ETHICS

by

P. H. NOWELL-SMITH

Professor of Philosophy, University of Leicester

PHILOSOPHICAL LIBRARY
NEW YORK
1957

CHAPTER 19

Freedom and Responsibility (1)

[1]

WE have now to consider the logic of the language which we use to ascribe responsibility, to award praise and blame, and to justify our moral verdicts. I shall consider five types of moral judgement.

He broke a law or moral rule.	(1)
He could have acted otherwise.	(2)
He deserves censure (or punishment).	(3)
It would be just to censure (or punish) him.	(4)
He is a bad (cruel, mean, dishonest, etc.) man.	(5)

It is clear that all these are logically connected. It is not just a fact about the world that we learn from experience that only bad men deserve blame or that it is only just to blame those who could have acted otherwise. Yet the items cannot all be treated as analytically connected; for we should then find that it was senseless to ask certain questions that obviously do make sense.

For example, the character-words used in (5) are partly descriptive; and it makes sense to ask whether a person who is consistently mean or dishonest deserves blame. To give, as most of us would, an affirmative answer would be to use, not to analyse moral language. It would also be a mistake to say that it must (logically) be unjust to blame someone who could not have acted otherwise, on the ground that this is part of what 'unjust' means. For (2) is a theoretical statement, while 'unjust' is a G-word contextually implying that no one ought to blame him. Nor does it help to say that we have insight into necessary synthetic connexions between the items on the list; for this is simply to say that we know them to be connected but cannot understand how. The connexions are of the quasi-logical kind that can only be understood by examining the conditions under

which the various expressions are used and the purposes of using them.

I shall start by considering the connexions between (1), (3), and (4). The connexion between (3) and (4) seems to be analytic. If a man deserves blame, someone would be justified in blaming him. Not necessarily you; for you may be in no position to cast the first stone or to cast any stone at all.

Now 'punishment' is a legal term and, in the case of punishment at least, (3) and (4) logically imply (1). A man can be justly punished only if he has broken a law, and the same applies, although naturally in a looser way, to moral censure. To deserve censure a man must have done something wrong, that is to say broken a moral rule. Now why should this be so? This question has already been partly answered in chapter 16. 'Punishment' is a complex idea consisting of the ideas of inflicting pain, on someone who has broken a law, in accordance with a rule laying down the correct punishment. But we have still to ask why we make use of this complex idea at all. Remembering that 'just' is a G-word, it is necessary to suppose that anyone who says that Jones deserves punishment must have a pro-attitude towards his being punished. But why should we wish to encourage the infliction of pain on those who have broken a law? The classical utilitarian answer is that it will either reform the criminal or deter potential criminals or both. Now, since laws and moral rules are devices for bringing about ends, we must have a pro-attitude towards reforming those who break them and deterring others. So, if it is a fact that punishment has these effects, this will explain the connexion between the infliction of pain and the breach of a rule.

But this simple theory will not do, if only because potential criminals would be as efficiently deterred by the punishment of an innocent scape-goat who was believed to be guilty as by that of a guilty man; and, whatever the effects might be, this would not be just. And we have also seen in chapter 16 how this simple theory can be amended. For we there saw that, although we might have a system of dealing with each situation as it arose, there were great advantages in having legal and moral codes. And it is because we have these codes that neither the punishment of Jones nor an adverse moral verdict on him could (logically) be called 'just' unless he has broken a law. Without the code we could still recommend people to inflict pain on Jones to stop him doing what he does, but the

peculiar force of 'just' could not be carried by any word. And, granted that we have rules, it is clear that the *purpose* of punishment and blame is relevant, not to the question 'Should Jones be punished or blamed?', but to the question 'Should the sort of thing that Jones did be prohibited by a rule to which a penalty is attached?'.

