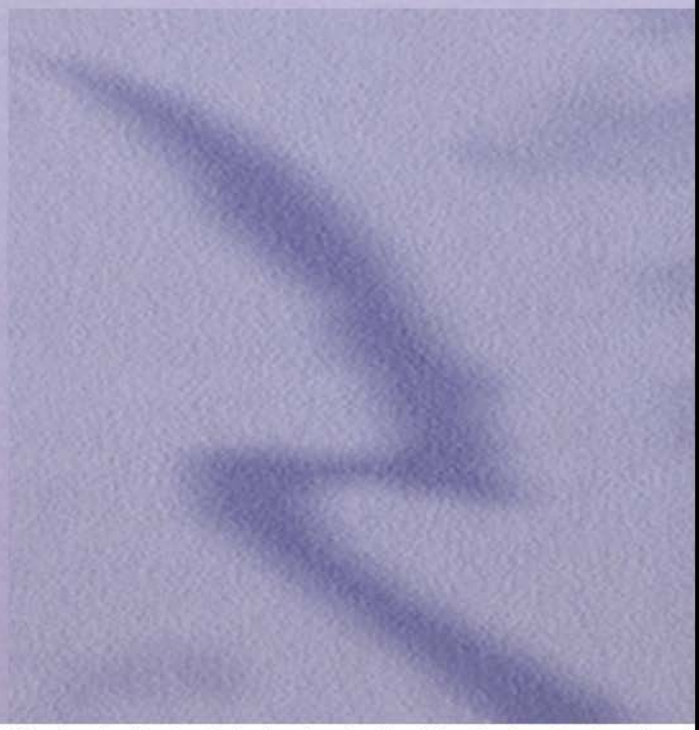
Weakness of the Will

Justin Gosling

The Problems of Philosophy their Past and Present



Weakness of the Will

The Problems of Philosophy
Their Past and Present

General Editor: Ted Honderich
Grote Professor of the Philosophy of Mind and Logic
University College, London

Each book in this series is written to bring into view and to deal with a great or significant problem of philosophy. The books are intended to be accessible to undergraduates in philosophy, and to other readers, and to advance the subject, making a contribution to it.

The first part of each book presents the history of the problem in question, in some cases its recent past. The second part, of a contemporary and analytic kind, defends and elaborates the author's preferred solution.

PRIVATE OWNERSHIP	James O.Grunebaum
RELIGIOUS BELIEF AND THE WILL	Louis P.Pojman
RATIONALITY	Harold I.Brown
THE RATIONAL FOUNDATIONS OF ETHICS	T.L.S.Sprigge
MORAL KNOWLEDGE	Alan Goldman
MIND-BODY IDENTITY THEORIES	Cynthia Macdonald
PRACTICAL REASONING	Robert Audi
PERSONAL IDENTITY	Harold W.Noonan
IF P, THEN Q: CONDITIONALS AND THE FOUNDATIONS OF REASONING	David H.Sanford
THE INFINITE	A.W.Moore
THOUGHT AND LANGUAGE	Julius Moravcsik
HUMAN CONSCIOUSNESS	Alastair Hannay
EXPLAINING EXPLANATION	David-Hillel Ruben
THE NATURE OF ART	A.L.Cothey

Weakness of the Will

Justin Gosling

London and New York

First published 1990
by Routledge
11 New Fetter Lane, London EC4P 4EE

Simultaneously published in the USA and Canada
by Routledge
a division of Routledge, Chapman and Hall, Inc.
29 West 35th Street, New York, NY 10001

Routledge is an imprint of the Taylor & Francis Group

This edition published in the Taylor & Francis e-Library, 2003.

British Library Cataloguing in Publication Data
Gosling, J.C.B. (Justin Cyril Bertrand)
Weakness of the will.—(The problems of Philosophy)
1. Man. Will. Philosophical aspects
I. Title II. Series
128'.3
Library of Congress Cataloging in Publication Data
Gosling, J.C.B. (Justin Cyril Bertrand)
Weakness of the will/Justin Gosling
p. cm.—(The problems of philosophy)
Includes bibliographical references.
1. Will—History. 2. Ethics—History. 3. Self-control—Moral and ethical aspects—History. 4. Will. 5. Ethics. 6. Self-control—Moral and ethical aspects. I. Title. II. Series: Problems of philosophy
(Routledge & Kegan Paul)
BJ1461.G67 1990
128'.3–dc20
89–49681

ISBN 0-203-40523-4 Master e-book ISBN

ISBN 0-203-71347-8 (Adobe eReader Format)
ISBN 0-415-03435-3 (Print Edition)

‘I must try to conquer myself’, said his wife, with the sigh natural to this purpose.
‘As you only have your own power to do it with, it sounds as if it would be an equal struggle.’

(*Parents and Children,* Ivy Compton-Burnett)

Contents

Explanations and Acknowledgments

The theme of the book is explained in the Introduction. Those who prefer to see the answers before they have read the questions may like to start with the epilogue. The bibliography contains reference to all works cited in the text, and in the case of foreign language works, I have, where possible, suggested translations. It also contains the titles of works not mentioned in the text, but which I have found useful, and therefore suppose that others might. An asterisk signifies the presence of a good bibliography.

There are two indices: a general one and an *index locorum.* Since examples form an important part of the argument, and are often easier to remember than the arguments they are intended to support, lists of examples are given at appropriate points in the general index.

The chapter on the Stoics is a slightly altered version of an articlle published in *Apeiron* (1987), and I am indebted to the editors for permission to use the material here. None of the other material has appeared elsewhere.

Earlier attempts at parts of the historical and of the straight philosophical sections have formed the basis of a graduate class at the University of Minnesota and of a talk at the Philosophy Colloquium there. I am indebted to Dr Graeme Marshall, Dr Norman Dahl, Dr Garrett Barden and Mr Christopher Taylor for discussion of various questions treated in the book. As always in such cases, their responsibility only extends to their time and charity.

I am also indebted to the University of Minnesota and Macalester College for inviting me during the winter of 1986, and giving me the time to iron out the more obvious errors in my approach; and to my colleagues at St Edmund Hall for giving me leave for that period. I am especially indebted to my colleague Dr J.Todd for taking on my duties during that period.

Justin Gosling

Introduction

It may help the reader to start with a brief account of the set of problems covered in this book, and the manner of treatment. The problems I shall deal with start from the thesis of Socrates that no-one can choose what they consider the worse course under the influence of pleasure, fear and so on. While Socrates voices his opposition to the view that emotions and desires could overcome knowledge, the basis of his opposition is that he thinks it is impossible that someone should deliberately act contrary to their judgment of what it is best to do. For this judgment must give what I take as reasons for acting in this case, so how could I for a reason act contrary to the reasons I have? A number of philosophers have thought that there is a problem about admitting such 'weakness', either holding it to be in some interesting cases impossible, or considering it possible, but needing careful distinctions to show it possible. Often the question has been treated as one about deliberate action under emotional stress, which explains the common term *akrasia* (lack of control) to describe the problem; but the problem has also been treated as one about deliberate wrong-doing, which might explain the expression 'weakness of will', and recently has been extended to questions outside the moral sphere to examples which seem to have the same basic structure.

Many philosophers hold views which should commit them either to finding no problem or finding weakness problematic, who nevertheless never broach the problem. I shall not spend much time on these.

Part 1 will be an historical survey of some of the main philosophers in the Western tradition who have considered the question of the possibility of weakness one worth trying to answer. Among these I shall concentrate on those whose questions have arisen from the Socratic starting-points. Descartes has problems, but they arise mainly in relation to his position about the unity of the mind and the relation of mind to

body. Fichte, Schopenhauer and Nietzsche have something to say about the will, and strength of will, but do not have problems about its possibility. The historical section, therefore, will not treat in detail of a number of figures, although the philosophical section will contain arguments which bear indirectly on what they say.

This history is a good example of a problem having its shape largely fixed by the philosopher(s) responsible for its inception. In this case this has led, as I shall argue in the second part, to misconstruing the problem, and in particular the notion of irrationality as applicable to weak behaviour.

Part 2 of the book will consist of a philosophical discussion of the problem. This is not, I hope, to say that the historical section contains no philosophical element; but I shall there be concentrating on interpretation and the interrelation between themes in the different philosophers. The points made here will in many cases be taken up and used and discussed in the philosophical section. In the philosophical section itself there will be less discussion of individual philosophers, with the exception of Davidson, whose views are used as a lead in to the problems of those chapters. Doubtless someone else could have been chosen, but Davidson's work has been the subject of many articles, and its influence seemed to make it a good place to start for placing my own position.

The discussion of weakness brings one regularly to the verge of other issues, such as freedom of the will, theory of action, the relation of folk psychology, so-called, to science, and so on. I have aimed to signal danger of trespass while exercising the strength of will to hold back.

The individual historical chapters can probably be read independently, though the later ones contain a certain amount of back reference to positions already discussed. The philosophical section should also be readable on its own, although it may sometimes be difficult for someone who has not read the historical part to understand the occasional remark. Also some of the moves made there might be more readily intelligible in the light of the discussion of problems in the historical section. It is my hope that anyone who has the perserverance to read the book right through will find that themes which develop through the historical treatment help one to understand the shape of the problem in contemporary literature, and appreciate some of the distinctions made there.

The requirements of the series to which the book belongs mean that in the historical section justice is not done to the secondary literature. In

the case of work on Plato and Aristotle this would be particularly difficult. I hope it is some compensation that the problems are put in a developmental context. So far as I know there is no literature on the problem in medieval philosophers. I shall be well content if Chapter VI encourages examination of those interesting texts.

The same requirements discourage extensive discussion of the by now fairly full range of contemporary literature. I have benefited from a wider range of writings than is mentioned in the text. The bibliography gives some indication of this.

PART 1

CHAPTER I

The *Protagoras*

In Western philosophy the so-called problem of weakness of will takes its start in the notorious Socratic paradox that people cannot knowingly choose the worse of two available alternatives. The best source for an early version of this thesis is Plato's *Protagoras* 351–8. There Socrates argues that the common view that our knowledge of what is best can be overcome by pleasure, fear, anger and the like, is wholly wrong. Agents always choose what they think best Those who know what is best *eo ipso* have correct beliefs about what is best, and since they cannot choose contrary to these beliefs, they cannot choose contrary to their knowledge. Apparent cases of people not being in control of their emotions, and acting against their better judgment, are really cases of people who, lacking the ability to distinguish illusion from reality, are victims of vacillating judgment resulting from the illusions of the moment. They always do what they think best, but their views on what is best have no sure grounding, and change with changing appearances.

As it stands, this hardly carries conviction, and has commonly been felt to fly in the face of the facts (see, e.g., Lemmon 1962). In what follows I shall first concentrate on expounding Socrates' views in this passage in so far as they bear on the possibility of deliberate wrong-doing, with some indication of his arguments. It will still seem pretty implausible. I shall then, in the next chapter, try to make it sound a more attractive position, partly by illustrating how Socrates could deal with apparent counter-examples, and partly by showing how he could challenge his opponents to produce a coherent alternative. At this point it should be easier to assess the extent to which Socrates' position is open to effective criticism, and the exteftt to which he has raised a genuine problem.

The Structure of the Passage

There is considerable dispute as to how seriously we are intended to take the argument for hedonism, and how we are to understand the thesis about the unity of the virtues, but I shall be interested in the passage solely for the light it throws on the thesis that one cannot knowingly do wrong. I shall engage in dispute with scholars only in so far as it seems helpful for clarifying the exposition.

Prior to this passage, Socrates has been arguing that all the virtues are in some important sense one. Protagoras has objected that courage at least is different from the rest Socrates goes off at an apparent tangent by raising three ordered questions. First, he queries the view that pleasure can be bad. Second, in order to settle this question he puts a further one: whether, as most people seem to think, knowledge can be dragged around by passion, and in particular pleasure. He recognizes that the denial that it can raises the third question: just what is the experience commonly described as being overcome by pleasure, if it is not what most people suppose? The answers to these three questions supply the answer to the background question as to whether courage is a virtue distinct from the rest.

The question of whether knowledge can be dragged around by passion directs the discussion to the further question of what it is that people value and pursue. Since the weak supposedly choose what they consider to be worse, it is pertinent to ask what thinking something worse/better amounts to, and what the object(s) of choice might be. Examples are proposed in order to convince us that all our valuing is based on judgments about overall pleasure. We can certainly make sense of saying that some pleasures are bad, but only in that if we take them together with their consequences they result in more distress than pleasure. The only factor which counts with us in favour of saying something is good is pleasure, and the only one which counts against is distress. This discussion employs a distinction between being pleasant considered in itself and being pleasant overall, and by 355a has yielded a use of 'pleasure', in equating pleasure with the good, whereby it refers to a life of pleasure free of distress.

It also requires that we be able to understand a distinction between 'good so far as it goes' and 'good overall' and between both and 'the good'.

In all this no distinction is made between the questions whether pleasure is the good and whether we all pursue/value pleasure.

Indeed,the view that pleasure is the good is argued not on the grounds that we ought to pursue it, but on the grounds that there is nothing else that we do value or pursue. Nor is any distinction made between what we all use as grounds for valuing something and what we all pursue. Thus at 354b5–*c*5 we find:

> Are these things good for any other reason than that they result in pleasures and release from and prevention of distress? or can you mention any other outcome to which you refer in calling them good apart from pleasure and distress? They would say 'No', I think. – I agree, said Protagoras.—*So* [my emphasis] you pursue pleasure as good and avoid distress as evil?

What we have, in fact, is a form of psychological hedonism whereby (i) we give something value just and only in so far as we think it gives pleasure/avoids distress; (ii) we consider one thing better than another (pursue it in preference to another) if and only if we consider that its pleasure value exceeds that of the other; and (iii) in all this we show that our final objective is a life of pleasure free of distress, because that would be best of all, and anything else is pursued as the nearest approximation.

Granted this, it follows directly that if someone fails to pursue the course which yields the maximum pleasure, then either they are unable to do it, or they do not know how to achieve that goal. Since we are assuming ability in the weak, they must be displaying ignorance. In fact Socrates does not move directly to this conclusion. Instead he argues first that he is now in a position to show the common opinion that we can be overcome by pleasure to be absurd. He claims that he is entitled to substitute 'pleasure' for 'good', or conversely, in the common description of deliberate wrong-doing, and that the outcome is absurd.

The details of this interlude are fascinating, but I shall not go into them. I shall just make two comments. First, the argument is nothing like so bad as it looks if we recognize that there is for Socrates an asymmetry between 'A believes that x is good' and 'A believes that x is pleasant.' The first is equivalent to 'A takes x as something worth pursuing', while the second attributes to A the belief that x has a certain property. The whole argument to this point has been that all our deliberating and valuing takes a certain form, that of judgments about pleasure outcomes. Second, it is an interlude whose function in the argument is twofold. First, it is easier to persuade someone to take

seriously an apparently implausible thesis if you can show them that on their own admissionstheir own thesis is absurd. This gives them a motive for taking an alternative seriously. Second, in discussing the results of substituting 'pleasure' for 'good' Socrates notes the objection that there is surely all the difference in the world between immediate and distant pleasure. This gives him a natural passage to a discussion aiming to show that weak people do not pursue what they think to be less pleasant/good, but are misled into thinking that the nearer pleasure is also overall pleasanter/better. It remains that the conclusion that ignorance explains bad choices follows directly from the earlier argument, and nothing in the substitution argument is needed for that conclusion.

Appearance and Knowledge

In developing the second substitution, Socrates recognizes that one difficulty people will have with it is that it requires us to say that the nearness or distance of a pleasure makes no difference to the question of pleasantness; yet the weak commonly fall for the near in preference to the distant pleasure. Clearly, Socrates can only cope with this by claiming that they must be duped into thinking that the nearer pleasure is in fact greater than the more distant ones. This is in fact what he proceeds to do, by drawing a comparison between the effect of physical distance on the apparent sizes of objects to sight, and the effect of temporal distance on the apparent degrees of pleasure attaching to courses of action. Most of us are familiar with the visual situation, and so do not convert how things look into judgments about their relative size. Socrates seems to claim that if we were not familiar with some measuring technique enabling us to compare the sizes of objects, then that conversion would be automatic, and our beliefs about relative size would change as we changed our relative positions. With pleasure we most of us lack the sort of measuring corrective which we have in the visual case, and so the appearance of greater pleasure conferred on an activity by its proximity is converted into a judgment of its greater pleasantness. Since, presumably, how pleasant something appears will be a function not only of how near it is, but also of how much we are currently enjoying ourselves or otherwise, our judgments can be expected to be in a constant flux and, by analogy with the visual case, most of them probably false. Only knowledge of the measurement of pleasure can save us from this situation, but, if we can transfer to pleasure the remarks on the phantasy case of everything in life depending on

judgments of size (356d–e), such knowledge will in fact render appearances powerless and bring peace and stability to life. (For criticisms of the visual analogy, see Taylor's commentary *ad loc.*)

This gives us the thesis that a certain form of knowledge is both necessary and sufficient for achieving our goal in life. Since everyone who lacks knowledge is a victim of appearances, and no-one who has knowledge goes wrong, this thesis does not distinguish between the weak and the vicious wrong-doer, at least in that first, they must both be ignorant, and secondly, as such, but nevertheless making judgments, must be judging by appearances only. Now the suggestion of the condition of someone lacking knowledge is that their judgments of relative value will change from moment to moment. This yields quite a good rival picture of the akratic as someone who at the time misjudges the value of the course chosen, and is only going against their better judgment in the sense of now judging something different from what they usually judge. It is not at all clear how to generate the unwavering villain.

In all this it is perhaps worth noting certain distinctions which Socrates accepts or rejects.

(i) Socrates accepts, and holds that we all understand, the distinction between something's being pleasant considered in isolation, and being pleasant overall, taking into account the action itself and its consequences. Our ability to weigh up pleasures in deliberation relies on our understanding this distinction.

(ii) Socrates denies that we are any of us capable of desiring anything but what we think best/pleasantest overall. So although we all know the contrast between short-term and long-term pleasure, there is no such thing as our wanting what we recognize to be pleasant in the short term but less pleasant than some alternative in the long term.

(iii) Socrates allows that we are sometimes attracted to/want what is in fact a short-term pleasure in preference to what is in fact a long-term greater pleasure, but explains the fact, consistently with (i) and (ii), in terms of our being victims of appearances, so that we erroneously judge the short-term pleasure to be overall pleasanter than the long-term.

(iv) There is a contrast which Socrates neither accepts nor rejects,

> but whose absence from the discussion makes many modern readers have difficulty in accepting his conclusion. This is the distinction between a probable and an improbable pleasure. Many people today take it for granted that the probability or otherwise of a pleasure must make a difference to how worth it is pursuing. While there are obviously risks involved in delaying one's pleasures, the distinction between the probable and the improbable is not the same as that between the near and the distant, and it is a distinction which finds no place in Socrates' exposition.

Socrates' Thesis

It should now be clear that the substitution argument is a counter-attack which does not itself make any contribution towards establishing the nature of *akrasia* or showing why knowledge cannot be dragged about, or whether any pleasure can be bad. If one omits it, the main argument proceeds quite smoothly. Its success does, however, depend on the main thesis, which is sufficient by itself to achieve Socrates' other purposes. This main thesis is extremely strong, and it is worth stressing certain important features of it.

The first is the streamlined view of evaluation and motivation. At 354b–c Socrates moves directly from the admission that the sole grounds for calling operations, wars and such like good are the pleasures they yield, to the admission that it is pleasure that we pursue as good. It might be thought that it would be open to someone to distinguish between what we recognize, when reflecting, as what makes projects worthwhile, and what in practice we pursue and are attracted to in the hurly-burly of life. Socrates rushes us past that possibility. Further, there is only one object of our pursuit, so that there is no room for effective conflict of desire. Three things that might suggest conflict can happen: I might, for instance, when faced with a snooty shop-assistant, say that I wanted to give him a piece of my mind but thought better of it, when the situation is that I could think of nothing suitably cutting to say, only some pompous idiocy. So in fact I wished I were clever enough to think what to do, but there was no specific option which I wanted but resisted. Second, I can recognize the lesser pleasantness of another course, and so, in a sense, ineffectively desire it. Third, I can vacillate: I eye the éclair, remember pictures of paunchy heart-sufferers, eye the éclair again, and so on. In this case I

want the éclair, then decide I do not, then think perhaps I do. All this Socrates can allow for. What he cannot allow is the simultaneous pull in different directions of desires for distinct objects not desired under a heading which yields an agreed relative assessment. Given my one objective, I shall always go for what I think will yield it, if my belief remains stable long enough to produce action. The result is that any reflection we do on the grounds of valuing things will, if it reaches the point of drawing attention to something which can be done here and now to achieve those values, lead to appropriate action.

As I have emphasized, the result of this is that if, in being overcome by pleasure, etc., I *choose* to indulge, then I must believe that the indulgence is pleasure-maximizing, or the best thing to do. Since the account is supposed to apply to being overcome by passions generally, this leads to the second point: Socrates is committed to the view that greed, fear, anger, love and the rest are forms of judgment of what is best/pleasure-maximizing or bad/non-pleasure-maximizing. This is not altogether implausible. Socrates is not alone in wanting to incorporate some element of belief into the passions. In his case the thesis would have to be that a person who is greedy thinks this food a good thing to get; who is afraid, thinks that this situation is best avoided; who is angry, thinks that this treatment is of a sort it would be best not to sit down under; who is in love, that this person is one it would be good to consort with. A person who chooses to act on passion thus acts in accordance with the judgment which at least partly constitutes that passion. (There will be more to be said about Socrates' view of the passions later.)

Third, it should be noted that the absurdity in holding that knowledge can be overcome by pleasure and the rest is brought out without anything particular having been said about knowledge. Obviously, it is assumed that if I know an indulgence to be bad and nevertheless choose to indulge, I at least believe it to be bad and nevertheless indulge. So if it is impossible for me to choose what I believe to be bad in preference to what I believe better, it is impossible to choose what I know to be bad in preference to what I know to be better. This contrasts with the later, *Republic*, view.

Finally, Socrates accepts very tight connections between (i) knowledge and true belief, (ii) acceptance of new information and change of belief and (iii) knowledge and the calming of passions. When at 356c *seq.* he turns to the explanation of how people are attracted to short-term pleasure, he draws his analogy between, the effect of the temporal distance of pleasures on attraction-appearances and the effect of physical distance on visual and auditory appearances. At 356d we read:

> So if our prospering depended on this, depended on doing and choosing large distances, avoiding and not doing small ones, what would emerge as the salvation of our lives? would it be the skill of measurement or the power of appearance? Did not the latter turn out to deceive us and make us often change this way and that about the same things and repent both our actions and our choices of large and small things, while the skill of measurement would deprive this appearance of its power, and by making clear the truth would bring the soul to steady rest in the truth and would save our lives? In face of this, would they agree that anything but the skill of measurement would be our salvation?

Note, to begin with, that we are only given two choices: the skill of measurement, which turns out to be knowledge, and the power of appearances. Since the power of appearances is characteristically in error and deceives us, it is useless for our purposes. So if anything is going to help, it has to be knowledge (or the skill of measurement). This, however, is apparently sufficient to remove the power of appearances and produce the truth. No room is allowed for relatively stable true (or false) opinion, though there might be room for a stage of progressive acquisition of skills of measurement Further, it seems that acquisition of the skill, and its application, bring belief automatically in their train. It is plausible to hold that if I know the relative pleasure measurements of two courses of action, then I have a true belief as to what they are; but Socrates seems to be drawing rather more from that apparent truism. He seems to suppose (i) that if I have the skill, I shall always exercise it; (ii) that if I begin to exercise it I shall always have the time or information to reach a true conclusion; and (iii) that applications of my skill will always carry my judgment with them. This last brings home the strongly 'intellectual' account of the passions to which he seems committed. For the analogy with the untutored viewer who is deceived about sizes is supposed to give us the condition of the putative weak person 'overcome' by passion. That person, then, must be in the condition of making untutored judgments about where maximum pleasure lies/what is best. Since proper tutoring removes the power of appearance by replacing false judgment with the truth, it must similarly remove passion.

This point is of some interest because it highlights what an extremely streamlined version of the human agent Socrates has. We have already seen that he hones down our motivational mechanisms to a single desire; but while this may be necessary it is not sufficient to produce immunity

to action against reason. One might, after all, claim that there are different, partially independent, routes to judgment. Thus, suppose I want to get safely over a deep ravine. Across it there is a rope bridge which swings alarmingly. I do stringent tests on the strength of the ropes and conclude that the bridge could safely carry three elephants at a time and cannot tip sufficiently to toss one down into the gorge. On approaching it, however, I cannot, as I might say, bring myself to believe in its safety; I am scared stiff; I stay this side. There is no division of motivation here: I just want to get safely across to the other side. There is, however, division of judgment. By one route to a sort of judgment I am convinced it is safe; by another I am not. The admission that measuring skill removes the power of appearances seems to suggest that conviction follows immediately on calculation. Emotions such as fear have to be judgments based on appearance, which will evaporate when we are faced with the facts. We are uncomplicated not only in motivation but also in our modes of acquisition of belief. It is this that makes Socrates' view so optimistic with regard to moral improvement

The picture irons out possible wrinkles in other ways too. Uncontrolled people always follow their judgment, but do not exercise their reason. It would in theory be possible to hold that even if emotions cannot make one act against one's judgment, they can prevent one from exercising one's reason. Surprised by something frightening or stung by an unexpected insult, I may react immediately without thinking. If this is a regular feature of my behaviour I may well be criticized for lack of control. But what is knocked aside is not my considered judgment, but even the beginnings of consideration. From the way Socrates talks it sounds as though he would not allow of this: if I once have the measuring skill about the one object of importance to me, it will be permanently in operation. At the other end, the passions will have to extend to include such things as tiredness, or sleepiness, which are responsible for as much supposed weakness as the more colourful cases, but which do not look very hospitable to the kind of judgmental analysis which Socrates wants to apply to the emotions.

Given all this, the question arises how anyone worth attending to could have cobbled such a position together. Some parts of it will not concern me. Thus, why Socrates thinks we have only one object of pursuit, or why he thinks that if we possess, a skill we exercise it and do so without error, while interesting questions, are peripheral to present concerns. More important is why he thinks we cannot act against our better judgment, and for that a new chapter is appropriate.

CHAPTER II

The Plausibility of Socrates

Socrates' position, as expounded in the last chapter, purports to tell us what form all human deliberation takes: it is all a matter of assessing where maximum pleasure lies. This in turn is based on an implicit view of the general form taken by every explanation of deliberate choice. If a person deliberately chooses, say, to go to the races, there must be a reason, which is their reason, for the choice. It may be for the fun of it, to gamble, to meet some friends—but whatever it is, it constitutes something which the agent takes as worth pursuing, and to which going to these races seems most conducive. In Socrates' terminology, the person's reason gives what that person takes as good, and the races are selected as what they consider the best way of achieving that good. So the general form taken by deliberate action is that it is the choice of the course which seems best for the achievement of the good in hand. Socrates, of course, holds the stronger view that with humans, at least, there is one good which they have as their objective in all deliberate action. This entails the weaker thesis that for every deliberate action there is some good being pursued, whose pursuit explains the choice; and this weaker thesis is itself sufficient to make action against one's better judgment puzzling. It is also a plausible thesis which has proved perennially attractive to philosophers. If it is true that I am going to the races deliberately, then it seems to follow that I must have a reason for going, something I take as my objective or good. If I have no reason at all for going, then while I may be going unthinkingly—inadvertently, in a dream, from sheer habit—it does not look as though I can be going deliberately. Further, it seems to be a test that we have the right account of my action that what I am currently doing is what I think most conducive to the achievement of my objective. Suppose I am deliberately walking down the street, and asked why, say that I am going to the races: 'But the race course is in the opposite direction.' To preserve the original account of my action we need some special explanation: I am

bad at geography, tickets for the races have to be got at a shop down the street. But these are ways of securing that I am doing what I think best for my objective.

This picture of deliberate action immediately creates problems for the supposition that people might deliberately choose something thinking it to be the worse course. If a woman indulges in chocolates, knowing them to be bad for her health, then the pleasure of eating the chocolates must be her objective in that choice—she cannot possibly be taking her health as her good. No doubt this seems at first glance to fly in the face of the facts, but since it is not at all obvious what alternative account of deliberate action is available, it is worth giving Socrates a run for his money. So first I shall illustrate how he might deal with the apparently intractable facts, and then sketch some of the problems facing anyone taking a non-Socratic line.

Socrates and Counter-examples

What Socrates has to do, to win conviction, is show how apparent cases of people being overcome by fear, pleasure or the like, so as to act against their better judgment, are really cases of people doing what they think best.

Suppose that I am frightened of dogs. I have been asked to lunch by a friend in a neighbouring village. I want to arrive on time with a good appetite for lunch. So I decide the best thing to do will be to walk, but by the shortest route. My wife then points out to me that that will take me past a house where a number of Alsatians are kept. Because I am afraid of them, I decide to take a slightly longer route and set out a little earlier. Clearly my fear has got the better of me and swept me from my originally chosen course. Equally clearly, Socrates can cope with such a case by pointing to the distinction between an historical account of my decision-making and the formal account of my final decision. The history is of someone who started with a goal and then, in the face of some unpleasant facts, altered course. But clearly what altered as my reflection proceeded, was my objective. This comes out in the final explanation of my choice: I take the longer route because my objective has changed from being to arrive with a good appetite, on time, with least delay, to being to arrive with a good appetite, on time, with the least delay compatible with avoiding alarming dogs. What I do is what I think will best achieve the revised objective.

This, of course, is about the weakest case of weakness. The more difficult sort would be where, having decided on and set out on the shortest route, I hear the sound of raging dogs baying from the path ahead, and in fear hurry down a longer alternative route. Now, surely, what has happened is that my fear has driven me from the pursuit of my goal? Yet why should Socrates meekly accept that account? There are at least three fairly standard ways of describing situations like the above: one is to say that I was overcome by fear; another is to say that I gave in to my fear; and a third is to say that when I heard the dogs it no longer seemed so desirable to arrive on time at all costs: it came to seem a better idea to avoid the dogs, even if it meant arriving a little late. Further, Socrates could point out that if we insist on taking the talk of being overcome or carried away at all seriously, then equally seriously we put in question the description of the action as deliberate. If we wish to keep that description, we find ourselves pushed towards either the second or third account above. But to say I gave in to my fear, while it suggests the choice was mine, leaves it puzzling, unless we fill it out by the third description. This is the only one which leaves us with a clear account of the behaviour as deliberate. But this is Socrates' description, and amounts to saying that my objective has changed; but I still do what at the time I think best It may indeed be that others think my initial objective better. I may myself, in my more reflective moments, wish that I could keep steadily to that objective. All this may give sense to calling that my better judgment. It remains that I only do what *at the time* I think it best to do. Apparent concurrent regret can be catered for along the lines suggested earlier (see Chapter I, p. 12). In short, such cases are never ones of my mind being set on one course and passion carrying me on another, but are always cases of my changing my mind. I may think my first judgment better, but acting against my better judgment in that sense is not acting contrary to what at the time I think best

It is sometimes thought that Socrates' thesis runs counter to the plain facts. So Lemmon:

> It is so notorious a fact about human agents that they are often subject to acrasia that any ethical position that makes this seem queer or paradoxical is automatically suspect for just this reason. Of Socrates we can say that as a plain,matter of fact he was just wrong—acrasia does occur, or in Aristotle's phrase, knowledge just is, however sad this may be, frequently dragged about by desire. (1962: p. 144)

It should now be clear that what Socrates is actually doing is raising the question of what the facts are. He does not question that circumstances arise which we commonly describe as being overcome by pleasure or fear; what he objects to is taking that description seriously as an account of the mechanisms of choice. His alternative account of the phenomenon is of judgment ungoverned by understanding.

Problems with Not Following Socrates

The Socratic objection to the ordinary picture of weakness comes from the Socratic account of deliberate action, and involves the claim that the ordinary picture, taken seriously, would undercut the claim that the action was deliberate. No doubt the result is unconvincingly clear, but the alternatives do not look very much happier. People experiencing the influence of passion are often alarmed by the strength of their surging anger, or jealousy or greed, and can hardly credit their own behaviour. But this whole way of talking shows their confusion. They try to think of themselves in two ways at once: both as agents struggling with ill-understood and alien forces, and as agents of the actions done under the influence of those forces, of whom those forces are an important and integral part. We are ashamed not just of giving in to jealousy or fear, but of being afraid or jealous. So the choices for the non-Socratic are either to abandon the idea that these actions are deliberate, or to accept a picture of a radically divided unified agent at least as puzzling as any Socratic paradox.

Socrates has been accused (see Austin 1961), along with Aristotle, of saddling us with the view that all deliberate wrong-doing, or supposed deliberate wrong-doing, is a matter of being overcome by passion or emotion. Certainly Socrates does not consider any other putative cases, and would presumably be entitled to feel that if he could cope with the most plausible, passionate, ones, the absurdity of the common person's view would become more obvious as we moved away from emotional involvement. This entitlement would admittedly partly come from his stronger thesis that there is one good that we all always pursue. This position excludes two possibilities: first, suppose a man who is attracted by two ambitions, one to be a person of intellectual integrity, engaged in academic pursuits, content with moderate[1] comforts and despising the vulgar show of wealth; the other to be someone so accomplished in deceit and sharp practice as to be

able to outwit all financial rivals, become a millionaire and so, without having to count the cost, indulge in every vulgarity that wealth makes possible. Socrates has to deny that these can appear to the man as two rival ideals of life which he cannot refer to any over-arching goal as a criterion by which to judge between them. Many would wish to allow for the possibility of goals which are in this way incommensurate, and certainly the possibility would admit strains in one's decision-making that Socrates would exclude. Secondly, suppose the woman mentioned earlier, who knows chocolates are bad for her health, but cannot resist. It is certainly plausible to accept a Socratic account of the case: faced with this enticing variety of mouth-watering continental mixture, all this health business did not seem so important to her any longer. But is that to say that she thought the good life lay with eating chocolates? Is it not rather that judgments about the good life have ceased to interest her for the moment? No doubt she is taking chocolates as the thing to be pursued just now—that is, she makes them her goal—but that seems a less heavily burdened state of mind.

Socrates certainly wishes to rule out these two possibilities, but two points should be noted: first, anyone wishing to take these options should consider the implications for our picture of ourselves as unified agents; secondly, Socrates' difficulty arises even if we abandon his streamlined version of both motivation and belief. He can, of course, accommodate people who acknowledge that what they are doing contravenes the requirements of some principle or goal called good; but if the claim is that at the time of deliberate action, the course deliberately chosen is seriously thought to be worse (not what is worth taking), or in some strong sense not what we really want, then the challenge is to produce a coherent account of deliberate action which allows for it. What is the force of the claim that they seriously consider something else better than what they are actually choosing, and consider the chosen action bad? In what sense can they be said really to want to choose some course other than the one chosen?

The Later Plato

While the Socratic challenge is one which remains to be answered, the Socratic position still remains unconvincing. Our calculations do not seem all geared to producing what we consider to be the good life. Sometimes our interest is in what we acknowledge to be short-term

objectives. Our woman with her chocolates has ceased to be interested in the wider questions of life, and her intellect, distracted from working out how best to secure her health, is engaged instead on the problem of the most delectable order of consumption of what is before her. She makes no pretence of concern with her health, nor of believing that eating these chocolates is conducive to her health, her overall pleasure in life, the Good, or any such long-term objective. We could only insist that nevertheless that is what she really believes if we were prepared to abandon our normal tests for belief in order to satisfy the requirements of a theory of action. By the *Republic* it is clear that Plato is no longer prepared to pay this price. In Book IV (435 *ad fin*.) he insists that desires cannot be collapsed into one desire. Thirst is a desire for drink, not a desire for the good. This allows for the possibility that our desire for a drink may come into conflict with our desire for what is good, as when we are thirsty but think that it would not be a good thing to take a drink. It leaves it open that either desire may get the upper hand, and Plato gives an example where a person's judgment of what it is good to do is overpowered by a disreputable desire. The example is of Leontius, who is overcome by a desire to gloat over the sight of some corpses while disgusted at giving in to so shameful an inclination. So Plato has abandoned the streamlined picture of human beings so far as motivation is concerned, and with it the rejection of the possibility of acting contrary to what we think best.

What is important for present purposes about the argument for division in the soul, is that Plato is arguing for a type of desire which does not manifest an interest in discovering and pursuing what is best, but is simply directed at what is taken as replenishment of a current lack. On the Socratic thesis, since every pleasure is as such attractive, it is presumably possible (see Woods 1987), especially in cases where we cannot tell where maximum pleasure lies, to have a conflict of desire of a sort. But they are always manifestations of the agent's interest in the best life. In the *Republic* Plato is acknowledging motives which do not show that interest at all, and which may even drive it out. He still thinks that we cannot act contrary to our knowledge of what is best, but the account of knowledge has changed beyond recognition. To know, one has to have undergone the philosopher's training. This leads to a grasp and appreciation of the structural beauty of the universe, a beauty which wins a response of intellectual love which distracts our desire from other things and concentrates it on this alone, bringing peace to the soul and order to our inclinations. This removes all chance of conflict. Desire cannot overcome

knowledge, because in the sage there is no turbulent desire to take issue with knowledge. It is, however, possible to havebeliefs about what is good, without being in the condition of a sage, and in that condition one may have unharmonized desires which get the better of one's judgment.

Clearly the thesis that one cannot knowingly do wrong has by now taken a quite different form. It is not dependent on the thesis that one cannot act against one's judgment of what is best and the associated account of deliberate action, but on a special account of the state of complete knowledge and its effect on the personality. In the last resort it is an empirical thesis about the effect on our other desires of a love and understanding of the structure of the universe.

Although at this stage Plato allowed for divided motivation, he gives no sign of allowing for vacillation of belief of the sort sketched in Chapter I (p. 15). It would have been possible for him. The spirited element of the soul in the *Republic*, which in Leontius is overcome by a disreputable desire, is at least an emotional response to what is pictured as noble or ignoble. This would allow for a conflict between what is emotionally and what is with the head viewed as noble; but Plato does not seem to have envisaged such a rift. The spirited part of the soul seems to be seen as always responsive to the findings of reason, in the way Socrates supposed true of all passion. It would, however, be rash to be dogmatic on the point. We do not get a close examination of the condition of someone on the way to full understanding, and it is possible that at that stage there could be residual lack of harmony between the findings of reason and emotional judgment. Further, although the argument for the distinction of the spirited element and reason is not in terms of reason being overpowered, they are portrayed as faced in different directions, with reason rebuking spirit Whether spirit always accepts rebuke or sometimes ignores it is left unclear. What is certain is that in the philosopher there can be no such conflict.

For the later Plato, therefore, there is no general problem about weakness of will: it occurs as the ordinary person of the *Protagoras* says it does. The cost is to accept some view of conflict of desires; in the *Phaedo* and the *Republic* there is much rhetoric about the warfare between soul and body or between parts of the soul, about the implication of which Plato expresses some unhappiness in *Republic* Book X (611–12). It is not an unhappiness that he anywhere succeeds in putting to rest. (For a fuller discussion see Gosling 1973: Chs 2 and 6.)

There are certain possibilities, ruled out by the *Protagoras*, which are

allowed for, even if not discussed, by the view in the *Republic*. I remarked earlier that in the *Protagoras* one gets the impression that Socrates wants to hold not just that knowledge cannot be dragged around, in the sense that I cannot act contrary to what I know to be best, but also that reason cannot be disabled by passion from operating: if I have the skill that constitutes knowledge of what is best, panic cannot disable me from exercising it. The *Republic* does nothing to discount this possibility. Further, the *Protagoras* (see Chapter I, p. 14) seems to require that my skill will always be in operation when appropriate. This rules out a fairly common kind of weakness. Suppose I have been invited to join a Mediterranean cruise by someone I do not know very well. It is something I have always longed to do, and this seems a golden opportunity. So strong is my determination to take it that I do not pause to ask the prudent questions about the company I shall be keeping. It is certainly allowed for on the *Republic* view that my desires may be so insistent as to prevent the use of reason on raising questions about the wisdom of indulging them, at least so long as I have not reached philosophical perfection.

While various possibilities are now allowed for, however, they do not receive any detailed discussion. Once the view of human nature which made *akrasia* seem impossible was abandoned, Plato lost his sense of urgency about the question. We do not, however, get any general account of deliberate action which makes it clear how to meet Socrates' difficulties; nor is it clear how Plato is going to deal with the apparent chopping up of the person which seems implicit in accepting the ordinary ways of talking about *akrasia.*

In all this, I do not wish to suggest that Plato abandoned all suggestion of Socratic paradox. As I have said, he still holds, in the *Republic,* that no-one can knowingly do wrong; and in the *Timaeus* (86d–e), and again in the *Laws* (73Ic, 733b) we are told that no-one is willingly (*hekon*) evil, unjust or intemperate. These retain the wording or apparent intent of the *Protagoras,* but the theses have changed. I have already indicated this in the case of the *Republic.* In the *Timaeus* and the *Laws* the interest has shifted from individual acts to states of character. I might do things which I thought wrong, but I would never willingly acquire an evil state of character, if I understood what it consisted in. The point here seems to be that such a state is clearly opposed to what I would recognize on reflection as desirable. This has little to do with deliberate acts.

Conclusion

It is time to draw together the threads of this discussion of Plato. The thesis found in the *Protagoras* is later abandoned. It seems likely that increasing distrust of the body, the senses and the desires associated with them, convinced Plato that sensual desires certainly could not be plausibly portrayed as manifesting any concern for or thought about what was the most desirable form of human life. Whatever the reason, by the *Republic* he is clear that desires are distinguished by their objects, and that they may conflict. Further, we can act contrary to our beliefs as to what is best, and in some sense contrary to our reasoning. We also, of course, in some sense do these acts for a reason. How does our decision-making go? There is no answer, no explanation of where Socrates went wrong so far as that is concerned. There is indeed a sense of 'better judgment' or 'judgment of reason' whereby these have to take the form of showing a concern for the good life or what is best for humans or what is best considering the universe at large. It remains that at the time of acting Leontius takes seeing the corpses as his goal/good, and does what he thinks best to achieve it. There is some awareness that the result of accepting the common view of *akrasia* is embarrassing for the unity of the soul, but the embarrassment seems largely metaphysical. There is little sense that if we take the rhetoric of violence and conquest seriously, we jeopardize the possibility of treating the akratic as acting deliberately and holding them responsible for their behaviour. The rhetoric of violence, however, especially with regard to bodily desires, strongly suggests some incapacitation of reason as regards the pursuit of long-term questions about the good life (reason is made the slave of lower desires in *Republic* IX), and this may have laid the foundations for part of Aristotle's later discussion. Anyone feeling the pull of the Socratic position, however, could be forgiven for feeling that Plato had not cleared up the question satisfactorily. There is plenty of room for a treatment by Aristotle.

CHAPTER III

Aristotle

Aristotle's main treatment of *akrasia* is in Book VII of the *Nicomachean Ethics*. Strictly, indeed, this is his only treatment of it, but it has to be interpreted in the light of his general account of action, and of various comments about *akrasia* in a number of different places.

Interpretation of Nicomachean Ethics *1145b21–47b19: the Innocent View*

There is considerable dispute about the interpretation of this passage, and a proper treatment would require a book in itself. I shall confine myself to reviewing some of the main problems so as to illustrate how they give rise to importantly different interpretations of Aristotle, and to clarify my own interpretation. The major problem is an apparent inconsistency in Aristotle's treatment, given a natural way of reading his introduction to it. Since the major variations of interpretation are most easily understood as reactions to this problem, I shall begin by outlining what I have called the innocent view. I shall then shown how it leads Aristotle directly into inconsistency. There then emerge certain obvious ways of trying to obviate the inconsistency.

It is worth noting that Aristotle might be expected to have some sympathy with Socrates. Following both Socrates and Plato he thinks that human beings are essentially rational animals, that is, beings characterized by a capacity-cum-inclination for discovering the truth. In practical matters the actualization of this capacity-cum-inclination will take the form of working out and pursuing what is the best way for a human to act. There will therefore be an air of paradox about cases where apparent actualization does not issue in appropriate action.

The problem is introduced at 1145b21 as one about the way in which people rightly understand (what they should do) when they act akratically. Socrates is cited as holding that it cannot be knowledge, since no-one can act with understanding contrary to what is best. Others hold that it is true judgment, not knowledge, against which the akratic acts. When at 1146b8 *seq*. we are given the order in which the problems are to be discussed, we are first to discuss whether the akratic have knowledge or not, and then what sort of knowledge. This is what is in fact done. At 1146b24 Aristotle argues that there is no point in insisting on the distinction between knowledge and true belief, since what gives rise to the difficulty is firmness of conviction. (The result of this is that we might as well face the difficulty as one about knowledge, but what we have to say must apply to firm conviction as well.)

We have now, therefore, to tackle the question of the way in which the akratic knows, with a view to removing Socrates' implausible opposition. While Socrates believed that it was quite generally impossible to act contrary to one's judgment of what was best, what he was immediately opposing was the view that someone's judgment of what is best can be overcome by desire or passion so as to make them choose a course they consider worse. It is the sort of knowledge a person has when in such a condition that is under investigation. To tackle the question of calm deliberate wrong-doing would not bear on Socrates' problem. So Aristotle proceeds to put aside situations in which no problem arises in order to isolate the situation which would be difficult and where we might need to specify the way in which the akratic knows. First there is the distinction between having knowledge without using it, and using it. No problem arises if use is absent, and so we assume some use. Then there is a distinction between a universal premiss (such as that everything sweet should be tasted), and a particular one (such as that this is sweet). There is no problem with the universal being used, so long as the particular is not. So we assume use of the universal—as indeed one might expect in discussing *akrasia*, since one would expect the akratic's principles to be engaged. Finally, the universal divides: one covering the type of agent, the other the type of thing to be chosen. We are now near the heart of the problem: given use of the universal that dry foods are beneficial for all human beings, and the particular that I am a human being and that such and such a food is dry, any one who is going to act against knowledge does not have, or at least does not use the knowledge that this food is of this sort. If you supposed that they were using it, then action against such knowledge would be amazing.

So far no mention has been made of the akratic, or how they know. We have only isolated circumstances where a problem would arise, unless we are content to attribute to the agent no more than having knowledge of a particular piece of information. At this point Aristotle introduces a distinction within having without using. There are a great many pieces of knowledge which most of us have, which are available for use on demand, but are not currently in use. If we are asleep, or drunk, or in a fit of madness, while we still in a way know these things, they are not at the moment available for use: only when we have recovered consciousness, sobriety or sanity will they be available again. Still, we are better off than someone who has never learnt or totally forgotten them. In a way we have them, but one remove back from the first way. This, says Aristotle, is the way the akratic know. While the passage is fairly dense, the picture seems to be as follows: the akratic are using the universal premiss, but if using any particular information must be having other relevant particular information in the manner of drunkards, etc.

At this point Aristotle envisages an objection to the effect that the akratic clearly know the information in question because they (often, or always to themselves?) express it orally. Expressing views, however, even if they are derived from knowledge, is no proof of knowledge; and various examples are given to show the inference invalid.

So far Aristotle has given conditions under which a problem arises, and specified the form of knowledge of the akratic: a way of having characteristic of someone in a state of desire or passion. He now turns to see how it works in practice. In practice what is important is not the agent's knowledge, but their judgments, and so we are told that there is a universal and a particular judgment, and that, as we would expect, when an inferential judgment results, the subject must assert the conclusion, and in practical cases act: if everything sweet should be tasted, and this is sweet, then if the agent can and is not prevented, they must do this at once. We now turn to the akratic case, and suppose a judgment preventing one tasting. In the context one would take that as picking up and reversing the previous example, and so it is natural to take the judgment as one to the effect that nothing sweet should be tasted. Assuming use of this, and use of the particular premiss that this is sweet, everything we have heard so far would lead us to expect that the agent must at once refrain. In fact Aristotle blandly says that as desire is present, together with the judgment that everything sweet is pleasant, desire drives the agent to taste, despite their judgment that they

should not. According to the earlier passage, given use of the universal, if the agent is to act against their knowledge, then they cannot use their knowledge of the particular premiss; but Aristotle goes out of his way to affirm use of the particular premiss in the akratic case. Then a few lines later, at 1147b9–12, he calmly goes back to the earlier view and says that the akratic either does not have or at least does not use the particular premiss. Our very natural interpretation, therefore, involves attributing to Aristotle a very blatant inconsistency.

The Inconsistency

There seem to be three main assumptions underlying the production of this inconsistency. First, that 'this is sweet' is the particular premiss which yields, with the universal forbidding one to taste, the conclusion that one should not taste this. One way of avoiding the inconsistency would be to make that false. Secondly, it is assumed that it is necessary and sufficient for my using my knowledge that this is sweet that (i) I should have before my mind some such sentence; (ii) I should understand what it means; and (iii) I should judge it to be true. If we could find different conditions for 'use' we might be able to absolve Aristotle of inconsistency. Thirdly, it is assumed that the different parts of the passage are only discussing one sort of akratic. If we can show Aristotle to have a variety of cases in mind, then the apparently inconsistent assertions might simply apply to different cases.

