientific ones may be a matter of indifference. In fact, in talking about ramifications, I am also talking about possible motivations, although they may also be reluctant consequences of someone's vision of things.

In the remainder of my remarks, I would like to discuss the ramifications of these two conceptions, accounting for them for the sake of simplicity as motivations. I hope that this will make my rather abstract

sketches somewhat fuller and more life-like, and that you might see these core conceptions actually at work in modern culture.

First, in the scientific domain. I am of course not neutral between them; and so the question that strikes me here is, what makes the first – let us call it the representation conception – so popular in our culture? I think an important part of the answer can be found in the prestige of the natural science model, which I discuss – and argue against – elsewhere.[1] Perhaps one of the key theses of the seventeenth-century revolution which inaugurates modern natural sciences is the eschewing of what one could call anthropocentric properties. Anthropocentric properties of things, but which they only have in so far as they are objects of experience. This was crucially at stake in the seventeenth-century distinction between primary and secondary properties. Secondary properties, like colour and felt heat, only applied to things in so far as they were being experienced. In a world without experiencing subjects, such properties could no longer be sensefully attributed to objects. They were therefore understood as merely subjective, as relative to us, not as absolute properties of things.

It was an important step in the development of modern natural science when these properties were distinguished and set aside. This distinction was a polemical instrumental in the struggle against the older conceptions of the universe as meaningful order. Such hypotheses, which explained features of the world in terms of their 'correspondences' against a background order of ideas, were condemned as mere projections. They concerned purely the significance of things for us, not the way things were.

This eschewal of anthropocentric properties was undoubtedly one of the bases of the spectacular progress of natural science in the last three centuries. And ever since, therefore, the idea has seemed attractive of somehow adapting this move to the sciences of man. We can see here, I think, one of the sources of that basic feature of the representation conception, its assimilating agency to things; or, otherwise put, its understanding agency by a performance criterion.

We are motivated to distinguish animals from machines which imitate their adaptive performance only if we take significance seriously, the fact that things matter to animals in an original way. But the significance of things is paradigmatically a range of anthropocentric properties (or in the case of other animals, properties which are relative to them; in any case, not absolute). So it can easily appear that a scientific approach to

[1] See volume 2, chapters 3 and 4.

behaviour which incorporates this important founding step of natural science would be one which gave no weight to significance.

Of course, we can admit that the distinction is in some way important for us. Things feel different inside to a human being, and to an animal; and there probably is nothing comparable in a machine. But all this has to do with the way things appear. It thus has no weight when we come to identifying the explanatory factors of a science of behaviour. Just as, analogously, we can admit that it really does appear that the sun goes below the horizon, but this must just be ignored when we want to establish a scientific theory of the movement of earth and the heavenly bodies.

Now if we follow out what is involved in a significance-free account, we shall come across themes that tend to recur in the modern sciences of man. We can see this if we return to my discussion above of our emotions, where I said that they incorporate in a sense a view of our situation. To experience an emotion is to be in a sense struck or moved by our situation being of a certain nature. Hence, I said, we can describe our emotions by describing our situation.

But this is only so because we describe our situation in its significance for us. We can usually understand how someone feels when he describes his predicament, because we normally share the same sense of significance. So someone says: 'imagine what I felt when he walked in just then and saw me'; and we can quite easily do so, because we share just this sense of embarrassment. Of course, it is a commonplace that between different cultures the sense of significance can vary, and then we can be quite baffled. Nothing comes across of how people feel from the simple narrative of events, or only something confused and perplexing.

So situation-description is only self-description because the situation is grasped in its significance. And in fact, we have a host of terms which operate in tandem with our emotion terms and which designate different significances a situation can have; such as, 'humiliating', 'exciting', 'dismaying', 'exhilarating', 'intriguing', 'fascinating', 'frightening', 'provoking', 'awe-inspiring', 'joyful', and so on. We can often describe our predicament with one of these, or alternatively we can give a sense of it by saying what it inclines us to do. Certain standard emotion terms are linked, as we saw, to standard situation-descriptions, as well as responses we are inclined to make. So fear is experienced at something dangerous, and inclines us to flee; shame at what is shameful or humiliating, and inclines us to do away with this, or at least to hide it; and so on.