The question 'What justifies punishment?' in fact conceals an ambiguity which is largely responsible for the dispute between those who answer it in terms of retribution for crime and those who answer it in terms of deterrence and reform. If we have in mind the judge's problem, the utilitarian answer is clearly inadequate; but if we are thinking about the legislator's problem, it seems very plausible. Each party has tried to extend their answer to cover both cases. But even if we are thinking about the legislator's problem, it would be an over-simplification to say that legislators either do or should decide what laws to have solely by reference to the purpose of having laws. There are two reasons why they do not do so, one bad and one good. The bad reason is that they are still to some extent in the thrall of the philosophical theory of Natural Law, which itself confuses the judge's problem with the legislator's. But there is also a good reason. It is desirable (on grounds of utility) that the law should be consistent and stable and that the penalty laid down for one offence should not be wildly out of line with those laid down for others. Consequently, unless we are to revise the whole legal code every time we make a new law, it is expedient not to consider the proposed law in isolation but to consider it as part of a system that we do not, on this occasion, wish to disturb.

Just as we might, but do not, live in a world in which every case was decided by an omnicompetent judge without reference to any general principle other than that of utility, so we might live in a world in which legislators decided what laws to pass solely by reference to this standard. But there are as good reasons for rejecting this system as there are for rejecting the system of judges not bound by laws. We know too little about the probable effects of any particular penalty and about the repercussions which a new law-cum-penalty is likely to have on other parts of the system. Hence even legislators do well to criticize proposed laws not only by reference to the purpose of having laws but by reference to the current system of laws, that is to say 'Justice'. The connexions between the justice of a punishment and its utility are thus exceedingly complex; but the fact that utilitarians have oversimplified

them is a reason, not for abandoning their theory or retreating into the asylum of intuition, but for revising the theory. It cannot be an accident that the punishments we call 'just' on the whole tend to reform and deter. And if in a particular case we find that they do not serve these ends, we tend to amend the law. On the Natural Law theory it would only be right to amend a law if we discovered that it conflicted with natural law. How we discover this is in any case a mystery, and it would be most remarkable if the discovery always went hand in hand with the discovery that the law fails to fulfil its purpose.

[2]

The most difficult and important of the items on our list of moral judgements is 'He could have acted otherwise' (2). The facts about its logical connexions with the others are tolerably clear. It is a necessary condition of all except (1) and it is also a necessary condition of (1) if 'He broke a law' is taken to imply that he broke it voluntarily. What is not so clear is what (2) means or why it should be a necessary condition of the other items.

A man is not considered blameworthy if he could not have acted otherwise; and, although it is often easy to decide in practice whether he could have acted otherwise or not, it is not clear how we do this or why we should think it necessary to do it. Let us first examine the use of 'could have' in some non-moral cases.

'Could have' is a modal phrase, and modal phrases are not normally used to make straight-forward, categorical statements. 'It might have rained last Thursday' tells you something about the weather, but not in the way that 'It rained last Thursday' does. It is sometimes said that it is used to express the speaker's ignorance of the weather; but what it expresses is not just this but his ignorance of any facts that would strongly tend to rule out the truth of 'It rained'. It would be a natural thing to say in the middle of an English, but not of a Californian summer. But, whatever it does express, what it does *not* express is a belief in a third alternative alongside 'it rained' and 'it did not rain'. Either it rained or it did not; and 'it might have rained' does not represent a third alternative which excludes the other two in the way that these exclude each other.

But these modal phrases are also sometimes used in cases in which

they cannot express ignorance since they imply a belief that the event concerned did not occur. It would be disingenuous for a rich man to say 'I might have been a rich man'; but he could well say 'I might have been a poor man' while knowing himself to be rich. The puzzle here arises from the fact that, if he is rich, he cannot be poor. His actual riches preclude his possible poverty in a way that would seem to imply that we could have no use for 'he might have been poor'. But this is only puzzling so long as we try to treat these modal expressions in a categorical way.

'Would have' and 'might have' are clearly suppressed hypotheticals, incomplete without an 'if . . .' or an 'if . . . not . . .'. Nobody would say 'Jones would have won the championship' unless (a) he believed that Jones did not win and (b) he was prepared to add 'if he had entered' or 'if he had not sprained his ankle' or some such clause.