The First Escape

The obvious way of saving the particular premiss from being the appropriate one for the universal is to change the universal. Since Aristotle does not specify the form of the universal we are not bound to the innocent version. We could free ourselves dramatically by making the universal 'no cake is to be tasted', or something of that sort. Clearly 'this is sweet' yields nothing with this universal premiss. There is room now for knowledge of intervening premisses such as 'all cake is sweet' and 'this is cake' to be had in a drunken fashion.

While this might give us a form *of akrasia*, it would be unfortunate if we had to say that Aristotle held that the universal must take some such form. It would mean that he felt that his account confined him to what

seems a rather limited range of possibilities. On the face of it, if we can have *akrasia* at all, we can have cases where desire for sweet things conflicts directly with a principle forbidding sweet things. This might make one prefer Aquinas' suggestion, in his commentary on this passage, that the universal has some such form as 'sweet things are not to be tasted between meals'. This has the advantage that the particular premiss has some relevance to the universal, while nevertheless being insufficient to yield a conclusion. It remains that Aristotle would not be able to envisage a conflict between desire and a general prohibition. Yet he was not averse to general prohibitions. He seems to have thought (1110a26–9) that there were some things one should never do, such as kill one's mother. Yet surely one might see one's mother as a hindrance to one's advancement and, if ambitious enough, while fully recognizing her as one's mother, despite the wrongness of matricide, want to kill her. It seems a resort of desperation to suppose that if one is to kill her one must be accepting as one's universal some such premiss as 'one should not kill one's mother between meals'.

The Second Escape

This might make one want to preserve a general prohibition on tasting sweet things as at least one possible sort of universal. In that case, if one is also supposing that throughout the passage Aristotle has one sort of akratic in mind, it will seem attractive to investigate the distinction between 'using' and 'having'. In a discussion in the *Prior Analytics* (67a33–65) using knowledge of a premiss is a matter of putting it together with other premisses to produce a conclusion. In that case there is no such thing as simply using a piece of knowledge: one has to be using it in relation to some premiss (and may be failing to use it in relation to another). The text might even suggest that this is how Aristotle is thinking. At 1147a31 *seq.* we find: 'so suppose there is a universal [judgment] present preventing tasting, and another, that everything sweet is pleasant, and this is sweet, and this one is actualised.' Here 'this one' would naturally be taken to refer back to 'another', in which case Aristotle seems to be putting together 'everything sweet is pleasant' and 'this is sweet' as jointly actualized, so emphasizing that 'this is sweet' is used in relation to 'everything sweet is pleasant'. It is not, however, being used in relation to 'nothing sweet should be tasted'. The perceptual knowledge is highjacked by desire into use with a premiss other than that of reason. The agent still in a way has the

knowledge available for use with the premiss for reason, and may even in drunken fashion give the appearance of using it, but the knowledge is at two removes from use, because of the influence of desire.

This certainly has the advantage of allowing the akratic directly to flout a general principle. There are at least three apparent difficulties with it. First, it is obscure how the akratic fail to use the particular premiss. After all, they know full well that nothing sweet should be tasted, and are supposedly aware of that fact at the time; they are also fully aware that this is sweet. How on earth, then, do they fail to conclude that they should refrain? Second, they clearly do, in fact, conclude that they should refrain. We are told that the one judgment says to avoid this; and that can only be because it has joined with 'this is sweet' to yield the prohibition; but then, according to 1147a25–8, the agent should act accordingly. In other words, the situation described seems clearly one where use as defined has occurred in relation to both premisses. Third, this interpretation seems to leave us with a situation where there is not really any conflict. Desire for sweet things simply drives out serious consideration of anything else: there is no struggle, no anguish, simply defeat of one desire by another. Yet Aristotle clearly thinks that both the self-controlled and the uncontrolled struggle: it is just that the outcome is different. Quite apart from this, it is to be hoped for plausibility's sake that Aristotle realized that people do take what they think to be the wrong course, aware that what they are doing is wrong.

The Third Escape

This might lead one to look at the third assumption. This, after all, is why people insist on interpreting various sentences in the way they do: they have to be made to fit the one kind of akratic. Freed of that assumption we can read the text more naturally.

To start at the end of the passage, we can now take the reference to the final proposition, which the akratic is said to know in an off-beat way, as a reference to the conclusion, not the particular premiss. This now fits nicely with the actual description where the akratic is said to draw the conclusion that this thing should not be tasted. Clearly, they use their knowledge that this is sweet in relation to both premisses, and it is their knowledge of the conclusion which is off-beat and therefore not acted upon. Desire is a factor which, by affecting the apprehension of the conclusion, prevents action from resulting even when the premisses have been used.

This is not to say that Aristotle does not also envisage another sort of akratic whose knowledge of the particular premiss goes wrong. He clearly does. But in failing to use their knowledge of the particular, they fail to draw the conclusion, and so are different. It is important to recognize that between the passage allowing such an akratic and that portraying the one who acts against the conclusion, there has been an opening up of the picture of the akratic. At first we seem confined to implausible pictures of quasi-drunkards, doubtfully aware of particular facts. At 1147a18–24 Aristotle has broadened the account of quasi-knowledge so that it is exemplified not only by lunatics and drunkards, but also by people who have started to learn, but have not yet sufficiently integrated their knowledge into their thinking to count as fully knowing. With this kind of model Aristotle is enabled to envisage someone who realizes at the time that what they are doing is wrong, but whose knowledge is off-beat in the sense of not yet being fully part of themselves, and so not always practically effective. Their failure to know the conclusion is not a matter of intellectual obfuscation, but a failure, through desire, to accept properly the conclusion of their practical reasoning. If it were only a matter of failing fully to understand the particular premiss, there would be no sense, in Aristotle's account, of genuine conflict.

The Final Escape

The arguments concerning this last view are extremely complex (see, for instance, Charles 1984 or Dahl 1984). Certain difficulties present themselves as obvious. The first is that in the early part of his discussion (1147a5–10) Aristotle does seem to say that an agent who is using knowledge that dry foods are beneficial for all humans, and that they are humans, but acts contrary to knowledge, cannot be using their knowledge that this food is dry. To suppose otherwise would be amazing. While Aristotle might have thought that akratics were amazing, it seems more natural to take him as ruling the possibility out. But then he is claiming that, given use of the knowledge of the other premisses, use of knowledge of a particular premiss has to give way. Second, the proposed broadening out of Aristotle's account of off-beat knowledge seems a mistake. To begin with, Aristotle does not say that these people know in a way: he is using them as examples to refute the general inference from saying to knowing, and so learners are not said to know in a way (which would hardly support his immediate point), but

simply not to know; and he can sum up the point by saying that the akratic's sayings are like those of actors, without having to attribute any even off-beat knowledge to actors: what they say may derive from (someone's) knowledge, but they do not have to have any kind of knowledge themselves.

It is true that the akratic must have failed fully to integrate their desires and their values; but it remains that their half-knowing is their epistemological condition when desire gets the better of them, and when they return to normal they are back in the condition of simply having or also using their knowledge, although they remain not fully integrated. The analogy with early learners would yield an account of the general condition of the akratic, but not of their epistemological state at the time of the akratic act.

One point in most interpretations, which both Charles and Dahl jib at, is that they seem to make Aristotle say that the akratic do not really understand that what they are doing is wrong. For their failure fully to grasp the particular premiss ensures that they either do not reach or do not grasp the conclusion. Since this is felt to be highly implausible, it motivates a hunt for signs of greater enlightenment. This would be shown if (i) Aristotle's akratic, or one of them, was able to grasp the particular premiss and (therefore?) successfully reached the conclusion; and (ii) apprehended the conclusion in a way which was, perhaps, off-beat, but did not show an implausible clouding of the mind, (i) seems to be achieved simply by inspecting Aristotle's description of the akratic, and (ii) by the broadening out of the account of knowing in a way and in a way not.

The best place to start, I think, is with the sense of 'use' in 'using knowledge'. As we have seen, there is reason for thinking that when this is applied to using knowledge of a particular premiss, this may require understanding the premiss, but does not consist in it, even together with considering it and assenting to it: the agent has also to put it together with other premisses in an argument. What, then, is the situation with the universal premiss? Aristotle seems to think that it is possible to use this without putting it together with a relevant particular premiss, in that he says that it is all right to use knowledge, so long as it is the knowledge of the universal that is used, and not of the particular. What could that amount to? If we made the conditions assumed in the first escape above sufficient, we seem to get an absurdity. For it would result that if I know that dry foods are good for all human beings, and you draw my attention to this fact, together with the fact that I am human and this is a piece of

dry food, then I should have to take it Not only would this be inconvenient: it is clearly not how things work.

A plausible version of 'using' would be that to use a universal premiss is to take it as one's starting-point in deliberation. If I am not hungry, I am not likely to be interested in food, and so am unlikely to take that fact about food as my starting-point. But if I am hungry, and also interested in health, I am. That will start me off on an enquiry about dry foods, but it might be abortive: all the food on the menu is wet. Eventually I might look for another universal giving types of food good for the human race, but until then, my use of the universal takes the form of its giving direction to my enquiries. (I have, of course, to use my knowledge that I am human—but that, as we shall see, is not the particular premiss.) I may, with luck, reach some particular piece of information about food which will crown my search with success, and then my universal, together with the particular, will yield a conclusion.

If this is right, then the fact that I understand, advert to and assent to a premiss is not sufficient for my using it, whether it be universal or particular. If it was sufficient, we should be able to infer from lack of use to failure in one of these three conditions. As it is, we are not in that position. Consequently, for all we know, my not using it is not sufficient to show that I do not fully understand it, advert to it or assent to it. In other words, failure of use does not show clouding of intellect, unless we can supply special reason for supposing that Aristotle thought it did. What he says about use leaves it open to Aristotle to say that the akratic understand fully the minor premiss, and the conclusion, and so understand fully that what they are doing is wrong. What desire has to do is not cloud our understanding of the facts, but ensure that we are not in a condition to use them.

It remains puzzling how, if we fully understand and acknowledge the facts, we fail to use them. It may, of course, be that Aristotle thought we just did; that it only seems a puzzle if we suppose a Socratic uniformity of motivation. There is, however, one point in the passage which has so far been totally ignored. It is noticeable that Aristotle does not keep spelling things out. Having declared that it is when use is in question that problems arise, he assumes use; having said that, given use of the universal, difficulties arise with use of the particular, he assumes use of the universal and becomes interested in the particular. These assumptions are carried over to his discussion of how *akrasia* works in practice. In short, we are expected to take the early distinctions as applying through the discussion. In that case, we should assume that the

universal premiss of reason always takes some such form as 'all Xs benefit from Ys'. So when we come to the example that causes the problem, the universal premiss should read something like 'No human being (diabetic, athlete in training) should taste anything sweet.' In order for this to reach application in any particular case, the agent needs to use their knowledge not only that this is sweet, but also that they themselves fall under the prohibition. At 1147a4–7 we have been told that, given use of the universal premiss and use of the particular premiss relevant to one of the universals in it, then anyone who acts against their knowledge must fail to use knowledge of the particular premiss relevant to the other universal. Since in the example of *akrasia* in practice we are told that the particular premiss relevant to sweetness is used, the akratic agent must fail to use their knowledge that they are the right sort of agent. The use of 'this is sweet' in relation to the universal premiss therefore remains ineffective. What desire does is commandeer the knowledge that this is sweet and make the agent 'forget that they are human'.

There are three obvious objections to this: first, that the agent clearly does use all that is required, for they are said to draw the correct conclusion; second, lack of use does not seem to explain how the akratic goes wrong, but simply to be a way of acknowledging that indeed they do—and this is what we will call their knowledge; third, lack of interest does not sound like a way of failing to know, and suggests that Aristotle does not think that there is real conflict in cases of *akrasia*.

These three are connected, but I will treat them in order. As to the first, Aristotle is clearly envisaging the akratic seeming to use their knowledge. When he argues that mere saying does not prove knowledge, the examples he has in mind are not utterances of single sentences, but of people producing proofs and arguments. So the fact that someone goes correctly through to a conclusion is apparently no proof of use of any of the premisses. This, however, may seem to weaken the notion of having and not having, and remove all force from the analogy with mad people and drunkards. This apparently does not suggest a lack of grasp of the premisses, and we seem simply to be left with the fact that they do not lead to action, to give us what is meant by lack of use. The suggestion that the agent loses interest not only suggests lack of conflict but does not seem to make it a way of knowing and not knowing.

The first point to be made here is that the distinction between use and lack of it is in context a difference between using and not using, at the

universal level, knowledge that, say, dry food is good for humans. Using this is making it the starting-point for deliberation. What we have, in fact, is a manifestation of *prohairesis*, a term commonly translated 'choice', and a technical term of Aristotle's for something not possible for animals or children because of their lack of rational development. He calls it deliberative desire (*orexis*) at 1113a10–11, and desiring intelligence or intelligent desire at 1139b4–5. Briefly, the point seems to be this: adult humans have a potentiality to use their reason on the question of how to live their lives (and for Aristotle, the potentialities of living things are tendencies). This can be described as a desire to live the best life for humans; but this desire can only by actualized by forming a conception of what is good for humans and trying to work out how to achieve it. Alternatively, it can be described as a potentiality to apply their reason to the question of how to live; but this rational potentiality can only be actualized by being applied to the pursuit of the good for humans. The actualization of *prohairesis*, therefore, is at the same time the actualization of a desire and the actualization of a capacity to reason. One can describe the potentiality as a desire of a certain sort, or as reason of a certain sort, but these will be two descriptions of the same potentiality. Use of one's knowledge in the practical sphere is a manifestation of one's tendency to use one's reason to work out how best to live and is *ipso facto* a manifestation of a desire, a desire of which only beings of a given degree of rational development are capable.The use of the universal premiss, therefore, is already the manifestation of motivation in favour of something considered good for human beings. The actualization of this deliberative potentiality, however, is complex, and is only complete if it reaches fulfilment in action: unless that were the norm, we could not attribute *prohairesis* to human beings. Granted that it is the norm, we can sensibly speak of abortive *prohairesis*. It can only be abortive if the agent fails to reach a decision as a result of it. One way for this to happen is for the agent to fail to use the knowledge of the particular circumstances so as to apply the principle to their own case. For this, however, some other potentiality than rational desire must have been actualized. This cannot itself be a potentiality to raise questions about how it is best for humans to live, and in fact sensual desire does not even manifest an interest in an object as a human pleasure, but only as pleasant to the agent. It does, of course, involve seeing, say, this sweet thing as good, in the sense of taking it as a goal; but its actualization will be antipathetic to the actualization of a potentiality to reason out what is good for humans or, what will be the

same thing, will run counter to that motivation (cf. *Nicomachean Ethics* X, 1175b). It will not mean that that motivation will cease to be felt, any more than the self-controlled person ceases to feel sensual desire, although that desire is not fully actualized; nor that the agent fails to understand the arguments they rehearse to the effect that they should not do it; but it does mean that their deliberative desire is not fully actualized.

Is this, then, to make the notion of off-beat knowledge unexplanatory? The answer is 'Yes and no'. On the 'Yes' side, it should be noted that Aristotle does not undertake to explain how the akradc can fail: that is a problem for Socrates, because he has gone wrong. Aristotle's aim, given that indeed *akrasia* as commonly believed does occur, is to specify the way in which the akratic knows, and with a view to finding a truth in Socrates' view. Granted that it is possible not to use knowledge, but have it, yet have it in a way such that it is not immediately available for use, we have a required sense of 'know'. Since the akratic are using their knowledge of the universal premiss, there must be a particular premiss that they are not using. Yet if it were available for use, why do they not use it? so their knowledge must take this form. On the 'no' side: Aristotle is not prepared to leave the answer so bare. One actualization inhibits another, and the general theory underlying this is presumably meant to explain how sensual desire might be expected to interfere with the actualization of *prohairesis*, and so the use of the particular premiss.

What, then, of the truth in Socrates? Given his view of sensual desire, and his homogeneous view of human motivation, Socrates could only envisage being overcome by pleasure as having our view of what is good dragged about by our view of what is good. Aristotle does not think that sensual desire does embody a view of the human good, nor does he think that in *akrasia* our view of the human good gets dragged around. It is an important fact about the akratic that they have good principles, and are even using them on akratic occasions. The trouble starts with the observation, for instance, that this is sweet. It is this that arouses desire. But desire does not commandeer the agent's knowledge of principles, nor prevent use at that level; so there Socrates is right. But reasoning in practical matters is fairly complex, and we can fail to apply our principles because our perceptual knowledge is dragged off for use with another premiss. The knowledge that this is sweet may be attached (or may not) to the universal that no human should taste sweet things, but the actualization of sensual desire ensures that we do not use our

knowledge of our humanity together with our knowledge that this is sweet to pursue our interest in the human good.

I shall turn to the question of just how Socratic Aristotle is, and to what extent he can cope with non-passionate cases of *akrasia* in the next chapter. I shall end this one by commenting briefly on the relation of this interpretation to escapes 1–3. It clearly has it in common with the first that it alters the form of the universal premiss, and with the second that it avoids some of the awkwardness of the first by adopting and expanding that view of 'use'. It has it in common with both the second and the third that it allows for a straight confrontation between reason/principle and desire, with the caveat that reason's universal premiss always requires the agent to be considering what is good or appropriate for agents of a certain category to do. While this leaves us with something more traditional than would suit supporters of the third escape, it still allows for clear understanding of both the particular premiss and the conclusion. The thesis is that denying clear understanding was not part of Aristotle's way out. On the other hand the passage reads as though it is a treatment of a single sort of weak akratic. It certainly contains no explicit indication by Aristotle that he has two sorts in mind. That is a view to which one might be forced by fear of attributing blatant inconsistency, but which does not stand on the surface of the text.

This still leaves open the question of just how far Aristotle moved from Socrates, and how far he can allow for calm selection of the worse course of the sort assumed by many present-day philosophers to be a familiar fact of life. That deserves a separate chapter.

CHAPTER IV

How Socratic is Aristotle?

The account that I have given in the last chapter is in many ways traditional. It seems to me that Aristotle, in the passage we have been discussing, is clearly facing a difficulty proposed by Socrates. It is true, as Aristotle recognizes, that Socrates starts from a premiss that it is quite generally impossible to choose what one considers the worse course, but his aim is to re-analyse the phenomenon of *akrasia* and reject the common description of it as knowledge overcome by pleasure. It is this re-analysis that Aristotle thinks leads Socrates to fly in the face of the facts; and so, accepting the supposed facts, he aims to give an account of the way in which we know when we are overcome by pleasure. Throughout Book VII of the *Nicomachean Ethics* Aristotle constantly repeats that the akratic are in a passionate state which is responsible for the kind of off-beat knowledge they possess (see, e.g.; 1147all–18, 1147b8, 11, 16–17; 1150a25–30; 1150bl9–22; 1151al–3), which is in turn only ever likened to that possessed by people asleep, drunk or mad (see 1147all–18; 1147b6–9, 12; 1152al4–15). He is dealing with the *akrates* proper, who is someone overcome by the bodily pleasures with which temperance is concerned (1148a). This raised the question of how far Aristotle has moved from Socrates. Some, for instance Robinson (1977), think that he has pretty well capitulated; others, for instance Hardie (1968), seem to think that while something fairly traditional is true of the special case of being overcome by pleasure, in his discussion of *akrasia* by analogy Aristotle clearly allows for calm deliberate wrong-doing. In what follows I shall first of all discuss how close to Socrates Aristotle is in his account of being overcome by pleasure, and then consider whether his treatment of *akrasia* by analogy shows that he does not accept even outside those cases that there is any problem about deliberately choosing what is believed to be the worse course. If he does not, then he has rejected Socrates root and branch. In that case, there

would be no need to find allowance for a wider range of examples in 1147a18–24, since they get admitted later in the analogical cases.

Being Overcome by Pleasure

Even here it seems to me that Aristotle has moved significantly away from Socrates, while moving away from the violence of Plato's picture. First, he abandons Socrates' monolithic account of human motivation and also, incidentally, does not even have a position where reason has a single criterion for settling matters: rather, it employs reference to a set of goods without a clear criterion for settling priorities on occasions. Sensual desire is a response to objects viewed as pleasant to the agent. These pleasures are necessary, but desire for them does not take the form of considering them good for humans to pursue, and in the akratic case the pursuit of them is actually considered bad by the agent. Secondly, he considers that this alternative motivation is sufficient to block the other on occasion by securing that the agent does not use, i.e. put into effect, their knowledge/view of what is best. It is true that he has a sense of 'using knowledge' whereby the akratic fails at some point to use some knowledge, but this, as we have seen, is hardly a concession to Socrates, who would be appalled at the suggestion that desire could secure failure to use. In Socrates' case, realizing the truth of the particular premiss would be sufficient to produce use and therefore action, given recognition of the universal; for Aristotle it is not, and so he does not require failure to understand on the part of the akratic. The obvious concession to Socrates is that in *akrasia* the agent's knowledge of principle is not enslaved. Their knowledge of principle is used, and is not removed by *akrasia*—although it might, of course, become corrupted by constant weakness. A stronger concession, perhaps, is contained in the view that akratic action must be the exception. This suggests that reason, or deliberative desire, must normally win, on penalty, presumably, of the agent failing to have that potentiality and so not counting as a rational agent.

This account of desire overcoming reason, however, modifies the violence of opposition found in the Platonic picture. There physical desires are portrayed as blind opposing forces almost external to the agent. For Aristotle they are actualizations of the human agent which cannot readily cohabit with other actualizations; but they are actualizations of their conscious tendencies and so lead to choices (but not *prohairesis*) made by them.

In modern discussions the word '*akrasia*' has become the word for what Aristotle would call weak *akrasia*, the case where the agent chooses the worse course against a better judgment made at the time. Aristotle does, of course, spend much more time on this than on any other, presumably because that is the kind of case which most obviously bothered Socrates. But he is well aware that the Greek word equally well covers cases where my desires run away with me so that seeing something pleasant, which in general I know I should not do, I do it without thinking, just because it is pleasant. Impetuous *akrasia* is just as much *akrasia* as the weak case. I suggested earlier that it is unclear whether or not Socrates would be willing to allow emotion/appearances to lead to action before reason got going in a person who had knowledge, but that the tone of the *Protagoras* suggested that he would not. Aristotle's acknowledgment of impetuous *akrasia* would in that case be another respect in which he departed from Socrates.

The Analogous Cases

While this allows that Aristotle has parted company with Socrates in important respects, it only gives an Aristotelian account of being overcome by certain bodily pleasures. So far no allowance is made for being overcome by anger, even, let alone by ambition and generally desires of a less obviously physical sort. How far Aristotle is prepared to go depends on how we are to take his treatment of the cases of *akrasia* by analogy.

Aristotle claims that when we speak *of akrasia* without qualification we are speaking of weakness with regard to those bodily pleasures and pains with which temperance and intemperance are concerned—a very limited set (1147b20–30). We can, indeed, speak of lack of control as regards temper or honour or wealth, but we have to add the qualification, as they are only called *akrasia* because of a similarity. The question now is: what does this amount to? It could be that Aristotle is being careful about ordinary usage, but in effect thinks that it is merely a linguistic accident that we have to add a qualification in these other cases. In fact, it seems a little more than that. He actually specifies some differences, and the example he gives of a word used sometimes with a qualification, sometimes without ('good', 'good doctor'), suggests that the similarity need not be very close. It will be necessary, therefore, to look at the examples and see what the differences of treatment amount to.

The evidence here seems to me less than conclusive, but I think the balance is in favour of saying that Aristotle is indeed interestingly limited in his approach. Before going into the details, it is worth repeating that for Aristotle and his readers the word '*akrasia*' does not have the connotations of struggle with temptation, or acting contrary to a current 'better' judgment, which it has come to have in present day Anglo-Saxon philosophy. The impetuous cases are just as much ones where the agent is not in control, or is overcome.

The analogous case which at first sight seems closest to the one of being overcome by pleasure, is that of anger. At 1147a15 Aristotle mentions anger as something which, like sexual desire, produces physical changes which can result in the kind of off-beat knowledge characteristic of the akratic; so one might expect the situation to be fairly similar. It is interesting, therefore, to look at Aristotle's treatment of it at 1149a24–bl6.

We might be inclined to feel that there is no important difference between lack of control of temper and cases where bodily pleasure is involved. We might indeed have a woman who is quick tempered, who, hearing a remark suggesting the intellectual inferiority of women, flies off the handle—the kind of example Aristotle has in mind. This is the analogue of impetuousness in *akrasia* proper (see 1150b 19–28). It is just as easy, however, to think of a woman who thinks that a hostess should be polite, realizes that the best way to interest her neighbour is to talk about the decisions of a recent equal opportunities tribunal in which he has been involved, but is finally carried away into rudeness by indignation at the chauvinistic stupidity of those involved. This is the analogue of weakness in *akrasia* proper, and is a familiar kind of struggle, not essentially different from other emotional temptations. The interesting thing about Aristotle's text is, first, that he does not mention this possibility, and secondly, that if he had considered it, he would surely have felt unhappy with his remarks on the difference between this case and *akrasia* proper. At 1149a25–b3 Aristotle claims that temper mishears reason, not waiting for the outcome, and involves a sort of syllogism; using the premiss that a certain sort of behaviour (say, contempt) should be opposed, when reason or perception points to an act of contempt, it leaps to avenge it. So temper follows reason, while desire does not; this *akrates* succumbs in a way to reason, the *akrates* proper does not. If we look back to 1147a35–b3, even a person overcome by desire is said in a way to succumb to reason, though only incidentally, because it is really desire which opposes reason: the

premiss that everything sweet is pleasant is not opposed to the premiss that forbids one to taste. The examples are importantly different. In the earlier passage the *akrates* does not use a rival practical syllogism: they have no premiss to the effect that pleasure should be pursued. So practical reason/deliberative reason does not play a part in producing the result. In the case of temper, however, there is a sort of practical reasoning using a (probably correct) premiss that a certain sort of behaviour should be resisted by persons of the agent's sort. In the case of the *akrates* proper there is opposition between desire and the agent's reasoning about what is to be done; in the case of temper, the temper itself is not opposed to any reasoning, but itself constitutes a sort of reasoning as to what the agent should do.

In all this Aristotle seems to be thinking in a *post-Republic* context. There Plato (see 435–45) contrasts bodily desires, which are directed simply to various bodily satisfactions, with temper or spirit (*thymos:* the word I have been translating by 'temper' in Aristotle), which embodies views as to what it is good or noble to do. Both are contrasted with reason, which is concerned with calculating what is best to do. Temper, however, is portrayed as responsive to reason, and there is no picture of temper overcoming reason. Whatever the final complexities, this is not an implausible initial account. Anger or temper does seem to be a reaction to some behaviour taken as insulting or offensive, of a sort a person ought to retaliate against. It is an emotion which seems to embody some view as to what a person should do. Thirst, by contrast, seems to involve no such general views, but to be directed simply at physical satisfaction. Further, reflection on the wider context of the occasion, on other goals, principles and ideals, may reveal that there is nothing to get angry about; and it is common in such a case for one's anger to evaporate. One's thirst remains obstinately unslaked by such considerations. In this way, temper may be made out to be responsive to reason; but reason here is the reasoning about and answering of the question what it would be good or best to do. Aristotle, of course, distinguishes, in a way in which Plato did not, between reason used in a theoretical and reason used in a practical context. But when used in the latter it is, again, a matter of considering how best to achieve some human good. In the case of *akrasia* we are dealing with a conflict between reason and desire, where obviously the reason in question is a matter of deliberating on what it is best-for a human being to do. The case of *akrasia* proper is not an example of conflict within one's reasoning—any such conflict is only incidental—because bodily desire

does not involve the use of any principle as to what should be done. With temper the situation is different. If Aristotle were to consider an example like our second one above, he would have to envisage some sort of conflict within our reasoning. It would be a matter of the agent reasoning that this was something a person should not avenge but in anger deeming it something a person should avenge. In fact, he does not consider any such possibility, and the reason is probably that he is taking a Platonic view of the responsiveness of temper to considerations of reason, and not allowing for the possibility of an alternative and semi-independent route to judgment of the sort envisaged in Chapter II. This makes it easy to imagine its jumping the gun, but also easy not to imagine the possibility of conflict. Any conflict would take the form of a divided view on whether this behaviour deserved retaliation. This is a very different, and more optimistic, view of anger from the one later taken by the Stoics; but the result is that anger is significantly different from desire, and in a way which suggests that the analogue of weak *akrasia* does not occur.

If this is so, then the treatment of uncontrolled temper begins to look significant. For it looks as though examples of it are not seen by Aristotle as ones where the agent does something thinking it to be wrong, but rather ones where he or she overhastily thinks it to be right. This makes it unlike *akrasia* proper in the respect which makes that a problem, though like it in that the action taken is in fact wrong, is not taken in accordance with proper deliberative reasoning and is contrary to what that reasoning would have yielded if it had taken place. This suggests that Aristotle may not have thought that the restriction of '*akrasia*' in its strict sense to cases where bodily pleasure is involved is merely a linguistic accident. It is also of interest that Aristotle does not, as we might, look upon cowardice as a possible case of *akrasia*. For an interesting discussion of this see Charles (1984:177–9).

Akrasia *of Honour, Money, etc.*

Temper is the most obviously passionate of the extended cases of *akrasia*, and the one one might have expected Aristotle to have thought to be much the same as the standard case. The other examples look much calmer and more calculative, and so more appropriate for showing Aristotle aware of that untroubled wrong-doing which many philosophers think to be a familiar feature of life. The sorts of people Aristotle seems to have in

mind in 1147b20–48bl4 are those who have an excessive interest in money or honour or family. It is of some interest that at no point does he suggest that such people suffer conflict and act against their better judgment, certainly not contrary to *prohairesis*. All the language is consonant with their simply pursuing these objectives more than they should. He does indeed talk of them going to excess contrary to right reason (*para ton orthon logon*), but that suggests only more than right reason would dictate: it does not entail that it is contrary to their correct judgment. The difference between these extensions and *akrasia* proper seems to be that these are concerned with genuine goods, not just necessities; with things people should make it their business to pursue. In other words, they have a rightful place in our reasoning about the way to live, and are characteristically pursued as good things for human beings to achieve. Bodily desire, by contrast, is aimed at best at necessities. Bodily pleasures should not be mentioned in premisses of the form 'It is good for a person to…'; that they do is a sign of a deformed nature, of wickedness; for bodily desires do not have a part of the good for man as their objects. It is characteristic of them to move the subject towards something irrespective of the human good. In the extended cases the agents' objectives are part of the human good, and the agents will, as in the case of temper, exhibit some form of deliberate choice. Once again, there does not seem to be any room for a conflict between reason and something other than reason: someone excessively devoted to his or her family, like Niobe, is not in conflict, but gives a wrong weight to a genuine good. A similar point holds with Satyrus (1148a32–b2). The metaphor of being uncontrolled (*akratic*) is appropriate because these pursuits are not brought under proper control or order, not because they overpower a rival operative motive. For these choices are all actualizations of deliberative desire: they manifest interest in a human good and desire to work out the best way to it. The nearest approach to conflict here would be for the agent to be in a dilemma. In this case there is not even the analogy of passion, only of excess, and perhaps a tendency to pursue the goods in question without proper consideration.

It looks, then, as though Aristotle thinks that the extended cases differ from *akrasia* proper precisely in lacking the element of acting contrary to what at the time is thought best, and so his acknowledgment of such cases is not an admission of deliberate taking of what is at the time thought to be the worse course, without the influence of passion.

This would not be to deny all sense of struggle in such cases. Dilemmas, too, can be painful. But the struggle would not be against, but within, reason's criteria. If the agent has been torn, to adapt the

example of the woman above, between the demands of politeness and the demands of upholding women's dignity, and has decided for the latter, then that is what she has decided it is best for a woman to do. Any suggestion that she has acted against reason, or her better judgment, must be based on the observer's assessment of what would be reasonable, or what judgment would be better. She has not acted from a motive in conflict with her deliberative desire.

Akrasia *and Irrationality*

The obvious way to read Socrates is as saying that it is impossible to choose contrary to what one considers the more reasonable course. Underlying this is a picture of irrationality whereby a course is irrational if it fails to be justified by the reason taken. Socrates can therefore be taken as saying that irrational action is possible so long as it is interpreted as a mistake on the agent's part; but to take y for some reason while acknowledging that the weight of reason is in favour of x rather than y is so totally irrational as to be impossible. To oppose Socrates, then, one has to show this kind of irrationality to be possible.

That the akratic are irrational has become axiomatic for many philosophers (for a fuller discussion of the points raised here see Chapters VIII and IX), but it is fairly obvious that one method of showing them irrational will not work. If a person considers that for health reasons it would be better to refrain from sweet things, then clearly, from the standpoint of that reason, it would be irrational to take something sweet; but equally, from the standpoint of the akratic, who takes the sweet thing because it is pleasant, it would be irrational to refrain. We need some further material to weight rationality on the one side. Charles (1984:133–5) suggests that Aristotle accepts a weak commensurability thesis, to the effect that in each deliberation there is a system of value in terms of which the rivals are assessed. This is required both to make the correct choice rational and the akratic one irrational.

I am not clear about the weak commensurability thesis. At *de Anima* 434a7–10 Aristotle writes: 'for it is the function of calculation to decide whether to do one thing or another; and it has to measure with a single measure; for it pursues the greater' (my translation). This seems to say no more than that the del iterator has to settle on a standard. If I am after health, that will supply me with a criterion for deciding about sweet things: they are less healthy than non-sweet ones; and a thing's being

pleasant will be irrelevant. But then we are in the position of the last paragraph. If, on the other hand, it is a condition for the choice's being rational, and for the akratic's choice being irrational, that there be some further system of value in terms of which health and pleasure are measured, then I think we could with equanimity drop the idea of the akratic's choice being irrational in that sense, for the reasons developed in Chapters VIII and IX. Nor, I think, is Aristotle committed to it, though it would need argument on a large range of passages. That Aristotle says that sensual desire has pleasure as its object and so treats it as good is non-probative. Every desire/pursuit counts as taking something as good simply in virtue of having it as its object, and that does not amount to taking it as having a value in common with objects of other desires, nor does it involve considering it as a human good.

This is not to deny that Aristotle thinks that the akratic is irrational, only that the sort of sense hoped for in the above moves is not one which Aristotle wished to apply. Aristotle follows Plato in thinking that the application of reason to practical questions involves putting questions of a certain form—in Aristotle's case some such form as 'What is it good for humans to do?' Failing to pursue questions this way is failing to behave as a rational animal about them. Animals other than humans cannot go in for this form of pursuit, and pursue unreflectively things which appear to them as answering to current desires. In *akrasia* potentialities are actualized which stultify the operation of deliberative desire, so that the agent acts like a non-rational animal, and certainly not as a rational one. Each case of akratic action is in this sense a piece of irrational behaviour, and the tendency to akratic behaviour is a tendency to being not even akratic, but altogether irrational. Irrationality here, however, does not involve taking a course unsuitable for the reason taken (though see Chapter VIII), and some might prefer to dub it non-rational.

Conclusion

On the question of whether someone can knowingly take the worse course through being overcome by pleasure, Aristotle has departed significantly from Socrates. Since he allows that we may be motivated by desires which do not manifest a concern for working out what it is best for agents in our category to do, Aristotle allows for a clash between this sort of motivation and that of reason. The nearest he will get to Socrates' view that our reasoning must always be effective, is to allow that adult humans must, in

being adult, develop their potentiality for reasoning out what to do, and this creates a presumption in favour of reasoning being effective: this is a condition of their being rational animals. There is no doubt, however, that desire for sensual pleasure can frustrate the operation of deliberative desire, and nothing in Aristotle requires that this be effected through a blurring of the understanding or belief.

When we try moving beyond sensual desire the break with Socrates is less clear. Aristotle does not allow for a struggle against anger in parity with the struggle in weak *akrasia.* The only analogy here is with impetuous *akrasia.* The difference comes because of another difference: that anger is an emotion embodying views on how it is fitting for someone in such and such a category to react. If my deliberation led me to think one reaction appropriate, my anger another, I should not be suffering from conflicting desires, but from a conflict of view about how it is appropriate for someone like me to react. If my anger has to be my reaction to perceived insults, then it ought to alter if I work to a change in my view of the supposed insulting behaviour. Presumably the same type of account would hold with other emotions which embody views about what it is good or noble for people of such and such a sort to do.

The other analogous cases are still further from the original. Once again, no analogue with the struggle of weak *akrasia* seems to be envisaged, nor is there any passion running away with deliberation before it gets under way. These people are like the genuine akratic in mat (i) unlike the intemperate, what they value are things one ought to pursue, but (ii) they pursue them not as they should, and perhaps (iii) they pursue them without proper consideration. They are unlike, however, even here, in that the genuine akratic in fact choose pleasures which they rightly think they should not, ones which the intemperate would pursue, while valuing something else; the analogous cases pursue courses which they rightly value, and what they do on each occasion is a good sort of thing to do, but they pursue them to excess. The prospect for conflict here is presumably like that with anger: at best deliberation would produce conflicting views—but that means dilemma.

Dilemmas, of course, can be uncomfortable, so a form of agonizing would be allowable; but it does not seem that Aristotle envisages someone deliberately choosing what they consider to be the worse course, but not under the influence of sensual desire, or possibly fear (see Charles (1984) again). The charge that Aristotle confines deliberate wrong-doing to cases of being overcome by passion seems well-founded.

CHAPTER V

The Stoics

After Aristotle the two dominant schools for many years were the Epicureans and the Stoics. While it is arguable that the Epicureans ought to have had problems about deliberate wrong-doing, there is no sign that they considered it a problem. The Stoics, however, did. The school was founded by Zeno of Citium in the generation after Aristotle and was to be influential for some 500 years. While the early leaders of the school, Zeno, Cleanthes and Chrysippus wrote a great deal—especially Chrysippus—no complete works have survived. The quotations we have are brief and the descriptions of their views are often derived from hostile critics such as Plutarch and Galen, or ones not altogether sympathetic, as Cicero, for instance, in *de Finibus*. Those from whom we have more extensive writings, such as Seneca, Epictetus and Marcus Aurelius, are more interested in moral exhortation than the psychological theory that underpins it. Even when they are interested in the theory, it is always open to question whether they are reflecting the views of Chrysippus or Zeno.

The Stoics were mocked by their critics in antiquity for denying, as Socrates had done, that agents can be overcome by passion so as to act contrary to their judgment. What amazed their critics was that they made passion (*thymos*) and desire functions of the reasoning part of us, thus apparently disallowing all internal conflict. The criticism echoes that of Socrates by Aristotle, that their view is contrary to the plain facts. They are portrayed as denying that there is an irrational, passionate part of us, which can on occasion get the better of the rational part of us. While some, such as Posidonius, are reported as weakening on this point, they were seen as unorthodox. In what follows I shall assume that there was a mainstream tradition, following Chrysippus, opposed to allowing the possibility of conflict, and so of *akrasia*; and that remarks in one Stoic author can probably be used to illuminate and fill out what is meant by

another. The question I propose to address is how justified was the bad press the Stoics received, in order to build a plausible account of their position on *akrasia.* Since a great deal of our information comes through the hostile reports of Plutarch and Galen, I shall concentrate on these, and it seems only fair to review them critically. The main burden of their criticism is twofold: first, that Chrysippus, in holding that emotions and desires are judgments, and so denying the obvious contrast between reason, passion and desire, ignores the familiar facts of conflict in our psychological experience; secondly, that in his account of emotions Chrysippus contradicts himself. Since neither Plutarch nor Galen is sympathetic to Chrysippus, and since Chrysippus himself (see Galen, *de Placitis Hippocratis et Platonis* (*PHP*) III.4, p. 192=*Stoicorum Veterum Fragmenta* (*SVF*) 2.902) was not found a model of lucidity, and so might easily have been misunderstood, I shall exercise principles of charity in his favour.

What riled his critics on the psychological front was Chrysippus' insistence that anger, fear, desire and so on were functions of the reasoning faculty, thus denying the traditional Platonic threefold division of the soul. The exercise of charity here will involve seeing what might lie behind this, and in what sense an irrational pan of the soul was denied. The contradictions with which I shall be concerned are also to do with passions, in that Chrysippus seems both to say that passions are judgments and operations of reason, and that they are judgmentless movements in disobedience to reason (compare *PHP* V.1, pp. 291, 294=*SVF* 3.461, and *PHP* IV.2, p. 240=*SVF* 3.462).

In what follows I shall first outline what I take to be the Stoic theory of action, since their position on *akrasia* arises out of that I shall then consider how the Stoics might cater for familiar examples of apparent internal conflict. At this point the influence of principles of charity will be obvious. The conclusion will be that an interpretation which fits well with their account of action and emotion fits well also with the view that various criticisms conceal interesting aspects of the Stoic position, a view which would make it far subtler than their opponents would suggest.

The Stoic position on *akrasia* cannot, of course, be completely disentangled from what they have to say on determinism and freedom. I shall not, however, be concerned with that larger problem. It cannot be disentangled because the account of action, which gives rise to the position on *akrasia,* seems to have arisen from concern to assert the responsibility of agents for their actions, and there is a problem as to

how this assertion is consistent with the apparent determinism. I shall confine myself to trying to see what the account of action is and how it gives rise to problems about *akrasia*. For this the best place to start is with the contrast between humans and animals, since that contrast explains why humans are, but animals are not, responsible.

Animal Behaviour and Human Action

Animals differ from plants in that not only do they grow, but they have the power of perception and of pursuit of objectives (*phantasia* and *horme*: see Origen, *de Principiis* 3.1.2–3=*SVF* 2.988). According to Diogenes, Chrysippus (Diogenes Laertius (D.L.), *Vita Philosophorum* VII.85=*SVF* 3.178) held that the primary objective of any animal is its own constitution and the consciousness of it An animal's power of perception is not one of neutrally recording the environment, but one of enabling the animal to learn and pick out what is helpful or harmful to the preservation of its constitution. Among perceptions some will leave an animal indifferent, some will attract it, some repel it The latter will lead to pursuit or avoidance. It is impossible for us to describe this process, barring cumbersome circumlocution, without using judgmental language. Thus observing a stalking cat we would say that it had *seen that* the fledglings were now hatched and was *hoping* to creep up on the nest without the parent bird noticing. Applied to humans this sort of description would also be supposing possession by the agent of concepts such as those of fledgling and hatching, and the judgment that they applied here; but the Stoics would not allow us to transfer the conceptual apparatus of the description to the epistemological equipment of the cat (see Seneca, *de Ira* 1.3.4–8), for that would make the cat a rational animal. The presentation (*phantasia*) or perception, however, is a presentation, in this case, of something as attractive which rouses the cat to pursuit. This combination of perception and power of pursuit is the basic attribute of souls/animate beings (see Arius Didymus, *Preparatio Evangelica* XV 20.1–7=*Doxographi Graeci* (*Dox.* Gr.), p. 471=*SVF* 2.821). But reason, and with it assent properly speaking, are not found in animals (or young children) (see Galen in *PHP* V.7.19, p. 340, Origen, *de Principiis* 3.1.2–3=*SVF* 2.988).

The possession of reason consists in the possession of a wide range of concepts and the ability to reason with them (see SVF 2.841). Clearly the development of reason will be a gradual process which the Stoics

thought would not be complete until about the age of fourteen (see D.L. VII.55). I shall not discuss whether we have to suppose that they thought there was some final point of achievement of maturity. What is clear is that they thought that a fair degree of mastery of language, and with it power of reflection, was needed for someone to be able to play a full part in adult life and so also be held responsible. In this account, learning a language is not a matter of learning a lot of labels, but of learning to think about things, and, in the case of action, acquiring the ability at least to work out what to do.

It might be tempting to think that humans have the same powers of perception as animals and the same power and tendency to pursue what is suitable to their natures, and all that the addition of reason does is give them the ability to name and relate their perceptions and work out which to take seriously, which not, in their pursuits. This might be suggested by D.L. VII.49: '*Phantasia* comes first, then the mind/thought, which can give expression to things, spells out in words what it is experiencing from *phantasia*.' But Diogenes goes on in 51 to distinguish between rational and non-rational *phantasiai*. The first are those of rational animals and are called thoughts; those of non-rational animals are nameless, but they are not thoughts (see also *SVF* 2.83). This is the point made two paragraphs ago, that with humans *phantasiai* involve the judgment that what is perceived falls under some concept, an element which has to be expunged from our description of animal *phantasia*. To have a rational *phantasia* is for things to look to be thus and so, and this look is what the mind articulates. Human *phantasia*, then, is not animal *phantasia* plus a power to describe: rather, animal *phantasia* is only analogous to human *phantasia*: it is as if animals were taking things to satisfy certain descriptions—but only as if.

It would be possible, however, for all this account tells us, to think of humans as differing from animals only in that they are able to classify what they perceive, and their perceptions take judgmental form; but for the rest, like animals, they will pursue what they perceive/judge as *oikeion* (fitting their natures). This would, however, leave a slight puzzle as to how reason was ever developed. Once I have noticed that the food on the plate is meat, my carnivorous *horme* would drive me to it, while when I was faced with nothing of interest, like a sated dog I would snooze. Further, this view makes no room for the play of deliberation. But just as the Stoics contrast rational with non-rational *phantasia*, so they contrast rational and non-rational *horme*, and for a rational being to live according to nature is to live according to reason, which becomes

the artificer of *horme* (see D.L. VII.86). Now what reason directs human pursuit to is both life in accord with human nature and life in accordance with nature at large. The natural thing for humans to aim at therefore is the exercise and following of their reason with regard to their life as part of nature. This means that they will have a tendency not only towards those things which preserve their constitution, considered in abstraction from their ability to reason, but also to think how best to achieve that in conformity with nature as a whole, or the will of Zeus. This attributes to humans both a capacity and inclination not, without more ado, to take what looks at first sight suitable to pursue as in fact suitable. Taking something as suitable requires assent to some proposition/presentation construing the act as suitable; and rational pursuit is assented pursuit, that is, pursuit of something understood and assented to as suitable for the agent to pursue. The tendency towards working things out puts a break on immediate pursuit of what is presented as appropriate. This dual aspect of the goal of human beings—that they pursue not just what is fitting to their constitution and that of nature as a whole, but also pursue the rational pursuing of it—rouses Plutarch to mockery (see *De Communibus Notitiis* 1072.E–F). It is, however, an essential part of the difference between humans and animals, human *horme* and animal *horme*, and introduces the need for an interpolated moment between presentation and *horme* in the human case (see Plutarch, *de Stoicorum Repugnantiis* 1057.A=*SVF* 3.177).

The Stoic treatment of assent and *horme* sometimes produces some strange sayings. Thus the agent is said to assent to a proposition and their *horme* to be directed at the predicate (see Stobaeus, *Eclogae Physicae et Ethicae* II.88.1=*SVF* 3.171). If John applies for a post as clerk, he must assent to some such proposition as 'It is suitable for me with my qualifications to apply for a clerkship.' The expression 'to apply for a clerkship' is the predicate being applied to John by 'it is suitable for…'. Inwood (1985: p. 64) suggests that it is natural, since what is pursued is what is designated by the predicate, to declare the pursuit to be directed at the predicate. Put like that, the claim sounds distinctly odd. Yet there is evidence in a different context that the Stoics were willing to move between what is designated and the designator. Thus we find Arius Didymus reporting Slobaeus' report of Zeno on cause as follows (*Dox. Gr.* p. 457.18=*SVF* 1.89) 'and the cause is a body, but that of which it is a cause is a predicate'. The word for 'predicate' is also the word for a charge in the legal sense, and the word for 'cause' also is taken from a legal context and can mean 'guilty'. 'What is guilty

is a body, and what it is guilty of is a charge' sounds not too odd, even in English, where 'charge' stands in for 'what it is charged with'. Yet if this is the explanation of Stoic talk of *horme* being directed at the predicate, it will follow that animal *horme* is directed at the predicate, too. After all, a cat is aiming to get the fledglings, so it should be natural to substitute the designator and say that it is aiming at the predicate. But the passages in question seem intent to say something special about rational *horme* as distinct from the non-rational, animal variety. It seems likely, then, that the explanation is a little more complex. Now in the rational case the pursuit is not just of something which is correctly designated thus and so: the agent conceives of it as so designated and pursues in accordance with his or her conception. Thus John's pursuit is given its form by his understanding of what is involved in applying for a clerkship, and that predicate as understood by him has a directing role in the pursuit. So it may be that the Stoic talk of the (rational) *horme* being directed at the predicate is an attempt to make this point that the *horme* is directed at an objective in accordance with a conception of it, thus bringing out something of the difference rationality makes to *horme*.