Now our ordinary description/explanation of action in terms of our emotions and other motives is based on our sense of significance. That is,

it either invokes this directly by using terms like the above, or it assumes it as the background which makes predicament descriptions intelligible as accounts of what we feel or want to do. The actions we are inclined to take are identified by their purposes, and frequently these are only intelligible against the background of significance. For instance, we understand the inclination to hide what is humiliating, only through understanding the humiliating. Someone who had no grasp of a culture's sense of shame would never know what constituted a successful case of hiding, or cover-up. We would not be able to explain to him even what people are inclined to do in humiliating predicaments in this culture, let alone why they want to do it.

What would it mean then to set about designing a significance-free account? Plainly what would remain basic is that people respond to certain situations, and perhaps also that they respond by trying to encompass certain ends. But if we had an account which really eschewed anthropocentric properties, and thus which did not have to draw on our background sense of significance for its intelligibility, it would characterize situation and end in absolute terms. In one way this might seem relatively easy. Any situation bears a great number, an indefinite number of descriptions. The predicament that I find humiliating is also one that can be described in a host of other ways, including some which make no reference to any significance at all. But of course the claim involved in this redescription would be that none of the important explanatory factors are lost from sight. It is that the explanatory relationship between situation and response can be captured in an absolute description; or that, in other words, the features picked out in the significance description are not essential to the explanation, but just concern the way things appear to us in ordinary life.

This kind of ambition has underlain various influential schools in academic psychology. At its most reductive, where there was a suspicion even of goal-seeking behaviour, as somehow tinged with anthropocentrism, we had behaviourism. Everything was to be explained in terms of responses to stimuli. But these were to be characterized in the most rigorously significance-free terms, as 'colourless movement and mere receptor impulse' in Hull's phrase. This school has passed its prime; the development of computing machines has shown how goal-behaviour can be accounted for in mechanistic terms, and so strategic action can now be allowed into the account. But the goal is to account for animate strategic action in the same terms as we explain the analogous behaviour of machines. And since for the latter case it is clear that the goal states have

to be describable in absolute terms, this must also be true of the former, if the account is to succeed.

It seems to me then that this ambition to follow natural science, and avoid anthropocentric properties, has been an important motivation of the representation view. It gives us an important reason to ignore significance, and to accept a performance criterion for agency, where what matters is the encompassing of certain, absolutely identified ends. But the drive for absolute (i.e., non-anthropocentric) explanation can be seen not only in psychology. It is also at the origin of a reductive bent in social science. To see the connection here, we will have to follow the argument a little farther, and appreciate the limitations on this transposition into absolute terms.

These are evident when we return to what I called above the peculiarly human motivations, like shame, guilt, a sense of morality, and so on. Finding absolute descriptions which nevertheless capture the explanatory relevance of situation and goal is in principle impossible in this domain.

To see this, let us contrast one of these motives, shame, with one where the absolute transposition seems possible, say, physical fear. This latter is fear of physical danger, danger to life or limb. Now the significance of the situation here can perhaps be spelled out in medical terms: something in my predicament threatens to end my life in some particular way – say, I am likely to fall, and the impact would be lethal. Here we have a sense that we could describe the impending outcome in physiological terms, terms that made no reference to its importance to me, as we might describe the death of a sparrow, or any other process in nature.

And so we might think of a disengaged, absolute account that might be offered by my behaviour; where we would be told that this fall, and the resulting physiological changes occurring on impact, constituted a counter-goal for me; something I strove to avoid. And my behaviour could be explained strategically on this basis. Here we have an account of behaviour which we could imagine being matched on a machine. This too might become irreparably damaged on impact. And so we might design a machine to compute the likelihood of certain possible modes of destruction possible in its environment, and to take evasive action. What would be left out of account, of course, would be the subjective experience of the fear. This would survive in the account only as a direction of behaviour, a bent to avoidance. It would be a mere 'con-attitude' towards falling, and quite colourless. That is, the specific experienced difference between avoiding something out of fear, and avoiding it out of distaste, would fall away. This would be part of the subjective 'feel' of the lived experience

which would be left outside the account. But we might nevertheless understand the claim here that all the really explanatory factors had been captured. We can presumably predict the behaviour of both machine and person, granted a knowledge of the situation described absolutely. What more could one want?

In fact, one might ask a lot more. But I do not want to argue this here. Let me concede the seeming success of the absolute transposition for this case of physical danger, in order to be able to show how impossible it is for shame. The corresponding task in this latter case would be to give an absolute account of a situation which was humiliating or shameful. This would be the analogue to the absolute description of danger above.