It is not so obvious that 'could have' sentences also express hypotheticals; indeed in some cases they obviously do not. If a man says 'It could have been a Morris, but actually it was an Austin', it would be absurd to ask him under what conditions it could or would have been a Morris. 'Could have' is here used to concede that, although I happen to know it was an Austin, your guess that it was a Morris was not a bad one. But 'could have' also has a use which is more important for our purpose and in which, as I shall try to show, it is equivalent to 'would have . . . if . . .'. It refers to a tendency or capacity. Consider the following examples:

(1) He could have read *Emma* in bed last night, though he actually read *Persuasion*; but he could not have read *Werther* because he does not know German.

(2) He could have played the *Appassionata*, though he actually played the *Moonlight*; but he could not have played the *Hammerklavier*, because it is too difficult for him.

These are both statements, since they could be true or false; and to understand their logic we must see how they would be established or rebutted. Neither could be established or rebutted in the way that 'He read *Persuasion*' could, by observing what he actually did; and it is partly for this reason that we do not call them categorical. But, although they could not be directly verified or falsified by observation of what he did, this might be relevant evidence. It would be almost conclusive evidence in the first case, since it would be very odd if a man who actually read *Persuasion* was incapable of reading *Emma*. On the other hand, his having played the *Moonlight* is only weak evidence that he could have played the *Appassionata*, since the latter is more difficult and also because he might never have learnt it.

In each of these cases, in order to establish the 'could have' statement we should have to show (a) that he has performed tasks of similar difficulty sufficiently often to preclude the possibility of a fluke, and (b) that nothing prevented him on this occasion. For example we should have to establish that there was a copy of *Emma* in the house.

Statements about capacities, whether of the 'can' or of the 'could have' kind, contextually imply unspecified conditions under which alone the person might succeed; and 'could have' statements can be refuted either by showing that some necessary condition was absent (there was no copy of *Emma*) or by showing that the capacity was absent. The first point could be established directly. How could the second be established? In practice we do this either by appealing to past performances or failures or by asking him to try to do it now. It is clear that neither of these methods could be applied directly to the occasion in question. We know that he did not read *Emma*, and it is nonsense to ask him to try to have read *Emma* last night. And the very fact that evidence for or against 'could have' statements must be drawn from occasions other than that to which they refer is enough to show that 'he could have acted otherwise' is not a straightforward categorical statement, at least in the type of case we have been considering. Whether it is possible or necessary to interpret it categorically in moral cases is a point which I shall examine in the next section.

It might be argued that the sort of evidence by which 'could have' statements are supported or rebutted is never conclusive; and this is true. The argument used is an inductive one, with a special type of conclusion. We might use an ordinary inductive argument to predict his future performance from known past performances or in support of a statement about an unknown past performance. But in this special case we know that he did not do the thing in question, because we know that he did something else; so we put our conclusion in the form 'he could have done X'.

Whatever the evidence, it is always open to a sceptic to say "I know he has always succeeded (failed) in the past; but he *might* have failed (succeeded) on this occasion". Now this sort of scepticism is

not peculiar to 'could have' statements; it is one variety of general scepticism about induction. It is *possible* that if I had tried to add 15 and 16 last night (which I did not) I should have failed; but it is also possible that if I tried now I should fail. Our use of 'could have' statements, like our use of predictions and generalizations, always ignores such refined scepticism; and it would be absurd to try to base either freedom or responsibility on the logical possibility of such contingencies. In practice we ignore the sceptic unless he can produce reasons for his doubt, unless he can say why he believes that a man who has always succeeded might have failed on just that occasion. If no such reason is forthcoming we always allow inductive evidence which establishes the existence of a general capacity to do something to establish also the statement that the man could have done it on a particular occasion. Nor is this practice due to the fact that (the world being what it is) we are unfortunately unable to find better evidence and must fall back on probabilities. Our practice lies at the heart of the logic of 'can' and 'could have'. For the sceptic is, here as elsewhere, asking for the logically impossible; he is asking us to adopt a criterion for deciding whether a man could have done something on a particular occasion which would make the words 'can' and 'could have' useless. What would be the result of accepting this suggestion? We should have to say that the only conclusive evidence that a man can do (could have done) X at time t is his actually doing (having done) X at time t. Thus the evidence that entitles us to say 'He could have done X at time t' would also entitle us to say 'He did X at time t', and the 'could have' form would be otiose.