Some Consequences: Propatheiai *and Inordinate* Horme

On this account, reason is something that a child acquires over a number of years; so it starts with non-rational presentations, which with the acquisition of language become things looking dangerous, attractive and so on. One might expect the Stoics to think of presentations as attractive or repulsive, and the growth of reason to be a matter of introducing a check to the immediate activation of *horme.* For a long time, however, there would be as it were an initial movement towards or away from what is so presented. Reference to such a movement is just what we find in the doctrine of *propatheiai*. The most detailed evidence comes in an excerpt from Epictetus, which Aulus Gellius (*Nodes Atticae* (*NA*) XIX.i. 14–21) affirms to be consistent with the views of Chrysippus and Zeno. According to Epictetus, even the sage, faced with the presentation of a falling building, will be moved for a short while before his reason dispels the appearance of danger. In the sage's case this movement amounts only to a temporary flutter, since the sage, if such exists, has reached that stage of understanding where false presentations of danger cannot deceive. According to Seneca (*de Ira*

1.16.7.=*SVF* 1.215), Zeno made a similar concession, saying that the sage's soul retains a scar even after the healing of the wound, so that he will experience hints and shadows of passions. What lies behind the talk of a wound and healing is probably something like this: when we are children we are like animals, and frightening presentations are reacted to fearfully—i.e. we recoil. As we learn to speak, these presentations are over the years moving from being non-judgmental to being judgmental. Seen from the standpoint of a fully developed rational nature, the early stages, when quasi-judgmental presentations are followed by immediate recoil, look like the operations of an organism so wounded as to be unable to function properly. As the child grows up it will falsely judge what is presented as evil to be evil, and so the appearance of a wounded reason persists. The philosophic development of reason brings about a healing of the 'wound', but the old automatic shrinking is never wholly cured.

These remarks of Epictetus and Seneca apply to the sage, the one who matches the Stoic description of the fullest development of reason. This development includes a properly absorbed understanding that there is nothing to fear, since all that happens externally is in accordance with the will of Zeus; and that our own *horme*, which is all that can really be bad so far as we are concerned, is under our own control. The implication is that presentations of the form 'It is fitting to…' exercise attraction or repulsion, and continue to do so on surprise impact even with the full development of virtue. It is this attraction or repulsion which may win too ready assent from the adult who, while having attained the age of reason, has not yet attained wisdom. For a presentation always presents something as an apparent truth, appropriate for assent—in this case an assent to a *hormetic* proposition.

It is perhaps worth dwelling on this last point, and the treatment of *phantasiai* (presentations) as judgmental. The point of using the expression 'judgmental' is not to suggest that the person having a presentation forms a judgment that such and such, in the sense that they conclude that such and such. That would be assent. But presentations come up for assessment by reason. As Origen puts it (*de Principiis* 3.1.2–3=*SVF* 2.988): 'The rational animal has reason too in addition to the power of presentation. Reason judges the presentations and rejects some and admits others so that it might lead the animal according to the presentations.' If the presentations are not at least 'lookings to be thus and so', there is nothing for reason to work on for rejection or admission. To a rational subject faced with a frightening presentation, it

looks as though there is a danger. Thus Epictetus (in Aulus Gellius, *NA* XlX.i. 15–20), first speaks of 'a sudden *messenger* of some danger'; and then contrasts the fool with the sage because the former 'takes things which *at first impression look* violent and harsh really to be so'. When I call presentations judgmental in form I am simply making the point that when anyone has a presentation this is a matter of something looking as though it is the case. In introducing the extra moment of assent the Stoics are making a distinction which Socrates in the *Protagoras* did not make. Appearances there have to be construed as what the Stoics would call assents. The Stoics are in effect distinguishing between 'It seems to me that…', which usually expresses one's judgment, and 'X looks…', which commonly gives how things appear without committal to how they are. If a tree looks larger than a mountain, then appearance as it were invites the judgment that it is, but a wise person will think before giving the judgment, i.e. the tree may look larger, but it does not seem to the sage to be larger.

The *propatheiai*, then, are traces of early pre-rational reaction to presentations, and these last are judgmental in form. The traces are at least in part the physical effects normally associated with the judgment, the shrinking or quivering or blenching. These effects will normally be associated because any judgment is a given state of the soul, which, being physical, has physical effects. In the case of any passion this judgment is false. Thus fear involves taking physical harm, say, as evil, whereas in fact the only evil is a soul out of tune with Zeus. Being in or out of tune with Zeus is always within our control. The sage does not pursue the standard goods as though they were goods, but pursues them conditionally on their being in conformity with the will of Zeus. The sage's soul is therefore always in a fit state. With the rest of us the false judgment that constitutes passion is a result of our reason being in an unfit (e.g. a flabby or overtaut) state. This state has physical effects, and these are not necessarily under our control. One might, therefore, further expect the Stoics to allow of behaviour resulting from anger or fear being beyond our control or that of subsequent *horme*. Once again, this is what we find. Galen (*PHP* IV.2, p. 240=*SVF* 3.462) has a long quotation from Chrysippus, to which I shall return, in which he compares someone under the influence of passion to a runner who starts to run intentionally but cannot just stop at will. So I may see a bull approaching my child at speed and, judging it to be dangerous, be afraid, and become riveted to the spot. I then judge that I should go to the child's help, but cannot move.

Why no akrasia?

It might seem that all this still leaves open the possibility of non-rational motivation which could in theory overpower rational *horme*, so leading to *akratic* action where passion overcomes the agent's rational decision. Some later Stoics, in particular Posidonius, do seem to have weakened on this point; but the main Stoic tradition rejects the possibility. It is not hard to see why they might. It is perfectly possible for a Stoic to allow agents to be responsible for actions to which they did not assent (e.g. ones done from forgetfulness or without realizing), so long as the agent is not disabled from assent. What becomes problematic are proposed cases where agents act intentionally under the influence of irrational motives and contrary to their rational preference. Stoic critics, such as Plutarch and Galen, were forever insisting that Stoics overlooked the most obvious facts of our psychological life (see Plutarch, *de Virtute Morali*, 446–7); but it is, in fact, arguable that it is not all that obvious just what the facts are and just how one should describe the proposed cases. Suppose we take the example, which was popular in the ancient disputes, of Medea torn between her desire for revenge against Jason, urging her to kill their children, and her judgment (of reason?) that that would be a wicked act. (For an excellent discussion see Gill 1983.) Critics of Stoicism tended to interpret this example in Platonic terms of the (real) agent, identified with her reason, being overwhelmed by passion. The oddity of this way of talking is that it risks undermining the supposition that Medea acts intentionally for revenge. Compare St Paul, Romans 7.14–20: 'I do not do what I would like to do, but instead I do what I hate…so I am not really the one who does this thing; rather it is the sin that lives in me.' If the murder of her children is, in fact, Medea's action, intentionally undertaken, for which she is responsible, then we have to admit that she opted for revenge, and it was reasons of revenge that determined her deliberations; she clearly did not assent to the findings of reason. This Gill persuasively argues to be Chrysippus' interpretation of the famous passage in Euripides' *Medea* (1078–9): 'I know what evil I propose to do, but anger rules my deliberations' (rather than 'anger overcomes the deliberations of my reason'). In practice, Medea could indifferently be described as acting out of anger, as giving in to anger or as being overcome by anger. These different expressions do not require different situations, but it taken seriously they suggest different analyses of what is going on. There can be dispute, in fact, as

to just how we should describe such cases, or, in other words, as to just what the facts are (see Chapter II, pp. 18–19).

In all this the Stoics are, I think, misrepresented as insisting on the unity of the mind; it is fairer to describe them as insisting on the unity of the *hegemonikon*, the principle of activity. This very word indicates a ruling principle. In souls generally it is 'life, perception and *horme*' (see Arius Didymus, *Epitomes physicae fragmenta*, 39 in *Dox. Gr.*, p. 471= *SVF* 2.821), and in humans includes reason and assent (see *SVF* 2.826, 831). What would it be to allow of a split *hegemonikon*? It could only amount to either no *hegemonikon* or two; for it is an attempt either, say, to have *horme* without assent, or to have two rival assented *hormai*, assenting to judgments of rival presentations. But in the latter case we have two rational subjects—a split Medea, one responsible for evil, one not In fact, the resistance to splitting her *hegemonikon* is a corollary of accepting that Medea intentionally opted for revenge.

It might seem that the view now just becomes dully truistic, but I think the Stoics might hold that their terminology saves one from a misleading way of describing matters, and it might be helpful to see to what extent and how they can cope with what seem to be familiar experiences of conflict. As critics brought out, this means dealing with cases both of uncontrolled and self-controlled action. As we shall see, the Stoics' treatment involves a controversial account of emotions and passions, but to start with I shall consider some cases of conflict without trespassing too far into that area. It is important to note that I shall be considering how the Stoics *can* cope, given their general account of action, not how they do. It seems to me that our evidence does not always allow us to be sure of the last. On the other hand some of the charges against them seem to be based on what some Stoics said, and it is then worth asking whether, if that is what they said, what is attributed to them might be consistently held by them.

Coping with internal conflict: akrasia *and self-control*

How, then, might the Stoics deal with cases of struggle and *akrasia*? If we return to Medea, we might ask what a Stoic could consistently allow. Plutarch (*Vin. Mor.* 450c–d=*SVF* 3.390) reports the concession that passion can drive out reasoning, and pounces on it as an inconsistency. Yet in one way the Stoics could obviously allow for Medea in her anger to reject the findings of reason. Using her reason she can spell out and

recognize that what she proposes is bad (i.e. that it contravenes principles of virtue); but the Stoics do not accept the apparent Socratic position that this recognition will result in rejection. Assent is required for that; and reasoning, at least about practical matters, does not compel assent. So reason may produce the presentation that the children should not be killed, while another presentation embodies the quasi-judgment that it is appropriate to kill them; but the former is rejected, the latter assented to. Socrates portrays us as all having one goal, and the operation of reason is always and only, in practical contexts, to work out the means to that goal. The Stoics portray us as faced with possibly conflicting presentations that it is appropriate to do A, B, C. We may now reason, i.e. ask ourselves questions about what it is best to do; and these reflections may result either in a further presentation, or in a modification of one of the original ones; but they will not automatically result in just one presentation, except in the case of the sage. There remains, therefore, the possibility of assent to a presentation other than that resulting from ratiocination; this assent to a presentation would be a passion, and would sweep aside the findings of reason. In this way, reason can be overcome by passion, that is to say, the agent may ignore the conclusions of reasoning and assent to an inordinate presentation. That is not to choose contrary to a rational decision.

So far we have minimum conflict. Reason is reporting results, but we get no indication of Medea being pulled two ways, except in so far as practical presentations always exert a pull (see above on *propatheiai*). Yet developing this last point can enable one to introduce a greater element of conflict by portraying Medea's struggle as a familiar case of the deliberative faculty being distracted towards contrary opinions. One might reject this as a portrayal of struggle, if one assumed with Aristotle (*de Anima* 434a7–10) that in deliberation we are always using some yardstick, so that vacillation in deliberation is always a matter of wondering what the yardstick brings out as right, never a matter of changing yardsticks. The Stoics, however, did not put such emphasis on deliberation, and the role of reason is to a large extent that of enlarging horizons. Using her reason Medea will recognize the appropriateness of revenge and how apt a form of it killing her children would be, but also how contrary to her natural parental role it is and so on. She will therefore become aware of conflicting yardsticks. Since the various presentations all take the general judgmental form, 'It is fitting to…,' they all call for, though they do not all receive, assent. As she moves towards her final decision

she attends first to one, then the other, seeming to move towards assent; but this is not yet assent; and some of the presentations will be ones which present what does, and what she recognizes to, conform to the will of Zeus, and so to be dictates of Reason. All this will give an air of struggle (and indeed she is struggling: to make up her mind), and one set of considerations can sensibly be called those of reason. But it is Medea who is struggling, not her reason/real self against irrational passion. None of this requires a split in her *hegemonikon*, and nothing in the Stoic position precludes this picture.

Medea's struggle could, of course, be more dramatic. It could be that she has no sooner decided on revenge than her heart misgives her and she resolves to resist, but has hardly done so before her mind turns to revenge again, as she gives her assent first to one then to the other cause. Here the appearance of struggle would be more marked as she first makes a move to prepare for the murder, then stops, and so on. This is the picture suggested by Plutarch's report (*Virt. Mor.* 447=*SVF* 3.459) that the Stoics portrayed *akrasia* as a succession of different choices. Obviously there will be a continuum of cases between those which clearly match this latter description and those which clearly match the previous ones.

The Stoics can also, however, cater for a stronger example, where the agent acts contrary to a current rational *horme*, and Medea might be like that too. For the Stoics give no guarantee that we shall be able to carry out our determinations. Although I have been using the word 'pursuit' to render '*horme*', 'pursuit' suggests the successful putting into effect of what we set ourselves to—and it is in pursuits that *horme is* typically manifested. A *horme*, however, may never issue in a successful outcome, since not even our bodies (see Epictetus, *Dissertationes*, I.i) are under our control. It is possible, therefore, for our rational *horme* to be directed ineffectively at an action. This allows for the possibility that as a result of assenting to a presentation of the appropriateness of revenge, Medea is subject to the physical effects which constitute the development of anger. Once these are set in train, all that her assent to the judgment that this murder shall be avoided can achieve is an ineffective internal effort to stop. Thus Chrysippus likens people overcome by passion to runners who cannot readily pull up. Although they start the movement, their bodies acquire a momentum of their own which is beyond their control (see Galen, *PHP* IV.6.35 (412 K)=*SVF* 3.462). So passion can even be said to carry Medea contrary to her rational *horme*, though this would not, of course, be a clash of

contemporaneous *hormai*; rather, what one *horme* started, a later *horme* would be insufficient to stop.

These are all of them different examples, all of which we might, in our careless everyday way, describe, given that Medea acts from anger, as anger getting the better of reason. All of them can be accommodated on the Stoic view, while preserving a clear account of the agent's responsibility. So what is so unacceptable?

It might be conceded that we do indeed need to look at cases more carefully, and that many can be adequately described from the Stoic viewpoint, but still be claimed that there are crucial examples where the Stoic position is absurdly in conflict with experience. The clearest examples come with self-control. As Plutarch says (*Virt. Mor*. 447B), a lover continues to love, even when he has reasoned that his love must be restrained, and still loves while he opposes his passion by reason. Life would be too easy if passion evaporated as soon as reason showed it to be undesirable. Yet the Stoics cannot allow for this. For the passions are not only themselves so rational that animals and young children are incapable of them, they are also *hormai* (D.L. VII. 110.18–20) and so, by the foregoing, rational *hormai*. Now a lover who controls his love by reason must be acting from a rational *horme*; so if the love he is controlling persists and is itself a rational *horme* we should end up with a divided *hegemonikon*. Here the picture of rapid vacillation is highly implausible. Since the struggle started, presumably, with the passion, which then had to be controlled, it is not clear how the runner analogy can be adapted for this case either. Once we have seen this point clearly where self-control is involved, it is not difficult to devise examples of lack of control where the same holds.

Solution to Self-control

There is no doubt that the Stoics have very fierce views on the passions—to which I shall return shortly—and in accordance with these views none but adult humans have passions. Further, what we might be tempted to call suppressed or controlled passions, feelings to which we do not give way, are denied the appropriate description. So clearly, the Stoics are using the expressions for passions in a far more restricted way than is usual—and this is not just a point about English, but holds for the Greek too. Consequently they will not be able to describe Plutarch's lover as controlling his passion. I shall discuss later how significant this

restriction might be, but it does not necessarily result in their being unable to describe the self-controlled lover, only in their being unable to ascribe the passion of love to him. The fact that they acknowledge *propatheiai* allows the Stoics to describe what is going on in this case. For while the sage will only experience minor perturbations, that is a sign of his sagacity. One less wise will not be in a position to quash so effectively the false presentation 'It is fitting to pursue her', not being able thoroughly to see through its falsity, and so nullify its attraction. Consequently the attraction will remain. But all this is something which appeals to the subject's judgment' it is a battle for assent; it is not some force set to overpower the judging subject.

In addition, the self-controlled lover might be suffering from the physical effects of earlier assent. At first he assented to the presentation of loveableness. Then he discovered that she was married. But by then the damage has been done, and he finds himself struggling with effects which are not entirely within his power. There is no need to make a choice here. Either or both accounts might apply, as the phenomena might be more complex than the single description suggests.

The Passions: Judgments or Irrational Motions?

The trouble with this interpretation is that it can make it seem increasingly obscure what the fuss is about. It is as though the Stoics wanted to sound interesting but succeeded only in rather tiresomely refusing to use the terms 'anger', 'love', 'fear' and so on of cases where the rest of us would use them, and confining them to occasions where these passions were acted on. The Stoics then seem excitingly to deny all sorts of obvious phenomena, not just *akrasia* and struggle, but also the fears of animals and children. When it emerges that they admit them, but by a different name, the whole episode looks like a piece of irritating histrionics, an attempt to get undeserved attention. I shall first look a little more closely at what the Stoics say about the passions proper, and then turn to the question of whether any important point is being made.

A report of the early Stoics (D.L. VII. 110–11=SVF 3.412) tells us that Zeno said passion was the irrational movement of the soul contrary to nature or excessive/inordinate *horme*; while Chrysippus said that the passions were judgments—love of money is taking money to be a good thing. Indeed, Galen claims that Chrysippus himself held both these positions, and thus contradicts himself. He quotes a passage (*PHP* IV.2.

=*SVP* 3.462) where, he claims, Chrysippus accepts a definition of *pathos* as an irrational motion, and stresses that by 'irrational' he means 'disobedient to reason' and 'putting aside reason' and 'in respect of this motion it is customary for us to say of someone that they are forced and borne along irrationally without the judgment of reason'. When one looks more closely at the passage quoted, however, there are points worth noting. First, it is clear that Chrysippus is discussing two definitions of '*pathos*' and claiming that they apply to certain phenomena, without, within the passage, endorsing them as definitions. Secondly, while allowing that in respect of such motions we speak of people being forced, he says 'it is not if they are borne along in error and mistaking something in reason that we use this description, but precisely in reference to the motion he sketches'. Who is 'he'? Since the definitions in question are those attributed elsewhere to Zeno, it would seem that in this passage Chrysippus is discussing Zeno's definitions, which differ from his own, and is explaining to what phenomena they apply. One might well expect him to explain the relation between his position and Zeno's. Zeno confines his account of passion to those movements which on Chrysippus' account (see Plutarch, *Virt. Mor.* 449c=*SVF* 3.384) are the result of the false judgment. Zeno could, therefore, talk of our being carried away by passion, but Chrysippus is at pains to point out that such talk can only refer to the irrational movements and not to mistaken choice. It looks as though Chrysippus' concern is to point out to his Platonist opponents that they have a choice: either they can be serious about their talk of being overcome by fear, anger and the rest, in which case they must have in mind the irrational, judgmentless movements covered by Zeno's definition; or they want to say that the agent chooses a course of action through fear or anger, in which case they cannot accept the Zenonian definition, but must be thinking of an erroneous judgment as to what should be done. That means accepting a Chrysippean definition. It looks, then, as though Chrysippus is not contradicting himself, but, rather, explaining the relation between his own view and that of Zeno. In the course of this explanation he allows an application for Zeno's definitions, but in a way which will give no joy to his opponents. In this and the following quotation in Galen we find Chrysippus using the disorderly motions allowed for on his own account as the result of erroneous judgment, to explain the phenomena commonly described as being overcome by passion. The result is to present his opponents with what we might call Chrysippus' fork: in talking of being overcome by passion, either you take the language of overcoming seriously, in which case you are not explaining an action or choice; or

else you are explaining an action, in which case that language cannot be taken seriously, and what is done must be the result of the agent's decision, not of the operation of some other 'part' of him or her. Our ordinary way of speaking, canonized in Platonic theory, simply introduces confusion in our understanding of ourselves and our responsibility. It might matter little whether we decide to define passion with Zeno or with Chrysippus; what we must not do is try to run both at once. In particular, if we wish to explain our actions as done under the influence of passion, we must accept a Chrysippean account. Similar points hold for the interpretation of such passages as *PHP* IV.4 (254).

Chrysippus' view of the passions does, of course, incorporate the irrational movements of Zeno's definition. Fear is a judgment that some unpreventable evil is coming. In fact, however, the only real evil for an agent is a *horme* contrary to the will of Zeus, and this is always within the agent's power to avoid. So there is never anything to fear. This passion, in fact, is always a false judgment that something such as pain, illness, loss of money, is an evil. It is therefore always a sign of a soul/reason that is not in good condition. Since the soul is itself physical its unfit condition/operation typically has physical effects, and we cannot ensure that later changes of mind will secure changes in these effects. But while reference to inordinate motions serves to mark out the judgments in question, it is the false judgments which for Chrysippus constitute the passions: the inordinate motions are their characteristic outcome. This view involves drawing a sharp line between what we might think of as judging that one should take something as dangerous and feeling afraid, judging that one should take something as an injury to be avenged and feeling angry. Only the first are called the passions of fear and anger. Why?

The argument is clearly not about ordinary usage. The Stoics' introduction of new terms reported by Plutarch (*Virt. Mor.* 449 A–B) indicates a consciousness of correcting that. It looks as though they thought that ordinary language showed a tendency to treat anger, where this is the angry stance one takes, and anger as the appeal of revenge-indicating presentations, as all part of the passion; and this is seriously misleading. If what I have said above is on the right lines, then there would seem to be three connected reasons for their view. First of all, the Stoics took a moral view that moral assessment should be attached to what an agent sets him—or herself to. That is what is under his or her control. The ordinary way of speaking of the passions confuses us into thinking that reactions not under our control somehow affect our virtue.

We cannot help a frightening presentation, nor a resultant trembling; but we can help agreeing that what is threatened is really evil and should be avoided. That is what shows us to be cowards or not Secondly, the ordinary view tends to treat passions as springs of action while treating both items within and items without our control as coming under passion. This leads to confusing talk of our being overcome, which makes it hard to make sense of talking of *an* agent who is responsible. In fact, in so far as what are commonly called passions explain our choices, it is assent to passion presentations that is in question; in so far as what is commonly called passion is out of our control, it does not explain any choices. Thirdly, treating the passions as springs of action and including under the heading of the passion reactive elements beyond our control, immediately invites a bewildering two-agent picture of the person. A Stoic might well have felt that Plutarch and similar critics had not thought through their talk of parts of the soul. Plato at least had the grace to be disturbed by it (see *Republic*, Book X, 611).

If this interpretation is right, then it is a mistake to ask of feelings of fear whether they count as fear or not. The answer is: 'in ordinary language, yes; in Stoic technical terminology, no'. The Stoics neither deny the occurrence of the feelings, nor their connection with what in Stoic terminology is the passion. The Stoics' critics should give their attention to the view of moral assessment, the claim about responsibility, and the question of how we are to think of agents. The Stoics could, of course, have signalled their points without such dramatic revisions in ordinary terminology; but it is those points which are important, not the terminological one.

It might be thought that this interpretation in effect accuses the Stoics of sleight of hand. It seems clear from Plutarch's criticisms in *Virt. Mor.* that the Stoics were against allowing an irrational part of the soul opposed to reason, and their stand on the passions was precisely to reject the irrational part. But the move seems empty, because it now emerges that the *propatheiai* are movements of an irrational part of the soul; and the movements resultant on false judgment are movements of an irrational part of *us*; it just happens that the Stoics refuse to call them passions. The victory is a hollow one won by verbal trickery.

It should be clear from the above that such a response would be less than fair. According to the Stoics the *hegemonikon* consists of presentation and *horme* and, in the case of rational beings, reason and assent as well (see Arius Didymus, *Prep. Ev.* XV.5=*Dox. Gr.*, p. 471). Now presentations are not under our control (see Origen, *de Principiis*

3.1.2ff.=*SVF* 3.178), and in *that sense* are non-rational. As we have seen, *hermetic* presentations present a course of action appealingly or unappealingly to the mind. If this amounts to an irrational part of the soul, then the Stoics have no objection to it. But since it is a feature of every *hegemonikon*, it cannot speak a divided *hegemonikon*. The important point on which the Stoics wish to insist is that these allurements are allurements to agents, trying to win their assent; they are not rebellious parts over-riding assent while at the same time being the agent's responsibility.

Summary

If the above is correct, then the Stoic position is not as crude as their opponents tried to suggest. So far as *akrasia* is concerned, the Stoics are determined to insist, in effect, that if the akratic's action is intentional, then they must know what they are doing, and do it for a reason or, technically, their doing it must be done from assent to some presentation expressed by 'It is fitting for me to…' But in that case the language of being overwhelmed or carried away is not to be taken seriously. The cost of taking it seriously is to make nonsense of portraying the behaviour as intentional and the agent responsible. On the other hand the Stoics can allow for familiar phenomena. They can allow, as with the comparison with a runner, that sometimes the result of an assent might be to put the agent beyond control by later assents; and they can give an account of familiar struggles both successful and unsuccessful. In this they are more nuanced than the Socrates of the *Protagoras*.

Some Criticisms

This is not to say that the Stoic position is beyond criticism. It is fair enough to complain that authors such as the Plato of the *Republic* and Plutarch owe us an account of how to take their talk of conflicting parts. It remains that there seems to be a continuum of cases from my being at first inclined to think a dog dangerous and then seeing that it is not, or deciding that it is, at one extreme, and my being seized by blind panic at the sight of a dog and running before I know what I am doing, at the other. We might be clear what we want to say at either end of the continuum, but the in-between cases do precisely fall in between. The

Stoics' opponents are doubtless in clanger of portraying every case as a battle lost or won—and if lost, where is their responsibility? The Stoics are in danger of portraying everything as clearly divided between the unproblematically intentional and the unproblematically uncontrolled.

This is, I think, connected with another feature of the Stoic position: the influence of reason on passion. They do not follow Socrates in thinking that judgment automatically follows extra feed-in of information. The different *hormetic* presentations are not all presentations about the same sort of long-term effects. The presentations of 'reason' may not embody any understanding of why what is feared is not evil, and so the attraction of running away is not annihilated. Nevertheless, the result of education over the years should be to come to such an understanding, and this raises the question of whether the simile of the scar in the account of *propatheiai* is not itself a scar trying to cover a wound in the theory. For the choice seems to be either to take the reactions to be the result of association with certain sights or sounds, or to be residual takings of those as signs of danger. The first sounds too weak, and too weak for what the Stoics were prepared to say; but the second seems to involve some dual level of judgment, even if fleeting. Further, if it is only fleeting, there must be a confident implication that fear of heights, say, can be overcome by extending one's understanding, and similarly with other emotional reactions. There is, of course, some truth in the position. Either extended knowledge of a thing's effect or a changed view of what is important can affect whether or not one is afraid and if so how much. Similar things apply to anger, indignation, pity and so on. But unless the residual reactions can all be accounted for as shivers or tremors simply associated, for example, with the sight of spiders—if they are residual takings of the spider as a menace—then the talk of a scar is covering up the fact that we are talking of different levels of judgment; and these judgments are not necessarily of the sort that say on one side, 'English spiders bite fatally', and on the other, 'English spiders' bites are not important enough to worry about'. Those, after all, are consistent. The judgments may be contradictory, so introducing conflict. I may have learned that English spiders do not bite at all, but the presentation retains the menacing bite-threat of my early training. None of this necessitates as yet talking of two *hormai*, but it suggests that the Stoics have not followed through quite to the end what their view is on the relation of extension of understanding through reason and the form of rational presentation. The more the residual reactions have to be portrayed as incipient avoidance movements in response to a

judgment, the more like *hormai* they will seem, and the more important it will become to try to portray them as mere blenchings and shiverings.

Comparison with Socrates and Aristotle

It may help to highlight some of the features of the Stoic position if one compares it with those of the Socrates of the *Protagoras* and Aristotle. There is a superficial similarity between the Stoics and Socrates. Both seem to hold that we cannot act against our judgment, and that emotions are to be intellectualized into forms of judgment, thus transferring them from the role of potential opposers of reason to that of expressions of one's reason. The similarities are, however, only superficial. To begin with, Socrates believes that we have always before us just one objective and the role of reason is to work out the means of achieving it; we are always and only responsive to judgments about the achievement of that objective. By contrast, the Stoics consider that we have a complex nature with many needs to be balanced, and that we start life with a tendency to recognize these different needs as they arise. At the start, however, we are given to false judgments about the importance of these needs, and no clear overall view of what is best. Education develops a tendency to want to understand the interplay of these needs in our own lives, and, if successful, in the economy of nature as a whole. We therefore slowly grow through a period of appreciating many conflicting calls on our assent to a theoretically possible stage where our overall understanding removes conflict. The development of reason is not, at least primarily, a having of calculative techniques, but a broadening of the understanding of the interplay of elements in our own motives and nature as a whole. For Socrates, every *phantasia* is a judgment as to where the best course lies, which will be followed unless ousted by a later judgment; additional information on where the best course lies produces automatic change of judgment For the Stoics, a *phantasia* is a presentation that it is fitting to do such and such, which does not become a judgment in the Socratic sense until assented to. Consequently the effect of further considerations of reason is uncertain. They will produce further *phantasiai* in addition to the original ones, but up to any given degree of reflection the result may just be conflict. While the Stoics agree with Socrates that the passions are expressions of our judgment what to do, they differ from him as to the form of that judgment, since for Socrates the form of judgment always concerns

what is best overall, whereas for the Stoics an agent may not have a conception of nature which enables them to make that form of judgment. Further, through their views on *propatheiai*, on the possible concurrence of conflicting presentations and on the independent physical effects of a passion, the Stoics can give a far richer and more convincing account of the phenomena of *akrasia*. Their view of the development of reason is nearer to that of Plato's *Republic* than to that of the *Protagoras*.

If we turn to Aristotle, it is easier to notice contrasts. Aristotle portrays *akrasia* as a conflict between the operations of reason and desire. Reason is displayed in the working out what to do in the light of some guidelines selected for the circumstances. The opposing passion results in some diminution of full actualization of deliberative desire. The Stoics, by contrast, do not have this deliberative picture of reason, nor the 'blind' view of sensual desire. The initial conflict will not be between reason and blind passion, but between presentations each of the form: 'It is fitting to…' These presentations will be sufficiently appreciated to exercise a pull on assent. Lacking Aristotle's emphasis on *prohairesis*, they have no presumption that deliberation will be effective, though they are confident that overall understanding will remove passion. There is no suggestion of our judgment being in any way off-beat. What Aristotle might portray as conclusions—'It is fitting for me now to…'—are fully grasped, but assent is always required to opt between them. But whereas Aristotle portrays desire as directed at pleasure, ignoring the question of what is good to do, which is reason's business, the Stoics portray desire as an assented *horme* to what is erroneously viewed as good. This presupposes a presentation presenting the *akratic* course as fitting, on all fours with the other presentations. All forms of conflict are possible until the stage of assent to a *hermetic* presentation; but, of course, one cannot assent to two incompatible presentations at once. So Aristotle's and Plato's mechanism for allowing for conflict is abandoned, and with it any favourable view of any of the passions.

CHAPTER VI

Aquinas and Others

Socrates had made deliberate action against one's better judgment impossible; Aristotle had explained it as always due to some epistemological malfunction; the Stoics inserted a moment in their account of action where the agent opts for/assents to a course of action. Aristotle had left it obscure whether the akratic agent could in the circumstances have done other than they did—whether passion could be resisted at the time, as distinct from the tendency to the passion being worked on over time. While it is not easy to reconcile the Stoic position with their general determinism, it seems clear that the insertion of assent is intended to secure responsibility for the agent, by leaving it always in some sense possible for them to assent to any of their current rational presentations; consequently reason can always win, in that the agent can always give reason the prize. This is sometimes seen as introducing some notion of will, of a power to enact or not enact what reason suggests. It is also worth noting that the Stoics abandon the Aristotelian distinction between reason's 'presentations', which are always of the form 'It is good for a human being to…' and those of desire, which do not present anything as good for beings of such and such a category to do. For the Stoics all these presentations are of the general form 'It is fitting for me now to…' Both these changes become incorporated into later treatments of action.

That no-one can knowingly do wrong was not, for the Stoics, as it was for Socrates, a main paradoxical battle cry. It came, rather, as part of their concern with the nature of the passions and the means for their control. Philosophers influenced by Stoicism in the early centuries AD were primarily influenced by the attitude advocated towards the emotions and to the ups and downs of life.They would take up the view that we are indeed responsible for our actions, but the special disputes surrounding supposed action against one's better judgment sank into the

background. St Augustine is, of course, famed for his discussion of his own practical difficulties with regard to weakness of will, but he does not, so far as I can find, treat its possibility as in any way a theoretical problem. Indeed, the discussion in *Confessions*, Book VIII, Chapter ix, for instance, shows a wildness of terminology appropriate enough to the dramatic description of psychological struggle, but not indicative of concern to tease out the strands of a theoretical problem. He talks at the same time as though there were more than one will, and as though the ineffectiveness of either showed that it was not the whole will—whatever that may mean. The talk of various wills suggests a veering towards Platonic division of the person, but without the Platonic contrast of reason and passion; the talk of the whole will, as though when whole it is effective within the psychological sphere, suggests a more Stoic view of the unity of the *hegemonikon* and the effectiveness of final assent. This last is tied to the view that the agent can always decide, in the internal forum, between temptations, and so is always responsible.

In the centuries which follow there is frequent discussion of a variety of problems concerning free will and responsibility, and these discussions, of course, contain material about the mechanisms of choice which make it clear that the philosophers in question should or should not have found the occurrence of deliberate wrong-doing problematic. The fact remains that it receives only the most cursory attention until the resurgence of interest in Aristotle in the thirteenth century. Since 'weakness of will' was a *topos* of Aristotle's, it became one among those who studied or commented on the *Nicomachean Ethics*. It is not clear, however, to what extent it was felt to be a real problem. The treatments by Albert the Great and Aquinas, for instance, read a little like dutiful rehearsals of arguments on a subject that does not really fire their anxiety. Aquinas' treatment is not, however, a mere rehearsal of Aristotle, but is rather an attempt to combine the main lines of the Aristotelian solution with ideas which became current in the centuries after Aristotle.

One development that is worth remarking is the separation of the question whether one can be overcome by passion from the question whether one can in general deliberately choose the worse course. Socrates seems to have denied the possibility of being overcome by passion simply because it would be a case of deliberate wrong-doing; Aristotle, by contrast, seems to have taken all deliberate flouting of one's better judgment as being due to the interference of passion, and does not even raise the question of whether without its influence there

can be deliberate choice of the worse course. The distinctions in the force of claims about knowledge, of course, make it possible, with his views on voluntariness, to agree to voluntary wrong-doing—but in ways which seem weaker than deliberate wrong-doing. For the Stoics, anyone who has not achieved the condition of a sage can opt to assent to any of a number of presentations, and there is no reason why any of these have to be turbulent. Calm decision contrary to one's reflective judgment is therefore a possibility. It does, however, become increasingly problematic as the agent's understanding approaches perfection. All actions contrary to the recommendations of reason are a sign of a failure fully to understand what is best. For full understanding would remove the appearance of appropriateness of any but the genuinely appropriate action. This led to problems for Christian authors who wished, in order to avoid attributing the genesis of sin too directly to God, to hold Adam and Eve and Satan fully responsible. One way was to make Adam perfect so far as understanding and aspiration are concerned (this was, of course, assumed of Satan). But while this makes it clear that they were responsible, it makes it a problem how they ever came to sin, given a Stoic-influenced view of the relation of understanding to erroneous choice. What makes it problematic is a view of rational choice which still makes even non-passionate deliberate wrong-doing a result of some failure of understanding. In Aquinas we get two separate questions: first one about 'incontinence', or wrong-doing under the influence of passion; and secondly, one about non-passionate deliberate choice of the worse course. The latter is at its clearest in his discussion of the fall of Satan.

How Aquinas Sees the Problem: Passionate Akrasia

Aquinas discusses passionate *akrasia* for two reasons: first, because Aristotle has a discussion of it and so any commentary on the *Ethics* must deal with it; and secondly, because there is a range of sins related to the various passions, and so his moral theology will necessitate a discussion of their seriousness and the sinner's culpability. Such sins result from disordered passion, but in a state of innocence human beings did not have disordered passions; so any account of the first sin must allow of non-passionate deliberate wrong-doing. Problems about the possibility of this do not, however, dominate the discussion of the passionate cases. There the concern is more about degrees of culpability.

Of course, these cannot be completely separated, since doubts about culpability will commonly take the form of doubts about whether the agent in such cases really knows that the action is bad; but Aquinas does not seem anxiously engaged with someone holding a Socratic style of thesis, but rather patiently sorting out what has to be said to make sense of the obvious facts.

As we have seen, Aristotle's treatment of the passionate cases concentrates on the way in which the akratic can be said to know that what they are doing is bad. If we ask whether at the time of their choice they could have resisted their passions and followed reason, and if so how, there is no explicit answer in Aristotle—one is left with views as to what he must have thought because he was an intelligent, sensible philosopher. Between Aristotle and Aquinas there have been philosophers insisting that the agent has to opt for one or other of the courses before them. In Aquinas there is indeed interest in the Aristotelian question of the way in which the agent has knowledge, but Aquinas is also insistent (*Summa Theologiae*. la.IIae. Q.77, art. 3 *ad tertium*) that it is in the will's power to assent or not to what passion inclines to. It is not immediately clear what this amounts to.

The problem is that Aquinas has a variety of ways of describing what is apparently the same situation. Thus in *Summa Theologiae* IIa.IIae. Q.77, there is a discussion of the cause of sin with respect to the sensitive appetite. In the response to article 3 we are told that the will has the power to assent or not to what passion inclines one to, and to that extent our appetite is said to be under our control, though the action of the will is itself impeded by passion. In article 7 we are told that in cases where passion does not completely remove the use of reason, reason can either push the passion out by diversion to thought of other things, or make it ineffective through its control of the movement of the limbs. The mechanism of distraction is expounded in article 1 and in *de Malo* Q.3, art. 9. In Q.156, on incontinence, the incontinent are said to fail through negligence, in not firmly setting themselves to the resistance of passion through the judgment of their reason.

The first of these ways of talking suggests a Stoic view, whereby the agent gives or refuses assent to what passion suggests. The second passage could just be saying the same if we take reason as having the function of assenting or refusing assent to passion, and in the latter case assenting to the entertainment of other thoughts. But in the third passage reason seems to be cast in the part of that, assent to which might serve to defeat passion. There seems here to be a not altogether comfortable

blending of Aristotle and the Stoic tradition. In Aristotle reason is concerned with what is good for beings of a given category, as opposed to desire which is just attracted to pleasure; and one might expect *voluntas* to be a rational desire, a concern with what is best for the agent. But in Aristotle's terminology, to claim that reason can always win would be to say that reasoning about what is best can always get the better of desire—but lacking some story of assent, it would become a good question why it does not. The agent is kept responsible because the failure of knowledge is not of the right kind to exculpate. It is true that in that case, according to Book III, whether or not to act should be within the agent's power, but we are not told how this might be so, given the account of how knowledge fails. In the Stoic account, on the other hand, assent, while being an operation confined to rational beings, is not confined in its operation to the presentations of reason, if that means the presentations resulting from consideration of what is best for humans. Aquinas' account of the incontinent not sticking to the judgment of reason must be claiming that they fail to stick by their judgment of what it is best to do, not that they fail to stick to their assent. In that case assent is not what reason does; rather, it is something either reason or desire may win.

The discussion of distraction is also somewhat obscure. The position seems to be that as one power of the soul is activated the operation of another must diminish. From this it would follow that if I am fantasizing lustfully, my fantasizing will diminish in so far as I succeed in thinking about a more respectable way of behaving, or engage in some activity exercising another faculty. Taking this to be an exercise of reason, it would be a case of reason repelling desire. But it is more plausible to think of this as an activity which reason encourages me to engage in. Presumably, then, the situation is that (i) there is always something to be done which will do away with the influence of passion; (ii) reason can always work out something effective to do; (iii) commonly, the agent will realize this, so that it can be said that reason presents this as something to be done; (iv) the agent can assent to this 'direction'. But we need some account of how the agent comes to reason, and here we need something like the story of assent. Not that this would yield us an account of why we assented to the proposal to reason, but it would acknowledge that reasoning is something we have to set about and need not: it is neither always operative, nor something which randomly and unpredictably starts up or fails to.

In discussing cases of passionate *akrasia* Aquinas speaks of reason or

the will or assent/dissent being hobbled by passion, thus suggesting some impediment; but also of the will having it in its power to assent or not to the inclination of passion. So far as the akratic's knowledge on the occasion goes, Aquinas seems to follow Aristotle's model, which is presumably what gives flesh to the talk of impediment or hobbling. Passion is not, as with the Stoics, an assent, but clearly it is looked on as making assent to anything else more difficult, and partly by interfering with the agent's knowledge of the particular premiss; yet this impaired knowledge, with the similarly impaired knowledge of the conclusion that the act, say, should not be done, is sufficient to leave open the possibility of assent.

Clearly, Aquinas is intent to hold to the point that at the time of choice the akratic could have overcome the influence of passion. Their case is not that of someone in whom passion has removed the use of reason. At the same time he wants to give its due to the talk of passion threatening to overwhelm reason. What is not clear is that he gets any further than simply preserving all these factors in his account. They are all things one no doubt wants to say, but the question is how they all fit coherently together.

Non-passionate Akrasia

One of the problems about the account of passionate *akrasia* is how to cash the talk of impediment or hobbling, especially as it seems that while reason, say, is only operating with a hobble on, it can nevertheless always remove the hobble, so that the agent returns to a normal state. Aquinas does not, however, seem to be worried primarily about how the passionate akratic does know; rather he is interested in distinguishing the seriousness of akratic wrong-doing from non-akratic, while still holding the akratic responsible. But he does seem to be committed to believing in two kinds of non-passionate deliberate wrong-doing: one where the agent has gone wrong in the matter of aims or principles; and one where in some way they know the course they are taking to be contrary to what they know to be right. This last Aquinas has to allow for because he is committed to saying that Satan sinned, but before sinning the angels could not be wrong about what is good (the distinction between real and apparent good gets no purchase in their case). Their sin, therefore, will not come from ignorance about the end, nor can passion be appealed to to give a sort of ignorance about the particular premiss, for the angels are not subject to passion.

As I have said, the question of non-passionate *akrasia* became acute for Christian thinkers. On the one hand, they wished to assert the doctrine that the creation is God's work, and therefore good: there is no independent force for evil in the world. Evil has to be seen as a falling-off of something good. This applies also to moral evil. Rational beings, like all animate beings, aim naturally at their own good; but in the case of rational beings this takes the form of pursuing their own good by the use of reason. For moral evil to occur in the first place, a rational being which has not so far erred must err. In the case of the angels there is no possibility of passion explaining this, and the intellectual power attributed to them makes it a good question how sin came about. They cannot, without qualification desire something evil, as that would suggest that God had created something with a love of evil. Yet since, if they were to blame, they must have known that what they were doing was wrong, they must in some sense have desired evil.

Anselm on the Sin of the Devil

In the medieval period this question is directly treated in the period before Aquinas by St Anselm, in a work entitled *de Casu Diaboli* (On the Fall of the Devil). This work is a short dialogue between a disciple and master, and the section which most concerns us begins in Chapter IV. Anselm wishes to show that the Devil must have sinned by wanting some good, but in some way wanting it inordinately. The primary aim cannot have been to lose something he had. His model is a miser who, by definition, wants to hold on to his money. If he is found to be spending his money the explanation cannot be that he wants to get rid of it, and so takes an interest, say, in health cures: if that were the case, he would hardly be a miser. We have to suppose that he tolerates the loss to achieve some other good. In sinning, the Devil lost the state of justice, and that could only have been done by willing something which at the time he ought not to have willed. Now (since he was in a good state) he must have willed to have whatever he had; therefore to sin he must have willed something he did not have and which at the time he ought not to have willed. But (since he cannot will evil as such), the only objects of his will were justice or some constituent of his happiness. Willing justice could hardly be sinful; therefore the sinned by willing some constituent of his happiness which he did not have at that time, which he ought not to have willed at that time, but which could at some time add to his happiness. Clearly, by

willing inordinately something more than he had received, his will overreached the bounds of justice. But when he willed what God willed that he should not will, he willed inordinately to be like God.

This creates a problem for the disciple in the dialogue. For God can only be conceived of as unique, such that nothing can be like him; so how could the Devil, with his clarity of intellect, want what he could not conceive of? To this Anselm's reply is that the Devil did not explicitly will to be like God, but did so in effect, in that he willed something of his own will without any subjection to the will of another—which is God's prerogative. Further, by willing what God willed that he should not will, he placed his own will above God's, and so willed to be greater than God.

It seems clear that Anselm has two main concerns: first, he wants to make it clear that the desire/will that led the Devil to sin was a desire/will for something good; and secondly, he wants the desire/will to be like God to be coherent, given the level of the Devil's intelligence. A human being might muddledly want to be a second unique omnipotent being, but an angel would smell a rat at once. On the other hand, he must have known that what he did was contrary to God's will, and that what is contrary to God's will is bad, in a way sufficient to make him culpable. This, however, leaves a problem: how did he fail to realize that what he wanted would only be to his good if acquired in accordance with God's will? We seem to need some account either of how the angelic intellect went astray without removing culpability, or of how the angelic will could freely pursue something realized to be a distortion of the good. The first of these routes is attempted by Aquinas, and the second seems to be favoured by Scotus.

Aquinas on the Sin of the Devil

Aquinas faces the problem most clearly in II *Sententiarum* ds 5 Q.I, art 1, where he considers the question of how the angels could sin. He first says that it is an article of Catholic faith that the angels did sin, and then says: 'but it is difficult to see how they sinned: for there can only be sin in the will if in some way reason goes wrong'. He then recalls Aristotle's views on knowledge of the universal anil particular premisses, and the siting of the necessary failure of reason at the level of the particular premiss, commenting on the fact that the rest of Aristotle's explanation will not apply to angels because of their lack of passion. He continues:

> 'but although in angels the judgment of the intellect cannot be hobbled in this way, for the simple reason that they have no such passions, still, it can be hobbled in so far as by considering one thing it is withdrawn from considering another. This is because its understanding is not of a plurality all at once except in the way in which everything is contemplated at once under the heading of a single term. It happens, say, that something should be chosen when one takes one condition of the matter into consideration, which should not be chosen when all the conditions are considered in relation to each other; and in this way error in choice, and so sin, was possible in the angels.

In his reply to article 4 he gives an example of a doctor who considers certain medication appropriate for a patient for the treatment of one condition, which is not overall appropriate because the patient also suffers from another condition—and presumably the medication would be fatal or very damaging because of this condition. The doctor, however, does not take this condition into account.

Applying this to the angels, we cannot, as we can with the doctor, suppose that they are ignorant of the other condition. Indeed, if we look at the description of their sin, they are said to know what is needed in order to know that what they are doing is wrong; so the explanation relies on their withdrawal of consideration. Satan was correctly aware of the value of his own excellence, particularly of his dignity, and had a correct desire to enhance it. He cannot, in fact, have believed that he could be, or therefore that it could be a good for him to be, equal with God—that would be an unacceptable failure of angelic intelligence. So what must have happened is that, attending to something which he rightly took to be a good, but withdrawing consideration from other facts, such as that his glory came from filling his part in God's creation, he refused to obey God, thus in effect claiming equality with God; consequently his valuing of his own excellence became disordered. This situation, whereby reason is confined to a partial consideration of the situation, is also called a hobbling, which suggests that that word may be operating as no more than a name for whatever explains reason's going wrong.

How, then, does the explanation go? Clearly the angels do not in a way have and in a way not have the requisite knowledge, as that notion is applied by Aristotle—certainly not if Aquinas' interpretation is right. For in his commentary on the *Ethics* (*Sententia Libri Ethicorum*, Liber

7, k. 3, n. 15) Aquinas interprets that as a condition in which the subject is two removes from use of the knowledge: when I am in a normal state there are many bits of knowledge which I am not using, but which I have in the sense that they are available for use on call; but when I am asleep, or dead drunk, these bits of knowledge are not available on call, but they will be when I come back to a normal state. Consequently, my epistemological condition is in the latter case superior to that of someone who does not have the knowledge at all, but inferior to that of someone who has it at only one remove. The angels must always have knowledge in the sense of its being available on call, as, indeed, must humans who non-passionately but knowingly do wrong. With humans, however, as we have seen, the usual explanation is that they have got into the way of giving a certain good a disproportionate value, which leads them to do some evil consequent on pursuing that good. Where they have gone wrong is in their way of thinking about what it is good for a human to do. This will not work with the angels, at least so far as their first sin is concerned, since for them to have gone wrong in their way of thinking they must already have sinned. So how is it that Satan failed to take into account those considerations which indicated that the only enhancement of his glory lay in not defying God? The only answer given is that being finite it is possible for him not to embrace everything at once in his understanding. In thinking that his glory is good, he does indeed, in the thought 'my glory', embrace everything that that concept embraces; but he may not advert to everything covered by the concept, and so not bring all his knowledge about his glory to bear, although he has it available for use.