But this we cannot do. The reason lies in the reflexive nature of these motives, which I noted above. I can give sufficient conditions of a situation's being dangerous in absolute terms, because there are no necessary conditions concerning its significance. The fall will be lethal, however I or mankind in general regard it. This is a hard, culture-resistant fact. But a situation is not humiliating independent of all significance conditions. For a situation to be humiliating or shameful, the agent has to be of the kind who is in principle sensitive to shame.

This can perhaps become clearer if we reflect that shame involves some notion of standards. To feel shame is to sense that I fail on some standard. We can only get an adequate account of the shameful, if we can get clear on these standards. But built into the essence of these standards is that they are those of a being who is potentially sensitive to them. The subject of shame must be one who can be motivated by shame. It might appear that in certain cultures a sufficient condition of the shameful could be given in purely objective terms. For instance, defeat in battle might be shameful for the warrior. But what is forgotten here is that defeat is only shameful for the warrior, because he ought to have been so powerfully moved by the love of glory to have conquered, or at least to have died in the attempt. And the trudge back in the dust in chains is only humiliating because he is – or ought to be – a being who glories in power, in strutting over the earth as master.

We can see the essential place of significance conditions here when we note that shamelessness is shameful. In other words, there are conditions of motivation for avoiding shame, viz., that one be sensitive to shame. The one who does not care, who runs away without a scruple, earns the deepest contempt.

It is these significance conditions that make it the case that we cannot attribute shame to animals, let alone to machines. And this makes the

contrast with danger. Just because this can be defined absolutely, we have no difficulty in envisaging animals as standing in danger in exactly the same sense as we can; and the extension to machines does not seem a very great step.

We can thus distinguish between motives which *seem* potentially capable of a significance-free account, and those which definitely are not. I emphasize 'seem' because, even for these, I have doubts on other grounds which I have no time to go into here. But they contrast with the peculiarly human in that these plainly are irreducible. These are the ones where the significance itself is such that we cannot explain it without taking into account that it is significant for us. These are the ones, therefore, where the variations occur between human cultures, that is, between different ways of shaping and interpreting that significance. So that what is a matter of shame, of guilt, of dignity, of moral goodness, is notoriously different and often hard to understand from culture to culture; whereas the conditions of medical health are far more uniform. (But not totally, which is part of my reason for doubt above, and my unwillingness to concede even the case of physical danger to a reductive account.)

This distinction underlies the reductive bent we see in much modern social science, towards accounts of human behaviour and society which are grounded in goals of the first type. The contemporary fad for sociobiology provides a good example. To explain human practices and values in terms of the goals of survival and reproduction is to account for things ultimately by explanatory factors which can be described in absolute terms. Survival, reproduction; these are conditions that can be predicated of animals as well, and could be extended analogously to machines, for that matter. The enterprise of giving a reductive account of culture in terms of these ends can thus appear as an answer to the demands of science, whereas anti-reductionist objections seem counsels of obscurity, or of despair of the scientific cause. The old requirement, that we eschew anthropocentric properties, is here working its way out, via the absolute transposition. The fact that it leads us into a blind alley in social science ought to make us reflect on the validity of this basic requirement, and hence of the natural science model. But that is a point I have sufficiently argued above.

But an absolute account would be a culture-free one, for reasons I have just touched on. It is the peculiarly human motivations which are reflexively constituted by our interpretations and therefore are deeply embedded in culture. The search for a 'materialist' account, as I interpret it here, is the search for an explanation in terms of ends which can be

absolutely described. But this would not only meet the demands of 'science', it would at a stroke cut through the intractable difficulties of comparative social science. It would give us a truly neutral standpoint, from which we could survey all cultures without ethnocentricity. We have seen above that this is an illusion, but we can also appreciate how powerful an attraction it exercises in modern culture.

III

In the remaining pages, I would like to make a few remarks about the ramifications of the two conceptions of a person for our views about moral deliberation. Here too, the first conception has its attractions. If we understand ourselves in terms of certain absolutely defined ends, then the proper form of deliberation is strategic thinking. And this conception sees the superiority of man over animal as lying in greater strategic capacity. Reason is and ought to be primarily instrumental.

The pattern is familiar enough, but its attractions are insufficiently understood. There is, of course, the sense of control. The subject according to the significance perspective is in a world of meanings that he imperfectly understands. His task is to interpret it better, in order to know who he is and what he ought to seek. But the subject according to the representation view already understands his ends. His world is one of potential means, which he understands with a view to control. He is in a crucial sense disengaged. To understand things in the absolute perspective is to understand them in abstraction from their significance for you. To be able to look on everything, world and society, in this perspective would be to neutralize its significance, and this would be a kind of freedom – the freedom of the self-defining subject, who determines his own purposes, or finds them in his own natural desires.