Capacities are a sub-class of dispositions. To say that a man 'can' do something is not to say that he ever has or will; there may be special reasons why the capacity is never exercised, for example that the occasion for exercising it has never arisen. A man might go through his whole life without ever adding 15 and 16; and we should not have to say that he couldn't do this. Yet a man cannot be said to be able to do something if all the necessary conditions are fulfilled and he has a motive for doing it. It is logically odd to say "Smith can run a mile, has had several opportunities, is passionately fond of running, has no medical or other reasons for not doing so, but never has in fact done so". And, if it is true that this is logically odd, it follows that 'can' is equivalent to 'will . . . if . . .' and 'could have' to 'would have . . . if . . .'. To say that Smith could have read *Emma* last

night is to say that he would have read it, if there had been a copy, if he had not been struck blind, etc., etc., and if he had wanted to read it more than he wanted to read anything else. Both the 'etc.' and the last clause are important; we cannot specify all the necessary conditions; and, granted that the conditions were present and that he could have read it, he might still not have read it because he did not want to. But if he did not want to do anything else more than he wanted to read *Emma*, he could not in these conditions be said to have *chosen* to do something else. He might have *done* something else, but not in the important sense of 'done' which implies choosing.

[3]

Libertarianism. Before considering why 'he could have acted otherwise', interpreted in this hypothetical way, is regarded as relevant to ascriptions of responsibility, it is necessary to examine the theory that, although the hypothetical interpretation is correct in most cases, in the special case of moral choice the phrase must be interpreted in a categorical way. It would indeed be remarkable if modal forms which are normally used in a hypothetical way were used categorically in one type of case alone; and I have already suggested that their logic is partly determined by the method that would be used to support or rebut statements which employ them. The thesis that 'he could have acted otherwise' is categorical is equivalent to the thesis that it could be verified or falsified by direct observation of the situation to which it refers.

It is essential to notice that the categorical interpretation is supposed to be necessary only in a very small, but very important part of the whole range of human choice. And this too is remarkable; for it implies that the words 'free' and 'choose' are logically different in moral and in non-moral cases. There is a sense of 'free' to which I have already alluded in which it is contrasted with 'under compulsion'; and in this sense actions are still free when they are completely determined by the agent's tastes and character. For to say that they are determined in this way is not to say that he is a Pawn in the hands of Fate or a Prisoner in the iron grip of Necessity. It is only to say that anyone who knew his tastes and character well enough could predict what he will do. The fact that we can predict with a high degree of probability how Sir Winston Churchill will

vote at the next election does not imply that he does not cast his vote freely. To be 'free' in this sense is to be free to do what one wants to do, not to be able to act in spite of one's desires.

According to the theory to be examined most of our voluntary actions are 'free' only in this sense which implies no breach in causal continuity. I choose what I choose because my desires are what they are; and they have been moulded by countless influences from my birth or earlier. But, it is said, *moral* choices are free in a quite different sense, and one that is incompatible with their being predictable. This unpredictability is an essential feature in the categorical interpretation of 'he could have acted otherwise'; for, if anyone could predict what I am going to do, I should not really be choosing between genuinely open alternatives, although I might think I was.

Professor Campbell puts the contrast in the following way: "Freewill does not operate in those practical situations in which no conflict arises in the agent's mind between what he conceives to be his 'duty' and what he feels to be his 'strongest desire'. It does not operate here because there is just no occasion for it to operate. There is no reason whatever why the agent should here even contemplate choosing any course other than that prescribed by his strongest desire. In all such situations, therefore, he naturally wills in accordance with his strongest desire. But his 'strongest desire' is simply the specific expression of that system of conative and emotive dispositions which we call his 'character'. In all such situations, therefore, whatever may be the case elsewhere, his will is in effect determined by his character as so far formed. . . ."