Presumably Aquinas would say something similar about the first human sin, though as the intellect is taken to be less powerful, the difficulty might seem less pressing. That something of this sort might occur is, of course, intelligible; what is strange is that this has to be resorted to as the explanation of what must have happened. For what it disallows is that Satan could have chosen to defy God adverting at the time to the fact that it was wrong to defy God. In fact, when considering Satan, rather than just giving examples of what he means by 'not considering', he says that it is possible for something to be choiceworthy taking one condition into consideration, which is not choiceworthy when all the conditions are considered in relation to one another. The thought may be something like this: if Satan had thought that this is against God's will, he would immediately have brought to bear his knowledge of why it is so, and that would immediately have

made him realize that his own glory could not be achieved this way, and it would then be unintelligible that in pursuit of his glory he should defy God. This would not be the case with human beings. They could presumably, being of inferior intelligence, realize that being like God is a good thing, realize that God has forbidden them to take a given means to become like God, but not understand why it is forbidden, and so not understand how what they are doing is incompatible with attaining the good they are after. There may therefore be room for allowing both of the explanations applicable to angels, and of an explanation which would cater for their doing something which at the time they thought was wrong. In this last case, however, we should have to postulate some failure of understanding. It seems as though, as with the Stoics, a proper understanding would remove any disproportion in the valuing of any good, at least in a person without disordered passions.

We have, then, the following possibilities of deliberate wrong-doing:

(i) The agent may be influenced by passion. Passion is such that in extreme form it removes the capacity to reason, and in that case no voluntary action occurs. In the akratic case, the one of weakness, passion has to fall short of this extreme, so that the agent is able to remove the impediment of passion, but does not. The impediment is an impediment not to action directly, but to the exercise of knowledge of the particular premiss of the syllogism of reason so as to result in its normal conclusion in action.

(ii) The agent may correctly take something as a human good, say developing their artistic talents, but give it disproportionate weight. They have not officially adopted an erroneous view of the human good, that, for instance, artistic development should take priority over all else; but if artistic aims are threatened, they might lie to procure their ends, recognizing that lying is wrong, but being prepared to do it for the end in view. In this case it is assumed that the agent has got into the way of giving undue weight to artistic considerations in their deliberations, and this deliberative disposition makes it 'difficult' for them to reason properly. They have their knowledge, however, not in the second—remove way of the akratic, but available for use.

(iii) This explanation could not be available for the first sin of the

angels, nor of humans, given Aquinas' view of the state of innocence. Something has already gone wrong if there is such a deliberative disposition, either in the agent or at least in the agent's educators. With the angels the assumption is that it is possible for them to consider some of the conditions which bear on appropriate action while failing to consider all the relevant conditions. This could presumably also hold in the human case even without postulating a generally distorted disposition: I could recognize the value of my artistic development, take steps to secure it and fail to take into consideration that my good, and so my artistic development in properly desirable form, requires that other activities also be given time, and more than I am allowing for. In this case, the agent has the knowledge available for use, and there is no need to attribute a tendency to downgrade it in use. What we do not, of course, get, is a case where the agent at the time thinks that what they are doing is wrong.

(iv) It may be that the peculiarities of the view of Satan's first sin come in part from views of the extent and form of his knowledge. This seems to require that Satan 'had' the understanding such that if he used it he would necessarily see that what he was proposing to do could not achieve the good he had set himself. In that case, given that human knowledge, even in a state of innocence, does not have to be supposed to reach such heights, we have the possibility that in the human case the agent might appreciate the value pursued in the one course, recognize that in fact it is forbidden, and so must be wrong, but not understand how the act forbidden could fail to lead to the good proposed. In that case we would be able to adapt the account of Satan's sin so as to allow for an agent in some sense knowing that what they were doing was wrong, but the limitation on their understanding would be an impediment to their putting their knowledge to use: not that they could not, but that the understanding they do have leads them to concentrate on one value and leave the other considerations on one side.

It seems, then, that while Aquinas does allow for someone deliberately doing what they think to be wrong, there are limits on what he is prepared to envisage. It is not clear, indeed, whether he is prepared to allow of the

possibility under (iv); but even if he is, he will do so only on the assumption that the agent's understanding is defective. He will not allow that someone might be uninfluenced by passion, fully understand that a given course is wrong, advert to that understanding at the time and nevertheless deliberately pursue that course. He could, of course, allow that someone might consider that a given course is educationally worse, or worse as a means of furthering some ambition, and nevertheless do it; but the assumption would be that they realized that on this occasion it was better for a human being to pursue the educationally worse course or forego that ambition. Underpinning this is the view of *voluntas* or will. To begin with, *voluntas* is a rational appetite. All beings tend to pursue their own good, that is the best kind of life for their species. In the case of human beings this is a rational desire for their own good: that is to say, they want to work out what is the good life and pursue what they work out. It may be that any given human being gets things wrong, and so pursues something other than the good life, but this is an aberration: the natural object of the will is the good. But *voluntas* is also what issues in chosen action. Thus the desire to taste something sweet does not directly produce the action of reaching for a cake; rather, the desire influences the will through hampering the operation of reason so that the will pursues something other than its proper object. On this view there is no room for anyone deliberately pursuing some objective except under the apprehension at the time that what they are pursuing is a good for humans. In particular, there is no room for an alternative form of motivation which can accompany full acknowledgment that a given course is good, but lead to the selection of some goal acknowledged to be bad for human beings and a consequent bad act. All pursuit of evil has to be explained in terms of some apprehension of a genuine good going wrong.

It is perhaps worth noting that this view does not entail that an agent must choose the better of two courses. It is perfectly possible, as with Aristotle, that two possibilities of a life should be without vice, and yet one of the two be a life with a fuller manifestation of the higher virtues. Since neither life would be turning away from God, but each developing different God-given gifts, or the same with different emphases, the choice in either case would not require any need to explain how something evil was in some way desired and willed. Consequently, if 'worse' means 'less good', it is perfectly possible to choose what one knows to be the worse course. What causes difficulty is the proposal that one can choose what one knows to be a bad course. It is to give sense to this that Aquinas' distinctions are brought into play.

Aquinas takes what might be called an intellectualist approach to the question. Deliberate wrong-doing is explained in terms of something faulty in the agent's reasoning at the time of acting. It is true that the agent turns away from something good, but they do not in any fullblooded way hate what they turn away from, but rather fail to take its goodness properly into account Scotus, by contrast, takes what might be called a voluntarist approach, and his insistence on the importance of the freedom of the will leads to interesting departures from Aquinas in his treatment of the Devil, and of deliberate human sin.

Scotus (*Ordinatio* II, dist. 6, q. 2: On the Sin of Lucifer) takes over from Anselm the point that aversion is parasitic on wanting: if the Devil was averse to obeying God, it cannot have been naked aversion, but must be explicable in terms of something he valued and wanted, but which obedience to God threatened. God may now be viewed as a threat to a valued end, and so hated. Of course, in Scotus' view the end is not only valued, but attracts the will because really valuable. Scotus also makes two distinctions which he believes to be built on Anselm. First, he contrasts what he calls 'friendship love' with 'love of desire'. This might sound like a contrast between selfishness and altruism, but in fact it is not. Scotus is contrasting loving something/someone in such a way as to want for them what is good for them, and loving things because they are good for something loved in the first way. Love of oneself is thus a special case of friendship love, which might better be called the love that is fondness. The second contrast is between love of oneself and concern with one's own happiness, and love of justice. It is the latter which makes altruism possible, and so friendship love for others than ourselves. Just as aversion is parasitic on wanting/valuing/loving, so love of desire is parasitic on friendship love; for the love of desire just consists in loving something because it is a good for something/someone loved in the other way.

The first two points are used, as they are by Anselm, to show that the Devil's sin must have arisen from a friendship love which made some other good look undesirable. On the other hand, this love had to be disordered. Since it is impossible for a love of justice to be disordered (since love of justice is at least a love of proper order), the love responsible for sin must have been a friendship love for himself.

The love of justice, however, not only makes altruism possible, but is a necessary condition for freedom. In his reply to the first objection

on the sin of Lucifer, Scotus argues that if the Devil had had only friendship love for himself, and so only concern for his own happiness, he would not have been free, and so would have been unable to sin. The point is that in those circumstances he would only have been capable of taking considerations of his own happiness as reasons for acting, and so his behaviour would have been determined by those. He would, I suppose, have had some ghost of freedom: if his reasoning either reached the conclusion that two courses were equally desirable, or was unable to determine which, if either, was more desirable, then he might have the ability, nevertheless, to opt for one; but there would be no question of not being determined by his desire, and that is apparently required for freedom. The love of justice introduces another (at least one) friendship love, such that the pursuit of what is good for the object of the first friendship love cannot be reduced to pursuit of what is good for the object of the second, nor vice versa. This may be necessary for Scotan freedom, but is clearly not sufficient: for an agent may be so constructed as to pursue the good of the first object of friendship love only if it is not detrimental to the second (but not conversely). Scotus is assuming that this is not the case, and that since we are able to act, we must have the ability to restrain either friendship love in favour of the other, and so be determined by neither. In other words, our wills are free. Since they are not determined by either love, it is possible for an agent to be fully aware that what they propose to do is incompatible with one good, but nevertheless do it. The only restriction on the will is that it choose the object of some friendship love or some desire (*concupiscentia*).

In order to hold this, it is obvious that Scotus must fail to hold what Aquinas seems to about the Devil's knowledge: that if he had taken into account all relevant factors of which he had knowledge, he would have realized that he could not achieve his objective by taking the chosen course, and so would not have chosen it Anselm faces the problem in terms of whether the Devil knew that he would be punished (*op. cit. cc.*22–3). He argues that certainly the Devil must have known that if he disobeyed God he would deserve punishment: it would need a low grade of intellect not to know that; but he argues that he did not know that he would be punished. This might seem strange, as it seems to suggest that God would not give him his just deserts. But it was possible for him to think that God would not wish to deprive of happiness so glorious a creature. Something of this sort must have been the case, since if he had believed that he would be punished 'he would not have been able, willing happiness and being in

possession of it, spontaneously to have willed something which would result in loss of happiness'. While this might succeed in escaping the difficulty of knowing that he would be punished, it is unsatisfactory so to limit the difficulty. The conflict, after all, is between a wish of the Devil's and a command of God's. Quite apart from the question of punishment, how did the Devil come to think that he could secure his own happiness by disobedience? He would have to think that God did not will the best for him, or did not know what was best for him, or could not deliver what was best for him. Of course, he might not understand how submitting to a given command would serve his happiness, but surely the presumption would have to be that it did? If so, then the basic form of Anselm's difficulty should recur: the Devil would be pursuing his happiness by a means which he should have realized would not achieve it

There are various possible responses to this difficulty. One could stress the distinction between knowing that disobedience would not lead to happiness and understanding how it does not, and then give a special role to understanding in the explanation of choice. Alternatively one might suppose that the Devil thought that God might well demand behaviour which redounded to his own glory but which, while not harming his creatures, did not give them the fullness of happiness of which they were capable. When friendship love attaches to some object other than oneself there is always the theoretical possibility of conflict between the demands of the one and of the other. Love of God too might require a willingness to forego some of one's own possibilities for happiness. In that case the Devil might have been in the situation of having to choose between two goods which were not so related that in the end the pursuit of the one could only be successful by achievement of the other. At this point Anselm's question would return: what are we to suppose the Devil believed would be the results for him of disobedience?

Clearly the Devil could believe that God requires him to give God's glory precedence over his own. Therefore obedience to God would be seen as foregoing one's own glory or happiness to some extent. According to Scotus it is possible for the will to moderate desires, and for independent desires to be immoderate. For the Devil's desire for happiness to be immoderate it is sufficient that either he desire it in a way which makes him see God's precedence as a threat, or desire its completion immediately, when properly it should be delayed, or that he desire that it be given him without his having satisfied the necessary conditions. In any of these cases he correctly sees that his objective is threatened by obedience to God, and if he fails to moderate his desire, as he might in the exercise of his

freedom, he might intelligibly prefer punishment, despite diminished happiness, or loss of happiness, to a state where he is not allowed to achieve happiness as he wishes, without any abatement.

Since he knows that he ought in justice to give precedence to God, he knows that he ought to moderate his desire for his own happiness so as to consider it less important than God's honour. The resultant inordinate desire leads to his (correctly) seeing obedience to God as an abandonment of what he wants. There is consequently a state of conflict between his desire for happiness and his desire for justice. There is no failure to take any factors into account. The failure is in his will, which refuses to moderate the inordinate desire, with the consequence that he knows that he is offending against justice and hates God as a threat to the achievement of his desire.

What is true of the Devil can also, in Scotus' view, be true of human agents. In his discussion of the Sin of Malice (*Opus Oxoniense* II, dist. 43, q. 2) he writes:

> If we give an affirmative answer to this question [whether it is possible for a created will to sin out of malice by willing something which is not presented to it as a true (i.e. unqualified) good, or an apparent or qualified good], it is easy to distinguish the sin against the Holy Spirit from other sins. For since the will is joined to the sense appetite, it naturally joins in its pleasures, and so in sinning by an act resulting from the inclination of the sense appetite for its pleasure, it sins from weakness or impotence, and this is appropriately said to be a sin against the Father, to whom power is attributed. It needs intellectual knowledge to act, and so when reason errs it does not will correctly, and its sin from reason's error is called a sin from ignorance against the Son, to whom wisdom is attributed. The third would be simply a sin of the will, from its own freedom…and that would be a sin of malice, and appropriately called a sin against the Holy Spirit, to whom goodness is attributed. Yet even if it is not assumed that a created will can will evil as evil, it is still possible to attribute to it a sin of certain malice when the will, from its own freedom, without either passion in the sense appetite or any error of reason, sins.

Scotus, then, feels no need to follow Aquinas in supposing some failure to take all factors into account. Believing that the desire for happiness

and the desire for justice do not collapse one into the other, and that we are not, as it were, programmed to give one precedence over the other in our choices, he can allow us full knowledge of the wrongness of our actions, and attribute our behaviour to erroneous willing. There is a sense in which this would not be willing what was thought to be the worse course: the agent is not willing what they believe to be the less good way of achieving their objective; they are, rather, pursuing an objective which they know to conflict with another to which they know that they ought to give priority. One consequence of this possibility of objectives finally conflicting is that there may come a stage when reason may have exhausted its contribution to the debate, and it is still left to the will to decide to which considerations it will give weight.

This notion of freedom is one to which I shall return briefly in Chapter XII. One point which people commonly find puzzling about it is its apparent irrationality: the will is, indeed, confined to taking goods as its objects, but the choice between them is entirely arbitrary. The objective selected gives the reason for the agent's subsequent action, but for the selection itself the agent has no reason. This looks like a freedom of indeterminism, with supposedly rational agents the victims of unpredictable influences, as mysterious to themselves as to observers.

Summary

To sum up: the medievals had to face squarely a question left uncertain by Aristotle, the question whether, uninfluenced by passion, a person could deliberately choose what they thought a bad course. The question reached its critical form in the case of the Devil, who lacked passion and was endowed with superhuman gifts of intellect. At the same time the tradition had absorbed the Stoic view that rational animals did not act immediately on presentations, but had the power of assent This opened up the possibility of rejecting the conclusions of reason. There was still some influence, however, of the other Stoic view, that full understanding would ensure assent to true presentations. This led to problems with regard to the Devil: did he fail fully to understand, or to use his understanding? or is even full understanding inadequate to ensure assent? On this there was division of views, with Aquinas tending to the former, Scotus to the latter, with the familiar problems attending intellcctualist and voluntarist views.

CHAPTER VII

The Post-Medievals

In the period from Descartes to the early twentieth century the sorts of problem that I have been discussing hardly feature in the works of major philosophers. Sometimes they have problems, but they arise from a different source; sometimes they should have them, but no problem is envisaged; sometimes their views on the explanation of action would hardly allow a problem to appear.

Descartes

Thus Descartes has an interest both in passionate *akrasia* and in deliberate wrong-doing. In *The Passions of the Soul* (especially Part First, Articles XLV–L) he addresses the question of what the power of the soul is in respect of its passions. He concludes first that the will cannot, in cases of violent passion, directly affect the passions, because

> almost all of them are accompanied by some disturbance which occurs in the heart, and so all through the blood and the animal spirits. Until this disturbance subsides the passions are present to consciousness just as sensible objects are so long as they act on the organs of sense…. [So in violent cases] while this disturbance is at its height, the will can do no more than refuse to give way to its effects, and check many of the movements to which the disturbance disposes the body.
>
> (Article XLVI)

The will can only affect the animal spirits through its action on the pineal gland; but this it is always free to, do, and by diligence it can separate the movements of the pineal gland, and of the spirits which

represent certain objects, from those which are normally joined to them in the excitation of certain passions, so as to acquire control. He concludes secondly that there is no soul that cannot, if it applies itself, achieve complete control of its passions (Article L).

Throughout the work Descartes clearly has a concern for the importance and also the mechanics of self-control; and in the course of it there emerges a picture of the will as a faculty whereby we can directly control the soul's own actions (e.g. its desires) but only indirectly the affections of the body, such as the turbulence around the heart and those bodily movements caused by the animal spirits in an emotional context, such as the tendency of the legs towards running in cases of fear. Weakness emerges as a failure of the agent to exercise their will in ways appropriate to achieving control. The concerns here, if not the manner of meeting them, recall those of philosophers we have already considered. There is also a sense of a problem about the struggle with the passions. Here, however, there is a contrast with earlier philosophers. Descartes is not puzzled as to how we can choose, under the influence of passion, what we take to be the worse course; nor, if we can, as to what sort of knowledge it is that we have; nor as to how we can reconcile talk of emotional influence with the attribution of responsibility. In his case the talk of internal struggle is seen as threatening his view of the soul as a simple substance with no parts:

> For we have just one soul, which does not contain any parts. The subject of sense impressions is the same as what is rational, and the various appetites are all acts of the will. The mistake of making it a play of different agents, usually opposing each other, arises from our not having properly distinguished the functions of the soul from those of the body; it is to the latter that we should attribute all the opposition to reason that we observe in ourselves. (Article XL VII)

He is therefore at pains to portray our sense of struggle without recourse to internal division. It has in fact two elements: first, the will and the animal spirits both operate in opposed directions on the pineal gland, and the tendency of the spirits is to cause a desire, of the will to work through such techniques as distraction to alter the movements of the spirits. The result is commonly a rapid succession of desires which is mistaken for contemporaneous conflict. Secondly, since the animal spirits initiate bodily movements which are not directly under control of the will, but may be in directions contrary to its wishes, the will is often

exercised in opposing them—but then the conflict is not within the soul (*ibid*.). He is not here at all concerned with the explanation of action and attribution of responsibility, but with preserving his metaphysical view of the soul. It is not, despite superficial similarity, like the Stoic concern for the unity of the *hegemonikon*. That, as we saw, was not a concern to claim that the soul has no parts (as a material thing, it does), but resulted from the fact that to divide the *hegemonikon* would be to have no agent or two. In either case we lose the possibility of attributing responsibility to one agent. Even here, then, Descartes is not developing the set of problems that I have been considering so far.

On the question of non-passionate deliberate wrong-doing Descartes wrote practically nothing. But in a letter to Father Mesland of 2 May 1644 (translation in Anscombe and Geach, p. 290) he writes:

> The only thing that stopped me from talking about our liberty to follow after good or evil was my wish to avoid theological controversies as far as I could and keep within the limits of natural philosophy. I grant you that whenever there is an occasion of sin, there is indifference; and I do not think doing wrong involves seeing clearly that what we are doing is bad—it is enough to see this in a confused fashion, or even to remember having previously judged it to be bad, without attending to the reasons that show it to be bad. If we saw clearly that it is bad, we could not possibly sin—not so long as we did see it this way; hence the saying *omnis peccans est ignorans*.

This has a strong Socratic flavour, but we get no exposition of what lies behind the view, nor of what is involved in seeing clearly. So while we might think that Descartes ought to have a problem about deliberate wrong-doing, we in fact get no treatment of it.

Leibniz

Leibniz also supports theses which commit him to a definite position on some of the issues earlier discussed, without apparently considering the traditional problems particularly problematic. Leibniz held a Principle of Sufficient Reason to the effect that for every event, whether a chosen act or unchosen event, there must be a sufficient reason. In the latter case the principle seems to amount to the principle that every event has a cause. In

the former case the principle holds that any agent that chooses a certain course must have a sufficient reason for the choice. Where the agent is subject to other influences than reason, as a human agent, the sufficient reason may be a reason or a cause; but with a purely rational agent such as God a sufficient reason always takes the form of a consideration showing that course to be best. Indeed Leibniz holds both (i) that if God chooses to do A, then God must consider it better to do A than any alternative (and since God is omniscient it will be better); and (ii) that if God considers A to be better than B then he will consider that there is more reason for doing A than B, and therefore there will be sufficient reason for him to do it; and being rational he will do it. Thus in the fourth paper in his correspondence with Clarke (*The Leibniz-Clarke Correspondence*, p. 39), we find: 'When two things which cannot both be together, are equally good; and neither in themselves, nor by their combination with other things, has the one any advantage over the other; God will produce neither of them.' In the fifth paper (*ibid*., pp. 56–7) Leibniz defends himself against a misunderstanding of (ii). He there distinguishes between moral and absolute necessity: 'We must also distinguish between a necessity, which takes place because the opposite implies a contradiction; (which necessity is called logical, metaphysical, or mathematical); and a necessity which is moral, whereby, a wise being chooses the best, and every mind follows the strongest inclination.' No-one has the freedom to choose to make a (Euclidean) triangle whose angles do not add up to 180 degrees; choice is between possibles, neither of which is necessary. But it is a mark of wisdom to choose the best of those possibles, and a failure so to choose would show lack of wisdom or lack of freedom. So rational beings always choose what they consider the best of the options. With human agents there are other factors which might operate, some of them, like the passions, conscious, others unrecognized; consequently their thinking something best may not always supply a complete explanation of the actions they choose. Further (*New Essays Concerning Human Understanding*, p. 190):

> I would not have it believed…that we must abandon those ancient axioms, that the will follows the greatest good, or flies the greatest evil, which it feels. The source of the little application to the truly good comes, in great part, from the fact that, in the affairs and occasions where the senses scarcely act, most of our thoughts are surd, so to speak…i.e. void of perception and feeling, and consisting in the bare employment of symbols…. Now such knowledge cannot move us; we need something lively in order to feel emotion.

The principle applied to God will remain, however, as one governing the rationality of choice.

If the earlier chapters have been on the right lines, only Socrates, of the ancients, held the ancient axiom without qualification. It was not uncommon to hold a view that the will could not pursue evil as such, but could pursue either of two goods, even though one were a lesser good. Not even Socrates was clearly committed to the view that a perfectly rational agent would be unable to act if faced with what it takes to be two equally valuable courses or, equivalently, that if a perfectly rational agent chooses a course of action, it must consider it better than all alternatives. These two principles stand out very starkly in Leibniz, strikingly unattractive, but have proved influential in less glaringly naked form in later writers. So while Leibniz does not give us an extended discussion of deliberate wrong-doing—his interest is in freedom and God's goodness—he certainly ought to find it problematic, and is committed to rather fierce principles on the explanation of rational action. In contrast, other philosophers of the period had views which would not allow any problem to arise. I shall take Spinoza and Hume as examples.

Spinoza

Like Descartes, Spinoza was much concerned about the power and control of emotions, but adopting a version of monism in preference to Descartes' dualism of mind and body, he cannot view the influence of the emotions as physically inspired as distinct from being a matter of ideas:

> By emotion I mean the modifications of the body, whereby the active power of the said body is increased or diminished, aided or constrained, and also the ideas of such modifications. N.B. If we can be the adequate cause of any of these modifications, I then call the emotion an activity, otherwise I call it a passion, or state wherein the mind is passive.
>
> (*The Ethics,* Part III, Definition III, Elwes' translation)

To be an adequate cause of an emotion, according to the first definition of Part III, I have to be able clearly and distinctly to perceive it, or fully understand it. To the extent to which that is not true, my emotion is a passion not an action of mine. Since knowledge of good and evil is itself an emotion (*ibid*., Part IV, Proposition VIII), it will also be a passion in so

far as we do not clearly and distinctly perceive it. So long as we are subject to passions we are not in control, and no one passion has, in virtue of what it is an emotion for, any special position of power. Control is a matter of our becoming adequate causes of our emotions, that is, of our perceiving them clearly and distinctly. So complete control would only come with full understanding, but would inevitably come with it. This just fleshes out an earlier passage (*ibid*., Part II, Proposition XLIX), where he rejects the contrast between the will and the understanding and portrays error (in each, since they are one) as a matter of inadequate ideas. Success of the will and success of the understanding are one and the same. There is no problem about our desire for good being overcome, and our experience of struggle there is just the experience of conflicting passions; there is no possibility of a fully developed understanding being overcome, since fully to understand is to be fully in control. Will and understanding go hand in hand; full understanding and passion are incompatible; conflict is only possible when our emotions are at least to some extent passions, when we have inadequate ideas of them or, in other words, when our understanding/will is underdeveloped. This intellectualized view of the passions, and the rejection of a separate faculty of will, which constitute a total rejection of Descartes, also make it impossible for Spinoza to develop anything like the traditional problems.

Hume

Hume argues that the traditional opposition between reason and passion is based on a mistake. Strictly speaking reason is concerned with truth and does not of itself constitute a motive to action. When we are faced with a set of facts, they do not determine what we shall do. For that we need some passion. If my reason tells me that the ice on a pond is thin and that the prime minister is about to walk on it, that is in itself unable to produce any action. Given a certain devotion or aversion to her, these deliverances of reason will doubtless affect my behaviour; but neither devotion nor aversion is contrary to the deliverances of reason. What influences the will is a person's passions. Sometimes these are founded on certain suppositions, and so the findings of reason come to influence the passions and affect the will. Indeed, Hume is rather optimistic about the influence of reason in this respect: 'The moment we perceive the falsehood of any supposition, or the insufficiency of any means, our passions yield to our reason without any opposition' (*A Treatise of Human*

Nature, Book II, Part III, section 3). Except in the cases where passion is based on a supposition either about the existence of objects or the sufficiency of means to the objective of the passion, no sense can be given to calling a passion unreasonable. Passions are 'original existences', they are just there and may be attached to any objects. Consequently,

> It is as little contrary to reason to prefer even my own acknowledged lesser good to my greater, and have a more ardent affection for the former than the latter. A trivial good may, from certain circumstances, produce a desire superior to what arises from the greatest and most valuable enjoyment; nor is there anything more extraordinary in this, than in mechanics to see one pound weight raise up a hundred by the advantage of its situation. (*ibid.*)

People tend to think that reason can oppose passion through a simple mistake. Reason, as a rule, produces no ripples on the surface of consciousness, and we are inclined to identify with it other actions of the mind whose operations are similarly indiscernible. In particular, since there are calm passions, whose operations we do not notice, we tend to call them reason:

> Now it is certain there are certain calm desires and tendencies, which, though they be real passions, produce little emotion in the mind, and are more known by their effects than by the immediate feeling or sensation. These desires are of two kinds; either certain instincts originally implanted in our natures, such as benevolence and resentment, the love of life, and kindness to children; or the general appetite to good, and the aversion to evil, considered merely as such. When any of these passions are calm, and cause no disorder in the soul, they are very readily taken for the determinations of reason, and are supposed to proceed from the same faculty with that which judges of truth and falsehood. Their nature and principles have been supposed the same, because their sensations are not evidently different.
>
> (*ibid.*)

Given these views, Hume can hardly be. expected to feel a problem about weakness of will. There can be no special problem of irrationality, or failing to act on reason. All we have is a conflict between two

passions, and there is nothing surprising in the fact that sometimes one sort, sometimes another should win. We should hardly expect, therefore, to find a long discussion of the 'problem' in Hume. We do not get one.

Philosophers and the Will

What is perhaps more surprising is that a number of philosophers who have a great deal to say about the will, have little or nothing to say about the problems I have been discussing. While the will plays an important part in Kant's moral philosophy, its role is first to be the vehicle of autonomy, and secondly the source of freedom, with an obscure relationship to the causal interconnections which rule the phenomenal world. It is open to dispute whether will can, in the same sense of the word, play both roles, but what is clear is that Kant seems to see no special problem about taking what one considers to be the worse course, whether under the influence of passion or not.

Fichte, Schopenhauer and Nietzsche also, notoriously, have a great deal to say about the will, and even about the weak-willed, but nevertheless have nothing to contribute to the problems discussed in this book. That is not to say that there is no connection. As will emerge, it is arguable that weakness of the will should be seen as weakness of just that, and not some sort of weakness of intellect, or mere lack of desire. What is said later about the will is therefore not unrelated to the theses of these philosophers; but I have not ventured into a discussion of those relationships. It would take one too far afield.

Once we reach the twentieth century, history, for the purposes of this book, is over, and I have only discussed individual philosophers in so far as I have considered it useful to do so for the development of the exposition of my own views.

PART 2

CHAPTER VIII

What is the Problem?

The Setting of the Problem

What in Anglo-Saxon philosophical circles is called the problem of weakness of will concerns what worried Socrates: the problem of how an agent can choose to take what they believe to be the worse course, overcome by passion. The English expression would not, or at least not primarily, bring this sort of case to mind, but rather such examples as dilatoriness, procrastination, lack of moral courage and failure to push plans through. The Greek word '*akrasia*', on the other hand, means 'lack of control', and that certainly suggests the Socratic sort of example. I shall have something to say later about these other sorts of weakness. To begin with, however, I shall consider the problem as it has developed from Socrates.

In the *Protagoras* Socrates seems mainly concerned to show that a person with knowledge of what is best to do is in control of their actions. The prime motivation is not with responsibility but with the value and power of knowledge. Although the concern is to show that passion is powerless against knowledge, this is achieved by first showing that it is impossible to act contrary to one's belief as to what is best—and knowledge entails belief. Knowledge is a matter of correct calculation about what is best. Passions are beliefs about what it is best to do now. The result is that while there is no conflict between belief about what it is best to do and passion, since all passions are beliefs as to what it is best to do, there is a contrast between reason/knowledge and passion; for passion is untutored belief, whereas knowledge involves calculative consideration of what it is best to do. Two things of later importance emerge from this: first, that Socrates disallows the possibility of being overcome by passion because he treats it as a case of acting deliberately against one's better judgment—and he disallows that; consequently he

makes no distinction between passionate and calm action against better judgment. Secondly, reason is allied to some form of long-term, overall consideration. For, although the judgments of reason and passion alike are to the effect that such and such is the best way to achieve the good, the person whose reason is trained necessarily raises and pursues questions of the context and effects of the proposed action, whereas the untrained may raise no such question, and so tend to take the obvious short-term advantage as being long term.

In addition, Socrates has some important theses about the relation of good to deliberate action. It is taken as axiomatic that in any deliberate action the agent takes something as a good, and acts in the belief that what they are doing is the best way to achieve that good—or at least that there is no better way. This is a form of the thesis that all deliberate action is done for some reason, with the terminology of 'good' substituted for that of 'reason'. In Socrates' case, however, it is given the strong form of holding that there is some one good that is the reason or goal of all deliberate action, in terms of which we are able to assess competitors for the title of best thing to do. It follows from this that we shall always choose what we judge best, since judgments that such and such is best are simply judgments that this is the way to achieve our good in these circumstances. It will seem plausible on this view to hold that someone will choose x in preference to y if and only if they consider it better to do x than to do y, and do not think that there is any z better than x to do. Socrates was a Leibnizian.

One might break this stranglehold in a number of ways. One might, for instance, deny that all deliberate action is aimed at something considered as maximizing the good, and claim, like the later Plato, followed by Aristotle, that desire is the pursuit of something simply considered as pleasant, without reference to wider questions of worth.This would allow for the possibility of longer-term considerations being swept aside by passion, but might leave it questionable whether without the influence of passion one could deliberately choose what one considered the worse course. An alternative route would be to claim that the agent might be unable to use one set of considerations to explain why another was not to be given weight. Medea might be unable to resolve the conflict between considerations of Jason's deserts and those of her duties as a mother. Her view of her maternal duties will support the position that it is better not to kill her children, whereas considerations of revenge will suggest that they had best die. What she lacks is an understanding which makes it clear that one or other of these

considerations does not bear. So long as she lacks it her judgments will be to the effect that such and such is best, given that these or these are the only relevant considerations; but she will lack any means of establishing that condition. Assenting to 'It would be best not to do it' could, then, be assenting to a conditional, not to a *hormetic* proposition; assenting to 'It would be fitting to do it' would not amount to claiming that it is best to do it, since perhaps choice can be made so long as the agent takes something as good, without having a view on whether it is overall better. Reason does not, until fully developed, produce a conclusion that such and such is better or best All it does is produce considerations which reinforce the presentation of such and such as fitting to do.

As it emerged in the Stoics this view involved each *hormetic* presentation presenting something as a human good, a view whose seeds are probably in Aristotle. This involves a view intermediate between that of Socrates, that there is one good allowing assessment of all courses of action, and a purely formal view that every deliberate action is done for some reason (i.e. there is an answer to 'what was your reason for doing it?') or, in Socratic sounding language, done for the sake of some good. I call this purely formal, because I am allowing that there may be no way of determining that certain reasons are proper reasons, or certain goods really good, and further that for the agent to have a reason/good it is sufficient that there was something that they were after in acting as they did. The intermediate position figures in many Christian philosophers who were inclined to hold that anything that attracted the will was in some way a good, and taken to be a genuine good. It therefore becomes a problem how Lucifer, and on some views Adam, ever came to sin. Of course, their desire for some good has to be inordinate, but the question arises how they come to choose a course which they know/believe to be worse.

Whether we take the full-blooded Socratic position, or an intermediate one, the agent, either when using reason or always, is concerned with the human good. *Akrasia,* therefore, tends to be seen not only as a failure to be rational, but a failure to be moral. It is viewed as backsliding or weakness, where these expressions have moral connotations. It would be possible to detach the notion of reason from consideration of questions of human goods, and simply contrast longer-term objectives with others. Then it might be plausible to ally reason to the former, in that their pursuit might be expected to require a deal of thought and calculation. Since we might be deflected from any pursuit

involving reason by other considerations, we can now see that if there is a problem about '*akrasia*' it arises more generally than in morals. It must be lurking over the whole range of intentional action. This indeed seems to be where it started with Socrates, and only got confined to moral contexts because of views on the relation of desire and good. This still leaves the portrayal of the problem as one of how an agent can act contrary to the determinations of reason. The *akratic* seem to act contrary to their reason, and yet for a reason; to act against their overall judgment as to what is best, but for something acknowledged to be a good. This seems necessary in order to satisfy our intuition that the *akratic* are irrational, but leads to a puzzle about the mechanism of the *akratic* choice.

Davidson is surely right to claim that this is the common theme that runs through presentations of the problem. While it seems plain that we do on occasion act contrary to our preferred course, it is not clear how we can portray that action as deliberate and at the same time claim that some other course was preferred. This is not to say that there is no special interest in the specifically moral cases, but that that interest in part comes from their being examples of the more general pattern.

So far the problem has been set in terms of reason: how people can, for a reason, act contrary to what their reason advocates. Faced with the apparent intractability of the problem so worded, one might be attracted to a different approach. When we describe deliberate action we can, indeed, say that it has to be action done for some reason. But if we look at examples where we offer these accounts of action it becomes plain that they cover for another mode of explanation. When we claim that Medea murders her children for a reason, viz. to take revenge on Jason, what is given as her reason simply specifies what it is that she wants that leads her to kill them: she killed them because she wanted to take her revenge. Clearly, at the time, she wanted to do that rather than follow her maternal instincts, or, in other words, her desire for revenge was stronger than her love for her children. It is natural to identify reason, as Hume suggests, with those desires which are directed to more carefully thought-out objectives, or ones conceived more in the round. It may be a mark of a rational animal that these desires should for the most part be effective; but once we put the matte* in this terminology there is obviously no problem about the possibility of reason so understood being overcome by passion: why should a shorter-term, perhaps more turbulent, desire not on occasion suffice to overcome a

calmer one? For that matter, why should any desire not on occasion overcome another? Being in the set designated the desires of reason does not guarantee a desire strength.

As with Socrates, so on this view, no distinction is made between calm and passionate *akrasia.* The idea that there is a distinction of interest to be made there comes from the failure to see that by 'reason* we mean 'calm desires' or 'tranquil passions'. There might, indeed, be something of more interest to be said about tranquil passions. It might be, for instance, that they are what are sometimes called higher-order passions or desires, desires, for instance, to have, or be the sort of person to have, certain lower-order desires or passions. It might interestingly be that the possession of such higher-order desires is connected with a certain degree of development of rational powers, and that might throw light on why there is a general prejudice in favour of thinking of the akratic as irrational. The fact remains that while on Socrates' view *akrasia* seems unproblematically impossible, on this view it seems unproblematically possible.

In what follows I shall start with the presentation of the question in terms of reason, following Socrates. This will lead to consideration of a popular contrast between overall or all-things-considered judgments and other more particular judgments. I shall then consider the statement of the problem in terms of desire, and the possibility of marrying the two forms of presentation. The outcome will be that on any of these ways of putting the problem, either *akrasia* is unproblematically possible, or it is unproblematically impossible. In no case is there any room for puzzlement. At this point I shall first question the importance of the contrast between reason and passion, all-things-considered and other judgments, and suggest that the source of people's puzzlement about *akrasia,* as a general puzzlement about deliberately acting against one's better judgment, should be located elsewhere. This will lead to a discussion of the relation of *akrasia* so construed to irrationality. I shall then return, in Chapter X, to *akrasia* in relation to passion, where different connections with irrationality will be found. After a discussion of specifically moral weakness, I shall return to some of the themes of Chapters VIII–X in Chapter XII in relation to the question of whether *akrasia* should not perhaps be considered to be weakness of *will*. Finally I shall consider whether other forms of weakness may be considered irrational in a related way, and shall explore some of the connections between traditional and not so traditional examples of weakness on the one hand and self-deception on the other. In the present chapter I shall

concentrate on approaches inspired by the way Socrates sees the problem.

The Socratic Problem

As we have seen, the Socratic denial *of akrasia* stems from an analysis of deliberate action combined with the supposition that the akratic's behaviour is an action. One might try to break the stranglehold of Socrates' arguments either (i) by conceding that in general there is a paradox in the suggestion that an agent might deliberately decide on doing what they thought to be the worse course, but claim that something odd happens when passion is involved; or (ii) by claiming that something goes wrong with Socrates' general account of action, so that even calm *akrasia* is possible and so any problems with the passionate sort must arise from special considerations. For these purposes I take 'Socrates' analysis' to refer to the purely formal account mentioned earlier, whereby if a piece of behaviour is to count as a deliberate action the agent must have some reason for doing it and do it for that reason.

Passionate Akrasia

Those taking the first position concede that some special explanation is needed for the occurrence of *akrasia,* and hope to find it in passion. They might, like the later Plato, portray the agent as a battlefield, or, like Aristotle, suggest that physiological changes turn the agent temporarily from an explicit good-pursuer into a particular-pleasure-pursuer, putting their beliefs to the service of this latter purpose. Either of these views might serve to cause problems for either the strong Socratic or intermediate position mentioned in the first section of this chapter. It is, however, the formal position which gives strength to Socrates' account of action, and here either view has to answer the Socratic challenge: did the akratic agent choose to do what they did? If so, there must be a reason/good they had in mind, which led them to choose so to act, and that is what the agent took as a reason or a good in so acting. So they cannot have acted contrary to their reason. No help is to be had from appealing to the presence of passion. Either passion rendered the agent non-responsible, and so no action was performed; or it did not, in which case it embodied the agent's reason for acting. I shall return to this question in Chapter XI.

Calm Akrasia

Blocked in like this by Socrates, one might hope to make some progress by being more sophisticated in one's account of reason and action. Socrates may be right that the introduction of passion is irrelevant. Where he has gone wrong is in his account *of akrasia.* The akratic is not, of course, someone who acts contrary to the reason they act on. Rather they act contrary to their considered judgment. If we can explain how that is possible in general there will be no special problem about passionate *akrasia.* So what we need to examine is the relation between reasoning about what to do, and one's reason for doing what one does. There seems to be a problem here because it is tempting to think that the conclusion of a piece of reasoning about what to do will be a reason for acting in one way or another. The reason one ends up with is the outcome of one's deliberation. It seems at the very best irrational, and at worst impossible, that a piece of deliberation leading to the conclusion that it would be best to refrain from further bets at the roulette table should result in the laying of a further bet for the thrill of it.

I shall approach this question through consideration of the views of Donald Davidson (see Davidson 1969 and 1982), partly because he makes a good stalking horse for the points I want to make, and partly because it is likely that most readers will be familiar with his work and developing from there will make it easier to locate my position.

Davidson

Davidson (see Davidson 1969) proposes the following definition of the *akrates*: A acts akratically iff A (i) does some y intentionally, (ii) believes that there is an alternative action x open to them and (iii) judges that, all things considered, it would be better to do x than to do y. The assertion that there are those who act akratically is apparently inconsistent with two supposedly obvious propositions:

P1 If an agent wants to do x more than he wants to do y, and he believes himself free to do either x or y, then he will intentionally do x if he does either x or y intentionally.

P2 If an agent judges that it would be better to do x than to do y, then he wants to do x more than he wants to do y.

P1 and P2 yield P4: If an agent judges that it would be better to do x than to do y, and he believes himself free to do either x or y, then he will

intentionally do x if he does either x or y intentionally. P4 seems inconsistent with P3, the claim that people behave akratically.

Davidson's strategy is to claim that the 'all things considered', which appears in the definition of the akratic, but does not appear in P4, makes all the difference. Not, of course, that the wording contains any magic (see Taylor 1984 and Schueler 1983); but it emerges that the correct analysis of the akratic's all-things-considered judgment shows it to be of a different form from that required for the truth of P2. He first points out that typical cases of practical reasoning, taken at face value, seem to produce inconsistent conclusions by valid arguments from true premisses, or at least ones accepted as true by the subject The last is enough to make what the agent is doing a doubtful candidate for the title of reasoning. An example would be someone who accepted both A and B:

A Anyone with heart disease would always better not smoke than smoke. I am a person with heart disease. Therefore I should always better not smoke than smoke.

B One would always better do what is necessary to make one's guests feel at ease than not To make this guest feel at ease I should have to smoke a cigarette. Therefore, I should better smoke a cigarette than not.

It would seem that if an agent accepts the premiss in each case, as surely they might, then they are committed to inconsistent conclusions—if, at least, this reasoning is to be interpreted in familiar ways—and they are entitled, not to say obliged, to detach the conclusions. Davidson suggests, however, that it should not be interpreted in familiar ways. Rather, the premisses should be read with a *prima facie* operator, Pf. Roughly, the universal premiss of A would read: For all x, Pf.(x would better not smoke than smoke, x has heart disease), where the comma stands in for something like 'given'. We can now see that from the first two premisses in A we can only infer 'Pf.(I would better not smoke than smoke, I have heart disease)'. At any rate, it is clear that on this analysis 'I had better not smoke than smoke cigarettes' is not detachable, nor 'I had better smoke a cigarette than not' from B, and therefore there is no contradiction. The general form of universal premiss is: for all x, Pf. (Ax, R_{1-n}) (*prima facie*, x should do A, given reasons R_{1-n}). Since it is the detached form that enters into P2 above, it is clear that that form of judgment cannot be yielded by practical reasoning. But all-things-considered judgments are just

conclusions of practical reasoning in which it is also claimed that all relevant reasons have been taken into consideration. In other words, they are Pf. judgments. Their conclusions are not, therefore, what Davidson calls *sans phrase* judgments, since those cannot be detached from these sorts of premiss.

This leaves a problem as to how we ever get to the *sans phrase* judgments. To explain this Davidson posits a principle of practical rationality, which he calls the principle of continence, which enjoins us always to act on our best judgment, i.e. our all-things-considered judgment as to what it is best to do. Since this is a principle of practical rationality, it is irrational not to follow it; but not impossible. This view succeeds, then, in securing our two intuitions: that *akrasia* is possible, and that it is irrational.

Pf. and Conditionals

There are problems about the interpretation of Pf. Its initial introduction is by reference to the expression '*prima facie*'; but Davidson also describes Pf. judgments as conditional. Further, it is necessary that a given form of conclusion be not detachable. Finally, the outcome should not only save our intuitions about *akrasia,* but also supply a plausible account of practical reasoning. A first attempt might go something as follows: suppose someone claims that the fact that they have promised to attend a wedding is a *prima facie* reason for saying that it would be better to go than not. A plausible rendering of this claim might be that they have promised, and if their promise is the only relevant factor, or is sufficient against all other considerations against going, then it would be better to go. Obviously, on this account, A cannot infer from 'I have promised to go to the wedding' and 'My having promised is a *prima facie* reason for the conclusion that it is better to go than not' that it would be better for them to go. They need to establish either that the promise is the only relevant reason or that it is sufficient anyway. '*Prima facie*' puts up the warning signs that there is more work to be done before the conclusion can be drawn. It does not, however, indicate that the work cannot be done. On the contrary, if the antecedent is satisfied, then the conclusion can be drawn. Nor is it to the point here to say that the conditions cannot or cannot be known io be satisfied. In discussing *akrasia* we are discussing the judgments of agents. The question is whether under certain circumstances agents should, given their beliefs,

detach, i.e. make *tout court* (and therefore *sans phrase*?), judgments. But on this account, if the agent believed that R was either in itself sufficient, or was the only relevant consideration, then on pain of irrationality they would have to detach the conclusion.

If these detached judgments are *sans phrase* judgments, this clearly will not meet Davidson's requirements, and that suggests a different account. Perhaps we should read 'Pf.(Aa, R)' as 'R supports Aa', where 'supports' is too weak to allow detachment, i.e. from the fact that R obtains and R supports Aa, it does not follow that Aa. It will now be true that even if all the relevant considerations support a conclusion and all obtain, nothing follows. On the other hand, it might seem rational to follow the principle of acting on all-things-considered judgments.

This better fits Davidson, and also the analogy with the probabilifying operator, even if it gets a little far from intuitions about '*prima facie*'. It does, however, raise problems about the plausibility of the account of practical reasoning. Part of the initial plausibility comes from the suggestion that much early deliberation should be seen in terms of *prima facie* judgments. Commonly we are consciously feeling our way at this stage and simply determining what would follow if certain considerations were the only important ones. But this all allows for the possibility that further consideration will lead the agent to decide that they are, or they together with some others which have the same conclusion; and in that case, the conclusion is one we should expect the agent to detach without benefit of any principle of practical rationality. The analogy with probability judgments helps to illuminate what Davidson is trying to say, but has no persuasive force as regards the thesis. In particular, there is no analogue of probability terminology characteristic of practical reasoning such that an analysis of it would illuminate what the form of these judgments is. It is at least as plausible to accept something along the lines of the first account of Pf., and allow that agents detach the conclusion. The form of the antecedent seems to secure that when satisfied it would not yield contradictory conclusions.

Part of the trouble here is the uncertain relationship between what we optimistically call deliberation and the display of the agent's reason for acting as they do. When I am wondering what to do today, there will almost certainly be some review of some of the possible activities which I think it worth pursuing, or to which I am attracted. While this may be important for getting going, and is a perfectly rational thing to do, it may only by courtesy be called a reasoning process (see example on p. 109). Suppose I decide to get on with my book. I might now indulge in some

exploratory thinking. The next section in the historical development is to deal with the Stoics, so it might be best to get out *Stoicorum Veterum Fragmenta* and start listing the relevant passages. On the other hand, perhaps it would be better to sketch out a picture of the Stoic development as I presently see it first, and then I shall have a better eye for the relevant passages. Yet again, given the time available, it might be better to spend it on the bibliography—and so on. I call this exploratory because I am once again reviewing possible courses, only this time I have some yardstick against which I might hope, though perhaps vainly, to measure them. They all have something to be said for them, but that recognition is nothing like so strong as the suggested reading forprima *facie* above. Still, since I do now have, though in vague form, a goal, some considerations which look like being in some weak form reasons pro and con, can be brought. At this stage, however, things are commonly 'concluded' in some such form as that I have used: 'so it might be a good idea to…' That something tells in favour of doing x does not entail that it is *prima facie* best, but at most that it might be worth pursuing the question whether to do it. When I have got a fairly precise objective, I might even make a *prima facie* judgment or two about what it would be best to do. Eventually, with luck, I shall, for some reason(s) decide on a given course. At this stage, only these reasons feature as an account of my practical reasoning'. The existence of the earlier stages may give some verisimilitude to some of what Davidson says, but to lay much stress on it would be like including an account of people's early fumblings towards discovery in an account of theoretical reasoning.

Detachment and Sans Phrase *Judgments*

The difficulties about the interpretation of Pf. lead one to the question of just what the relation is between being detachable and being a *sans phrase* judgment. It seems clear that a *sans phrase* judgment is what is involved in P2. If we can understand what is required there we should get some illumination. If we look at P1 and P2, it is obvious that we need interpretations of 'A judges that it would be better to do x than to do y' and 'A wants to do x more than he wants to do y', such that the second might plausibly be thought to follow from the first, and also plausibly be thought to entail that if he believes himself free to do either x or y, then he will intentionally do x if he does either x or y intentionally. While it is obvious that we need these interpretations, it is not obvious what they are.