Now I believe that the attractions of this freedom come from more than the sense of control that accompanies submitting nature and society to instrumental reason. They are also of spiritual origin, in a sense which is understandable from our Western religious tradition. In both its Greek and Christian roots (albeit a deviation in this latter stream), this has included an aspiration to rise above the merely human, to step outside the prison of the peculiarly human emotions, and to be free of the cares and the demands they make on us. This is of course an aspiration which also has analogous forms in Indian culture, and perhaps, indeed, in all human cultures.

My claim is that the ideal of the modern free subject, capable of

objectifying the world, and reasoning about it in a detached, instrumental way, is a novel variant of this very old aspiration to spiritual freedom. I want to say, that is, that the motive force that draws us to it is closely akin to the traditional drive to spiritual purity. This is, of course, highly paradoxical, since the modern ideal understands itself as naturalistic, and thus as quite antithetical to any religious outlook. But I believe that in this it is self-deluded. This is one place where Nietzsche had more insight than most modern philosophers; he saw the connection between the modern scientific ideal of austere truth and the spiritual traditions of self-denial that come to us from the ancients. From this point of view, it is not surprising to see a modern naturalist like Hobbes denouncing vainglory with the vigour of an ancient moralist.

The analogy is that, in both cases, we have a place to stand outside the context of human emotions in order to determine what is truly important. In one case, that of the tradition, this is seen as a larger order which is the locus of more than human significance; in the modern case, it is an order of nature which is meant to be understood free of any significance at all, merely naturalistically. And this is by no means a minor difference. That is not my claim. Rather it is that beyond this difference, something of the same aspiration is evident in both. And this is linked with my belief that the aspiration to spiritual freedom, to something more than the merely human, is much too fundamental a part of human life ever to be simply set aside. It goes on, only under different forms – and even in forms where it is essential that it does not appear as such; this is the paradox of modernity.

But whatever the motive, this first conception of the person grounds a certain view about moral deliberation. Our ends are seen as set by nature, and thus discoverable by objective scrutiny, or else as autonomously chosen; but in either case, as beyond the ambiguous field of interpretation of the peculiarly human significances. In the light of these ends, reason is and ought to be instrumental. Utilitarianism is a product of this modern conception, with its stress on instrumental reasoning, on calculation, and on a naturalistically identified end, happiness (or on a neutral, interpretation-free account of human choice, in terms of preferences). The stress on freedom emerges in its rejection of paternalism. And in rationality it has a stern and austere ideal of disengaged, disciplined choice. This is by no means the only fruit of this modern conception, but it has been one of the most widespread and influential.

The alternative perspective, which I have called the significance view, has arisen in the last centuries as a reaction to the first. It objects to the first as a flight from the human, and sets up a completely different model

of practical deliberation. Rather than side-stepping the peculiarly human emotions, and turning to instrumental reason, the main form this deliberation takes is a search for the true form of these emotions. Typical questions of this kind of thinking are of the form: what is really, that is, properly, shameful? What ought we to feel guilty about? In what does dignity consist? And so on.

This deliberation, of course, takes place in a modern context, one in which no larger order of more than human significance can be just assumed as an unargued context. And this gives it its tentative, exploratory nature. Those who hunger for certainty will only find it in the first perspective, where the ends of man are thought to be defined by a naturalistic science.

I believe that both these models of the person are current in modern Western culture, and that most people operate with a (perhaps inconsistent) combination of the two. It is on the level of theory that they are sorted out, and become exclusive alternatives. But this does not make them unimportant. Theoretical models with their inner coherence have a great impact on our thinking even where – perhaps especially where – they are not fully conscious or explicit. I have tried to demonstrate elsewhere something of the baleful effects of the natural science model in social science. Here I have been trying to dig deeper, into some of the sources of that model. I have been looking for these in a conception of the person, which is also the background of modern views about practical deliberation. I have tried to contrast this with an alternative conception, which I believe is its chief rival in modern Western culture.

Some of my reasoning here has been perhaps too tenuous to be fully convincing. But I believe that a deeper examination must show that the struggle between rival approaches in the science of man, that we have been looking at here, is no mere question of the relative efficacy of different methodologies, but is rather one facet of a clash of moral and spiritual outlooks. And I believe that we can only make even the first halting steps towards resolving it if we can give explicit recognition to this fact.

PART II

PHILOSOPHY OF PSYCHOLOGY AND MIND