Suppose we start with the second, the plausibility of P1. What interpretation of 'wants more' would make this plausible? Take the following example: Alice claims that she was free to go on a trip to New York, or stay at home and visit her uncle in hospital. She wanted, indeed wanted more than anything else, to go to New York, but stayed at home and did her duty. She believed she was free to go to New York, i.e. believed that she could have gone, if she had wanted. Now that 'if' gives an unfulfilled condition and the whole seems to mean: she could have, and the only reason why she did not was that she did not want to. But that cannot be true since she did want to. Then it must be that she did not want to as much as she wanted to do whatever she did. But she wanted to go to New York more than anything else. Then there must be another sense of 'want more' in play, according to which she did not want to do anything else more than stay at home and visit her uncle. An obvious candidate is one whereby A wants to do x more than y iff A is decided in favour of doing x rather than y. In that case, in thinking herself free she thought she could have gone and that the reason why she did not was the fact that she was decided in favour of staying at home rather than going. It is implausible to suppose her mistaken in thinking that she was so decided. So we may suppose she did, and therefore wanted, in this sense, to stay at home more than she wanted to go on the trip. It follows that in this sense she did not want to go on the trip more than she wanted to stay.

This might persuade us that if she believed the two courses, x and y, open to her and intentionally did y, she did not want to do x more than y; and if this is taken as an arbitrarily selected example to illustrate the general, then it illustrates the equivalent of PI for this interpretation of 'wants more'. This is some reason for accepting the interpretation.

We now have to insert this interpretation into P2, which becomes: If an agent judges that it would be better to do x than to do y, then he is decided in favour of doing x rather than doing y. Now often a judgment using the general form 'It would be better to do x than to do y' serves to give expression to a decision in favour of doing x rather than y. In this case P2 could be made even truer by confining 'judges that it would be better to do x than to do y' to these cases. This at least yields interpretations of PI and P2 which makes them look true and also yield the apparent inconsistency with 'There are akratic actions.' The result is that a *sans phrase* judgment that it would be better to do x than to do y is the expression of a decision in favour of doing x rather than y. It will no doubt be a decision taken on some ground or for some reason. Alice

decided it was better to stay at home because her uncle needed her. But this does not make it conditional: there is all the difference between deciding that if her uncle needs her she had better stay at home, and deciding that she had better stay at home because her uncle needs her.

Suppose now that Alice had decided that her uncle's welfare was a more important consideration than her own enjoyment. She would then argue that if her uncle needed her, it would be better to stay at home, and then confirm that her uncle did indeed need her. On pain of irrationality she now ought to detach the conclusion. Once again, the fact that the conclusion is based on certain reasons will not make it a conditional conclusion. Nor, however, is it clear that it will turn it into a *decision* in favour of doing x rather than y. It is not clear, in other words, that a detached conclusion and a *sans phrase* judgment are the same. On the other hand there may be a lingering feeling mat it is going somehow to be difficult to prevent Alice's detached conclusion being tantamount to a decision. After all, her conclusion is reached after considering all that she takes to be relevant reasons, and if it is in the light of these that she judges it to be better to do x than to do y, what reasons is she left with for doing y? To get clearer on this it will be necessary to consider further what is meant by such expressions as 'all-things-considered judgment' and 'best judgment'.

All-things-considered and Best Judgments

In a trivial sense even akratics do as they think best. Given that they have a reason for what they do and act for that reason, then it must be assumed that they think at least that no other action better satisfies that reason. If they act against their better (best) judgment, therefore, further conditions must be put on the application of 'best judgment'. 'All-things-considered' seems well suited to play this rôle. It suggests, after all, some over-arching use of reason, which will surely make it better from the point of view of rationality. Thus Davidson's account has an all-things-considered judgment take into account all relevant reasons $\{R_1\text{-}R_n\}$, where the akratic's reasons $\{R_j\text{-}R_k\}$ are included in $\{R_1\text{-}R_n\}$. We get the impression that the akratic's reasons are somehow outweighed by the rest of the set. Yet it is not easy to see how.

It might help, to begin with, to look at two examples where one might naturally use the expression 'all-things-considered'. First, suppose an engagement is cancelled and I unexpectedly find that I have a free

afternoon. I could sort out my desk and tidy up my filing system, which stands much in need of doing; or I could take my son to a football match, which I have promised to do sometime; or I could go to a highly spoken of film which is on a restricted run; or I could take the opportunity to finish off a paper, which it will be a relief to get behind me. None of these things has to be done this afternoon, nor do I think that any one has priority over any of the others. In the end I say, 'Well, all things considered, I think I'd better (or just 'will') sort out my desk.' Let's suppose that this list exhausts the good ways I can think of of spending the afternoon. *Ex hypothesi* the reasons for doing the different things are equally good—that is, I do not consider that one is more important than another. For each course of action there is a reason for doing it, but three reasons each of which tells in favour of doing something else. So for each, there is more reason for not doing it than for doing it It is highly implausible to suggest that I think that all these reasons taken together somehow rationally support the sorting out of my desk. The most that 'all things considered' can mean here is 'now that I have considered everything'. While the reasons available supply a reason for clearing my desk, they do not supply a reason for preferring that to the other courses of action, no reason for supposing that course better. This kind of example gives no teeth to the notion of a better judgment over and above its being the judgment which gives expression to my decision.

Let us now take a different situation. I have decided to apply to some university to take a course in modern languages. I decide that the sort of course I want should, ideally, give a good competence in speaking, reading and writing the contemporary language, together with a sound grounding in the history of the language, its literature and the main countries where it is spoken. I recognize that in some of these respects Bath is better than Bristol, in others that Bristol is better than Bath; but in the end I conclude that all things considered it would be better to go to Bristol. Here 'all things considered' embraces just the requirements that I have decided on as relevant. You might know that I like *pâtisserie,* and ask whether I realize that Bath has the better *pâtisserie* and have I taken that into account? The answer will be: certainly not; that has no relevance to the question of whether it is better to go to Bath or Bristol. It may, of course, explain my final choice of university; but it will give no support to the conclusion that it is better to go to that one.

The difference between these two examples is obvious. In the one we get no limitation put on relevant reasons, with the result that we can give no force to the idea that there is more reason for one course than the

others, or that reason supports one over the others. (The case might be different, of course, if one course of action satisfied several of the desiderata; but even then, only if we assume equipollence of each reason.) In the other we do get a limitation, with the result that many considerations which might influence my choice get ruled out as irrelevant, as no reason for supposing that it is better to go to one university rather than the other—although one or more might be the reason why in the end I go where I do.

So far this might all seem grist to Davidson's mill. All that is needed is to interpret the point about Pf. as saying that in practical reasoning the agent only ever reaches conclusions to the effect that such and such an action is better in a given respect. That is not the same as deciding in favour of that action; so we should need some explanation of how the decision is to be generated. It clearly could not be generated rationally in examples like the first, since *ex. hypothesi* there is no reason for any preference for any given respect. Further, even if we extended the list of reasons in the second example so as to add to the 'relevant' reasons the further consideration about the *pâtisserie*, they would not, in combination, support any course as better, since the different reasons support the conclusion that their chosen course is better in different respects: they do not combine in the same form of assessment (unless, of course, I have changed the requirement to 'best university with a *pâtisserie*'). On the other hand it seems plausible in this example to put reason on the side of the 'relevant' reasons, leaving the *pâtisserie* as irrelevant To begin with, they seem to evidence a clearer manifestation of rationality in that they show the agent using reason over a broader range. Secondly, they form the bulk of the reasons the agent considers, so that in some sense there seems to be more reason on that side of the debate. Thus it may be significant that Davidson seems to speak as though the akratic's reasons which finally determine their action form a minority subset of $\{R_1\text{-}R_n\}$. A combination of these two points might make it attractive to suppose that a rational agent will tend as such to follow a principle of aiming to manifest maximum rationality, and following that will, in this sort of case, normally produce a decision in favour of what reason favours.

One trouble with this is that it tends to encourage thinking of the akratic as someone who chooses a relatively narrow or short-term course in preference to something considered in a broader more long-term context. Suppose, however, we consider the following example: Arthur has two visions of his future life. Sometimes he sees himself as

an unworldly academic, writing books which will win him the admiration of his colleagues and the plaudits of future generations, but not a great deal of money, to the blandishments of which he is enviably indifferent. At other times he hankers after a different future, in which he outsmarts all his contemporaries in his dealings on the Stock Market, lives with his wife in a splendid manor house and establishes a wealthy dynasty whose members will hold positions of power in the country. After much cogitation, Arthur decides on the scholarly path. Well on into his second year of academic life he gets into conversation with some stockbroking friends. Tempted by his old yearnings he agrees, against his better judgment, to challenge one of his friends to a race to double an investment, on the understanding that if he wins he will join the firm. He despises his weakness, but now the die has been thrown.

This is not a moral dilemma, but one self-image pulling against another. Neither involves more use of reason than the other; neither is longer term than the other. Yet the case has just the same degree of puzzlement (or lack of it) as the standard akratic case. If one asks what makes the judgment against which Arthur acts his better judgment, the only answer seems to be that it is the judgment in accord with the set of values by which he is currently trying to live. That is to say that it is Arthur's announced ordering that determines that reason lies on that side, or that that is the better judgment. There is no way of making it more rational. Once this is seen, the whole idea of a principle of continence loses its charm. If it is not to be a purely arbitrary principle whose sole function is to produce action of some sort, one needs some way of determining some side of the dispute as that taken by reason. Only so will it seem a principle which it is rational to adopt except, of course, on the grounds that some action is more rational than none. Even so, it seems odd to declare it always more rational to abide by the initially preferred course.

If the principle of continence is abandoned, then we seem to be left with the following: a judgment that it would be better to do x than y may either (i) express a conclusion from certain considerations that in a given respect x would be a better thing to do than y, or (ii) express a decision in favour of x over y. In the former case, there seems no particular reason why the agent should decide in accordance with the conclusion. In the latter case the decision to do x will doubtless have a reason, and one in accordance with which x is the best thing to do. In this case the reason for doing x is not simply one which determines the respect in which x is the best thing to do, but is also what the agent takes as his

or her reason for deciding what to do. The fact that a person is trying to live in accord with certain values does not seem to entail that those are taken as the reason for every choice, even when they are relevant to the situation—though I shall return to this question. Granted that, there seem to be two ways of setting up the case of *akrasia*: first, the akratic derive from some reasons whose relevance they acknowledge, that it would be better to do x than y, but for some other reason decide to do y. On this account, *akrasia* is unproblematically possible. Alternatively, the *akratic* take R_i as their reason for deciding what to do, i.e. x, and decide to do y taking reason R_{ii} as their reason for deciding. This is inconsistent, i.e. the suggestion is inconsistent, and so on this description *akrasia* is unproblematically impossible. To say that on the first account it is unproblematically possible is just to say that the description involves no contradiction. It may turn out on other grounds to be impossible. As things stand, however, it is not even clear that there is any sense in which the behaviour is irrational, except in the very weak sense that it flouts a set of reasons acknowledged to be 'relevant', and declared by the agent to be important.

Later Davidson

In his later paper (Davidson 1982), Davidson seems in some ways to modify his earlier approach. In particular he makes use of an analogy between individuals and groups of individuals more or less politically organized. When we think of a person simply having conflicting desires, they are like a group of people with conflicting sub-groups. Neither action advocated by either party to the dispute is of itself rational or irrational, though each is rational or irrational in relation to the objectives of either sub-group. If the main group is to act as a whole other than accidently, it will have to accept some mechanism for settling such disputes, and the existence of some such mechanism seems necessary if we are to consider the group as in any sense an agent. Let us suppose that a set of arbitration principles and an arbitrating body are agreed. All parties are agreed on the reasons for such a system, which might be felt to constitute the rational organization of the group. Since they are geared to the preservation and working of the group as group, this doubtless yields a sense of 'all-things-considered' as applied to judgments by the arbitration body using the arbitration principles. On any given occasion, however, one party may feel so strongly about its cause that it will flout the decision of the arbitrator.

It will have a reason for doing what it does but (i) it will not be an all-things-considered reason, considering the group as a whole and the arbitration principles, and (ii) it will not be a reason which tells against acceptance of the arbitration mechanism. Indeed, the sub-group may continue to accept the need for that, although what it is doing is damaging to that system. We could strengthen the analogy with an individual by bringing this closer to Aristotle's likening of the akratic to a city with good laws which it does not apply. For we could consider a group which accepted certain settlement principles, but as a group decided not to apply them—for instance, a group which accepted principles of fair trial but deprived a particular war criminal of their protection. In this case the community would have a reason, outrage, for their action, but while leading to flouting of the laws it would not be a reason directed to abandoning the laws. Applied to the individual, this encourages us to think of the akratic as having principles which lead to 'better judgments', which have the role of settling conflicts, but acting contrary to the resultant better judgment. They will have a reason for their action, but it will not, of course, be an all-things-considered reason, nor will it be a reason against the arbitrating principle.

As he expounds the view, Davidson speaks of a second-order principle to the effect that one should follows one's better judgment. This has obvious affinities with the principle of continence of the earlier article. It might seem to share with it the problems about determining what a better judgment is. The analogy with group interaction, however, suggests the possibility of freeing the view from that difficulty. Setting up an arbitration mechanism does not involve fixing on something independently identified as a better judgment, and declaring that it should be followed. Rather, it is a matter of instituting some goals or principles as the ones to have the deciding role, and so constituting the judgments in accordance with them as the agent's better judgment. So the earlier criticisms of the notion of better judgment could be accepted, but do not bear on the new view.

We do, however, retain the advantages of the earlier view. It seems to leave us with no problem about the *possibility* of *akrasia*; on the other hand it leaves it with an air of irrationality. For the weak-willed person acts contrary to what the arbitrating principle enjoins for a reason, but for one which gives no reason against the arbitrating principle, the reason for which they still accept. It is, then, of interest to see just what form of irrationality is involved.

The first and weakest is on the surface of the description. It seems less than rational to dismiss from application a principle whose

relevance is acknowledged, without a reason either for changing the principle or modifying it in this case. Given that the arbitrating principle is, as in the case of the group, the principle of rationally settling conflicts, then the action is an abandonment of a principle of rational settlement.

In the case of the group the analogous principle seems necessary for it to count as a group agent, or in a political context to count as a state or *polls*. One might, therefore, hope that something stronger of this sort might apply to the individual: that such a principle is necessary for someone to count as a person or a rational agent. Just as in a state constant flouting of the arbitrating mechanism would be the same as its ceasing to operate, and so the same as the state ceasing to exist, it might be that the supposition of constant *akrasia* undermined the supposition that we were dealing with a rational agent at all. In that case the connection between *akrasia* and irrationality would be somewhat stronger; not only would each case be irrational in the weak sense, but each is also the sort of case which can only be tolerated as an exception if the general attribution of rationality is to be preserved.

Yet in the case of the individual this seems highly implausible if we try for an analogue of anything stronger than just an arbitrator. That, after all, would amount to no more than saying that a rational individual must be able to decide. If we try to build in the acceptance of arbitrating principles, even granting that they may vary from agent to agent, the claim seems implausibly strong. It might at first seem that thinking of someone as a rational agent requires interpreting their beliefs, desires and pursuits in a maximally consistent way: *akrasia* always introduces a measure of inconsistency, and so threatens rationality. But this seems either trivial or false. If the claim is that taking someone as a rational agent requires discerning some over-arching principle of organization of priorities which is generally effective, it seems false. I may have a job, belong to a choir and a football club, be married with children, like to be respected by the neighbours, enjoy a discussion group, etc. These provide various reasons for doing things on different occasions, and at times they conflict. It seems quite extravagant to suppose that I have, even over limited periods, principles by which I decide on priorities, or rather that if I do not I do not count as a rational agent. One could, of course, make it what one meant by 'rational agent' that I should, but that becomes a rather unexciting failure of rationality.

Alternatively, it might be argued that what characterizes a rational agent is the ability to pursue complex objectives which require

conceptual understanding and deliberation for their pursuit. Thus, if I have decided to run for political office I must have some understanding of what political office is and of the workings of society which bear on it Further, such pursuits require the ability to give priority to certain calls on one's time and to decide against the satisfaction of various competing desires. In other words, rational agency absolutely requires the operation of arbitrating principles.

There is some plausibility to this; but even granting it, it is not strong enough. All that this says is that if an agent decides on a pursuit of any complexity, men this will supply, for the time of the pursuit, a means of arbitrating between competing calls on their attention. Nothing follows from this about the agent having any means of arbitrating between the claims of this pursuit and any rivals. The initial choice between running for political office and accepting promotion in my firm may not have had the help of any higher-order principles of selection. It might have been like the situation considered earlier of unexpectedly being faced with a free afternoon. On the face of it, it seems that many people have a variety of acknowledged likings, obligations, fears, together with a tendency to acquire new ones; sometimes they give priority to one, sometimes to another, without consistency with some overall canons of selection. One may deplore this fact, but for the present the point is merely that it does not seem sufficient to deprive them of the title of rational agent. If that is so, to count as rational an agent does not have to have a set of principles which determine which parts of their deliberation give their better judgment We are left with saying that one part is given that status just by the agent's say so. In that case, either the agent takes the reason which grounds that judgment as their reason for acting on this occasion, in which case, if the action is intentional, they act for this reason, and cannot be acting against it; or they do not, in which case there seems to be no problem in their acting against it. So either *akrasia* is impossible, or unproblematically possible, without being in any obvious sense irrational. For even the weaker sense of 'irrational' considered earlier depends on there being rationally accepted arbitration principles, and these, as we have seen, are no more than an option.

The upshot of all this is that we are still no nearer to an account of 'belter judgment', judgment of reason' or such like, which will show the akratic to be in any interesting sense irrational. 'All things considered' sounds promising, but it emerges that it really only seems promising because of the insertion of 'relevant' before 'reasons' when we expand

it to 'all reasons considered'. For now it seems that we can get the akratic reason outweighed by the agent's other relevant reasons. Trouble begins when we ask what makes a reason relevant. If we take one familiar way of talking of relevant reasons, then usually the akratic will consider the akratic reason irrelevant for determining what it is best to do. If, in that sense, we include the akratic reason as relevant, it ceases to be akratic: instead we have a new determination of their goal, which becomes, say, the best university which also has a good *patisserie.* If, on the other hand, everything which the agent might recognize as an inducement for taking any course is allowed in as thereby relevant, without, indeed prior to, the agent taking any set as determining what they are after, that set of reasons does not yield any best course. In short, it is the agent who has to make reasons relevant.

Desire and Akrasia

As I remarked earlier, an alternative way of describing *akrasia* seems to be in terms of conflicts of desire rather than entanglements of reasons. One reason for this move might be the hope that things become clearer in this terminology. For the talk of reason we substitute talk of a set of desires geared to working out longer-term ordering. On the other hand, any reason for a decision is itself a desire. Acting against reason for a reason now becomes acting according to certain desires contrary to some of the desires in the reason set. But anyone who decides on x in preference to y must want x more than y. So every akratic action is just a case of one desire or set of desires overcoming another desire or set of desires. Put like that, there is no mystery. If, on the other hand, we try supposing that the akratic want to do x more than y, but do y because they want to do y more than x, we get a straightforward contradiction. The only way now of avoiding the contradiction would be to distinguish different forms of the claim that A wanted to do x more than y. One might, for instance, distinguish between considering that doing x is more in conformity with one's ideals than doing y, and viewing the doing of x with more pleasure than the doing of y, and distinguish both from deciding in favour of x over and against y. In that case, so long as we gave different interpretations for each case, it would be possible to want to do x more than to do y and at the same time want to do y more than to do x; but then everything becomes obviously unproblematic again.

If, like Davidson, we do not look on reason and desire as alternatives

giving a different view *ofakrasia,* we shall be committed to one version of what it is to want to do x more than to do y, the version whereby A wants x more than y iff A is decided in favour of doing x over doing y. In that case the move to desire cannot provide a different model. If we try to keep them separate, then the move might suggest a different picture of the human agent. Instead of the Socratic rational pursuer we might have the reactive Humean agent, at any one time a collection of desires of varying strengths awaiting arousal by some informational input. Such a view would require an account of how strength of desire is to be measured, and some argument for supposing that desires stronger by that measure always produced action. If that can be done, a genuinely different picture of the agent will be supplied. If it turns out that the only measure is success in producing action, then we do not seem to have a genuine alternative to the reason model. (For further discussion see Chapter XII, pp. 173–8.)

All this makes it very puzzling why anyone ever thought that there was a problem about *akrasia.* The same seems true if we approach the question through certain analyses of evaluative language (see Chapter XI, pp. 153–5): once we are clear, there is no problem, just the possible and the impossible. But this would certainly be rather surprising. In the next chapter I shall take another look at Socrates and see whether something more interesting can be teased out.

CHAPTER IX

Akrasia and Irrationality

Reflection on the last chapter suggests that we need to reconsider the connection between practical reasoning and reasons for acting, and the sense, if any, in which acting against reason is both possible and irrational. It certainly seems plausible to think of intentional action as behaviour done for a reason, and to display the reason in argumentative form. If I get up from the deck chair to mow the lawn, my reason can, rather ponderously, be spelt out as follows: I thought it would be a good thing to get the lawn mown; that will only happen if I get out of this deck chair and fetch the mower from the shed; therefore…If we accept a Socratic view, the full display of my reason for acting will always correspond with a fairly full display of the deliberative preliminaries. This is because there is nothing else that the human agent is interested in apart from the overall good. Consequently, any consideration of what it is best to do will start from that point There may, of course, be some explorations which come to nothing, and so disappear from the final account; but the starting point of deliberation and the basic reason for acting will always be the same, with the intervening steps which are approved by reason filling in the gap between reason and action. Once we drop that monolithic picture, even if we postulate that practical reasoning will always start from the consideration of some human good, we make room for the possibility of deliberative preliminaries resembling those when one suddenly discovers that one has a free afternoon (see Chapter VIII, pp. 109–10). In that case one's early practical reasoning takes the form of considering various options in the possibly vain hope of finding some decisive considerations in favour of one or other. One does not, however, start with a principle of decision; and in the end there may be an arbitrary choice of the good which will then determine what is to be done. In such cases the full display of the agent's reason for acting will not reach any further back than the good

finally settled on. The various things I might do on this afternoon are all acknowledged with sincerity as equally good things to do for a variety of reasons, for judging between which I have no criterion. This important part of my practical reasoning does not figure in the display of my reason for doing what I do. It is what I adopt as my reason for deciding how to arrange my afternoon which takes pride of place as the premiss of the argument by which my reason for doing as I do is schematized.

In the traditional examples of *akrasia* it is assumed that that preliminary stage has been completed, and the agent has settled on the considerations that are to decide the question. This would be like the case of deciding between Bath and Bristol as universities satisfying certain criteria (see Chapter VIII, p. 110). Even here, of course, we are not guaranteed a result. The criteria chosen may fail to yield a clear preference, and we are back in a situation like that above. Let us suppose, however, that they do. This suggests two possibilities of failure. First, I might claim that, all things considered, I think it would be best to go to Bristol, and mean by this that I consider that the criteria I have chosen tell in Bristol's favour; but then I decide that it would be better to go to Bath because it gives a better grounding in contemporary spoken language, which is one of my chosen criteria.

The problem is what to make of such an agent, or, rather, of such a description. For it seems difficult to retain all parts of the description of me. How can it be true that I accept that set of criteria with a given ordering of importance, consider that they bring out Bristol as better to go to, but consider it would be better to go to Bath in virtue of its better satisfying one of the criteria? If one could lessen the degree of self-consciousness there might be some hope. Otherwise something has got to give: either I do not really accept my official ordering of priorities, or I am not really convinced that they support Bristol, or some extraneous consideration is coming in to influence my decision, and it is not true that I think that it would be better to go to Bath.

The second possibility is that I do introduce an extraneous consideration, such as Bath's possession of a good *pâtisserie.* This would not, of course, affect my judgment that Bristol was a better university to go to, nor make all those considerations irrelevant to the choice of university. On the other hand it would ensure that they failed to give the reason adopted for my action. Suppose, extravagantly, that, discovering a good *pâtisserie* in a town with no university, I decide to give up university education after all and look for another job; then the

display of my reason for going where I go will contain no reference to all those university considerations and the reasoning about them. Not that they were insincere, but that they are finally unproductive of action. In short, to accept reasons as relevant to some question of action is distinct from taking them as the reasons on which to act. As one might expect, there is a frequent overlap, but the distinction remains. This is, of course, being stipulative as regards the use of 'relevant reason'. One could decide to say that the reason(s) an agent decides to act on is thereby promoted to the position of relevant reason(s) for determining what is to be done. In that case there will be a difference between, say, a reason being relevant to determining which of two universities it would be better to go to and a reason being relevant to deciding what to do. A reason will only be relevant in the latter sense if the agent takes it as a reason for determining what to do. The impossibility of *akrasia* comes when acting against reason is taken as failing to act on the reason on which one acts. The possibility seems obvious when we think of acting against reason as intentionally doing something other than what some piece of deliberation brings out as best

Socrates and Animals

If one makes this distinction, it does not immediately clarify everything. The discussion has been conducted in terms of intentional behaviour and action, on the tacit assumption that this is peculiar to humans. Purposive behaviour, at least, is common to all animals, and it is not clear why the Socratic picture cannot be transferred to animals in general. If a cat has learned that there are some fledglings in a nest down the garden and sets out to catch them, then it takes catching the fledglings as its good and does what it thinks best to catch them. No doubt with humans we think that they could articulate what we give as their purpose, whereas the cat cannot. On the other hand, humans commonly do not, and it is not clear what the importance of the possiblity of articulation is. Now our cat has a relatively long-term project in mind as it sets off down the garden, and is alert to all sorts of signals which might require a change of plan. Suppose that on its way it passes a rubbish heap on which some bits of kipper have been thrown away. It pauses, looks towards the nest, then back at the rubbish heap. After some hesitation it makes for the rubbish heap and the kippers. Has it shown weakness of will? It has certainly failed to keep to its original

purpose, and we might contrast it with cats which are not so easily diverted. Given its hesitation, we might describe it as vacillating. But it seems a straightgforward case of change of direction, or change of mind. It is, in fact, precisely as Socrates wants all humans to be.

Even Socrates assumed that if any problem arose it arose with regard to rational animals, and that the appearance of *akrasia* came only with the possession of reason. Later philosophers have mostly taken it for granted that *akrasia,* real or apparent, was a phenomenon only found in rational animals. The general assumption has been that this brings in the possibility of calculation and so the possibility of a conflict between reason and other desires. As we have found, it is not easy to see how this might generate any puzzlement, except through muddle. It may, however, be worth wondering what difference the possession of language might make to the interpretation of action, and how it might give rise to possibilities not there in the simple case.

Language and Akrasia

The only access we have to the cat's purposes is general observation of cats' pursuits and learning abilities, knowledge of this cat's background, together with observation of its behaviour on this occasion, and the environment within which it is operating. With humans we have all these things, but humans also learn to speak. In doing so they acquire the ability to articulate their purposes. In this they not only acquire the ability to put into words the correct interpretation of their actions: they are in general expected to know what they are about and be able to express their knowledge. This is only a *general* expectation. Sometimes I may feel in my bones that it is important to make some point in a discussion, but not be clear why I want to make the point; and sometimes I may act in uncertainty as to whether I am wanting to save my child a distressing experience or assert my superior ability to organize their life. The general presumption is, however, an important element in the characterization of someone as a rational agent. This becomes clear if we try carrying through the description of someone who never has any idea what they are up to. They come down to breakfast in a daze, have no idea why they reach for the milk, cannot tell why they make the remarks they do, and so on. If it was just at breakfast, we could cope with the situation. But suppose this has to be the account of their life. If they do hazard accounts of their actions and must always be wrong, then there is a question as to how we

justify the supposition that they are talking the language of reasons and purposes with any understanding. If they simply do not ever have any idea what they are up to, then we are not faced with a rational agent at all. Note that what is causing the trouble here is not that our supposed agent is always acting contrary to the findings of reason, but that they can never give the reason of any action. To count as a rational agent a person has to be a generally reliable immediate source of information as to their purposes. Consequently, their declarations give us a means of access to understanding their behaviour which is not available in the case of other animals. There must in general be a tally between their actions and their sincerely stated intentions.

Given this, the attribution of *akrasia* might seem to be an incipient threat to the characterization of the agent as rational. There is no threat from a case where the agent, say, has a general policy of not betting, but occasionally decides to have a flutter; nor from a case where the agent starts out sincerely with the idea of going to a concert, works out detailed plans, and then abandons them when invited for a drink. In neither case is there any rift, without change of mind, between stated intention or declared preference bearing on the action, and the action. This is what makes them less than the standard case of *akrasia.* What marks out akratic cases is the apparent occurrent rift within the agent, and the apparent need to give importance not only to the admitted reason for action but also to some declaration of preference for doing something else. The trouble starts when we are no longer content to say that one thing is in general their preference, but on this occasion they are making an exception; nor that in this case they are not serious about the stated preference; nor that they have simply changed their mind. It is the desire to hold that the agent in some way at the time has a preference for doing something other than what they choose (and so in some sense prefer) to do, which arouses Socrates' resistance.

One attractive move to make here is to suggest that the akratic agent always has a second-order desire or preference, in accordance with which they would rather not be the sort of agent who has or acts on the sort of reasons for which they are in fact acting. This preference can continue through akratic action, but is not the agent's preference in the sense of the reason for which they finally decide to act. While this is an attractive move, to which I shall return (see Chapter XI, pp. 157–60), for the moment I shall leave it aside. For a more primitive level of rift is possible, and one which is influential in most descriptions of *akrasia.*

If we return to the example in the previous chapter (p. 111) of a man

torn between two self-images—one of himself as a successful, money-despising academic, and one of himself as a financially successfu dynasty founder—it is clear that there is no simple ordering contras between a desire which does and one that does not involve views on the desirability of the other. It is part of each to prefer not to be the sort of person who has the other desires. The problem arises simply because we want to keep the assertion that the agent takes the one objective seriously while acting on the other. What motivates this is the agent's declaration.

We may get a similar case without the paraphernalia of long-term views. Suppose we have a woman who is rushing towards a cinems pushing past people with the words 'Excuse me, I must get there in time for the beginning.' She has just seen the notice of a film which she has heard well spoken of and has taken a snap decision to see it She has a clear sense of urgency as she enters the foyer of the cinema. Then hei eye catches a coffee stall. 'I'll just get a cup', she says, in a state of fluster, realizing that she will miss the beginning of the film.

Left like that, the case need raise no problems. She might have decided not to bother about the beginning—she would not miss much. She might simply have changed her mind about what was important. Certainly to see the film from the beginning ceased to be her reason for doing what she was doing and was ousted by the purpose of getting a cup of coffee. The problem starts in so far as it seems that she is not simply paying lip-service to the desire to see the film, and has not simply changed her mind about that and decided on coffee instead. Those accounts can seem to oversimplify the situation. In some sense her declared purpose on entering the foyer is still alive, and that is why she is in a fluster. On the other hand, it is clearly not so alive as to be the reason for her action in getting a cup of coffee. We want to place the situation somewhere between saying that the declared purpose gives the reason for her behaviour and employing one or other of a variety of ways of discounting the declaration. These may take a variety of forms: she may be lying; or it may be characteristic of her to be distractable, however great the flurry with which she makes her declarations; or she may, on seeing the coffee, simply have changed her mind about the importance of seeing the beginning of the film. There are characteristic criteria for one or other of these ways out holding, and the problem arises when it seems that there is a lot of evidence about her, and none of it tells in favour of any of them. Yet it is not clear whether there is any space to be occupied between them and seeing the beginning of the film being her reason. It is tempting, therefore, to think that closer

examination must in the end produce a chink in the description, a difference of some sort apart from the outcome.

The motivation for thinking this is strong for the following reason: if we can find no difference, then we have to admit that even in the most favoured cases we cannot tell genuine from apparent articulations of goals—the difference only shows up in the outcome—and this goes for the agent as well as the onlooker. Alternatively, we can tell genuine cases, but the connection between articulation of goal and intentional action is accidental: if it happens most of the time, that is just a lucky fact of life. In either case we have the agent's sincere declarations having an accidental connection with their intentional behaviour, which is enough to cause problems for our view of them as rational agents at all. The attempt to insert cases between the unproblematically impossible and the unproblematically possible, therefore, immediately arouses unease, because if one admitted their existence it is not clear that one could retain one's characterization of the agent as rational: the very case would have to be one in which they failed, without explanation or self-understanding, to pursue their declared goal. It is as though they were not in control—but that must not be taken too seriously, in case we have to say that our film-goer did not decide to take a quick cup of coffee. At the same time, cases seem to occur where our only ground for discounting the agent's words *ispostfactum.*

In all this it should be noted that the strain is not between reason and either passion or short-term goals, but between seriously declared intent and intentional action. Of course, this strain may also occur where the seriously declared goal is some long-term objective on which reason has worked, and the reason decided on some relatively short-term one; but that is incidental. Indeed, although the traditional examples have been couched in these terms, it seems clear that what is at issue is how serious the agent can be taken to be about reasons which are not the ones on which they eventually act. In Aristotle's terminology: how far can use be supposed? Often apparent opponents in fact agree that there is no space between the case which is obviously ruled out, and various ones which create no difficulty.

Passionate akrasia *and Irrationality*

If the above is correct, then one ground for the feeling that akratics are irrational has nothing to do with acting against reason or better judgment,

but is rather related to the fear that the proposed akratic is failing, in circumstances where the condition should obtain, to satisfy the condition on rational agency of knowing what they are about. It would, however, be wrong to suggest that there was nothing in the traditional feeling that failure of reasoning was somehow an important element in the irrationality of *akrasia*. This emerges most evidently in the traditional examples of passionate *akrasia*. One characteristic of a rational agent is to be able not only to articulate goals, but also to formulate and pursue fairly long-term ones, and subordinate other goals to their achievement. As the Stoics seem to have felt, it is not easy to see how they could acquire a full mastery of language without such abilities. There is a presumption, therefore, that they will go in for bits of thinking which might with various degrees of justice be called deliberation, and that these will result in the adoption of certain reasons for action. If a person always acted spontaneously, without thought reacting to their immediate apprehension of their situation, then the proposed description would again be problematic. At the very least it would be a description of a very immature agent, and we should have qualms about holding them responsible at anything above a very primitive level. If we consider the Aristotelian account of the passions, it seems clear that they are, on that account, tendencies to behave unreflectively on interpreting one's situation in a certain way. It is the form of motivation characteristic of children, of those *en pathei*. This is the motivation of the passionately akratic, aggravated by the fact that for Aristotle desire does not even start from some view of what is good for such and such a category of person to do. It is not a form of motivation whose manifestation is a prime example of rational agency. Further, on the akratic occasions, it succeeds in ousting deliberation as the effective producer of action. Since its general effectiveness would remove someone's claim to rational agency, even its occasional effectiveness will not only be an example of rational agency not being manifested, but also will seem to pose a threat to rationality: it shows that the agent cannot be relied on to perform in accordance with deliberation. Of course, in these cases it will also be the case that the agent fails to act in accordance with an apparently sincerely adopted reason; but that, as we have seen, is a different point.

Overall Rationality

It might be felt that there is yet another way in which the akratic are failing to display rationality. In taking someone to be a person, we are

committed to interpreting their behaviour in the light of their desires and beliefs. On any particular occasion our taking this desire or belief as explanatory relies on the attribution to them of others. There is an assumption of a range of desires and beliefs, and also of consistency on the agent's part. Thus, if a man helps an old lady across the road, we might interpret this as an act of kindness. This attributes to him a desire to help the old lady, and a belief that helping her across the road is what she wants. This is probably done by transferring our beliefs about others to him. If we discover, however, that he has a deep distaste for old ladies, and thinks that they should be exterminated, then his behaviour seems to display inconsistency, and we should be inclined to look around for some other explanation consistent with his known preferences. Perhaps she has not yet changed her will in a desired direction, and so is worth preserving for a time and treating in a way which might persuade her to the change. The inconsistency in both believing that old ladies should be exterminated and that they are worth pleasing for their own sake is thus removed. There is no inconsistency in thinking both that old ladies should be exterminated and that their money is worth having. It seems, then, that interpreting behaviour as action requires an option in favour of maximum consistency; and this is consistency over their beliefs and desires as a whole. There are various ways of accommodating apparent inconsistencies so as to preserve the general characterization of rational agent, but the akratic seems a particularly flagrant example of inconsistency. It is not as though they had temporarily forgotten some aim to which they normally give importance, or some belief which would bear on the present case, or anything of that sort. On the contrary, they are supposed to act in full knowledge that what they are doing is inconsistent with those desires and beliefs which we rely on to give consistency to their lives and justify our interpretation of their behaviour as rational action. Their behaviour is a direct affront to their rational status. Akratic behaviour, therefore, has to be seen as an anomaly, a wart on the face of rationality, something that could only become widespread at the cost of obscuring the face altogether.

While there is some attraction to this view, it has, I think, to be treated with caution. As I have expounded it it seems to suggest that our title to the description 'rational agent' relies on a fairly close approximation to an agent with a set of consistent beliefs and objectives consistently pursued, and this seems to me far too strong. On the other hand, it does seem that we do, in the way illustrated, tend to search for consistency if possible.

A minimum truth seems to be this: to consider some entity as a person we have to be able to attribute to it (or at least to developed members of its species) a range of capacities which include such things as ability to speak, to want to associate with others, to pursue and plan for long-term objectives, to learn a variety of facts and skills, to remember in a variety of ways, and so on. To get purchase for these attributions we need a number of fair spans during which the subject's behaviour can be displayed as manifesting these capacities: they are none of them capacities which can be manifested in isolation in a moment. Once we have grounds for attributing a number, of course, we can take an isolated occurrence as manifesting one—but then we are taking for granted what in its turn could not be displayed in a moment If we are to attribute an interest in getting into Parliament, we shall need a fair amount of behaviour which can be seen as consistent with, if not required by, that objective, given various beliefs. To attribute the beliefs we shall require a good deal of behaviour that can be displayed as showing interest and success in acquiring knowledge, i.e. is at the very least consistent with this combination of interest and ability, much of it required. So we certainly need stretches of consistency, and no doubt also some stability of capacities. If we had from moment to moment to interpret behaviour in terms of changed desires and beliefs, we might as well give up the pretence that we are dealing with a rational agent or person at all.

On the other hand, this seems a far cry from requiring any overall consistency. It is doubtless true that for there to be a person there has to be a full range of beliefs and desires, and also that there is no knowing which may become relevant to the interpretation of any particular piece of behaviour; this is a good deal less than claiming that all are relevant and their consistency important. Indeed, there are going to be limits on consistency. In the example given, the initial interpretation was questioned in the light of information about typical earlier behaviour and expressions of view. It was inconsistency with these that led to the behaviour being re-interpreted. It is, however, important to note that the concept of a person requires some change and development on the part of the being so designated. Suppose we have an entity to which are attributed at a given time a range of beliefs, aspirations, skills and memories. Twenty years later we examine this entity to find not a single alteration: it is in all respects in the same state as twenty years earlier. It is not clear that such lack of change shows that at last we have irrefutable identity of person. On the contrary, it raises considerable doubt as to whether we have a person at all. Perhaps we can save the

supposition by hypothesizing coma on a dramatic scale; but short of that kind of abnormal condition the signs are that we simply do not have a (normal) person. Further, the kind of abnormal condition required is one which is not just unusual, but involves postulating a suspension of the normal condition of a person, i.e. of a certain range of changes and developments. Consequently, while some stability is required, with a tendency to consistency over periods, too much would be too much. The kinds of change required by the attribution of learning, maturation, interest in their environment and such like, will put limits on how strongly we can require consistency. Even in the example given, while the earlier history arouses our suspicion, it might turn out that this is the turning over of a new leaf, or even a flash in the pan. While there are limits to the frequency of change that can be tolerated, they are pretty extreme, and some change seems necessary. It is, of course, inconvenient if people change frequently, as it makes interpretation of their actions more doubtful. On the other hand, it would be too strong a view which made action out of character irrational, because inconsistent with previous behaviour. We have to allow for people changing not only beliefs, but attitudes and values as well, in ways which leave later beliefs inconsistent with earlier ones. This is not irrational or undesirable, but very much what we should expect, not to say want, from a rational agent.

This, of course, might be accepted, and the consistency requirement be modified to the claim that any inconsistencies must be serial. What becomes troublesome is much concurrent inconsistency. I am not quite sure what to make of this, partly because of the possible ranges of concurrence. It certainly seems possible for people, for quite considerable periods, to be torn by incompatible ideals for themselves, as with the man vacillating between dreams of academe and dreams of wealth (see Chapter VIII, p. 111). Once again, while such circumstances make a person less than perfectly predictable, they hardly seem to threaten their claim to be rational agents, nor, as they vacillate in their behaviour, are they clearly irrational, though their condition may well be uncomfortable.

In the light of all this, it does not seem that the akratic's inconsistency with other, perhaps more commonly operative, aims, is going to support a charge of irrationality any more than it will with either vacillating or out-of-character behaviour. As with the last, persistent *akrasia* may put in doubt the possession of the other aims, but even here one needs to be careful. If someone claimed to be a sincere Christian and claimed

weakness on every single opportunity for the practice of Christian virtue, then we might well start doubting the sincerity of their Christianity (but one has to be careful with ideals which are self-proclaimedly inachievable). It is not so clear that if they were consistently weak in one particular area, that would throw doubt on their sincerity about the standard to which they failed to adhere. In fact, what is special about *akrasia* is, once more, the fact that at the time of the action the akratic is unwilling to disavow the reason which would count against doing what they do, and that explicitly. But the inconsistency here is between one sincere declaration of value and the acceptance of a reason for acting which is incompatible with that value. That is a return to the problem as viewed earlier in the chapter.

Conclusion

Davidson was right to diagnose the (or, rather, a) root of Socratic worries as lying not in the area of ethics but in that of the analysis of rational action. Agreeing that akratic action is in some way a failure of practical reason, he has tried to offer an analysis which allows for the possibility of *akrasia* while explaining how it is irrational. This depends on an account of all-things-considered or better judgments and relevant reasons. I have argued that that account fails, and no such account looks likely to succeed. We seem left with *akrasia* being perfectly possible, but in no obvious sense irrational.

In the present chapter I have offered a different account of what underlies Socrates* worry, which depends on supplementing a plausible account of purposive behaviour with special features belonging to a language-speaking rational agent. In the light of this I have suggested one explanation of the feeling that akratic behaviour is irrational which does not rely on any notion of all-things-considered or better judgment. I have also suggested why one might feel that there is something to the view that akratic behaviour involves a contrast between reason and some other consideration. This has nothing to do with reason delivering all-things-considered judgments, but simply with an expectation that a rational being will tend to pursue long-term objectives which require some use of reason and deliberation, and so will be expected in general to be able to carry out the actions decided on by deliberation. In other words, there is an expectation that many of the reasons which an agent takes as reasons for acting will be reached as a result of reasoning or

lead to a fair amount of reasoning. In so far as akratic behaviour flouts this expectation it seems to threaten the supposition of rationality. Finally, I have tried to show that general considerations of what is required for interpreting behaviour as that of a person or rational agent do not support anything stronger.

CHAPTER X

Passionate *Akrasia*

The Position So Far

So far, since Chapter VIII, I have been treating the question of *akrasia* as one about the coherence of claiming that an agent can deliberately choose what they consider to be the worse course, without any reference to passions. Obviously, someone who so acted under the influence of fear might be an example, but the assumption has been that the set of cases is far wider than the traditional one. The passionate examples are, nevertheless, worth attention. To begin with, someone puzzled by the sorts of case discussed in the last two chapters might feel them to be intelligible because they always or often require the assumption of the influence of some emotion or passion. In particular, one might feel that passion has its part to play in explaining the feeling that the akratic are irrational, since there is a widespread feeling that there is some opposition between passion or emotion on the one hand and reason on the other. Yet as we have seen in the historical sections, this line of thought gives rise to two related problems: first, how one can reconcile the talk of a conflict between reason and emotion with the description of the behaviour as deliberate; and secondly, how such talk is to be reconciled with belief in a single agent responsible for both sorts of behaviour.

In this chapter I shall first discuss the apparent conflict between reason and emotion as it might seem to arise in the traditional examples. This will lead on to a discussion of ways of being out of control, which in turn will lead to a discussion of what it is to control one's emotions and other states. The concentration will still be on traditional examples, but will serve to loosen up one's ideas of the relation of reason to emotion and to sharpen one's view of the sort of internal division required by talk of self-control, or incompatible with attribution of

deliberate choice. Even so, the traditional examples are curiously limited, and I shall move on to discuss briefly the variety of emotions. The question of whether the puzzle cases of Chapters VIII and IX require appeal to emotions will be delayed to Chapter XI. What will, I hope, emerge, is that there are special problems about passionate *akrasia,* and that there are special contributions which these cases make to the idea that the akratic are irrational.

Reason and Emotion

There is a common feeling that there is an opposition between reason and emotion, but consideration of examples suggests that this is more complex than it might at first seem. The most obvious example that leaps to mind to support the idea of opposition is that of intense emotions such as rage, panic, lust and such desires as extreme hunger or thirst. In such cases the agent seems incapable of reasoning altogether—though even so the position is hardly that of Chrysippus' irrational motions, in that the agent's behaviour is commonly adapted to the situation and their interpretation of it; so while one's panic reaction, for instance, may be a condition in which one cannot sit down and calculate the best way to deal with the situation, it remains a manifestation of some intelligence on the agent's part. The opposition to reason shows not in the opposition to manifestation of intelligence, but in the difficulty created for reason to operate in selecting the best way of achieving what is taken as a desirable goal.

This contrasts with a case where a woman, say, is very frightened that her estranged husband is plotting to abduct their children. So far from interfering with her ability to reason about means, her fear may sharpen her wits to an unusual degree of acuteness. What it may well interfere with is her ability to wonder about whether she is getting things out of proportion. She cannot be got seriously to consider questions about other values: all her thinking is concentrated on imagining possible dangers and working out how to combat them.

It is not clear how the first kind of example fits into Chrysippus' scheme, but the second seems a clear case of the agent reasoning in a fearful way. If her reason has been overwhelmed, the analogy is not with someone who has been kidnapped and carried off against their will, but rather with someone who has been bludgeoned into agreement by a more forceful personality. Of course, between these extremes we find other cases. Thus

the woman might be able, but only with difficulty, to give her mind to considerations other than her children's safety; or at times, when thinking out a clever system of barricades against her husband, she may experience feelings of rising panic which make it hard for her to think clearly at all.

In these cases one is tempted to talk of struggle, but it is not clear that even here the analogy with the kidnapped agent carried away by force is a very happy one. It is not simply that taken seriously it might make difficulties for attributions of responsibility. After all, one might hold that while on any given occasion the agent is seriously overpowered, still, we can do something about getting into the position of being overpowered, and so can still be blamed. It is rather that the analogy suggests a lucid, unwilling victim, carried away by passion; whereas the examples suggest as much someone struggling not to be transformed into an agent who only thinks of her children's safety, or one who can no longer think sensibly at all. This suggests a different tie between *akrasia* and irrationality from those considered in the last chapter. In so far as the akratic are considered to be acting from passion, then in one way or another their responsiveness to discussion is lessened, and so they are acting in a way neither under the control of deliberation, nor likely to respond to it Yet their behaviour is an, in some way, intelligent response to an appreciation of the situation they are in; and in this respect it is unlike losing one's balance, say. Sometimes the situation is that of Aristotle's impetuous akratic, who is presumably distinguished from the merely spontaneous agent just by some degree of potential unresponsiveness to discussion; sometimes, as with Aristotle's weak akratic, the agent is actually unresponsive to discussion and ungiven to deliberation; sometimes, like neither, they are very capable of deliberation, but not open to discussion of the starting-points of deliberation. In this last case the sense of irrationality is less strong.

In cases of passionate *akrasia,* the presumption that is threatened, so that an air of irrationality is given, is that rational beings should be capable of and open to discussion. The person who fails to act rationally in this sense may nevertheless act perfectly rationally in the sense that they may act in a way very appropriate either to achieving what they are aiming at, or achieving what they would anyway have decided on calmly had they not been under the influence of passion (see the example on p. 149 of this chapter). The point is not that they act contrary to reason, but that they act not under Us control or even potentially responsive to its control. In addition, of course, there may be conflict, and passion may lead an agent to do what at the time they realize is not what they would decide to do if they were deciding calmly. Given a

general prejudice (not necessarily in each case a good prejudice) in favour of calmly decided action, such passionate behaviour might be held to be against one's better judgment. This would be a different sense of 'better judgment' from those considered in the last chapter. It is the judgment one would reach on reflection.

In the previous chapter the akratic was felt to be irrational either because of failing to carry out their declared intention, or because of failing to have any overall plans. The first introduced conflict with the assumption that rational agents know what they are about, the second with the assumption that rational agents can and do carry out deliberated projects. Passion can, of course, enter into such cases. The point being made here, however, is that it introduces suspicion of irrationality without conflicting with the conclusions of reason. The lack of control is not a matter of reason's conclusions being ignored, but of an agent not operating as a (potential) calculator.

Lack of Control

This still leaves a contrast between being in a reasonable and being in an unreasonable state, and this contrast can suggest internal conflict between two factors struggling for control. This is precisely what the Stoics found objectionable, and what might anyway seem puzzling.

It seems fairly clear that there are cases where we consider that we lack control, which are not in the required way puzzling. If we take Chrysippus' example of someone running downhill, where their continued running is out of their control, one is inclined to agree with him that this is unproblematic. Not that it is uninteresting in the context of the relation of mind and body and our control over our behaviour; nor, consequently, in the context of the question of the unity of the mind/ body/person; but since the continued running is not held to be intentional, the kind of unity of the agent required for the attribution of intentional behaviour is not under threat. More interesting, not least because it does not get considered in literature on *akrasia,* and yet is not very far from Chrysippus' example, is the case of sleepiness. Suppose someone is organizing a conference and attending the evening session, given by a very distinguished speaker in a very hot room. It would be most offensive to drop off to sleep, but a, heavy day of administration and a stuffy atmosphere combine to make the eyelids droop. The subject is seized with a strong desire to sleep, but from politeness struggles desperately to stay awake.

At first sight it is not clear why this should be thought different from Chrysippus' example of the runner, except that in that case the agent is supposed to be unable to pull up—but then the story could be adapted to make control possible. It would still be for some reason unnatural to speak of a desire to keep running, though it is not clear what the reason is. It may be that in the case of sleepiness we tend to think of the body as asserting its needs/wants, so that sleeping presents itself as something to pursue and not just to submit to. It is as pure a case of physical desire as one is likely to meet. What removes the pains of hunger seems to be something we have to find out; what we have to find out here is that the movement of the eyelids is something that we can resist. Yet for all that this is a case where we can give in to temptation, or be overcome; where we can resist and control ourselves; it does not seem to give rise to disturbing division in the agent, or any air of paradox.

At least part of the reason for this can be seen if we turn to a different case, such as panic or sudden uncontrollable rages. If we take these descriptions seriously we usually hold the agent not to be responsible. Nevertheless, they can seem worrying. The reason seems to be that in both cases we have emotions whose manifestations are also manifestations of assessments of the situation by the agent. If I panic when I knock down a pedestrian while driving, and put my foot on the accelerator instead of stopping, that shows my view of the importance of my reputation and/or of the horrors of prison or public exposure in court; if I let fly at some unfortunate subordinate who takes it on himself to give some permission that it is my privilege to give, I am showing my appreciation of the dignity of my position. These sorts of view, whose betrayal is characteristic of such emotional outbursts, seem indicative of my personality, in a way in which drooping eyelids are not Sleepiness is not on any given occasion an indicator of character. Consequently, while emotional outbursts can look like a change of personality, sleepiness does not. It is this that makes the influence of strong emotion already somewhat disturbing, even when there is no evident conflict involved. Both for the agent and the onlooker it can raise a puzzle about what sort of person the agent is, and how open to 'reasonable' relations with others. It can be as though a different personality had taken over.

It is this relationship of emotions of various kinds to the personality of the agent which makes the traditional examples of *akrasia* traditionally disturbing. For in the traditional examples we are presented, apparently, with two aspects of the personality at odds with one another. It is not just that the agent seems torn between two

objectives, like someone who cannot decide whether to spend a present on a book or a bottle of wine: rather, they seem to be being two sorts of person at once. In the case of the book or the wine it is a matter of calmly deciding which calmly to pursue. The agent is engaged in a piece of pre-practical-reasoning practical reasoning. In the traditional examples they are hovering between being calm deciders and acting on what seems a different decision-making system altogether. Usually this is portrayed as a conflicting desire, but as we saw earlier (see Chapter I, p. 15) these emotional examples may not involve conflict of desire in the sense of conflicting objectives, and the trouble may come from the agent's finding that they retain at one level a belief that conflicts with what they acknowledge at another to be the evidence.

The talk of levels, here, of course, already suggests the operation of partially independent systems, and it is this that causes the trouble. With sleepiness one can hope to get away with talk of the body's demands, to which the agent gives way; in the present case the choice looks like being between talking of the passions' demands to which the agent does or does not give way, or talking of the agent giving or refusing to give way to the agent—which sounds bizarre. Yet the first way of talking is only unproblematic if we can take the passions in question as not themselves an expression of the person and one of their characteristic mechanisms of decision. If we could do that we could perhaps treat these cases like sleepiness, and then resist the Stoic moves. After all, while we cannot guarantee to control our bodies, the case of sleepiness shows that we sometimes can, and can be held responsible. The same might be true with emotions, in which case there would be no need to make them manifestations of reason: all we should need is to say that sometimes emotions, like sleepiness, are beyond our power to control, sometimes within our power; in the latter cases we can be held responsible for acts which flow from giving in to them.

What gives the Stoic position plausibility is not just that actions done from passion are chosen, but that our passions are manifestations of our personalities: it is not just the actions, but the attitudes, belief and motivation that we cannot disavow. What the Stoic position tends to blur over is that many of what we call passions cannot just be passed off as bodily reactions to presentations, but are the operation of a system of belief and choice which is not automatically sensitive to that of reflection. Our emotions cannot be written off as not really us, nor do most people wish for their extinction; but our emotional tendencies are tendencies to non-reflective behaviour which bring home to us that we

are not just reflective agents. It seems that however clear our grasp at a rational level, we sometimes find ourselves acting on a different understanding of the situation. A Stoic might now claim that that fact showed that our 'rational* understanding was not as clear as we had claimed; but this seems to trivialize the position that knowledge will dissipate the emotions, by making it a condition of 'true' knowledge that such dissipation occurs; alternatively it makes the substantial claim that a form of understanding is available which will have this effect.

Yet all this talk of different levels is verging on dualism of agency while trying to avoid the consequences of dualism. What leads to the desire to do the latter is not a concern for responsibility, as the example of sleepiness shows; rather the point is that whereas when I give way to sleepiness I am not acting from an intention, when I give way to anger I do act from the intention, say, to take my subordinate down. As the Stoics might put it, my anger gives the form of my decision-making. At the same time this can suggest an apparently unrealistic picture of our powers of control.

Control

This last point might become clearer if we consider some examples, which will also, I hope, loosen up our ideas of control. In the case of sleepiness control is partly a matter of using the muscles of the eyelids to prevent the eyes from closing, but partly a matter of doing things such as moving about, making and drinking cups of coffee, and generally doing things which will make falling asleep less likely. The first of these is the sort of measure which can give purchase to the notion of an act of will, as I exert myself against the natural inclination of the lids. There is no particular difficulty about thinking reflectively except in so far as actual sleep will rule it out There is no tendency to be thinking in a different way altogether.

If we turn to the example of fear, there is, of course, some analogy. I may breathe deeply to relax my tension; deliberately refrain from twisting my handkerchief; make myself do something which takes my mind off what I am afraid of. There are also, however, factors for which there is no analogue in the case of sleepiness. Suppose we have an ambitious woman who feels her position threatened by the admission of some silly mistake. She is frightened at the prospect of some colleague discovering and revealing it. Her fear might abate, of course, if she

realized that her colleague had in fact not discovered, or would not have the courage to reveal what he or she knew. More interestingly, it might lessen if she came to consider either that the revelation would be less damaging than she had originally thought, or that the damage, though accurately assessed, was being given more importance than it deserved in the context of her ambitions. There are also two further ways in which her fear might be brought under control. First, she might come to think that her fear comes not from the threat to her ambitions, but from an insecurity which, while it might fuel her ambitions, leads to an inhibiting and quite unreasonable perfectionism. Secondly, she might come to rethink her ambitions and become attracted to a different set of values in life; and in so far as she accepts them it ceases to be alarming that her previous ambitions are threatened. In neither of these two cases, of course, is there necessarily an automatic dispelling of fear; but in so far as she can make the new thoughts her own, they will erode the power of the old ones, which underlie her fear. There is no analogue to these forms of struggle for control in the case of sleepiness. Further, what is going on is an attempt at self-understanding or self-transformation.

It is these features, which differentiate fear, say, or anger from sleepiness, which also lend colour to the view that the passions, or some of them, are either expressions of, or responsive to, reason. Certainly many such feelings are in various ways reactions to supposed information, and their influence waxes or wanes according in part to what is believed. It would be an oversimplification, however, to try, without more ado, to characterize all the above examples in terms of the influence of belief. For this can suggest, though it does not necessitate, the view that these emotions are just applications of a different learning system from our language-based one—with its articulated marshalling of evidence—even if the operation of the latter might influence the former. This would be a picture influenced by the Stoic view of us starting out as animals with an animal capacity for learning, which is expanded and influenced by the use of reason. At least three considerations might, however, give one pause.

First, if we return to the woman above, one possibility is that her fear is not straightforwardly based on an assessment of the threat to her career, and would not be simply responsive to reasoned argument to the effect that revelation of her stupidity would not harm it. What underlies her reaction might be the nightmare of contempt. The recognition of this might well help lessen her fear, but it would not be new information about the supposed threat, but an understanding of herself and just what

it was that she found threatening. The fear of contempt is one that a person often prefers not to acknowledge, thereby increasing its influence. The very disproportionate violence of the fear can then itself be alarming. However that may be, her fear is not just based on belief.

Secondly, her fear might be increased or lessened, not by information, but by stories, or even infection. Thus someone recounting and laughing at the stupidity of a third person might intensify her horror of discovery, but not by adding new information relative to her own case. Or a friend confiding a similar deeply felt fear might thereby fuel hers. Alternatively, stories of how someone else, by a combination of courage and cunning, concealed their shortcomings, might, with no added basis in fact, give her heart. In no obvious sense is this the operation of learning, however primitive. Indeed, so far from being obviously primitive, it would seem that a rich vocabulary might be helpful in developing fears in these ways. Language can extend the power of imagination as well as that of learning.

Thirdly, she might change her view of what is important. This is not acquiring new information, but altering her way of assessing the significance of information. Nor is it easy to say how this is achieved. It is not just by feeding in more information. No doubt, in some sense it involves thought and the use of reason, as she questions the value of her life-style and indulges in imaginative portrayals of alternatives. Of course, the outcome might well be described in terms of belief about what is important in life. This is a belief, however, which is not in any clear way based on evidence, and the influences on it do not follow the learning pattern of the simple model.

As has been said, most of these examples can be spoken of in terms of belief. The point is that different forms of belief enter in in a variety of ways; they vary in the ways in which they influence the 'emotion', and in the ways in which they are themselves influenced. In some cases, as, for instance, where imagination or contagion come into play, other factors than belief are influential. Such a picture lessens the attraction of the view that a full understanding will bring emotional control in its train, though it leaves intact the view that typically information will affect emotional reaction.

It will doubtless be noted that in the above example what the woman feared would not normally be considered to be anything dangerous. Some might want to deny the description 'fear' to such a case, I cannot myself see the point of such legislation. The woman will certainly suffer if her stupidity is revealed, and she has an unreflective reaction to avoid

that at all costs. There is, however, an interesting difference between this and what might be considered the primitive case of fear as a reaction to threatened physical pain: no amount of reflection on the unimportance of pain will lessen the torment of a kick on the shins; but in the woman's case it is her view of what is important that makes revelation hard to bear, so that a change there will, at least in the end, remove the suffering from what was originally feared.

This simply brings home that even what we might be tempted to think of as 'one' emotion, fear, covers an interesting variety of reactions with different relations to 'reason'. When we venture into a wider range of passions or emotions, the variety extends. Thus affection can be dented by new information, e.g. about lack of reciprocity, or depravity; but there is not the same expected tie to information as with fear, and affection is very vulnerable to apparently irrelevant factors such as peer contempt (for further discussion of some of these differences see Gosling 1965, 1969).

I shall return later to the variety of phenomena brought under such headings as 'emotion' or 'passion'. For the moment it is enough to glimpse it and recognize that it is implausible to suggest a single picture of how best to exercise control. One final point on control is perhaps worth making. In the above, no distinction has been made between control exercised on particular occasions and control of one's propensity to emotion. If I am afraid of dogs I might, on any occasion of meeting one, take a variety of steps. Some of these may be 'physical', such as deep breathing, or making myself walk slowly past, crooning 'nice doggie'; others might be 'psychological', such as reminding myself that this dog has never yet bitten anyone, and will in fact run away if I bark back. Or I might phantasize about myself as the brave tracker coolly ignoring the howling wolves. None of this directly attacks my propensity to fear, but is geared to controlling the situation here and now. To deal with the propensity, I might resolve on a practice of accustoming myself to dogs and learning about them; or I might reflect on the suffering as imagined with the suffering as probable in reality in the hope of getting it into proportion. While in this case there is little room for a general re-ordering of my priorities, in others, such as that of the woman considered earlier, reflection on one's values, with a resultant change, might gradually effect a diminution of fear. All this would be controlling the emotion in the long term. When someone is called self-controlled, this might refer to their ability to face down occurrent emotions, or to their having attitudes which result in their not

being subject to violent onsets of emotion. While different means tend to be appropriate to the different cases, there is also considerable overlap, as we have seen. What should emerge from consideration of control is that indeed some complexity of the agent is supposed by our ways of describing internal struggle, and more than is allowed for on a Stoic or Socratic view. It is not, however, a duality of agent. Rather, there are conditions of the subject when it is, or is on the way to becoming, a non-agent.

The Variety of Emotions

So far most of the examples I have taken have been those of fear or anger. Traditional examples have also included sexual desire and either hunger or greed, and thirst or love of alcohol (which is what being given to drink seems to amount to). All of these might count, in Aristotelian terminology, as *pat he*, as counting among the things that come over us rather than among those that we do—even though they may lead to doings. They all have it in common, first, that they are directional: they direct the subject towards some objective, the avoidance of some danger, exacting of vengeance, or the satisfaction of some desire; secondly, that they all display a gradation, from forms which leave the subject open to a normal range of considerations, to forms where the objective of the *pathos* dominates the subject's thoughts, or where the subject becomes out of control and ceases to be responsible. It is these two characteristics which make them obvious examples in discussions of *akrasia*, for the first enables them to supply the subject with a reason for choosing a course, and the second gives rise to all the talk of struggle and the suspicion of irrationality. To start with, therefore, it might be worth noting two points: first, not all of these would naturally be called either passions or emotions, and secondly, not all reactions which would naturally receive these titles show these characteristics. There is no obvious reason to be disturbed by these facts. One could, after all, coin a term to classify just those reactions which have the characteristics necessary for the situations which raise Socratic puzzles. As a first move, however, that might prove pre-emptive. I propose, therefore, to take the two points in turn in order to get a broader view of the territory. I hope it will emerge that the route of rapid definition would have led to a restrictive view.

Someone experiencing fear, anger or sexual desire might naturally be said to be experiencing emotion, and 'passion' would not be too strained

a term, either—though it sounds slightly odd with fear. But neither 'emotion' nor 'passion' seems at all a natural term to use for hunger or thirst; and while a glutton or a dypsomaniac might be said to have a passion for food or drink, this seems to come from the use of 'passion for' to refer to love. 'Emotion' would sound distinctly odd. Part of the reason for this might be that hunger and thirst are normally felt physical discomforts, at whose removal the related desire is aimed, whereas anger and fear typically take their start in beliefs, which evidence might change. Sexual desire gets classed as an emotion in so far as it is part of sexual attraction: the case is less clear with masturbation. Yet however close the examples might be for giving rise to akratic puzzles, they will diverge as regards what can plausibly be said about control. Improvements in self-understanding, which commonly bring control in the cases of fear and anger, are quite ineffectual in assuaging thirst. But that suggests that in the former cases lack of control may rest on unreconciled divergences of view, or various forms of rational failure which would suggest possibilities as regards irrationality not available in the other cases.

On the other hand, many reactions which might be dubbed emotions fail to show the common characteristics. Thus some are turbulent, but non-directional. A child told that it is going to be taken to a fair may get very excited. Their excitement is not a motive to any particular directed behaviour, but rather results in a failure to settle to any directed behaviour. In this case, of course, there is delightful action in prospect which cannot be immediately indulged, but this is not a necessary feature. Someone who has just seen their favourite football team win the FA Cup may experience a similar need to let off steam. Yet for all the failure of direction, which probably accounts for an unwillingness to talk of chosen or even intentional behaviour motivated by it, the situation has much in common with the standard cases of *akrasia.* Certainly, an excited person may find that they realize that it would be better not to be displaying excitement, but have difficulty in controlling themselves similar to that of an angry person. Nor is it just a difficulty of physical control: they find it hard to get their minds running on anything but what they are excited about, if that even. So there are non-directional emotions, which seem to raise difficulties of control. At the other end, there are directional emotions which do not seem to lend themselves to uncontrolled extremes. Someone who feels grateful is oriented to behaviour pleasing to their benefactor, and so their gratitude supplies a reason-explanation for their intentional behaviour. To be overwhelmed with gratitude, however, is not, as with fear, to be out of

one's mind with it, but to consider the gift or service so great as hardly to know what would be adequate acknowledgment of it. As gratitude grows it does not tend to remove the subject from a state where they can respond to discussion. It may nevertheless sometimes account for a person's doing what they realize they had better not.

In addition, there seem to be emotions which are neither directional nor turbulent. Nostalgia, for instance, or awe felt when contemplating the stars at night, are not emotions which yield explanations of intentional behaviour; but nor are they ones which are characterized by turbulence. Nostalgia, however, can certainly have an enervating effect, and in this it is similar to depression, which does not always get allowed as an emotion. Nostalgia probably qualifies because it is a state in which a person may be liable to tears and maudlin sentiment—liable to get emotional. Depression tends to induce a dulling of feelings. Yet it can yield akratic-style situations. It is common for a person in a state of depression to realize that they ought, say, to be filling in their tax return, but 'simply cannot bring themselves to get down to anything'. Because depression is not, like fear or anger, a condition in which the subject has a desire to achieve something, the imagery tends not to be that of control. Instead, people talk of having to drag themselves out of their depression, force themselves into action, and so on. But just as in extremes of anger a person is by over-excitement incapacitated so far as rational discourse is concerned, so in the extremes of depression a person may get so withdrawn from active interests as to get beyond the reach of influence by 'reason'. A deepening depression produces a drift away from rational agency. What is true of depression is true of apathy, which few would intuitively classify as an emotion.

Reflection on this range of possibilities suggests two further examples: (i) one of the features of akratic cases is the agent's absorption with some aim, such as revenge or escape. Now suppose I have become intrigued by some logical problem. I am sure that there is a proof, but as I turn it over and over in my mind I find to my frustration that I keep meeting a dead-end. I cannot keep my mind off it. I know, as I sit at lunch, that I ought to attend to my companions, but the problem keeps coming back, and I get abstracted. I am not by any ordinary account influenced by any emotion, but I am showing a failure to do what I think best, and the explanation in many ways parallels that of the emotional cases, (ii) Suppose we have a man who is very keen on committee procedure. Two things might make us say that he is emotional about it: first, of course, that he gets very excited in his

insistence on everyone following correct procedure. But even if he shows no particular sign of excitement, he may come up with all sorts of explanations of why it is important not to brush them aside, which are patently unconvincing or out of proportion to the occasion ('Maybe it would do no obvious harm in this case, but it is the sort of practice which led to the Nazi take-over in Germany', at a meeting of the local vegetable-growers association). One common connotation of the word 'emotional' is to suggest a condition where a person shows some break-down in their operation of rational processes.

Unity of the Agent Again

As remarked earlier, the worry about the unity of the agent arises from the imagery of struggle, which suggests opposing forces. One of these forces is reason, the other some passion. Now in itself this does not give rise to problems unless one holds an ultra-Socratic thesis that no strength of emotion could ever render one unable to think calmly. Most people think there are cases of extreme emotion where the person is no longer responsible for their actions. But then such actions are not, either, intentional, but like those of someone thoroughly drunk: intention may come in earlier in explaining how they came to be in that state, but not at the stage of drunken action. The presumption with akratic action is that that stage has not been reached, so that the akratic can be said intentionally to take the akratic course. But if they have decided for the akratic course, where is the struggle? A review of the examples given above suggests that it is a red herring to portray the struggle as between the dictates of reason and the push of passion. The struggle is to retain or gain a state where one is open to consideration of what to do, and able to pursue and act on the outcome—in other words, to remain reasonable. The picture is not that of a person running downhill and unable to stop themselves: they, after all, may remain perfectly rational, but try as they will cannot work out a way to stop. By contrast, agents who are akratic in one way or another find it difficult to be open to reason, and those in the more extreme condition of being out of control have lost that capacity.

This makes it tempting to talk of different, semi-independent systems of desire and belief, which normally observe a given hierarchical structure, but not in the akratic case. This is one attraction of Davidson's analogy with a state, where there are constitutional mechanisms for setting up rules of behaviour and officials which and who are normally

recognized as having authority. Sometimes, however, sectional interests may cause official action contrary to those rules. Some sub-group, with its own structure of purposes and beliefs, has managed to influence the decision procedure, by-passing the constitutional canons of decision; or it may even by-pass official decision procedure, facing the state with a *fait accompli*. The first is an analogue of *akrasia*: its being official action is analogous to its being intentional, while the pressure ensured that the decision-procedure was unconstitutional, which is analogous to the agent not being open to reason, but deciding in accordance with other desires and ways of assessing the evidence.

The problem with such talk is to make out just what it is intended to explain and how it does so. For present purposes it has to explain how talk of internal strife can be reconciled with saying that whichever side wins the result will be intentional action by the same agent. This is achieved in the present analogy by the fact that action is taken through the official channels but not operating the official canons of decision. The struggle is between the pressure group and any who attempt to get the proper canons attended to. The trouble is that in the model the agent is made up of groups of *homines*. When we transfer the analogy to the individual *homo*, the *homines* become *homunculi*. If they are to be of help, the *homunculi* must not have the complexity of structure of the *homo*. We must not be able to re-raise with regard to the 'sub-group' the problem with which we started with the original agent, as to how/why it opts for one consideration over others as the one to act on.

There are two obvious ways of avoiding this. One is to point out that the traditional akratic, at least, has to be described as using two lines of reasoning, and perhaps, sometimes, two ways of assessing the facts; these two do not cohere into a single way of reaching a decision, and the agent is attracted to each. On the other hand the agent considers one set of considerations as the one that deserves respect, and would normally be expected to act on them. So we have two reason systems, and the normal one is ousted by the other. But in this case the talk of systems looks like being no more than a dimly picturesque way of describing the situation we started with. It throws no light on the talk of struggle, or how it is to be reconciled with talk of intentional action by the one agent.

The second way is to treat the systems as *sub-homunculi* whose interaction explains the choices and behaviour of the *homo*. In normal circumstances the *homo* acts in accordance with a decision system responsive to a deliberative system; other systems of attraction and

revulsion are responsive to the deliberative system and only affect the decision system through the filtering of 'reason'. Sometimes, however, the sub-system by-passes the deliberative system to affect the decision-system directly. This is what happens in *akrasia.* Since the normal operation is subject to the outcome of reflection, we may describe normal behaviour as in accordance with reason, and akralic behaviour as caused, though done for a reason. This latter language preserves our sense that something has gone wrong, while nevertheless the action is intentional.

This seems to be elaborate paraphernalia explaining nothing. What we start from is the apparent fact that agents having, after reflection, reached a conclusion as to what to do, then do something else, and that deliberately for a reason that on reflection they had rejected. In other words, the eventual decision was made in accordance with some reason/way of assessing the facts other than the reflective one, and yet the agent made the decision knowingly for that reason. The proposed explanation simply states this situation over again, leaving it as puzzling as in the original why the conclusion reached is overturned for a reason previously rejected and without any reason for overturning the earlier rejection.

It may, of course, be that the talk of systems is intended not to explain, but to supplant, in a terminology which no longer has puzzles, the talk of deliberation, reasons and decisions. There just are different action-producing mechanisms; but one, the 'rational' one, either usually, or perhaps also desirably, produces actions and controls the influence of other mechanisms, but in the akratic case does not Given that in rational animals the 'rational' system usually predominates, and this is what makes them count as rational, the akratic case has to be the exception on pain of the agent ceasing to be a rational one. Given that we want to associate with rational animals, we might get a little disturbed by threats of descent into irrationality. What happens in the akratic case, however, is quite straightforward: a cause operates which does not usually operate in these animals.

The puzzle now is the relationship between this explanation of what goes on to the common talk of deliberate choice and intention and related understanding of agents. That would take us beyond the scope of this book, which is concerned with an apparent problem internal to the language of intention and explanation by emotion. What is relevant is that it seems to throw no light on how accounts of agents choosing are consistent with saying that they acted against their better judgment

Conclusions

It is time to draw some of these tangled strands together. A review of traditional treatments of *akrasia* suggests that they have been influenced by a number of factors. A tendency to think in terms of a possible conflict between something and possibly morality, certainly reason, has led to concentrating on cases of temptation where there is a conflict of desire. The most natural examples for this are emotions like fear and anger, and desires like those for food, drink and sex. These yield dramatic cases of conflict, and it is tempting to lump them together. It is then easy to think of fear and anger as setting the pattern of emotions or passions, and of these as being a set characteristically opposed to reason. This has at least two effects: first, one tends only to go to the examples of emotions which are sufficiently like the two models and this makes one ignore interesting differences; and secondly, one tends to think only of what one might naturally call emotions or passions. This last has the result of encouraging one to ignore interesting cases. The word 'passion', in English, suggests love or anger, not really fear, and certainly not sorrow or pity. 'Emotion' may suggest a wider range, but still, as its etymology might lead one to expect, they are cases where we think of a person being moved. So all the examples that come to mind are likely to be ones which show a certain degree of excitation. They direct the mind away from depression and apathy, although these last, as we have seen, give rise to situations of a generally akratic structure. They differ, however, from traditional examples in that the traditional examples suggest the operation of a system of motivation and learning independent of any informed by reason, whereas these cases suggest an absence of motivation making difficulty for any interested thinking. The examples of the obsessive logician and the committee addict suggest two further points: first, that trawling what would naturally be called emotions and their near relatives is trawling too narrow an area; and secondly, that a further way in which this whole set of cases introduces irrationality *intoakrasia* is through the attendant malfunction of rational procedures.

Consideration of these examples makes it possible to be a little clearer about the senses in which the passionate akratic ('passionate' now being clearly a term of poor art) is irrational. To begin with, it is perhaps worth repeating that it is not necessary to envisage any opposition between the motive on which the akratic acts and the achievement of what reason recommends. Suppose a man sees his child

fall into a swift-running river. He remembers being told that in these circumstances the chances of saving a person are greater if you can stay on land and throw a life-belt or reach out with a stick. He tries to control his rising panic, but despite a rapid, but inadequate look round, takes off his shoes and jacket and leaps in. He has taken what he knows not to be the best course; but if it is a conflict of desires, then desires are to be differentiated by something other than what they are for: his objective is simply his child's safety. If we wish to speak of two desires, we should have to incorporate into the distinction of the desires the different systems of assessment of the situation that are in operation. Standing back from theory for a moment, we might be inclined to put the blame on conflict of belief. He seems torn between believing that his child's safety will be better achieved by calm consideration of the situation, and believing that immediate action is required; or possibly between believing that it would be best secured by looking around for a stick or a life-belt and believing that the best way is for him to jump into the water. If he is acting against his better judgment, it is not that he is wanting something other than what his better judgment is aimed at, but that he is in some way failing to believe what his better judgment declares true. Problems about the seriousness of the declarations of belief parallel those about the seriousness of declarations of intent discussed in the last two chapters. But any irrationality cannot come from going counter to the objective that his better judgment puts before him, but from a failure to have his decision guided by the preferred 'rational' mechanism. The operative mode of assessment is called irrational by contrast. It may for all that be, on any occasion, sound: our man might save his child, and it might have been the only way, but his behaviour would still have been irrational in the sense of not reason-guided—that is, he was not in a state where he was capable of or open to discussion, understanding full well what he was doing. He was not, as Aristotle might say, using his rational faculties.

In this case the agent is reacting to the situation in a sense rationally, but with a different mode of assessment from the preferred 'rational' one, and so is operating irrationally. In the example where I am distracted by my intriguing problem, there is no similar contrast between 'rational' and 'emotional': what I am engaged in is a highly sophisticated reasoning operation, but so intriguing as to distract me from the pursuit of other goals. Yet in so far as control is a matter of carrying through deliberated objectives, I am not manifesting control. Nor is it true that I change my mind about what to do: I keep finding

myself back on the problem. The situation may not be like a person with rising panic, but it is somewhat like the woman discussed earlier, whose fear of having her children abducted sharpened her wits, but left her finding it difficult to concentrate long on questions other than how to thwart her husband.

The example of the man over-keen on committee procedure highlights another way in which emotion tends to be associated with irrationality. We do not necessarily, just because we are under the influence of emotion, lose our command of language. We are, however, liable to get a little wild in our use of it and the reasoning capacities which go with it. Thus the woman fearing her children's abduction will tend to notice remarks which could by some wild stretch of the imagination be construed as showing complicity with her husband, and use her imagination so to construe them. She is liable to indulge in wild distortions of evidence, and be struck by words or actions which could relate to her husband, which would never strike her in other contexts, or not in such a way. What makes this behaviour irrational is not that it fails to subserve her proclaimed end, nor just that it is guided by some non-deliberative mechanism, but that it is purporting to be using reason while flouting rational canons.

One common use of 'irrational' is to declare an action not to measure up to what is required by some reason for action. It is common that the akratic's action is in this sense irrational, in that there is usually some reason, sometimes on some ground canonized as Reason, by reference to which the action chosen is irrational. Commonly, the passionate akratic's action will be in this sense irrational, and the reason to which it fails to conform be the reason(s) supplied by deliberation, or the standard they generally recognize as the one they want to rule their lives by. In this sense, of course, the action recommended by reason would be irrational by reference to the reason on which the akratic acts. I have been trying to show that in the case of passionate *akrasia* a different notion of irrationality is also operative. The sense here attaches not to the action assessed in relation to some reason for action, but to the mode of decision-making, contrasting it with a 'normal' deliberative one. 'Mode of decision-making' has to be taken a little generously, because in cases such as apathy that is just what is lacking. The agent is still failing to operate according to 'normal' rational procedures, but not in this case ruled by non-rational procedures. In more standardly emotional cases, moreover, the passionate akratic is liable to show irrationality in an actual abuse of standard reasoning procedures.

What is disturbing about passionate *akrasia,* both for the agent and the onlooker, is the apparent withdrawal of the agent from a condition in which they can enter reliably into interpersonal dealings. At the extremes, this disturbance is well based: very widespread dominance of these forms of motivation would be a move towards ceasing to operate as a person at all. The apparent centrality of control, however, should perhaps be treated with caution. Fear of emotional display may not in all cases be rational.

CHAPTER XI

Moral Weakness

Résumé

In Chapters VIII and IX I considered supposed problems which arise for *akrasia,* where this is taken simply as the agent taking what they consider to be the worse course. 'Worse' here does not carry any moral connotations. It is sufficient that the agent has some considerations to which in the context they give greater weight, but which nevertheless they knowingly ignore in the particular case, and for some reason choose a course other than that indicated by these considerations. This has every appearance of irrational behaviour, but I there argued that it is fruitless to try to explain that notion of irrationality by reference to all-things-considered judgments, even with the help of some principle of continence enjoining on us to do what all things considered we think best. The air of irrationality does not come from the fact that somehow reason is, or is predominantly, on one side of the dispute, and choice goes against it. Rather it is that the agent seems decided in favour of one side of the dispute, but mysteriously chooses the other. All the traditional examples, in terms of a conflict between reason and some more limited motivation, get their bite from the fact that it is assumed that the longer-term, 'rational', considerations are in some way taken by the agent as the considerations on which to choose. But the same bite recurs where there is no contrast between wider and more limited judgments. The puzzlement about *akrasia* arises from the contrast between apparently sincere declaration of intent and actual performance. This is not to deny that longer-term pursuits have an importance. An agent's general ability to indulge in them and carry them through seems a requirement for their counting as rational agents—but that is a different point.

In the last chapter I argued that special problems arise in relation to emotional *akrasia.* Of course, one finds there the conflict between

declared intent and actual performance, but in addition there are features which suggest disturbance of the operation of the agent's deliberative faculties. It is these cases which give purchase to the picture of dual agency, whether at the time, if we consider the imagery of struggle, or serially, if we follow that of personality change under emotional influence. What seems to underly the Stoic insistence on the unity of the agent is the following: treating someone as a responsible agent is supposing them either (i) to be open to the influence of discussion or (ii) to be in a normal state at least to the extent that their emotions and so on can in the longer term be modified by changes of attitude and belief which can be brought about without recourse to clinical treatment Any refusal to count a dipsomaniac responsible for their behaviour is based on a view that, their approach to drink is not influenceable in normal ways by considerations of ambition, family welfare or whatever. The behaviour is certainly well directed, and indulged for a reason, but the motivation operates out of reach of the influence of other reasons for action. A certain cohesion and power of mutual effect is assumed among the beliefs, desires and attitudes of a rational agent. If we have a well-established case of multiple personality, there is a more general break in this cohesion, in that we seem to have two distinct sets of independently operating character traits at distinct periods of the 'individual's' life. The kind of unity of the agent required is that condition where in general deliberation is possible and possibly effective, either immediately, or eventually through the influence of reflection on the 'emotions'. This is compatible with various sorts of partitioning of the self.

Prescriptivism

Passionate *akrasia* was discussed because I think that while it displays the structure of the more general case of choosing to take the worse course, it has features of interest peculiar to it. Similarly, while that more general structure is exemplified in the special case of moral weakness, there are features of this which deserve separate discussion. Although the discussion of moral weakness has received impetus in this century from R.M.Hare's prescriptivism (see Hare 1952: p. 169; 1963: Chapter 5), I do not propose to discuss that doctrine in detail. Briefly, prescriptivism is the doctrine that certain terms have both a descriptive and prescriptive element in their meaning; and among these some, such as 'good' and 'ought', are primarily prescriptive. There are secondary

uses of these expressions where the prescriptive element is moribund, but if someone sincerely asserts a non-conditional sentence such as 'It is good to take a holiday from time to time', or 'One ought not to irritate one's elders', using the words 'good' and 'ought' in their primary senses, then they are committing themselves to prescriptions in favour of or against holidays or irritating one's elders. If the sentence bears on some actions immediately doable by the agent and appropriate to do, then the agent will proceed to act, unless prevented. This seems to give rise to the old Socratic paradox, but as Hare points out, since there are many gradations from using the terms with their full primary meaning, it is possible to accommodate a range of cases where the agent in some sense does something other than what they think they ought

This view obviously involves theses about meaning, and ones which are very complex. For present purposes, however, it seems not to matter whether or not Hare is right. Suppose he is. Then someone who asserts 'One ought to pay one's debts' with its full primary meaning, and sees that it applies to a current situation where they are being asked to repay, must on Hare's thesis be taking the obligation of debt repayment as their reason for deciding what to do. 'I ought to pay this debt now' becomes in effect the expression of a decision to pay it now, and failure to do so would give us an example of just the puzzlement discussed in Chapters VIII and IX. Suppose he is wrong. It remains that whether or not such cases are examples of using such expressions in their primary sense, we do on occasion use expressions of the form 'It would be best to x' or 'I ought to y' to declare our decision in favour of x or y. It is such occasions that give rise to the puzzlement discussed in Chapters VIII and IX. It seems to me, therefore, that the shape of one's disquiet about *akrasia* retains the same form whether or not the prescriptivist thesis is right. (For an extended discussion of prescriptivism, see Dunn 1987.)

Purposivism

Matters are somewhat similar, I think, with Charlton's purposivism (Charlton 1988), which he puts forward to take the place of prescriptivism or emotivism. The basic idea is to see thinking it good to do x as taking the doing of x as an objective in a fairly weak sense which need involve doing no more than indulging in some exploratory thinking about doing x. To think that it would be better to do x than not to do it, however, is to take the doing of x as an object of pursuit, the

not doing it as an object of aversion; and so the deliberate refusal to do x becomes, to put it mildly, problematic. It becomes problematic precisely because the agent is being supposed to take the doing of x as their objective in doing what they do. Thinking that it would, all things considered, be better to do x, or that it would be pleasanter not to do x, both fall short of thinking that it would be better to do x than not to do it, and so no problem arises here. It is therefore perfectly possible for us to think one course, on the whole, better, but deliberately choose another; but impossible for us to choose a course which we think it would be better not to take. (See Charlton 1988: pp. 87–91.)

Clearly the impossibility of *akrasia* comes from the fact of the connection between thinking it better to do x than not and taking x as one's objective. Pseudo-akratic cases of acting contrary to what one thinks it, all things considered, better to do can be allowed for, but there does not seem to be any room for what Hare allows: not quite genuine or full-blooded uses. Charlton can, indeed, allow for apparently sincere declarations that something is best; but the test of whether the agent really thinks that best seems just to be whether, if they act intentionally, they act in accordance with the judgment. If purposivism is right, therefore, the problem for *akrasia* arises where I have claimed that it should: it is a question of how close a subject can come to thinking it better to do x than not to do it, so as to seem to think it, without actually taking x as their objective. There will be the allied question of whether an agent can think that they think it better to do x than not, while being mistaken. If purposivism is wrong, it remains that sometimes people use declarations that it is better to do x than not, to signal that they are decided in favour of doing x rather than not. The structure of the problem remains the same, therefore, whether or not purposivism is true.

Morality

Although it may not matter whether or not either of these, or various other, views on the analysis of 'good* or of judgments that it is better to do x than not, is right, there do seem to be certain questions which arise about moral weakness which have not been automatically covered in discussing the possibility of deliberately taking the worse course, or the special points about passionate *akrasia.*

There are many senses of the word 'morality', and in many of them there is no problem at all about an agent thinking that something is morally

wrong, and proceeding to do it. It may, of course, be that one could show that in some sense of 'moral' it is rational to be moral. In that case we should doubtless have a sense of 'irrational* in which the akratic is irrational. Whether that is a puzzling situation or not will depend on the case made for rationality. One might view moral rules, for instance, something as follows: it is useful to have agreed standards of conduct, without having to resort either to the use of force, or to rely on constant reference to the variable moods and tastes of individuals; our eschewing of force will mean that we have to find standards which other reasonable people can see as having a chance of success with other reasonable people, and so on; and doubtless underlying this will be some notion of an optimal outcome from the acceptance of the standards, in some sense of 'optimal'. Such standards could be said to have a rational basis partly in that they are supposedly the result of acceptance by unforced, reasonable people, and partly because of some games-theory like notion of rationality based on the conception of optimal outcome. It is fairly obvious that this is too weak to create any problem for *akrasia.* A person may simply have no interest in being rational in that sense; may indeed think it weak-kneed folly. There are stronger versions, which attempt to show that whatever a person's interests, so long as they are of the longer-term sort needed for them to count as rational agents at all, then it can be shown to be in that person's interests to be moral, and so rational for them to be. These simply lack conviction when they pass beyond the point of claiming that it is for the most part in their interest to behave in moral ways. In some uses, however, the term 'moral* gets restricted so that a given view or code is only an agent's moral one if it is accepted by them as a guide in the conduct and ordering of their lies. This creates a presumption that when a situation engages a moral view or principle, that is what the agent takes as the guide to how to act in that situation. In any such sense of 'moral', therefore, a problem will arise if one tries to say that the agent considered something morally wrong, but decided to do it. If they thought it was *morally* wrong, then they must have accepted reasons for guiding their lives which they acknowledged ruled such action out. But if they in fact did it, does that not show that they did not accept those reasons as their guide?

Obviously one has to be careful here. It might be true that a person has something as a major goal in life, but allow themselves the occasional relapse; or someone might, say, have teetotalism as an objective, but think that it can only be achieved by gradualism: that an attempt to be immediately teetotal would put impossible strains on the

system. In such cases failure to live up to the ideal does not impugn the genuineness of the ideal. But then one has to suppose, what can obviously be done in the second case at least, that the ideal itself is not viewed as requiring on each apparently relevant occasion behaviour which conforms to it. The point is, however, that some such special explanation is needed if an agent is to be held to be sincere in holding a moral view which applies to the present case, while deliberately doing something which fails to conform to that ideal.

In senses of 'moral' which carry this condition, action becomes the test of sincerity of belief. At the same time, as in the cases discussed in Chapters VIII and IX, we seem to get cases where we are unwilling to say that the agent is insincere, or has abandoned their moral view, or changed their view of the situation, but nevertheless knowingly chooses to flout the principle. There can be similar rifts between practice and declared intent.

In these cases it is normal that a view only counts as moral if it constitutes some ideal of the sort of person one is trying to be, or some code of conduct which requires one to have or not have certain desires. In other words, it is usual that the desire to be moral is a second-order desire. In this case, to be second order is not to be a desire the satisfaction of which will require some ordering of the pursuit of sub-objectives: after all, desires of fairly low complexity have this characteristic. To satisfy my desire to have a certain woman come out to dinner with me, I might have to pursue the objective of going to the bank, give that priority over my desire for tea, suppress my usual preference for shabby day clothes, etc., etc. In so far as the pursuit of the main objective requires the selection and ordering of other pursuits, it could, of course, be said to be, on any operative occasion, of a higher order relative to them. This is a matter, however, of their being operative. Put another desire in the driving seat, and they might reverse their relative order. Some desires, however, are desires at least in part to have or not have other desires. These need not be anything which we should naturally call moral. For instance, I might admire someone for their refinement of taste and wish that I were the sort of person who appreciated caviare and claret, instead of having my vulgar preference for beer and pickles. This sets up a disgust, or even shame, for the tastes which I nevertheless indulge. My desire to be a gourmet leads to a constant disapproval of my actual eating habits. This is not just a function of being operative or not. In the case considered earlier (Chapter VIII pp. 111–12) of the man caught between two self-images,

each consists partly of the desire not to have the desires characteristic of the other. Many views which would normally be allowed as moral also have this feature. What are called moral views typically carry the implication that it would be better to be the sort of person who did not have, at least in any strength, certain desires. This can result in shame in having, let alone indulging, them.

In the case of my admiration of refinement, I may recognize that I am unable to tell claret from burgundy, and do not much care for either; I may even recognize that I shall never change, whether or not I give up the struggle; but none of this suffices to dissipate my shame that I have the coarse tastes I do. Something similar seems possible in moral cases: that I may persist in disapproving of my actions as I fail to be the sort of person I aspire to be. This enables us to get clear about what is felt to be a common and puzzling feature of *akrasia*: that the agent even at the time of the act deplores what they are doing. In ordinary cases of conflict of desires the agent may, of course, regret, more or less wistfully, having to forego the satisfaction of some other desire, but that seems less dramatic than the overall self-disapproval found in cases of moral weakness. Yet how can that disapproval be sincere, when I obviously persist in my wrong-doing?

We can at least say here that the agent fails to make the moral principle or ideal that which determines their choice in this situation. That was not, therefore, the reason for their action. Yet there still seems to be a distinction between this case and one where all we have is that the pursuit of some objective leads to leaving some desire or desires unsatisfied. For in the present sort of case the agent seems to be passing unfavourable judgment on themselves as an agent. Yet how can someone be set against themselves in this way?

It has sometimes been felt that the distinction between first—and second-order desires is what helps here, and, indeed, helps one see how moral weakness is possible without the puzzlement discussed in Chapters VIII and IX. A first-order desire is directed to some objective to which some action here and now is relevant. A second-order desire is directed towards being a certain sort of person with or without certain desires, or, at least, such as not to indulge them. Second order desires only lead to actions which are themselves aimed at reforming or consolidating character, and have a programming role. They are not directly engaged in the market place of day-to-day life. They may lead to practices which will either lessen the strength of the undesirable desires, or inculcate desires which will combat them. When, however, I

find myself wanting to lie to conceal my stupidity, then while my second-order desire may be affronted by the realization of the sort of person I am, it is not in direct conflict. Wanting to be a truth-loving person is not having a love of the truth, though it may lead to the acquisition of such a love. The latter is what is needed to generate conflict on any particular occasion.

It is, of course, true that people are given to desires and aspirations of this general sort, and that commonly what are called moral views embody some ideal of the sort of person that the agent wishes to be. This in turn gives some sense to this judgment being the agent's better judgment: it is their judgment as to how it would be better for them to be. Consequently failure may well be accompanied by self-disapproval, self-disgust or remorse. It is not at all clear, however, that the relevant feature of such views creates any presumption that a rational animal holding them will act on them, still less that failure so to act is irrational. 'Better judgment' carries no connotation of rationality, though it does, of course, suggest that the judgment in question puts the agent's proposed action into a wider 'all things considered' context. To this extent it might be felt that a better judgment involves necessarily the exercise of rational faculties in considering the pros and cons of having or indulging other desires. Any resultant sense in which a person acting against such a judgment could be said to be irrational, however, would be profoundly untroubling.

The move to second-order desires, then, does not seem to throw light on why, if, we feel the akratic to be irrational. But nor does it help explain how the desires do not really conflict, so that both can be, so to speak, alive and kicking at the time of action in any way that is not equally possible in the case of first-order desires. I may want to go to sleep and feel very thirsty; whichever I indulge, the other is undimmed. If I am of an ascetic turn and decide to resist both, I shall not thereby extinguish either. Similarly, my desire to be the sort of person who has refined tastes can co-exist with a desire for beefburgers and beer. On the other hand, if I decide to take a drink, it is my desire to quench my thirst, not my desire to sleep, that explains my action; and similarly, if I go for beefburgers, it is that desire, not the desire for refinement, which explains my behaviour. It is not clear how a second-order desire can in some way be more serious, while remaining inoperative, than a first-order one.

The illusion might come from the fact that it is not clear how one goes about satisfying a desire to be a person of a certain sort, so that

typically such a desire does not engage directly with action. Two points might be made here: first, this situation is not unique to such second-order desires—any desire might on occasion fail so to engage; secondly, it is not generally true of second-order desires. I might well believe that the only way to become a person of refined taste is never to indulge my vulgar ones, and to get into the habit of eating gourmet menus. In that case we would get direct conflict. Failure to act on the second-order desire would then be just as puzzling or unpuzzling as in any other case.

It seems, then, that the distinction between first—and second-order desires does nothing to explain the possibility of *akrasia*, nor to throw special light on its irrationality. What it does do is explain the special air of self-condemnation which seems to characterize moral weakness. That we should be able to act in the face of such condemnation only becomes puzzling, however, in so far as we are trying to attribute some seriousness greater than lip-service, but less than acceptance as a reason for acting on this occasion, to the agent's second-order desire. We are back in Chapters VIII–IX.

Moral Weakness and the Unity of the Agent

While second-order desires may not help with the problems of Chapters VIII and IX, they do bear on the points made in Chapter X about the unity of the agent. I claimed there that worries about the unity of the agent arose in connection with passionate *akrasia* because so many especially of the traditional examples of passions tended to be indicative of the personality of the agent In so far as moral views embody a desire to be a certain sort of personality, they commonly embody a desire to be a different sort from that evidenced by certain of their emotions. In some cases, perhaps commonly, for instance, in relation to jealousy, the desire may be to be the sort of person who does not have that emotion at all. In others, as, say, with affection, the desire may be merely to be the sort of person who is not quite so prone to it, or so ready in its expression. Just because my moral view involves a desire to be a certain sort of person, I can get the feeling of being two sorts of person at once: the one shown in my ideals, and the one displayed depressingly in my emotional reaction. Even if I refrain from acting on the undesirable emotion, I still find myself thinking along the undesirable lines, and so feel I should like to disown myself.

Self-control and Integration

Whether we think of morality as a code for the regulation of social behaviour, or as an ideal of the fully human life, there will be some tension between morality and passion, at least to the extent of there being moral reasons for having passion under control. As we have seen in Chapter X, passions have relations not only to judgments, but also to general views on what is important and good in life and to imaginative portrayals of possibilities. On the other hand, there is the potential for a gap between the heart and the head, between what we consider in our reflective moments to be valuable and what is emotionally important to us. It therefore becomes a matter of interest to the moralist how adequate control is to be achieved. Since many passions are reactions based on views of what is important and such like, there has been a hope that there might be some view of what is worthwhile in life such that if it is sincerely held and fully grasped, then all the subject's passions become reactions based on the same overall view, which is also that held on reflection. In this case there would be full integration not only of one's passions with one another, but also with one's moral view. Lack of control only arises from lack of integration. In theory, given unlimited plasticity of human nature, there might be an indefinite number of overall views which would variously suffice to produce integration. There remains the possibility that there are certain desires and aspirations which, at a suitable level of description, remain inextinguishable, or at least only suppressible at a cost elsewhere. Thus it might be that desire for affection can only be denied all satisfaction at a cost of an increased tendency to resentment and hostile behaviour, and that this will hold, whoever is in question. In that case a moralist would not be able to avoid an investigation of human nature; and some moral views, for instance one which advocated complete independence of affection, would necessarily fail to achieve integration and be left with a permanent risk of *akrasia.*

This, of course, is supposing that the earlier comments on control of emotion are correct. There is a view that all passions short of mania are subject to control by the will. Consistent exercise of will-power is all that is needed to bring emotions under control. As in other areas, practice makes perfect. I shall return to this possibility in Chapter XII. If the earlier views on control are correct, one can understand the Aristotelian hope that correct views on life will produce an emotional integration which removes any liability to passionate *akrasia.*

Conclusion

The outcome of this discussion is that while there are special points of interest about moral weakness, it is not in any special way problematic, nor does there seem to be any special and interesting sense in which it is irrational. On the other hand, moral views of the sort we have been considering will be liable to give rise to conflict and the risk of passionate *akrasia.* The desire not to have or indulge certain passions or desires sets the stage for conflict whenever such passions or desires are aroused; whereas other desires will only conflict with them by chance on occasion. This feature comes from moralities embodying the sorts of second-order desire that have been considered. It is not confined, of course, to moralities. Many ambitions which on most definitions would hardly count as moral (an ambition to succeed where Hitler failed, for instance) will nevertheless involve a desire not to have other desires, which might be seen as weaknesses. It is nevertheless a feature of such moralities and may help explain why *akrasia* is often felt to be an especially moral problem. Similarly (but again, not exclusively), moralities tend to generate self-condemnation not only for actions, but for the liability to various passions. This gives a special edge to the experience of moral conflict as distinct, say, from the struggle to control one's temper in order not to make a rash move in a chess game.

As I have said, neither of these features is peculiar to morality, but it is not difficult to see how they might be thought to be so. If we think that what makes a being a moral agent is their capacity to use their reason on the question of how best to conduct their lives and to act on those views, it is perhaps natural to attribute a rational desire for the good, which may indeed get distorted so as to attach to such improper objects as the emulation of Hitler, but whose proper object is the proper organization of the agent's life. In this way the will might emerge as a special moral faculty. Moral weakness then becomes of special interest, because the functioning of the will is a necessary condition of a subject's being a rational agent, and so moral weakness threatens the agent's functioning as such an agent and so, perhaps, as the sort of being it supposedly is. That, again, is a matter for Chapter XII.

CHAPTER XII

Willing, Trying and Wanting

The Will

The treatments of weakness of will so far considered have tended to have what might be called a strong rationalist air. *Akrasia* is a matter of something going wrong with our reasoning, or else comes from the fact that the agent does not really believe that the action taken is not the best. Even when we drop the suggestion that the akratic is always acting against what is in some sense a more considered (and so rational?) judgment, and see the case as one of acting contrary to an apparently sincerely expressed intention, the impression is given of a less than model rational agent. Will plays no part in the discussion. The nearest we get to it is in talk of relative strength of desires. Yet why should one not recognize that the akratic is not irrational, but simply, as the English suggests, weak? But weak not in intellect or apprehension, but in will. The akratic know full well that what they are proposing to do is take the worse course, but take it because too weak of will to do anything else.

Strength of Will

While this may have an air of moving us from the excessive intellectualism of many philosophical treatments, and returning us to good common sense, it is not without its problems. After all, if the agent was too weak of will to resist temptation, and a stronger will was needed to pull off the better act, why should they be blamed? If I am physically too weak to lift a weight, it is not my fault if I fail; so why does the same not hold if I am too weak of will, suffering, as it were, from debility of spiritual muscle? How can I have deliberately taken the worse course if I was too weak to take any other? And if I *could* always have taken the

other, in what sense was my will too weak? One could, of course, accept that at the time my will was too weak, but claim that I am to be held responsible because that is a condition which I can improve. I am weak of will because of lack of character training. Just as, if I went in for regular physical training, I should be able to lift weights that are at present beyond me, so, if I went in for training of the will, I should get better at resisting temptation. But to this two comments are in order. First, it does nothing towards showing how my weak acts can be said to be deliberate; but unless that can be done, it looks as though the language of weakness is in a similar case to that of being overcome by passion: it is not so much explanatory as in need of explanation. Secondly, the picture of will-exercise loses its attraction once one has reflected a little on what makes it plausible.

The plausibility of talking of training the will comes from some familiar possibilities. If I am frightened of the dark, I might attempt to overcome it by going out deliberately after dusk, or making myself go up to bed without turning the lights on. I might succeed in largely or even wholly overcoming my fear, and have doubtless shown great strength of will in tackling the problem. But what has happened? The muscular imagery of exercise might suggest that a toning up of the will on fear would leave it in good shape for coping with one's greed, one's sexual drives or one's inclination to linger in bed on a cold morning. Life is not so easy. This success might, indeed, convince me that I do not have to accept my other weaknesses as inevitable; but while there may be analogies, coping with one emotion or desire is not the same as coping with another. Further, it is not clear that I have been strengthening my will. After coping with my fear of the dark I am not thereby any better off when it comes to controlling my irritation. Again, ordinarily speaking, I needed strength of will, determination, to make myself go out alone on the first night, and probably the first few; but as I learn to cope, less girding of the loins is required. My control might have come in one or more of the various ways suggested in Chapter X (pp. 138–42), combined with persistent practice of night walking. It is not clear that I am training or exercising something called the will, unless, of course, doing all these things is just what constitutes exercise of will.

Philosophical views of the will seem to have arisen in something like the following way. In development from Socrates and Plato, Aristotle held that human beings are by nature rational, and so have a desire to fulfil their nature by the rational pursuit of their welfare. But they also have desires which do not manifest a concern for their good. The

actualizations of these desires impede each other; and Aristotle's concern is to explain how, while knowledge of what is best is still present, actualization of desire for pleasure can impede the full actualization of rational desire. If we ask whether, and, if so, how, the akratic could at the time have overcome their desire for pleasure, Aristotle is silent. We do, however, have a rational desire for good in potential opposition to other desires.

The Stoics introduced a new element, insisting that an important characteristic of rational animals is the ability to assent or not to what any desire, rational or otherwise, presents as fitting to pursue. In Christian thought this was emphasized as the freedom to choose; but this freedom attached to the rational desire for good, which thus became a power to choose between goods, controlling other desires. The will thus becomes a rational power, and as its object is the good it is a moral power, always, in akratic cases, able to have its way over other inclinations, typically seen as relatively short term.

This development makes it easy, though not necessary, to think of the will as a special moral power to enforce moral judgment against the vagaries of passion and desire; and it is a short step from that to thinking of it as something to be trained by exercise. In fact, however, no such picture is necessitated by the view. What we have is the following: (i) a desire for the good, always present in a normal person; (ii) an ability, in a normal person in a normal condition, to assent to the conclusions of reason or not to assent; (iii) an ability, in a normal person in a normal condition, to take steps which will diminish the influence of any contrary desire or passion—what these steps are will, as we have seen, vary from case to case. If we now ask where talk of strength of will comes in, the answer is not simple. In general, a person of strong will is one who persists in their chosen course in the face of certain sorts of difficulty, in particular difficulties which are viewed as burdensome. I may persist in my attempts to solve a problem which bristles with difficulties, but, if I find the challenge stimulating, do not thereby show strength of will, nor do I need to make any special effort of will. The position would change if I found the problem boring or frustrating. My persistence must also be self-motivated: I do not show strength of will by persisting under threat

So far, strength of will consists in persisting, of my own free will, in doing something burdensome, and taking whatever steps are appropriate to ensure persistence. It involves a willingness to take a variety of steps according to circumstances and the sort of difficulty involved.

Sometimes a special effort of will will be required to keep one's eyes open, for instance, or make oneself go out after dark. What does such an effort consist in? Usually one's eyes are open during daylight hours, and we have to think to close them; on this occasion one has to think to keep them open, and exercise muscles which we do not usually have to exercise during the day, as well as think to do other things which will make it less likely that the eyes will droop unexpectedly. In the case of fear of the dark I have to spot, and counter, a tendency to think of reasons for not going out now, and put out of my mind thoughts of Jack the Ripper or bats; finally, I have to be prepared to go out, trembling, into the night. These are specific things I have to be prepared to do of my own free will, and the doing of them in these circumstances constitutes an effort of will. While we have to suppose that I have the ability to take these courses at will, there seems no need, indeed no room, for some special strengthened power, brought into operation on each occasion.

While the view of will that we have been considering does not seem committed to any ghostly muscle ripe for exercise, it does postulate a special desire for the good, usually allied in some way to the special exercise of reason about how to run one's life. Quite apart from whether there is such a desire, at least in all people, it seems unnecessary for cases of effort of will. Similarly, it is unnecessary for generating the puzzles about *akrasia*. So if we wished to claim that weakness of will was to be explained not in terms of some failure of belief, but as weakness of *will*, what would the point amount to? It is not, apparently, like suffering from some muscular defect. It seems to amount to the claim that akratics fail, when they could if they chose, to do something which is unpleasant or boring or…They have a motive, but fail to persist. Not only do they have a motive, but they seem set to act on it, and perhaps even claim it as their reason for doing something, but then do something else. We are back in the situation of a lack of tally between declared intent and action.

Trying

It might be felt that this still leaves something out: the element of effort or trying, which is so crucial to strength of will. That is what we fail to do when we are weak: try. Appeal to trying might seem to help if we can establish a number of plausible theses about trying. First, it seems plausible to claim that if I successfully poison my rival, then I have

made a successful attempt on his or her life. So my having succeeded is not incompatible with my having made an attempt, or tried. Since we are usually more interested in what people have succeeded in doing, if someone only mentions that they tried, it is common to infer that they did not succeed, on the assumption that if they had, they would have mentioned it. The assumption is important, however, and would not be made in all contexts. In any case, it does not affect the fact that the person did try. If we ask what my trying consisted in, it seems to have consisted in, say, looking for an opportunity to put a powder in his wine and then putting it in. The putting it in is something I did which can be described in terms of its effects as poisoning. But then what I did was stretch out my hand and unclench it, with the result that the poison dropped into the glass. My successful dropping of the poison into the glass was a successful attempt, which consisted in these actions. But now it looks as though we should say that it was this which constituted my attempt to poison this man: both 'dropping the powder in the glass' and 'poisoning this man' are ways of describing what I did, my attempt, in terms of its effects. If we pursue this line of thought we shall see that our actions are all of them bodily doings or refrainings which are tryings and, often, described in terms of effects outside the body. So what we commonly think of as actions (or at least the largest and most interesting set) are tryings which have at least some physical effects.

On the other hand, not all tryings are actions, for we may sometimes try to move an arm, say, and totally fail, that is, there may be no physical effects. Thus in the case of Landry's patient (see James 1890 [1950]: pp. 490–2), when he was unhindered he moved his arm, but when held down totally failed, although he thought he had succeeded. To him it was the same as when he succeeded; but when he succeeded he tried. Therefore he tried when he did not succeed. Similarly, in cases of sudden paralysis a person may try to move a limb and there may be no contraction of muscles or movement of the body. While every action is an event of trying, and every trying that is an action precedes and causes a contraction of muscles and movement of the body, there are tryings which are without effect and so are not actions. (For arguments for this sort of view see Hornsby 1980: pp. 33–60.)

It might now be tempting to develop this sort of view. What we are interested in in *akrasia* is intentional action, and trying is integral to all intentional action. What a weak person does is fail to inhibit or produce the bodily movement as required for satisfying their intention: the trying is ineffective. But trying admits of degrees, that is, it is possible to try

more or less hard. So weakness is a matter not of something going wrong in practical reasoning, but of not trying hard enough.

Response on Trying

The first point to examine about this position is the view that events of trying occur even in the paralytic cases. For it is this which could give rise to the possibility of claiming that this is the location of the will, and that strength of will is a matter of such trying being hard enough to be effective.

Now the whole argument is conducted from within our common ways of speaking. These first of all convince us that all tryings are events. Consequently, since we should admit that the paralytic tried to move their arm, we have to admit that an event of trying occurred in that case too. Yet the fact that the event in question is indiscernible should give one pause. It is not a good idea to postulate events to accommodate one's speaking habits. Since it anyway seems rather puzzling just what the paralytic's trying consists in in the proposed example, it might be worth asking just what the concession that they tried amounts to. Why are they said to try?

In the unproblematic cases people are said to try in virtue of doing something intentionally. In successful cases, this may be what they succeed in doing; in cases of failure it will be something else which, while not bringing success, was done in the hope of success. Often the imperative 'try" is used to encourage people to do things in the hope of producing the effect mentioned in what they are encouraged to try doing ('Try to persuade your mother to come'). But it is worth noting that the imperative can be used to encourage experiment ('Try pressing the red button and see what happens') and discovery as to whether one can do something ('Try saying "ah"', 'Try to move your arm'). In the latter cases what we are told to try to do is commonly something which we can normally do without any problem: we know perfectly well what we are being asked to do, and we know we can usually do it. If this were not the case there would be problems. Suppose you tell me to try raising all the hairs on my head on end. I have no idea how that is to be done. So I do not know what would count as trying—though I know what would count as succeeding. It is not just that I cannot do it I cannot, perhaps, sing a high G, but I know the sort of thing to do to approximate. Not so with the hair. But with the saying 'ah' or raising my arm, I am being encouraged to do something I 'know' I can do, so I make no bones about responding to the request and am surprised that on this occasion I fail to do anything. What explains the

use of 'try* is that context of normal expectation and intention. There is no call to postulate any mysterious event of trying. When I realize that I am paralysed, and accept my incapacity, I cease to expect to be able to respond to the request to move my arm, and shall only be irritated by being exhorted to try: there is nothing that counts as trying. Of course, when I find that I can move it again I shall be willing to respond to such exhortations again, and my trying will be my success, complete or partial. But when I try, and fail completely, my attempt to discover whether or not I can raise my arm simply leads to the conclusion that I can do nothing; that whereas usually I can raise it, now *nothing* happens—not even an event of trying.

As one might expect, there is not in these cases either, anything that counts as trying harder, such that if that happened one might pull off the intended effect. I can, no doubt, try again, i.e. see if this time things have returned to normal so that I can raise my arm on request. But there is only either doing it or failing to do it. In general, trying harder is something one can do only in cases where there is something one can do differently and in a way more likely to succeed. If I am trying to solve a problem, I can try harder by exploring more avenues or experimenting with further methods which I have not so far tried; if I am trying to raise a weight up to a roof, I can try harder by looking for a more efficient set of pulleys, or just pulling harder on the rope I have. If I am trying to win the jackpot on a bar machine there is nothing to distinguish trying harder from just trying again and again. In short, turning to trying does not give us anything that was not already available with strength of will. We display this either by keeping on trying or trying harder.

The Will Again: Ability to Ignore one's Strongest Desire

Yet while it seems fruitless to turn to the will as what is weaker in the hope of explaining where the akratic fails, the feeling that we should not just concentrate on failures in practical reasoning seems right. In some sense, also, it seems to be a matter of the will. The important point here is the one first adumbrated by the Stoics, that if a given entity is a rational agent, then it must have the capacity, in general at least, to consider whether or not to follow this course or that, and whether or not to bring thinking to an end and do something. This requires the capacity to consider whether or not to indulge various desires, or act on certain reasons; and that supposes that in some senses of 'strongest' they are

not, when in a normal state, in a condition whereby once they realize what satisfies their strongest desire or reason, they act.

To count as a rational agent it is not necessary always to exercise this capacity, nor, indeed, always to possess it Sometimes we act spontaneously, without excercising the capacity; sometimes we are in a condition where once we have worked out how to satisfy some desire, we have to act. A dipsomaniac, for instance, may be in a condition where they can indeed work out means to acquiring drink, but (i) cannot think of anything else but how to get it and (ii) will be 'driven' to act once they work out how to acquire some. But then anyone in such a condition is not capable of acting as a rational agent. For this reason, although their behaviour is very intelligently purposive, the clearer we are about the attribution of mania, the more hesitant we are likely to be about calling the behaviour intentional. Since akratic behaviour is supposedly intentional, because deliberate, the akratic must be supposed to have possession of their rational faculties, and so must be in a condition where they have whatever the ability/freedom is that goes with that fact. Consequently, their behaviour must be a matter of will, i.e. they do the akratic act because they choose to, of their own free will. What the notion of freedom is which is required for attributions of rational agency is beyond the scope of this book. What does seem clear is (i) that if we measure strength of desire by intensity, then a person in a normal condition does not necessarily, if they act intentionally, act in accordance with their strongest desire; (ii) that if we measure strength of desire by reference to what on reflection, and even for the most part in their lives, they think would be the best way to behave, such a person does not necessarily act in accordance with their strongest desire if they act intentionally; (iii) that their intentional behaviour is not necessarily behaviour in accordance with what they consider to be the most reasonable course (almost whatever that might plausibly be taken to mean). When in a normal state an agent is capable, if they will, of acting contrary to their strongest desire. They show strength of will when it is burdensome so to act and they persist in trying or try harder. To do this is to do something more, which they can, if they will, the doing of which might have a greater chance of success.

Objection 1: Reason

This point, that an agent may act against their strongest reasons or desires, meets resistance for a number of reasons. To begin with, it may

be felt that if all the agent's reasons for acting tell in favour of a given course, then it is unintelligible to suggest that if they act intentionally they will do anything but act in accordance with those reasons. Here, as we have seen, we need to be careful what is being claimed. We may mean that every inducement which the agent has to act tells in favour of a given act. In that case, much depends on the strength of 'has an inducement to…' If we are assuming that the agent is engaged in deciding what to do; and that the inducements are those considerations which they find incline them one way or another; and that they consider them to be between them sufficient to determine that the act in question should be done; and that they find that they all incline to the same action; then doubtless there is a presumption that if they act intentionally they will do what those reasons support If we weaken those assumptions, the presumption falls. At the other extreme, if the inducements do not fall into a set determining a goal, but pull against each other, then, as we have seen, there is no such thing as what those reasons as a set support: we wait on what the agent decides to give greater weight to—but they are not determined to that weighting by the reasons since, *ex hypothesi,* the reasons determine nothing.

Similar things hold with thinking it better/best to do x. Given a criterion of assessment, it will doubtless be possible to give reasons for thinking that it is educationally best or politically best to do x; but unless education or political success is taken as the agent's reason for acting, no presumption follows as to what they will intentionally do. On the other hand, if the agent is supposed to think that it *is just best* to do x, it is difficult to know what to make of this other than that they have decided in favour of doing x.

So far we have been considering what Pears (1984) calls the forward connection between evaluation and action, the Leibnizean view that if a rational agent considers a given course best, they will follow it. This last point takes us naturally to the supposed backwards connection, the other Leibnizean view, that if a rational agent chooses a given course, they must think it best After all, if I have decided in favour of doing x, that means that I prefer x to any alternative; but that surely means that I think it better to do x than any alternative. Otherwise my opting for doing x would be an unexplained leap: I would be doing it for no reason.

Attractive though this may sound, it is surely a mistake. If ever I am vacillating between alternatives it may be assumed that I have some reason(s) for the one course, some for the other. Given the one set of reasons, the first course is best, given the second, the second. But, as we

have seen, commonly no sense can be made of one course being better overall. This does not mean that I do not have a reason for doing x, but it does mean that I have no reason for thinking the reasons for doing x better than the reasons for doing y. Sometimes, of course, I may have precisely the same reason for doing x as for doing y: they may be alternative, but equally good, means of achieving the same goal (alternative ways of mating in one). In this case the reasons I have for acting simply fail to determine one course as better than the other. I might, of course, decide to toss a coin; but the result does not provide a reason for supposing the favoured course better: I might just as well have chosen one arbitrarily, but may get some self-deceptive comfort from pretending that the outcome of the toss gives me a reason for that course. No doubt in such a case I have got to do something, perhaps on pain of being out of time and losing the match; but a reason for doing one or other of two things, however pressing, is not, of course, a reason for doing one rather than the other. Yet either move is perfectly rational, and does not lose its rationality by having a co-equal.

It is in fact quite common for there to be nothing to choose, in the agent's eyes, between two courses, and this is a circumstance with which rational agents have to be able to cope if they are not to be subject to repeated paralysis. They have a reason for whichever course they choose, but no reason for preferring the one they choose. This sounds objectionable because it is tempting to suppose that if they have no reason for their choice they must have no reason for what they do—but that temptation should be resisted. It might also be felt that if they have no reason for their choice, then an element of indeterminacy enters in—and how can we be held responsible for indeterminate wobbles in our choice mechanism?

It is certainly true that neither choice is determined by the reasons which the agent envisages in making up their mind. Consequently, the choice is arbitrary. This is not to say that their choice comes from a wobble, something beyond their control. It is not as though they did not know what they were doing, and to their surprise found themselves moving the knight—they could so easily have found themselves dancing a jig. By ordinary criteria they knew they could have done either, and decided in favour of one but with no reason at all for thinking it a better move. There may, of course, be 'a reason' why they choose to move the knight rather than the rook: they might have a dither-saving rule always to take the knight move in such cases; or they might take a special delight in the jumping, slanting knight moves, so that, other

things being equal, they will choose them; or they might have a general tendency in favour of moves involving movements of the right hand, and in this case the rook would have required movement of the left. But none of these is a reason for thinking it better to move the knight. We have in all cases moved into a different pattern of explanation, and none of them consists in showing the agent to think that move better than the other.

Objection 2: Desire

This set of difficulties has come from the reason end. Another comes from the desire end. As we have seen (Chapter VIII, pp. 100–1), when we talk of someone doing something for a reason, this is just another way of saying that they did it because they wanted whatever it is that is specified in the description of the reason. The various considerations which the agent takes seriously can also be given as various things they want This all points in the direction of a situation where, if we only knew the agent's desire-state better we should realize that their behaviour was determined by their desires. Our actual explanations often do leave things undetermined, but that may only be the measure of our ignorance. The ignorance may be solely psychological: a more complete picture of people's desires might yield either the influence of unsuspected ones (as in the second explanation above of why I prefer the knight move), or greater strength in those whose influence is acknowledged. Or one might consider that these folk-psychology desire explanations only work at all because of some underlying truths about the physical system, that there is either type/type or token/token identity between a desire and a physical state of the desirer: once these are established we shall tighten up our desire language, or recognize that it is an on-the-whole makeshift for day-to-day life which can always in theory be replaced by a proper causal explanation. In either case, we should hope to be freed of the indeterminate cases, and the supposed special freedom of the rational agent.

Clearly, there are big questions here whose adequate treatment would require a thorough discussion of such questions as freedom of the will and the relation between mentalistic explanations and physical ones. Since problems about weakness of will arise within our day-to-day explanations of human behaviour and attributions of responsibility, I shall stay fairly close to those starting-points. It may emerge that those forms of

explanation are incoherent, or that they need development to ward off charges of incoherence, or that they are based on presuppositions which in the light of well-established findings are not tenable. As we approach that more interesting country I shall become unsatisfyingly sketchy.

The starting-points of the objection I am considering are, like those of weakness of will, from within our day-to-day explanations of rational action (see e.g. Pears 1975). The first point, as noted, is that rational action is action done from desire: I explain my own or someone else's action of writing a letter by reference to a desire to make sure of booking a holiday. The agent is doing A from a desire to achieve B. The claim is that it was that desire that gave rise to the action, and this is a standard causal claim about the presence of conditions sufficient to produce the effect. It converts into: if the agent had not done A, there would not have been a desire to do B.Kenny (1975: pp. 112–13) objects to this on the ground that the conversion does not work. From the truth of 'I put the money on the counter because I wanted a drink' it does not follow that if I had not put the money on the counter I would not have wanted a drink: that would be too cheap a way with quenching one's thirst. While this may be fair enough given Pears' formulation as it stands, it is fairly clear that Pears thinks of the expression 'because A wanted to achieve B' as claiming not just the presence of a desire to achieve B, but of a desire of degree D, with the further, potentially falsifiable claim, that a desire of degree D (or a set of desires jointly achieving degree D) is sufficient to produce A. While the reformulation might itself be open to objection, so formulated the equivalence holds.

The talk of degrees has itself given rise to objection. Charlton (1988: pp. 126–34) rightly points out that we have various ways of talking of strength or degree of desire. Sometimes it is by reference to the felt intensity of a craving, sometimes to the degree of pleasure envisaged, sometimes to the relative value overall given to the course desired. Each of these is different, and it is simply false that by any of these measures we always act on the strongest desire or set of desires. Pears can, I think, avoid this objection, at least in that form. He could make either of two responses. First, he might claim that there are different sorts of desire with different modes of influence and so different measures of strength. Any might on occasion be opposed by others. When we are explaining action we are always in the last analysis referring to the total desire-state of the agent prior to the act, and this will, of course, involve a claim about the combined strengths of the various desires and aversions operative at the time. Alternatively, he might claim that these

assessments of strength have nothing to do with the sufficiency of the desire(s) to produce action, but only, say, with their power to impress the subject's consciousness. This is at best a rough guide to their effective power. That could only be discovered by deeper delving.

It remains, however, that whichever defence he takes, he owes us an account of degrees of desire. Whichever way he takes, we have moved away a little from our starting-points. There is no 'folk' notion of degree of desire such that when we explain an action A by reference to a desire to achieve B we are claiming some degree of desire which could in theory be shown to be causally sufficient for A.

This last might be contested. After all, the discussion of Davidson's P1–P3 (see Chapter VIII, pp. 107–9) revealed that we can indeed make sense of saying that in giving the reason(s) for which an agent chooses to do A, we must be giving what they want most. Thus, if I indeed wrote a given letter because I wanted to book a holiday, then booking a holiday was what I wanted more than (in preference to) doing anything else.

The trouble with this is that in so far as it is true it is truistic, and it seems that 'I wanted to book a holiday more than anything else' is made true simply by the fact that that was the reason why I did as I did. Since success is the only measure of strength here, we cannot use strength to explain success.

Of the ways we have of talking of degrees of desire, then, one cannot be claiming any causal role for degrees, and the others fail individually and, so far as we can tell, collectively, to play the required part. This might make one suspect that the proposed model of desire explanation is itself moving rather far from the starting-points. This is, I think, the case, and it happens because it is assumed that 'want' operates in much the same way in a variety of contexts.

Varieties of Desire Explanation

One feature of the view outlined is that it treats desires as dispositional preconditions of the *psyche* which are usually, at least in part, accessible to the consciousness of the agent. An explanation by desire, which is an explanation by reference to rational agency, appeals to the presence of such a precondition as sufficient for bringing about the action. An agent who is conscious of a desire will, in action, be commonly conscious of the act as issuing from that desire. Since operative desires are usually identified via their objects—what they are desires for—we get an

explanation of why the agent's self-explanation is so useful: first, the desire occurs inside the agent, and so is more accessible to them; and secondly, it is identified by the agent as having a certain object, and so comes in explanatory form. Further, the agent's consciousness of their desires is often what is needed for the desires to be effective and is often, through deliberation, a factor in bringing desires to bear on situations (see Pears 1975).

I find this picture very obscure, not least the supposed internal identification of desires via their objects, and the picture of deliberation; but what concerns me is what makes it a plausible story to start on. There are at least four ways in which we talk of being or becoming conscious of desires, which lend colour to Pears' talk of internal impressions. First, we are most of us familiar with the experiences of thirst, hunger, sexual desire, jealous yearning to deliver the final put down, attraction to the thought of becoming an MP and so on. It is plausible to speak of these as internal impressions, sometimes of physical sensations and some sense of what would relieve them, sometimes of a tendency to think of certain objectives with accompanying pleasure. It is also plausible to call what we are aware of desires, and to think of these as dispositions to think favourably of achieving what answers to the desire, on some fairly generous interpretation of 'thinking favourably'. It is not at all so clear what would be meant by the agent's feeling their action issuing from such a desire. If I am thirsty and decide to take a drink, no doubt I reach out for my glass because I am thirsty, because I want a drink. But in what sense do I feel my reaching out issuing from my thirst, the desire I am aware of? This suddenly seems to be putting a different picture on things. In the sexual case, of course, I may be conscious of various reactions issuing from my desire, but then they are not usually thought of as actions. In the thirst case I may feel *that* my action is apt for the achievement of my objective, and this might supply a sense to its issuing from my desire, but that sounds a different story altogether. The same goes for the other examples: what we are conscious of as an internal disposition is not what we feel our actions to issue from.

There is a second set of examples where we tend to think of the agent both as aware of their desires, and, indeed, of their desires as influenced by, and even made more effective by, thought, and at the same time as feeling that what they do issues from their desire. Thus a person in some serious sense overcome by fear or anger or jealousy may be very conscious of their emotion/desire; it may be that information leads to

more effective achievement of what is desired, and that what they do can be explained as done because they wanted to escape, teach someone a lesson, or put down a hated rival. To say that they feel that what they are doing issues from their desire now sounds quite attractive. But in so far as we are serious about it, their behaviour looks a doubtful candidate for intentional action, and 'because they wanted to achieve B' has different connotations from those it has in the standard case of explaining rational action. In particular, if I am, say, in a panic, and try claiming that I ran because I wanted to, the claim lacks full conviction.

Thirdly, we sometimes think that a person's failure in some project can be put down to the fact that they did not really want the objective in question. Thus, suppose a man is taking a course in accountancy, so as to be able to increase his income sufficiently to enable him to live in the comfort to which his parents have accustomed him. He fails his examinations, and his failure is explained by the fact that he never really wanted (sufficiently wanted) to be an accountant. That is not to call in question that being an accountant was a (subsidiary) goal, which well explains all his activities of signing on for the course, attending classes and so on. What is lacking is not desire to be an accountant in that sense, but enthusiasm for accountancy. Did he have it, he might have felt his activities issuing from his desire in the sense that he felt eager and enthusiastic to do them. Yet, while this might be giving desire in this sense some causal role, it is not a sense of 'desire' which gives the rational explanation of his action.

Finally, there are standard examples of explaining rational action. Here the agent writes a letter, say, to book a holiday. Since they are in general expected to know what they are about, they are likely to take the view that that is why they are writing it; and given the sort of explanation being offered, if they are satisfied with their performance they will probably feel that what they are doing is suitable for achieving their objective. If one is in poetic mood, one might describe this as feeling their action issuing from their desire; but one should not let one's poetic instincts mislead one into assimilating this notion of 'feeling one's action issuing from…' to any that might be appropriate to the previous examples. Crudely put, when 'want' is used in contexts where it introduces the explanation of rational action it introduces the plan or principle to which the action is supposedly appropriate, thus enabling us to understand what the agent is after. It is not an appeal to the influence of prior psychological states on analogy with the first three examples above. No doubt if I write a letter because J want to book a

holiday, it is a fact about me that I want to book a holiday, i.e. that is what I am after in writing the letter; but if it is then claimed that my action resulted from the fact that that desire was stronger than any other, the claim is either truistic and unexplanatory, or explanatory and dubious. It is unexplanatory if the sole measure of strength is effectiveness. If it is felt to be explanatory it will doubtless be because in these contexts 'I want to book a holiday more than anything else' seems to be true just in case I am decided in favour of booking a holiday, in which case I must think it best to do that. In that case I shall only want to do x more than anything else if I think it better to do x than anything else. That takes us back to all the earlier problems about 'more reason' and 'better'. As the example of being faced with a free afternoon shows (see p. 109), there are situations where it is just false that the agent thinks the course they are taking to be the best. Nor is this uncommon. Once I have decided to tidy my desk there will be different ways of approaching the task between which there will be nothing to choose. Then, as I consider the details I am likely to come up against aspects of the project which I had not originally reckoned with. To clear my desk effectively I shall have to decide what to do with various papers on it; but to deal with them properly would take more time than I can afford if I am to tidy up the desk. Clearly I shall have to make decisions on priorities, but they may be *between* reasons, again, not *for* reasons—though, once again, what I finally decide on will have a reason. So if 'want more' means 'consider to have better reason for', while it might serve to explain the choice, it seems that it is not generally true that such an explanation of one's choices/actions is available. So (i) my strongest desires do not determine my choices and (ii) my choices do not show my strongest desires by any interesting test of strength.

Lack of Determinism Again

This lack of determination by desire is at three levels. First, there are cases where, given the desire on which they act, the agent can see nothing to choose between one course and another. Secondly, there are cases where the agent has to 'decide what they want', or make a decision about what considerations to take as determining their action. Finally, it comes at a crucial point in the attribution of rational agency. For one thing that is being attributed is the capacity to deliberate and weigh up the pros and cons. Only rarely is it clear that now everything relevant has

been investigated. Commonly, there is no clear point at which it is not worth pursuing one's thinking further; yet if one is to act one must be capable of bringing it to an end. Suppose I want to increase my capital and am trying to choose a suitable investment. I start making enquiries about a number of firms. I realize that I cannot investigate every firm; that with regard to the three that I have selected there are many more things I could enquire about which might change my judgment. I decide that I have gone as far as I can, and decide on a particular share. Why did I decide to limit my enquiries to three firms? Why did I stop my enquiries then? My desire to increase my capital hardly suffices: I agree that I might, if I continued, find a better investment How might I explain it? I might say 'why not?', but that is hardly an explanation. Or I might say that I had to stop sometime: if I went on forever I would never invest my money at all, and so would fail to achieve my goal. That may be true, but hardly explains why I should stop now rather than next week, why I should limit myself to three firms rather than five. Nor is it any use appealing to a desire to stop sometime, which has a similar failure to home in on this particular time. Nor does it help to say that there did not seem any very good reason for continuing: there was no very good reason for stopping then either. This simply illustrates a common feature of decision-making: that its determination is not the result of either reasons or, if they are different, desires.

Finally, it is worth noting a point about 'because I wanted to. It was noted earlier (p. 177) that my panic behaviour is not something I can convincingly claim to have done because I wanted to, although clearly done from some desire. It might, therefore, seem that this expression serves to claim that my action was a paradigm of rational agency: I was not forced, I was in possession of my faculties, it was my choice. It is now interesting that I put down this action to my desire, and cite it as cause. In the light of earlier objections it may be that we should limit the set of desires that are appropriate to cite here, but the fact remains that our ordinary accounts of rational action are open appeals to the cause of the action.

This would be building rather a lot on the occurrence of an expression. It is true that 'because I wanted to' commonly serves to reject any suggestion that I did not act of my own free will. Suppose, however, that we return to the example of my looking for a suitable investment to increase my capital. If I am asked why I stopped at a given point, I might well say 'because I wanted to'. But that cannot be converted into 'because I had a desire to stop', as though my stopping was not for any further reason, but

from a desire for stopping itself. It may be, of course, that I was overcome by a sudden yearning to stop; such things do come on one from time to time; but that is clearly a special case. In the example described I had no particular desire either to stop or to go on, but, as I said, I could not go on forever, and this seemed as reasonable a time as any other to stop, so I stopped (so I stopped?).

It seems, then, that if we are working from our day-to-day explanations of our actions there is no appeal, open or covert, to causation by desire; but there is an implication that the agent's choice is not determined by desires or reasons, although it is typically for a reason. While in the above I have worked from Pears, that is because his position is a good starting-point for distinctions I wanted to make. The distinctions are, however, quite commonly ignored (see e.g. Stitch 1983; Mele 1987).

Objection 3

This still, however, leaves some uneasiness. After all, it was admitted above that if I did A because I wanted to achieve B, then it is a fact about me that I wanted to do A, and one that explains my doing A; and unless we can, as it were, ground this fact in the agent, we shall be left with some mysterious indeterminism. It may be that we should not conflate this talk of wanting with its occurrences in other contexts, and it may be that the explanation does not lie with the sorts of psychological factors discussed earlier. It remains that there is an implication of the existence of something about the agent which accounts for the action, and perhaps we should look to neurophysiology to supply the answer, to give us the structures which underlie our appeals to wanting.

This is far too big a subject to embark on with any thoroughness. It often, however, takes the form of supposing the following: (i) the account of action-explanations sketched above looks dangerously Ideological; (ii) ideological explanations should, where possible, be supplanted by non-teleological, causal ones; (iii) the way to do this is to show that corresponding to the attributions of belief and desire—the menialistic notions which enter into action-explanations—are certain physical structures which are responsible for what happens. This may be the assumption that corresponding to the type **belief that the ice is thin** is a type of physical structure, and similarly with the **desire not to get wet**; or it may be that for each occasion where A's belief that the ice is thin interacts with their desire not to get wet, there are corresponding

physical states of the organism which are what that belief and desire are respectively in this case. The underlying motivation is often a desire to avoid a dualism of mind and body and a belief that allowing in teleological explanations untied to physical counterparts would give dualism a foothold.

My interest at this point is simply to lessen resistance to the combined point that action-explanations are not explanations appealing to causation by desires, and are such that they commonly fail to explain why one action was done in preference to another. I simply want to argue that allowing in explanations of this general sort is no threat to monism, which does not require that there will be physical states which in any interesting sense correspond to various beliefs and desires.

It is a commonplace of theories of action that the explanation of action appeals to the interplay of belief and desire. What I wish to argue is that this is an interplay within the structure of explanation, and that nothing follows about the interplay of anything within the structure whose behaviour is being explained. Suppose we have a computer that has been designed to play chess. It will have been programmed so as to react to inputs about the state of the board, and to react in the direction of mating the opposing king and defending the home king. These two 'computer-desires' give rise to sub-desires, such as the computer-desire to prevent the opposing knight reaching a certain square. What action these will lead to will depend on the input of information, on what computer-beliefs are acquired. When we are watching the computer at play we shall want to understand its moves, and this will be a matter of seeing them displayed as appropriate, given the computer-desires and computer-beliefs. Failing any such explanation of a reasonable degree of plausibility, we shall suppose that something has gone wrong and examine the hardware, given confidence in the software. As with humans, however, such explanations will often rest with showing that the move could reasonably be taken to serve the objective(s), given the input of information. Given two equally effective moves, they will not explain why the one was preferred to the other. This is because, *ex hypothesi*, there is no such explanation. But why should we suppose that corresponding to this structure of interplay in the explanation there is any corresponding structure of interplay within the computer? In any given case there might be distinct parts which interact. But why should it not be just that there are two facts about the computer—that it has been programmed with a certain set of orientations and that there has been a given input—with a resultant change in the state of the system,

but with no possibility of specifying some part of this change as what the desire amounts to, and another as what the belief amounts to? All we have is the present total state, which is at once the state of orientation to be expected of a computer so programmed after such an input, and the state to be expected of a computer after such an input if it is so programmed. There is no call to tie bits of the computer's 'psychology' to bits of its physical state. If we wanted to explain the physical 'correlate' of the desire to block the knight's advance, we could only go back to explaining how the program was appropriate to the game of chess, and how the construction of the computer was appropriate for embodying the program, and how that, together with the intervening history, all led to the present situation. It would indeed be very surprising if this could not be done, but this monism is under no threat from the availability of explanations in terms of computer-desires and computer-beliefs, nor does it turn those explanations into covert appeals to identifiable elements in either computer psychology or computer physiology. They do, however, presuppose general facts about computer structure. The computer-beliefs here, of course, are mainly the information input considered as accepted. There is no call for the implausible belief that this move is better, but only that it is appropriate.

The purpose of this will be obvious. Day-to-day action explanations are in terms of reasons, and enable us to understand what the agent is up to in the light of the information they have. As with the computer, there will be a history of how and when they acquire the information, and of how and when they come to have the desires; but the explanations in terms of these factors are not appeals to interactions within the *psyche,* nor within the physiological structure. Once again, it might turn out that there are correlates to be found, but the acceptability of the explanations does not require it, nor does their acceptability pose any threat to monism: one might even suppose that there is something about our neurophysiology which explains the availability of such explanations in our case.

This move is only defensive. There may turn out to be features of what we suppose of rational agents which for some reason fit ill with the ideal of neurophysiological explanation. Humans look to be a little more complex than most computers to date, and it may be that the sort of freedom from determination by desire/reason that I have suggested is required for rational agency causes difficulties. What cannot cause difficulties is the mere fact that there are many cases where what is done can be explained, but why it is preferred cannot; nor the fact that

day-to-day explanations are not appealing to prior conditions sufficient to produce the behaviour. These explanations do not answer that sort of query. Explanations of moves are not explanations of movements.

The outcome of all this is that if we are operating from within familiar modes of explaining action there is reason for supposing that such explanations operate on the assumption that actions are commonly not explained by reference to the strongest desire or best reason in any familiar sense, although the explanations are usually of the form that A did x because they wanted to achieve y, or that A did x, and their reason was to achieve y.

We can therefore rest with the account of weakness of will given on p. 170. The weak agent fails to do something which they could do if they chose and which would lead to success, but which is in one way or another burdensome to do. As it stands, this can sound too sharp. For it seems to make a very clearcut distinction: either the agent could or they could not If they could, why did they not? if they could not, why upbraid them? But of course it is not like that There are cases where it is obvious that by any normal criteria the agent was able to take these steps; at the other end there is a degree of difficulty that will outpace their abilities, and they plainly cannot; but in between there is an area of varying degrees of grey, and as it becomes darker it becomes less certain that the agent could take the necessary steps. Degrees of difficulty are approximations to impossibility relative to any given capacities. They consequently lessen our confidence in the attribution of credit or discredit. Degrees of difficulty are also measured by degrees of burdensomeness. There is always the background question, in addition to 'What can the agent do?', the question 'What is it reasonable to expect of them?' These degrees of difficulty can operate at two stages: problems, for instance, can vary in difficulty, and a weak person may give up too soon, without any suggestion of difficulty in taking a next step which, if anything, might lead to success. In cases of passion, however, the doubt about ability may come even at the stage of taking the next necessary step.

Finally, it is worth noting that talk of strength and weakness is most appropriate where there is a context of difficulty. In the cases of so-called clear-eyed *akrasia* there seems little call to talk of strength or weakness. We are left with a situation where an agent has reasons for which they are serious about doing x (which they could do if they chose), and do y deliberately. There the problem is one of seriousness rather than strength.

Epilogue on Wanting and Belief

At the risk of tedium, it may be worth underlining some of the points about wanting. 'Want' is ambiguous, and so attributions of desire constitute different sorts of claim about a subject according to the sense of 'want' or 'desire' at issue. Sometimes the claim that A wants x is the claim that A looks with pleasure on the attainment of x; sometimes that A has a craving for x; sometimes that on reflection A thinks it would be good to have x; sometimes that x is what A is after. In line with the first three are different senses of the claim that A wants x more than y: A looks with more pleasure on the attainment of x than on the attainment of y; A has a stronger craving for x than for y; A on reflection thinks that it would be better to have x than to have y. It is possible for A to have a stronger desire for x in one sense, for y in another; and the different measures of strength of desire are measures of strength of desire in different senses of 'desire'. In none of the first three senses of 'desire' does strength of desire determine action. If we did discover some psychological or physiological state which did determine action, it would not be the measure of strength of desire in any of these senses, though it might yield a valuable technical sense and accompanying measure. None of these senses is in play in the apparent truism that intentional action is a function of the interplay of beliefs and desires. For that apparent truism we need the use of 'want' whereby it serves to introduce the subject's objective or reason, and in this use it is not clear what sense can be made of the notion of degrees of strength of desire. Belief and desire here function as parts of an explanation of what is done as an exhibition of the subject's abilities to learn. For this the subject has to be able to identify goals, be oriented in its behaviour towards them, capable of adapting its pursuit of them in the light of experience of what is conducive to them, and so forth. Since we want to be able to interpret behaviour as manifesting these abilities even when we think mistakes have been made, we need a term to indicate that some input of information has affected the subject's orientation without committal to whether the subject is right or not: 'belief plays this role. Belief, however, is itself relative to forms of learning and as we have seen (Chapter I, p. 15) it may be futile to ask what the subject 'really' believes: in the context of exhibiting its ability to weigh up evidence and give good engineering advice it believes* the rope bridge is safe; in the context of wanting to reach the other side of the gorge it does not believe it is safe. There is nothing which is its real belief state, though on any

particular occasion of behaviour its behaviour will be explicable in terms of some given belief/desire. Further, there are difficulties about separating belief and desire. As the example of the parent panicking on the river bank suggests (pp. 148–9), one can as well talk of conflict of desire or conflict of belief; and when one is talking of desire in the context of intentional behaviour, 'doing as one wants' assumes the normal condition of operation of evidence assessment and not that typical of passionate desires. The mode of acquisition of belief is integral to the type of desire.

When it is claimed that it is a condition of being a rational agent to know, in general, what one wants (see Chapter VIII, p. 122), it is the fourth, not any of the first three senses of 'want', that is in question. This is not to say, alas, that it may not be true of the first three that language users with the relevant vocabulary are expected to know what they want in these senses. But for the point at issue it is the fourth sense that is important, and knowing what one wants is more akin to knowing in what direction one is facing than to knowing what psychological forces are influencing one. Our expectation that language users will have this knowledge of themselves does not come from any supposed special access to their inner states, but is part of attributing to them the ability to deliberate and work out what to pursue, articulately. This attributes to them the ability to decide what they want (i.e. what to pursue); and there is a difficulty about attributing to them the ability to decide what to pursue but not in general know what they have decided or what they take to be the case. The fact remains, of course, that human beings are rather complex, and knowing what one is after is not at all the same as knowing either why that objective seems worthwhile, or why certain facts about its pursuit strike one as significant. It remains interesting what form these further explanations take, but not obvious. In particular, if subconscious desires are appealed to, which might 'surface' on examination, one will have to ask the sense of 'desire' in which they are desires and are now surfacing. In any given explanation the word 'subconscious' might be functioning as a red flag, so that what we have is not a desire, in some familiar sense, which is surfacing, but an explanation which has some, but not total, analogy with some (but which?) more familiar desire explanation.

CHAPTER XIII

Varieties of Weakness

The Relativity of 'Weakness'

It should be clear that 'weakness of will* is used in a specialist sense in philosophical contexts. That is not to say that people's treatments of the philosophical problem(s) are uninfluenced by the pre-philosophical uses of the expression; but it is important to try to keep them free of each other. One point to note is that this expression, like '*akrasia*', is commonly used as a term of moral opprobrium. What someone counts as weakness is therefore liable to vary in contrast to what they admire as strength. Another is that the usual connotations of 'weakness' direct the mind in certain directions, and discourage consideration of certain examples, and these bear an uncertain relation to the philosophical examples.

As to the first point, what a Christian may admire as meekness, requiring great self-control, Nietzsche might despise as weakness, as also, for different reasons, might Hume. What one person thinks of as showing great strength of will, someone else might consider the obstinacy of the insecure, betraying a basic weakness of character. The result is twofold: first, that there is a general sense that weakness of will must be a moral phenomenon; and secondly, that according to their moral views, people will be inclined to select certain paradigms of weakness.

On the second point, in some contexts 'weakness' may connote no more than lack of determination in the pursuit of goals or in the assertion of our own will. It might suggest giving in to love of comfort, or greed, but not so readily anger, which is strong and assertive; and giving in to some inducement to pursue short-term goals over against longer-term considerations. Certainly a man who thought his family and its financial security important, but risked all in the hope of wreaking

complicated but long-delayed revenge, would not leap to mind as an example of weakness. Yet he might be a perfectly good example according to a given way of setting up the problem. The expression is used in the philosophical literature to describe a problematic situation, but it is not clear that the choice of examples is not also governed by common connotations of the expressions. At the same time some situations covered by the expression in common use have an interest that is concealed because they do not conform to the pattern of those from which the problem started.

It does seem to be features of the expressions used that have helped direct philosophers' attention first to cases of being overcome by passion, and also to cases of moral weakness. Yet in so far as puzzles have arisen they do not seem confined to ethics, nor to combats with reason. Whether they should be seen as confined to what I have called passionate *akrasia* is another question. In Chapters VIII and IX I argued that if a problem arises about deliberate pursuit of what is thought the worse course, it does so because of an apparent rift between sincerely declared intention and deliberate action. It might seem clear that the passionate cases give convincing examples of this, but that it is not so obvious what to make of the so-called clear-eyed cases. The presence of passion, after all, seems to explain how the agent may come to be thrown off their originally chosen course. When passion is absent there seems no explanation. Yet the review of various 'passions' in Chapter X suggests that the line between passionate and non-passionate cases may not be as clear as one might hope.

In practice many different kinds of behaviour are commonly described as showing weakness which do not feature, or not largely, in traditional accounts. Consideration of some of them suggests a criss-cross of overlapping problems about rationality and intentional failure, which opens up the field beyond the confines determined by the model of standard passionate cases. In the next section I shall consider examples of what would commonly be called weakness and argue that they have a different relation to intentional action from the more traditional cases; that they attract some different charges of irrationality; but that at one crucial point they share a problematic feature. I shall approach the differences and similarities by considering whether appeal to passion is at any point needed to explain weakness.

Varieties of Weakness

The traditional examples have been ones where someone can be described as giving in to temptation, where the tempting object arouses some desire which might in extreme form render the subject out of control. So fear, and desires for food, drink and sex have been obvious examples. In all these cases the temptation supplies the agent with an objective to pursue, so that the weak action can be seen as having a purpose deliberately chosen, and at odds with another (better?) judgment. As we have seen, some passions may lack this feature. In my excitement I may realize that I ought to be considerate to my defeated opponent and knowingly give offence, without having any purpose which excitement is driving me to. This may make one hesitate about saying that I deliberately offended them, or even, perhaps, intentionally; but since I knew, and was not compelled, it is hard to say that I offended them unintentionally. Such a case would not normally get considered because my behaviour, though not unintentional, is not, either, done for a reason, or, in Socratic language, for the sake of some good. Consequently it does not fit the model from which the problems of *akrasia* started.

Procrastination et al.

There is a kind of example which certainly fits the general description of failing to do what one thinks best, and which is surely one of the commonest forms of what would naturally be called weakness, and that is procrastination. Suppose a woman has promised her lover that she will break with her husband, but is firm that he must not discover this from anyone else: she is going to tell him herself. She has not lost all affection for him, and she knows that it will be a difficult occasion. She thinks of doing it when he comes home that night, but he seems to have had a very hard day, so she puts it off; but the next day is their wedding anniversary, and it would be cruel to celebrate it in that particular fashion; the day after, she is feeling very irritable herself, which might lead her to put things in a way she would later regret—and so on.

One feature which distinguishes this from the traditional case is that while she accepts that she must break the unwelcome news some time, there is no time which is the time when she must break it Consequently we do not get the traditional format whereby she is failing to do what

she recognizes she ought to do now. Yet something strange is happening. Seen in terms of reason, it is not either that she is acting against what she thinks best for some 'irrelevant' reason, or that her action on any given occasion is not in conformity with her declared intent. As to the first, her reasons are on each occasion ones which she considers reputable and to the point. As to the second, while she has a determination to tell her husband some time, she forms the intention, and may express it (though not to her husband), not to tell him now. What is strange is her facility for finding plausible reasons for putting off the evil day. It would be an exaggeration to describe her as overcome by fear, but her fear seems to give her both an unusually quick eye for circumstances which might provide an excuse for delay, and a tendency to give these considerations an importance which they might not have had were no unpleasant interview in prospect There result two irrationality charges: the first, that there is a lack of consistency between the canons of importance being used in this set of circumstances and those used in the rest of her life, without there being any acknowledgment of change on her part; the second, which becomes stronger the longer the process goes on, that she is deceiving herself in thinking that she is serious about the succession of excuses.

'Deceiving herself, of course, is an expression which simply makes it clear that the situation is puzzling: while we know that it cannot be true on exact analogy with deceiving someone else, we want to say something stronger than just that she is being disingenuous, but are unsure whether there is any space to be filled or how to describe it. One feature of this and many other cases of self-deception is that the agent's attention gets drawn to reasons which will supply a plausible justification of her action to critics, and distracted from the business of justifying it to herself—though she may half convince herself (but what is that?) and even in the end, by repetition, come to believe it. In the present case her failure is in part a failure to realize the change in her modes of assessment, together with an unwillingness to decide precisely when she will do the deed. While she would commonly be described as weak, and is, from fear, failing to do what she thinks should be done; and while she might well be described as acting irrationally; she does not fit the standard pattern.

A somewhat similar case may occur without procrastination. Someone in the flush of possessing a credit card may know that they should not overspend, but be 'unable' (or unwilling?) to control their expenditure. Each time there may be a good enough reason for using the

card, and it is always possible to hope that matters will improve by the time the day of settlement arrives—especially as that is indefinitely postponable, so long as a partial payment is made. None of the expenditures has to be passionate, nor does it have to be true of any that it ought not to be indulged. Nor does it have to be true of this month's set that they overstep the limit. Unlike the previous example—where one required act was being deferred—here we have a pattern to be controlled, and what is deferred is consideration of a method of control. Weakness is shown in failing to get down to an overview of expenditure and a plan for its regulation. So long as there is no clear view, there is no reason even to think the pattern excessive until one month's required payment becomes unpayable.

While there may be, there is no requirement of, self-deceit here—simply unwillingness to face a dreary task. Whereas in the previous case there is in some sense a determination to do what is each day deferred, in this case, while that might be so, it might be that there is no more than a general recognition that it would be a good idea sometime. This secures that they are not merely feckless or irresponsible, but there is some puzzle about the seriousness of their committal to review, given their propensity to defer. But it is a definite puzzle, for often the person seems to wish they could get down to it, and there is no obvious difficulty about the first steps, or even about accomplishing the task: just an in-built resistance to it. Somewhat similar circumstances occur with student work problems, or in the aftermath of bereavement: people suffer from paralysis of the will. There is then, as with passionate *akrasia,* room for dispute as to whether the metaphors should be taken seriously. If so, the agent is not in the normal condition of responsiveness; if not, then the capacity to respond is there, though it does not come easily. As with depression, the agent is teetering on the edge of not operating as a normal agent.

Depression is itself, as we saw in Chapter X, a case of some interest and one which at least in a practical way much exercised the Medievals (see Aquinas on *Acedia*). As in these last examples, there is not a counter-purpose, as with fear, but a lack of interest in anything so energetic as decision or action. This makes it unnatural to describe the agent's failure as deliberate, or even intentional, although it will commonly be true that the agent is knowingly failing—a feature which only gets modified in the self-deception examples.

In some of these sorts of case something which could with a little generosity be called an emotion is cited as explaining the failure to act

appropriately. It is not at all clear, however, that that is required. The woman failing to break the bad news to her husband may have been afraid, but neither procrastination nor self-deception seems to need such feelings. If I am an indecisive person I shall not need fear to procrastinate: however important I may think a decision, I am temperamentally averse to committing myself, and have the same propensity to give unusual importance to reasons for not making a decision now. A Sartrean woman who thinks she ought to withdraw her hand from her escort's grasp, but detachedly watches it happen as though the hand were just an interesting object, not part of her, is not in any familiar sense under the influence of emotion, but she is failing to do what she thinks to be best, and the same strange explanation by reference to self-deception seems in order. Similarly, appeal to depression is appeal to a quasi-emotion; but if I am accused of failing to act through lethargy it is not clear that that supplies a diagnosis of the cause rather than just labels my behaviour as unsatisfactorily inactive.

In these cases, then, we seem to have examples where passion is not, in any familiar sense, necessary to explain weakness; and even where it does seem to have that role, since it is not a purposive sort of passion, it does not seem to yield deliberate, or even clearly intentional weakness. At the same time the agent does seem to show signs of retreat from the normal state of responsiveness to discussion. It may still be, however, that it is because of a feature of the examples that they do not have to be explained by appeal to passion; that when we come to the traditional cases there are differences which require appeal to passion to explain what is going wrong.

Passion and Deliberate Failure

One apparent difference is this: in neither the procrastination nor the depression case do we seem to have an active interest at the time of the action in doing what is thought best. It is therefore easy to feel that the agent did not really at the time think that course best In the other cases the supposition is that the agent is really at the time serious about what they then fail to do or refrain from, and we need some explanation, therefore, of why they fail. Passion seems to supply such an explanation; but when we remove passion we lose all explanation along with it.

The difficulty seems to be exacerbated if we posit that the judgment that passion sweeps aside is that of reason. For either we shall be

tempted to think that that is a judgment that the balance of reasons is in favour of what one fails to do, or that it is a judgment that what one fails to do would be better to do than what one does. That, of course, takes us back to Chapters VIII and IX, and the question of just what sort of judgment that is. The apparent extra difficulty seems illusory.

We are left with the question of whether, if an agent deliberately does something out of line with their sincerely declared intention, we need passion to explain the fact. This, it seems, is what distinguishes the 'purposive' passionate cases from the ones discussed earlier: that the agent both apparently has an interest in what they do not do (whereas the depressed, for instance, have little interest in anything), and an interest and purpose in what they do do (which makes their action deliberate). The interest in what they do not do seems to make it necessary to explain their failure. Here passion seems a good and perhaps necessary candidate. Where it takes the form of being a purposive reaction by the agent to an appreciation of the situation which has some independence of the operation of reflection, it will also give hope of making the action at least intentional.

Yet if the examples in Chapters VIII, pp. 111–12 and IX, pp. 124–5 have any plausibility, they both seem examples where appeal to emotion in any familiar sense seems strained. One could always, of course, escape this by insisting that emotion is involved. The advantages of such a move, however, are obscure. Not that it can be proved wrong. As should have become clear, the words 'emotion' and 'passion' do not stand for well-defined categories; nor, indeed, is it clear that words such as 'fear' and 'depression' operate as well-devised terms of theory. An argument about how to extend or confine the words would have to be part and parcel of developing a good way of talking about the phenomena. Before one gets deceived into supposing more similarity than is plausible, however, it is worth asking how like one is prepared to make the man's itch to be a successful financier, or the woman's sudden interest in coffee, to rage or depression. Presumably someone interested in asserting the possibility of clear-eyed *akrasia* wants to hold that we do not need to appeal to anything as disturbing as these feelings in order to have deliberate choice of the worse course. As I have already suggested in Chapters VIII and IX, part of the trouble here is that in insisting on the seriousness of the agent about the reasons for the course they in fact reject, one is constantly tempted to talk as if they in fact accepted them as their reasons for acting, when, *ex hypothesi,* they did not. There seems to be to this extent a similarity between the cases of

passionate and non-passionate *akrasia,* that with both we seem to come across cases which we are unsure how to describe. In the one case we are unsure how far we can push the talk of turbulence while retaining that of deliberate action; and in both we are unsure how far we can push the talk of seriousness with regard to a set of reasons which are not taken as the reason(s) on which the agent acts.

Once we become unsure about 'emotion' and 'passion' being well-defined terms of theory, and contemplate the many examples of uncertainty, it comes to seem a somewhat arbitrary matter whether we allow for clear-eyed *akrasia,* since the notion of being clear-eyed is itself only as clear as are those of emotion or passion. Since it receives its interpretation by contrast with them, this becomes tainted with all the uncertainties attaching to them. If we reserve 'passionate' for cases where the agent seems beyond the reach of influence by beliefs acquired by evidence assessment (as the man in panic on the river bank), then many of the other examples will be clear-eyed. If 'passionate' extends to cases like the woman afraid that her husband will abduct her children, then we are still left with the other examples being clear-eyed. If 'passionate' extends to all cases where we can say of the agent that they did as they did because they wanted something to which it seemed appropriate, then clear-eyed cases have been ruled out, but only, of course, by depriving 'passionate' of any teeth: any acknowledgment of something as a consideration in favour becomes a passion.

As emerged in discussing Aristotle, it is not always clear what someone is wanting to secure in securing clarity of eye. Is it that the agent should understand what they are saying? Or that they really believe it? Or that they put it to practical use? Someone in a panic may well understand 'The best way to save the child is to stay on the bank; diving in is not staying on the bank; ergo…' As we have seen, the question, whether the man in the example *really* believes it, is futile, or is answered by both 'yes' and 'no'. In a panic what becomes clear is that he does not operate in the normal rational way: he cannot discuss the matter calmly. If he believes that it is better to stay on the bank, it is not as a result of a mode of belief-acquisition which is part of his current mode of operation. What is anyway certain is that he does not use that set of considerations in deciding what to do. In a high degree of panic, of course, the agent may neither understand nor believe the view that another course of action would be better. Even in cases where something we might call emotion is involved, however, nothing so violent is required, and certainly understanding of, and by some criteria belief in,

the view that the proposed course is the wrong one is quite possible. The question that arises is not to do with understanding or belief, but with the agent's seriousness about the reasons on which they do not act.

Corresponding to the various ways of being weak are various ways of being strong, and it becomes even less attractive to think of these as all being strength in some one respect A person shows strength (of will? of mind? of character?) in being prepared to face up to and act in uncomfortable situations without self-deceit; in being prepared to persist in a project in the face of discouragement; in overcoming and restraining their rising anger, and so on. In all these cases the 'strong' agent is manifesting 'normal' responsiveness to discussion and rational methods of assessment of evidence. The cases of weakness, in so far as they are puzzling, show varying deviations from that norm. The norm, of course, is not the usual, but the standard of rational agency. It has to be noted, however, that 'weak' and 'strong' have now become defined by reference to that norm. As I remarked at the beginning of the chapter, there are examples of mulish obstinacy which might by these criteria display weakness but which would be surprising candidates for that description in day-to-day discourse.

Conclusion

What I hope emerges from this chapter is first, that 'weakness' as used in discussions of *akrasia* has a different scope from what it has either in day-to-day discourse or that more specialist discourse of some given moral view; secondly, that nevertheless many examples which would commonly be labelled 'weakness' give rise to related forms of puzzlement, even though they are not obviously examples of deliberate wrong-doing; thirdly, that they give rise to their own forms of irrationality; fourthly, that passion is not required at least to allow for these; fifthly, that even with cases where the agent is serious about thinking this is the occasion to do x, because of R_i, but does y because of R_j, it is not clear that passion is required in any interesting sense. Nevertheless, all *akrasia* is puzzling just in so far as it threatens absence of a normal state of responsiveness and capacity for rational agency. It is this which makes it plausible to suppose that passion is always required: some explanation of a partial failure of normal operation of the agent is required, and passion is a ready-to-hand candidate.

CHAPTER XIV

Epilogue

It is now time to try to draw together some of the main threads of the previous chapters. As I have remarked before, the problems about *akrasia* start, in Socrates, with the view that deliberate mischoice is impossible. The deliberations of reason cannot be disobeyed, because one's judgments of what is best cannot be disobeyed, and the deliberations of reason determine what these are. Reason always exercises itself, in practical matters, on the question what it is best to do. This already sets the problem of *akrasia* as one about the possibility of a conflict between desire or passion on the one side and reason and calculation on the other. It also ensures that the question becomes one about the possibility of deliberate wrong-doing. This remains true in the later Plato, and also, though with a different picture of practical reason, in Aristotle. In both these authors, however, a distinction is made between desire for food, drink, sex and such like, and certain passions, particularly anger, in that the latter do, while the former do not, consist partly in views of how it is good to behave. *Akrasia* is the result of conflict between desire and reason, not between anger and reason. This contrast is abandoned by the Stoics. These desires and the passions are all assents to presentations about what is fitting to do now. Since in most people their reason is ill-developed and weak, the results of their reflections will not be full understanding, but presentations about what is fitting, to which a person may or may not assent Being overcome by anger receives the same account as being overcome by hunger or greed. There is nothing impossible in acting against the results of cogitation. Two things are, however, impossible: first, acting wrongly when one has attained full understanding; and secondly, acting deliberately contrary to an assented hormetic presentation. The Stoics therefore introduce two elements: the generalization of *'akrasia'* to all passions, and the requirement of assent. It remains that the phenomena referred to as

akrasia are all to be seen in terms of the rule of passion, which is antipathetic to (proper) understanding.

In the Middle Ages the need for assent becomes tied to the view that the akratic acts from choice, and so can always overcome passion. The requirements of theology, however, make it important to be able to say that in some sense those not under the influence of passion can also deliberately take the wrong course. What was viewed as problematic was the possibility that people should choose what they realized to be simply bad, and the concern was to show that what they chose was always a misunderstood good. There is no sense that there is a problem about their thinking it to be the *worse* course. No-one seems to have been attracted to the view that one must always choose what one thinks best. Indeed, in so far as it was held that there are courses of action, both of which are permissible, though one better than the other, it was allowed that it is not only possible but legitimate to choose the less good course. There emerges a difference between those who predominantly saw deliberate wrong-doing as needing some explanation in terms of failure to take all one's knowledge into account at the time, and those more ready to allow that a sinner might freely choose what they knew to be wrong, with the proviso that what was chosen, although wrong to choose as chosen, belonged to a category of goods for the agent.

Between the Middle Ages and the twentieth century (the point at which, for the purposes of this book, history ends and the present day begins) interest in these problems diminishes. In some cases the earlier problems perhaps should have surfaced, but did not; in some, as with Leibniz, new elements were introduced, in particular in regard to the relation of rationality to thinking something best; in others, the general approach to passion, desire and action ensured that no problem was visible.

In the twentieth century *akrasia* became a topic of discussion again for two types of reason. First, R.M.Hare's prescriptivism, worked out primarily to expound views about moral judgments, seemed to have the result that it is impossible to act against one's judgment that such and such is best to do now, if that is a genuine judgment using the expressions in their primary sense. While this was worked out for elucidating the nature of moral judgments, it does not seem a long step to apply it to non-moral ones. The second source was consideration of intentional or deliberate action. Here it seems plausible to say that what marks this out from other behaviour is that the agent does it for a reason. It now seems to be a problem how an agent who thought there was *more* reason for doing x than for doing y could nevertheless deliberately choose to do y. This can easily get extended to include the moral cases.

Since morality tends to embody overall judgments about the way it is best to live, that way into the problem will at least incline one, although it does not require one, to see akratic cases as ones of acting against some wider all-embracing judgment. The Davidsonian route through consideration of intentional action also inclines one to see *akrasia* in terms of going against what is judged the more reasonable course. So although both routes will allow or encourage one to see the problem, if any, as not confined to ethics, they both encourage one to see it as a conflict between shorter—and longer-term, or narrower and more all—embracing sets of considerations—between reason and something else. Consequently there is a feeling that if it is possible, still *akrasia* is certainly irrational.

Irrationality 1

This notion of irrationality has given rise to some problems. One simple model is to make rationality relative to reasons. Thus if R_i is satisfied by the doing of x, but requires the non-doing of y, then doing y is irrational relative to R_i, doing x rational. The trouble about this, as has been remarked, is that *ex hypothesi* in the akratic case, since the action is deliberate, while the agent's action will be irrational relative to R_i, there is always an R_j relative to which it is rational. We need, therefore, a method of showing that the agent's rationality somehow requires adherence to R_i. This leads to a search for an account of all-things-considered, or better judgment. It seems sensible to follow one's better judgment, since the assumption is that the judgment is better by being more reasonable, or better based, which for present purposes means that in the agent's judgment it is more reasonable or better based: The question is: what does that amount to? An account of all-things-considered judgments might seem to answer this in terms of the bulk of the agent's reasons, or the agent's reasons taken all together, supporting the judgment; but this turned out to be illusory. On the other hand, if it means what the agent decides on after considering everything, it is too weak, since that, after all, might be the supposedly akratic act. If we make the agent's better judgment what they judge to be best we have a choice: either they judge it best in a given respect or set of respects, in which case there is no obvious irrationality in acting against that judgment; or they judge it just best, in no particular respect—but it is doubtful what that means; or judging it best is being decided on doing it, in which case there seems no special air of rationality. The hunt for

a way to give teeth to the notion of better judgment or judgment of reason seems a wild goose chase; but in that case the contrast between reason and something else as a general feature of *akrasia* collapses, and one is left with the agent's declaration, that some set of considerations gives the relevant or important ones, as what makes the judgment on the basis of them their best judgment. It emerges that what gives the akratic an air of paradox is their apparent seriousness about reasons on which they do not in fact act, although they act as they do deliberately.

Irrationality 2

This is not to deny that in various ways there is an air of irrationality about the akratic, only that it has been mislocated. To begin with, with language users there is a presumption that they can work out what to pursue, articulate their goals or principles and in general know what they are about. A puzzle about *akrasia* is that the agent seems to satisfy our best criteria for declaration of purpose while deliberately doing something else. If we can add that the agent seems genuinely puzzled about their own behaviour, the problem deepens. While it is obvious that the agent does not take the declared reasons as the ones on which to act, it is not clear just what is going on. The resistance to allowing such a case comes from the sense that we should have to take the agent as not knowing what they are about in the way required for counting as a rational agent in possession of their faculties. This unease retreats in so far as we can account for what happens either as not deliberate, or as a change of mind or…This kind of case does not depend on one set of considerations being more long term, more embracing, the result of reflection or whatever, and the other not. There may be no distinction on these grounds.

Irrationality 3

The claim so far has been that the traditional examples, although dressed up as reasons against something else, in fact get their bite from being examples of reasons taken seriously but not acted on. There is, nevertheless, an air of irrationality about someone who seems never to act on more embracing reasons. For anyone to count as a rational agent it does seem that they have to display some tendency on occasion to pursue relatively long-term projects involving a certain amount of planning, even if not

necessarily on the akratic occasions. The closer we get to someone supposedly never planning beyond the immediately accessible in the immediate environment, the more doubtful it becomes that we have a rational agent at all. Not that they are out of control, and so akratic, but that they are not in control. Since it is easy, though erroneous, to think that passion is short term, concentration on the traditional examples would seem to bear out the view that a tendency to *akrasia* is a tendency that threatens rational agency. If we drop passion but stick to the short term, the same threat will persist. It is not, however, that the actual choice or action is irrational, but that the manner of choice is such that if generally characteristic of an agent it would put that agent's rationality in question.

Irrationality 4

When we consider what I have called passionate *akrasia* different problems arise, and different charges of irrationality seem to come to bear. In the case of turbulent passions, the turbulence is a sign of the agent not being in a state to respond to discussion either of the end, or, sometimes, even of the means: they seem at least on the verge of operating on a way of assessing the situation, and the best means to deal with it, which is not affected by rational assessment of the situation or evidence. Here to be in a rational state is to be in a state where one is open to reflective assessment and discussion. One may, indeed, respond spontaneously from gratitude, affection, fear or whatever, while still being in such a state; but some of these emotions may reach a degree at which the agent seems to have passed the point of being capable of responding to rational considerations in the required way. In such cases, therefore, the influence of passion seems to threaten a change from a rational to an irrational agent.

Irrationality 5

In sub-turbulent cases there arise other charges of irrationality, some of which no doubt also arise in turbulent cases. Thus:

(i) Commonly, such examples show the agent not merely liable to be struck by features of the situation which would otherwise pass them by—in which there is nothing

particularly irrational—but also liable to manhandle the standard ways of assessing the evidence (cf. jealousy, fear of failure).

(ii) In many cases (the examples were with procrastination) the agent indulges in self-deception. Sometimes these will be examples which fall under (i), but in the example given in the text the doubt that arises is whether the reasons apparently seriously adduced on each occasion are not being a cover for avoiding the unpleasant.

(iii) In cases such as self-pity, depression and lethargy people tend to get into a state where they are dead to emotions of the familiar sort, and are disinclined to take that interest in life supposedly characteristic of rational agents.

These different failures of rationality do not make the action or choice irrational. Those under (i) and (ii) are failures to operate one's rational faculties properly, and under (iii), perhaps, at all.

One cannot list sub-turbulent types of case to which the different failures of rationality are confined. Nor, it emerged, can one confine these failures to so-called passionate cases. It might seem that this is false for those under (iii), but it is not clear in what sense of 'passion' or 'emotion' self-pity, depression and lethargy would all count.

Weakness of Will

While these very different charges of irrationality may arise in different cases, it would seem to be a mistake to suppose either that there is some form of irrationality which is exhibited in all cases of *akrasia,* or that each case must exhibit some form. For some of the explanations of the feeling that *akrasia* is irrational do not justify the conclusion that the akralic act displays irrationality, but only that if generalized the agent would not count as rational. Further, accounts of practical reasoning will have no light to throw on the problem of *akrasia.* What we do need to suppose in all cases, if we are to dub the behaviour deliberate, and not simply out of control, is that the agent could have refrained or acted had they chosen, with the addition that some burdensomeness or unpleasantness attached to the course the taking or avoiding of which would have shown strength or weakness. It is right to think of *akrasia*

as a matter of the will, so long as one does not hold bizarre notions of the will. In relation to passionate *akrasia,* especially when we are dealing with the turbulent cases, there is a point in thinking that reason is what is in danger of being overthrown; but it is not the judgments of reason, or the outcome of practical reason, but the state of being able to indulge in and adapt to the suasions of reflection, that is under threat But this is equally a state in which the agent is able, should they so decide, to act contrary to the pull of passion: reason and will go together.

Moral Weakness

Despite the history of the subject, this book has shown very little concern with moral weakness. This is because, although in its beginnings the problem did seem to be whether anyone could be coherently described as deliberately choosing to go against what they recognized as required by morality, the difficulty got its force from a more general question about going contrary to the determinations of reason, and got its special moral air from views about the connection of morality with reason. Questions about giving in to temptation become special cases of the wider question. They will only arise essentially in ethics given certain views of the nature of moral judgments: that they have a certain overriding, directive role. At least three points, however, are important in relation to morality. First, on most views of the nature of morality, it will become of practical concern to wonder how people can be brought to integrate other objectives with their moral views, and in particular how and to what extent general emotional control can be achieved.

Secondly, moral views are likely to aggravate the practical difficulties of control in at least two ways: first, they are likely to embody views on the undesirability of certain motives and emotions and desires, so that what in other circumstances might just seem a tiresome distraction comes to seem an affront to one's ideals, and possibly more attractive for that Secondly, the more perfectionist the ideals the greater the added disgust and despair at being the sort of person open to temptations against them—unless, of course, the moral views contain their own comforts for these conditions.

Thirdly, discomforts about how to reconcile talk of weakness, in its general interpretation, with belief in the unity of the agent, are intensified in the moral cases in so far as on most views of the nature of morality a moral view will embody views of the sorts of desires and

motives one ought to have. These last seem to be integral to the sort of person I am, but so, importantly, do my moral views. So it seems that I am the sort of person to want not to be the sort of person I am.

Weakness and Folk-Psychology

The problems about *akrasia* start from our day-to-day ways of explaining behaviour, which are pre-theoretical, at least in the sense that they are what we start with prior to the development of any systematic and testable general theory of human behaviour. The presumption has been of sufficient similarity between the pre-theoretical stances of Ancient Greek, Latin, Modern English, German, etc. It has become fashionable to describe this as folk-psychology or, if one thinks the dissimilarities significant, folk-psychologies. It is then attractive to poke fun at folk-psychology and look with longing to the day when it can be supplanted by something better and more scientific; or, if not supplanted, at least underpinned, so that we understand the realities which make folk-psychology as reliable as it is.

Since the problems arise within it, whether or not they have a solution can only be decided from within it. A theory (not yet available) which supplants it will not solve these problems, but will show them not to be worth bothering with. On the other hand, if we are either to solve them, or show folk-psychology not worth the keeping, perhaps by showing them insoluble, then we have to take the trouble to get folk-psychology right

The first problem is to discover how we are to do that, not least because there is some problem as to what folk-psychology is. One possibility is that it is to be found in the use of a set of terms in our ordinary pre-scientific language for explaining behaviour of, roughly speaking, a purposive sort. This presumably lies behind the common view that actions are explained by the interaction of beliefs and desires: we often do, after all, explain why a person does something by reference to something they wanted and something they believed about what was needed to get it. This structure of explanation is familiar, and so much a part of our day-to-day thinking as to merit the expression ' folk-psychology'. When we learn English, say, we learn to operate these explanations and so to accept, at first uncritically, the psychological views implicit in them.

What is not clear is what sort of view we have got. As we have seen, it is not true that all behaviour explained by reference to some belief-informed desire is deliberate or even intentional. Panic behaviour

satisfies the condition, without being either. So that structure, so described, is not sufficient to ensure that deliberate behaviour is being explained. Nor is it true that all desires have an action-explaining role. What makes one fail to see this may be that when one operates with the noun it is easy to be deluded into thinking that it is univocal and determines a homogeneous set. In fact, it is just a noun to correspond to various occurrences of the verb 'want'; and, as we have seen, this is ambiguous. In some of its uses it has no role at all to play in explaining deliberate behaviour. In so far as it is required for the behaviour to be deliberate, 'A wants…'serves to introduce what A takes as a reason for doing what they are doing. The assumption is that we are dealing with a being capable of learning how to pursue things, and we are supplying the object of pursuit. This, of course, is common to uses of 'want' explaining the behaviour of non-rational agents, or the non-responsible action of rational ones. In the responsible cases 'because they wanted…' carries further implications about the nature and condition of the agent.

When we turn to belief, that word may also, of course, be ambiguous; but what emerged in the examples was something different. 'Belief' seems in the uses in question to be a word for indicating that the being we are considering has received certain information and adapted to it in accordance with the expectations of the explanatory model. But it transpires that with humans at least there are different ways of processing and adapting to information, which lead to cases (see the examples of the rope-bridge and the anxious parent on pp. 15 and 148–9) where the question 'What do they really believe?' seems futile.

What is badly needed in this area of 'folk-psychology' is some careful consideration first of the way that that belief/desire pair operates which is integral to the attribution of deliberate behaviour. This will involve special attention to the senses of 'want' and any differences there might be when this is used to introduce goals generally and when it is used to introduce them in the context of deliberate behaviour (when I fly in panic, do I run away because I want to?). Also we should have to ask just what we are claiming of an entity when we claim it believes something. One difficulty here is that ordinary language is not a corpse awaiting dissection, but is liable to be changing in the course of and as a result of the discussion.

It will already have become clear that in one sense ordinary language does not embody a theory: the expressions variously used in explanation have uncharted ambiguities. It does seem, however, that we take for granted some conception of a rational agent as a being capable of thinking out what to do, responsive to considerations put to them and

presumed to know at some level why they do what they do: a conception which requires us to consider them as not in general determined by strength of desires or reasons in taking their decisions. Those with a programme of eliminating folk-psychology would have to see more clearly than anyone does as yet just what this assumption amounts to. Further, it will be important to investigate what sorts of question are answered by our familiar explanations, in order to see whether they are better answered by some alternative.

I have tended to take examples from what is commonly thought of as the area of practice. The same assumptions seem to underlie our view of rational enquirers: scholars are expected to know what their problem is, what they hope to prove, why the evidence adduced is thought relevant and so on. The expectation is defeasible, but had better not fail too widely. One question at issue, then, is whether this way of viewing ourselves is dispensable. On the face of it, it lies at the base of our co-operative lives and relationships, including intellectual ones, and not only our quasi-speculative attempts to predict which way the cat will jump next. Even ordinary conversation is conducted on the assumption that people for the most part know what they are trying to say.

As I have said, the expectations are defeasible, and things go wrong in a variety of ways. There is some pre-scientific terminology available for talking about ways of going wrong. Among this is the terminology of emotions. But 'emotion' as used pre-scientifically is a very ill-defined term (see Gosling 1965), and it is doubtful whether there is any very good sense of ' theory' in which this terminology can be said to embody a theory. No doubt the way people talk is interesting, but it does not clearly yield anything systematic. As we have seen in discussing passionate *akrasia* (see, e.g. pp. 17–19) people are not at all sure just how they want to talk in some cases, nor what the talk precisely commits them to. People reflecting on such descriptions may be tempted, for instance, to take seriously the talk of being carried away by passion, and develop a view of akratic action on that basis. But this is canonizing one dramatic way of describing experience, and removing it from its context of dramatic description to a serious part in an historical play. This may be armchair psychology, but how many folk are committed to it? Not, it seems, English folk in virtue of having adopted English. We may here be encountering gestures towards a psychology, but different folk adopt different gestures, and perhaps the same folk different gestures according to circumstance.

Given that a rational agent in a normal state is such that (i) they are capable of responding to reason, and (ii) know what they are about to

some degree—a knowledge which is manifested in their sincere declarations of intent or preference—the proposed akratic puzzle cases must be ones where there is some occurrent threat of breakdown of that norm. They therefore call for some explanation, and one outside the pattern offered by the declared intention or preference. For this role emotions have seemed well suited, because they seem to be a familiar condition of such threat. Consideration of examples suggests a messier picture. Even if we stay with fear, say, it is not clear that the interference envisaged is all of a sort (see the examples in Chapter X, pp. 133–4). If we move to lethargy, the explanation seems simply to be a name for the fact that the agent finds action difficult. Something similar is probably true of depression: the explanation might be like explaining someone's tetchiness by reference to stomach-pains: it becomes interesting whether they are those of indigestion, ulcers or stomach-cancer. Similarly, one might expect more revealing diagnoses of depressions, which might earn them varying feeling-labels and constitute differing explanations. In this area we seem to have untidy gestures towards psychology, with a likelihood of finding diagnoses which suggest that some cases are not really akratic, because not responsible, which were previously thought to be.

All these explanations 'explain' actual, or in akratic cases incipient, deviations from a norm. That norm, however, does not appeal to views about mechanisms which result in behaviour: it supposes that human beings have certain capacities, can be interacted with and understood in various ways as outlined earlier in the discussion of explanation by desire. Only when we understand more adequately than we do at present what this supposed norm is, and the kinds of understanding, explanation and interest supposed to go with it, shall we be in a position to tell what would count as replacing it or better doing what it does.

So long as we remain at a level prior to developed theory, the situation seems to be that all the cases of weakness must be to some degree puzzling, because of the agent's apparent failure to satisfy the norm, without our being willing to declare the complete failure that would go with loss of responsibility or control. Something has gone wrong in that the normal tally between sincere declaration of objective and actual choice has in one way or another broken down. The situation is felt to lie between a clear case of normal operation and a clear case of failure of rational agency, and we find ourselves wanting to avail ourselves of the language suitable to each extreme, while realizing that it sits uncomfortably in the intervening territory.

It seems probable that whatever refinements of psychological explanation emerge, short of a total change of our view of rational agents, some such grey area will remain. That change would involve replacing a conception of ourselves as beings capable of a variety of long-term pursuits, knowing for the most part what they are, with some ability to reason out ways of achieving them and normally open to thinking about considerations relevant to them. That is not all there is to us, but the assumption that we at least have these properties seems to underlie a good part of our mutual dealings with and interest in each other, which give rise to familiar, non-scientific, areas of puzzlement about explanation of each other. The puzzlements and explanations, however, are relative to the underlying assumptions.

Then in some cases folk use bits of professional psychological terminology in an unprofessional way. They talk of sub-conscious desires, obsessions, complexes and such like. Is that ordinary language? And is it folk-psychology? Of course, there are worthwhile questions as to whether professional psychologists are too uncritical of pre-scientific terminology, and as to what the relationship is between their investigations and those of neuro-physiologists. Questions about the relation of folk psychology to either of these, however, are bedevilled by uncertainty as to what the folk animal is. As it appears in the literature it has a tendency to be a hybrid born of mixed idioms. Like other offspring of mixed parentage it is liable to prove mulishly intractable. The discussion in the text has been confined to problems as they arise in our pre-scientific attempt to talk of deliberately taking what we think to be the worse course. Discussion of various sorts of irrationality also raises the question of just what sort of explanation we are giving in claiming, say, that jealousy distorts a person's assessment of evidence, or panic drives out reason. Those are interesting sets of questions to which the present discussions give rise, but which they do not solve. The desire/ belief pair that does seem integral to the explanation of action operates in the context of interpreting the behaviour as the manifestation of some intelligence and learning. It is non-committal as to the mechanisms at work.

Practical Reasoning

Some approaches to the question of *akrasia* make it tempting to think that a clearer view of practical reasoning will help us to understand the phenomenon: something seems to have gone wrong in the application

of reason to the question of what to do. While this seems to be mistaken, the discussion showing it to be mistaken suggests that the attempt to give canons of practical reasoning might be doomed anyway. As we have seen, at the level of selecting reasons to apply in deciding how to spend today, the project of finding canons of reasonable selection does not look hopeful; and this difficulty keeps recurring (see pp. 106–7 and 178–9). Of course, some of an agent's thinking will be deductive: doing this will mean taking someone else's umbrella, and taking someone else's umbrella is theft, therefore…Indeed, it seems that an agent's decision for a reason can always be presented in deductive form.

More importantly, other bits of reasoning, often those establishing the 'particular premiss', will be inductive, showing the best means to an end. Sometimes the reasoning is about whether this particular action counts as theft, say, and, whether this is a legal or a moral argument—but especially if the latter—there may be no test of whether the argument is successful, even though there are fairly good tests of relevance. At other times things get muddier: I may be set on proving myself a trendy liberal; I wonder whether it is more liberal to leave my daughter to conform, as she wishes, to the gender norms of her contemporaries, or to insist that she take courses which will equip her to break free from the constraints of these norms. I shall doubtless decide on one course or the other, but it is totally obscure what would count as a good or decisive argument to settle the matter, and trial and error will hardly help. There is a great deal of thinking on practical matters which can only with flattery be called reasoning, but is not on that ground either avoidable or unimportant. So far as *akrasia* is concerned, however, it does not matter how messy the reasoning is. That is irrelevant The problems arise once the agent has decided on R as the reason to act on, and takes doing x as the best way to satisfy R. Faulty reasoning does not make for more akratic action.

Conclusion

Many of the arguments about weakness of will have been negative, and repeatedly, as the discussion has approached interesting areas, I have taken the line that that belongs to a different book. What does, I hope, emerge, is that central to the problems about *akrasia,* and giving rise to the sense(s) of irrationality surrounding akratic action, is the conception of ourselves as rational agents, where this is not primarily the

conception of ourselves as long-term pursuers, but as having a range of capacities of which that for long-term pursuit is just one. I cannot claim to have given an adequate delineation of this conception, but it does seem to me to involve thinking of ourselves as in certain ways not determined by reasons or desires in our decisions to act, to deliberate and to bring our deliberation to an end. I have tended to make this point provocatively, but I hope that reflection will suggest that there is nothing surprising about it: it only becomes offensive if one either starts with a contrary prejudice, or confuses arbitrariness with a form of indeterminism which entails lack of control. This freedom, whatever on analysis it amounts to, is needed if we are to be capable of strength or weakness of will. What it amounts to, once again, would be better discussed in another book.

What experience does not allow us to do, however, is to think of ourselves as unalloyed rational agents. To begin with, we do not know everything about the influences on our choices, or on our deliberations about them, and discovering our distortions is doubtless galling as well as instructive. While such distortions arise in the context of akratic behaviour, very obviously in the passionate cases, they are not confined to such cases, and I have therefore ducked the problems of the relationship between the availability of explanations of their occurrence and the applicability of 'rational* explanation in terms of belief and desire. In some ways more troubling for our self-conception is the occurrence of turbulent passion, with its threat of turning us from rational agents to ones which in a way know what they are about, but in any ordinary sense are out of control. Neither the term 'passion' ('emotion'), nor the terms for the various passions are precise terms of theory. Between them, however, they roughly cover a range of phenomena of central importance to human life. Strip a person of all emotional reaction and you get something inhuman; picture someone always and only acting under the influence of emotion, and you picture something inhuman. One has to try to avoid, however, contrasting them as a set with reason. There is no general incompatibility between long-term calculation and emotion, nor between acting on some emotion and being in a rational state. The strains and uncertainties in our descriptions rise with the temperature of our emotions. These are areas where we do not understand ourselves. But that merely indicates the shape of problems, it does not solve any.

Bibliography

Entries marked with an asterisk contain useful bibliographies.

Texts

Plato

Protagoras, ed. J.Burnet (Oxford Classical Texts), Oxford, Clarendon Press, 1909. Translations: *Plato*: *Protagoras,* by C.C.W.Taylor, with commentary, Oxford, Clarendon Press, 1976; *Plato*, *Protagoras*, by Benjamin Jowett, revised by Martin Ostwald, with commentary by Gregory Vlastos, Indianapolis and New York, Bobbs-Merrill, 1956.

Aristotle

Nicomachean Ethics, ed. I. By water, Oxford, Clarendon Press, 1894. Translations: *Aristotle*: *Nicomachean Ethics*, by Terence Irwin, Indianapolis, Hackett, 1985. *Aristotle*, *Nicomachean Ethics*, by Sir David Ross (The World's Classics), Oxford University Press, 1980. *De Anima*, with an Introduction and Commentary by Sir David Ross, Oxford, Clarendon Press, 1961.

Stoics

Most citations are from *Stoicorum Veterum Fragmenta* (*SVF*), ed. H.von Amim, Stuttgart, Teubner, 1979. Galen's *De Placitis Hippocratis et Platonis* (*PIIP*) has a translation by P.DeLacey accompanying a new text in *Corpus Medicorum Graecorum* (Berlin, Akademie-Verlag, Books 1–5, 1978; Books 6–9, 1980). Euripides, *Medea*, in J.Diggle (ed.), *Euripidis Fabulae*, Oxford, Clarendon Press, 1984. There are translations in Loeb editions (The Loeb Classical Library, London, William Heinemann, and Cambridge, Mass.,

Harvard University Press) of Cicero, Epictetus, Aulus Gellius, Plutarch, Seneca and Sextus Empiricus which are quoted in *SVF*. Some citations are from H.Diels, *Doxographi Gracci* (= *Dox. Gr.*) Berlin, Reimer, 1879.

The Medievals

St Augustine's *Confessions* can be found in M.Skutella (ed.), *Sancti Augustini Confessiones,* Teubner, 1981. There is also a Loeb edition. Anselm's *de Casu Diaboli* is not, so far as I know, in translation. It can be found in *S.Anselmi Opera Omnia*, ed. Francis Schmitt OSB, Edinburgh, Thomas Nelson, 1961. For Aquinas I have used 5. *Thomae Aquinatis Opera Omnia,* ed. Roberto Busa SI, Stuttgart-Bad Constatt, Frommann-holzboog, 1974–80. There is a translation of the *Summa Theologiae* by the English Dominican Fathers, Thomas Gilby OP (ed.), London, Eyre and Spottis-woode, New York, McGraw-Hill, 1963–75. For Duns Scotus I have used *loannis Duns Scoti Opera Omnia,* ed. P.Carolo Balic, Rome, Vatican City Press, 1950. A somewhat free translation of probably all the relevant passages can be found in *Duns Scotus on the Will and Morality*, Allan B. Wolter, OFM, Washington DC, The Catholic University of America Press, 1986.

Post-Medievals

For Descartes, see Charles Adam and Paul Tannery (eds), *Oeuvres de Descartes*, Paris, Léopold Cerf, 1897–1910. Translations: E.S.Haldane and G.T.R.Ross, *The Philosophical Works of Descartes*, New York, Dover Publications, 1951; G.E.M.Anscombe and P.T.Geach, *Descartes*: *Philosophical Writings*, London, Thomas Nelson, 1963. For Leibniz, see C.J. Gerhart, *Die Philophischen Schriften von G.W.Leibniz*, Berlin, Weidmann, 1890. Translation: H.G.Alexander, *The Leibniz-Clarke Correspondence*, Manchester University Press, 1956. (Translations are my own, unless otherwise specified; *The Leibniz-Clarke Correspondence* contains Clarke's slightly amended translation of Leibniz.) For Spinoza, see C.Gebhardt *Spinoza Opera*, Heidelberg, Carl Winter, 1926. Translation: R.H.M.Elwes, *The Chief Works of Benedict de Spinoza*, New York, Dover Publications, 1951. The Hume passages can be found in L.A.Selby-Bigge, *Hume's Enquiries*, Oxford, The Clarendon Press, 1902.

General

Audi, R. (1979) 'Weakness of Will and Practical Judgement', *Nous* 13. (1982) 'Self-deception, Action and Will', *Erkentniss* 18.

Austin, J.L. (1961) 'A Plea for Excuses', in *Philosophical Papers*, Oxford, Clarendon Press.

Bratman, M. (1979) 'Practical Reasoning and Weakness of Will', *Nous* 13.
Broadie, A. and Pylus, E. (1982) 'Kant and Weakness of Will', *Kant-Studien* 73.
Burnyeat, M. (1980) 'Aristotle on Learning to Be Good', in Amélie Oksenberg Rorty (ed.), *Essays on Aristotle's Ethics,* Berkeley and Los Angeles, University of California Press.
*Charles, D. (1984) **Aristotle's Philosophy of Action,* London, Duckworth.
*Charlton, W. (1988) **Weakness of Will*: *a Philosophical Introduction*, Oxford, Basil Blackwell.
Cordner, C. (1985) 'Jackson on Weakness of Will', *Mind.*
*Dahl, N.O. (1984) *Practical Reason, Aristotle, and Weakness of the Will,* Minnesota, University of Minnesota Press.
Davidson, D. (1969) 'How is Weakness of Will Possible?, in Joel Feinberg (ed.), *Moral Concepts,* Oxford, Oxford University Press. (1982) 'Paradoxes of Irrationality', in R.Wollheim and J.Hopkins (eds), *Philosophical Essays on Freud,* Cambridge, Cambridge University Press. (1985) 'Replies to Essays I–IX', in B.Vermazen and M.Hintikka (eds), *Essays on Davidson: Actions and Events,* Oxford, Clarendon Press.
*Dihle, A. (1982) *The Theory of Will in Classical Antiquity,* Berkeley, University of California Press.
*Dunn, R. (1987) **The Possibility of Weakness of Will,* Indianapolis, Hackett.
Gallcp, D. (1964) 'The Socratic Paradox in the Protagoras', *Phronesis.*
Gill, C. (1983) 'Did Chrysippus Understand Medea?', *Phronesis.*
Gosling, J.C.B. (1965) 'Emotion and Object', *Philosophical Review.* (1969) *Pleasure and Desire,* Oxford, Clarendon Press. (1973) *Plato,* London and Boston, Routledge and Kegan Paul.
Gould, J.B. (1970) *The Philosophy of Chrysippus,* Leiden, E.J.Brill.
Grice, P. (1985) 'Davidson on "Weakness of Will" ', in B.Vermazen and M. Hintikka (eds). *Essays on Davidson: Action and Events,* Oxford, Clarendon Press.
Hardie, W.F.R. (1968) *Aristotles Ethical Theory,* Oxford, Oxford University Press.
Hare, R.M. (1952) *The Language of Morals,* Oxford, Clarendon Press. (1963) *Freedom and Reason,* Oxford, Clarendon Press.
Heinaman, R. (1986) 'The Eudemian Ethics on Knowledge and Voluntary Action', *Phronesis.*
*Hornsby, J. (1980) *Actions,* London, Routledge and Kegan Paul.
* Hurley, S.L. (1989) *Natural Reasons,* Oxford and New York, Oxford University Press.
*Inwood, B. (1985) *Ethics and Human Action in Early Stoicism,* Oxford, Clarendon Press.
Jackson, F. (1984) 'Weakness of Will', *Mind.*
James, William (1890 [1950]) *Principles of Psychology,* Vol. II, New York, Dover.
Kenny, A. (1975) *Will, Freedom and Power*, Oxford, Basil Blackwell.
Lemmon, E.J. (1962) 'Moral Dilemmas', *The Philosophical Review*.
Lloyd, A.C. (1978) 'Emotion and Decision in Stoic Psychology', in John M. Rist (ed.), *The Stoics*, Berkeley, University of California Press.
McTighe, K. (1984) 'Socrates on Desire for the Good and the Invohmtariness of Wrongdoing: Gorgias 466a–8e', *Phronesis*.

Mele, A. (1987) *Irrationality*. New York, Oxford University Press.
Mortimore, G. (ed.) (1971) *Weakness of Will*, London, Macmillan.
O'Shaughnessy, B. (1980) *The Will.* Cambridge, Cambridge University Press.
Peacocke, C. (1985) 'Intention and *Akrasia*', in B.Vermazen and M.Hintikka (eds), *Essays on Davidson: Actions and Events,* Oxford, Clarendon Press.
Pears, D. (1975) 'Sketch for a Causal Theory of Wanting and Doing', in his *Questions in the Philosophy of Mind,* London, Duckworth. (1984) *Motivated Irrationality,* Oxford, Oxford University Press.
Pylus, E. and Broadie, A. (1982) 'Kant and Weakness of Will', *Kant-Studien* 73.
Robinson, R. (1977) 'Aristotle on *Akrasia*', in J.Bames, M.Schofield and R. Sorabji (eds), *Articles on Aristotle,* London, Duckworth.
Rorty, A.O. (1980) *'Akrasia* and Pleasure: *Nicomachean Ethics*, Book 7', in AméZlie Oksenberg Rorty (ed.), *Essays on Aristotle's Ethics,* Berkeley and Los Angeles, University of California Press. (1980) 'Where does the Akralic Break Take Place?', *Australasian Journal of Philosophy* 58.
*Santas, G.X. (1979) *Socrates,* London, Routledge and Kegan Paul.
Schiffer, S. (1976) 'A Paradox of Desire', *American Philosophical Quarterly* 13.
Schueler, G.F. (1983) *'Akrasia* Revisited', *Mind.*
Stich, Stephen P. (1983) *From Folk Psychology to Cognitive Science*: *The Case Against Belief*, Cambridge, Mass., and London, MIT Press.
*Taylor, C.C.W. (1976) *Plato: Protagoras,* Oxford, Clarendon Press. (1980) 'Plato, Hare and Davidson on *Akrasia*', *Mind.* (1984) 'Reply to Schueler on *Akrasia'*, *Mind.*
vermazen and Hintikka (eds) (1985) *Essays on Davidson: Actions and Events,* Oxford, Clarendon Press.
Vinci, T. (1985) 'Comment on Doxastic Incontinence', *Mind.*
Watson, G. (1977) 'Skepticism about Weakness of Will', *Philosophical Review*.
Wiggins, D. (1980) 'Weakness of Will, Commensurability and the Objects of Deliberation and Desire', in Amélie Oksenberg Rorty (ed.), *Essays on Aristotle's Ethics*, Berkeley and Los Angeles, University of California Press.
Woods, M.J. (1987) 'Plato's Division of the Soul', Dawes Hicks Lecture in Philosophy, in *Proceedings of the British Academy*, LXXIII.

Index Locorum

In addition to authors' works, the *Stoicorum Veterum Fragmenta* are also indexed.

General Index