. . . (On the other hand) "in the situation of moral conflict, I, as agent, have before my mind a course of action, X, which I believe to be my duty; and also a course of action, Y, incompatible with X, which I feel to be that which I most strongly desire. Y is, as it is sometimes expressed, 'in the line of least resistance' for me – the course which I am aware that I should take, if I let my purely desiring nature operate without hindrance. It is the course towards which I am aware that my *character*, as so far formed, naturally inclines me. Now, as actually engaged in this situation, I find that I cannot help believing that I *can* rise to duty and choose X; the 'rising to duty' being affected by what is commonly called 'effort of will'. And I further find, if I ask myself just what it is I am believing when I believe that I 'can' rise to duty, that I cannot help believing that it lies with me, here and now, quite absolutely, which of two genuinely open possibilities I adopt; whether, that is, I make the effort of will and choose X or, on the other hand, let my desiring nature, my character as so far formed, 'have its way', and choose Y, the course in the line of least resistance."[1]

Now it is certainly true that many determinists have paid too little attention to the concept of 'trying' or 'making an effort'; but I think that there are certain difficulties in Professor Campbell's account of moral conflict and, in particular, in his attempt to construe 'I could have acted otherwise' in a categorical way. The first point to which I wish to draw attention is the question of method.

(1) Campbell insists that the question whether choice is 'free' in a contra-causal sense must be settled by introspection.[2] But is this so? To doubt the findings of his self-examination may seem impertinent; but the doubt is concerned, not with what he finds, but with the propriety of the language he uses to describe what he finds. The universal negative form of statement ('Nothing caused my decision,' 'No one could have predicted my decision') does not seem to be a proper vehicle for anything that one could be said to *observe* in self-examination. That I know introspectively what it is like to choose may be true; but I cannot be said to know introspectively that my choice was contra-causal or unpredictable; and this is the point at issue. He represents 'I can rise to duty' as a report of a mental event or, perhaps, a state of mind, not as a statement about a capacity, and 'I could have . . .' as a statement about a past state of mind or mental event. But, if this is really so, it is at least surprising that, in this one context alone, we use the modal words 'can' and 'could have' for making categorical reports. The issue between determinists and libertarians is an issue about the way in which expressions such as 'choose', 'can', and 'alternative possibilities' are to be construed; and this is surely an issue which is to be settled not by self-observation but by logical analysis.

There are many other phrases in Campbell's account which give rise to the same doubts about the propriety of the introspective method. The phrase 'conative disposition' is embedded in a large and complex mass of psychological theory and its use implies the acceptance of this theory; so that one could hardly be said to know by introspection that one has a conative tendency to do something.

[1] *Mind*, 1952, pp. 460–3.

[2] *Scepticism and Construction*, p. 131.

And phrases such as 'determined', 'contra-causal', and even 'desiring nature' take us beyond psychology into metaphysics. To say this is not to condemn the phrases; perhaps metaphysics is just what is needed here. But a metaphysician is not a reporter; he is an interpreter of what he 'sees'; and it is over the interpretation that the disputes arise.

(2) A more obvious difficulty – and it is one of which libertarians are well aware – is that of distinguishing a 'free' action from a random event. The essence of Campbell's account is that the action should not be predictable from a knowledge of the agent's character. But, if this is so, can what he does be called *his* action at all? Is it not rather a *lusus naturae*, an Act of God or a miracle? If a hardened criminal, bent on robbing the poor-box, suddenly and *inexplicably* fails to do so, we should not say that he *chose* to resist or deserves *credit* for resisting the temptation; we should say, if we were religious, that he was the recipient of a sudden outpouring of Divine Grace or, if we were irreligious, that his 'action' was due to chance, which is another way of saying that it was inexplicable. In either case we should refuse to use the active voice.

The reply to this criticism is that we must distinguish *In*determinism from *Self*-determinism. Choice is a creative act of the 'self' and is not only unconstrained by external forces but also unconstrained by desire or character. But the difficulty here is to construe 'self-determinism' in such a way that the 'self' can be distinguished from the 'character' without lapsing into indeterminism.

If we could construe 'self-determined' by analogy with other 'self'-compounds, such as self-adjusting, self-regulating, self-propelled, self-centred, self-controlled, and self-governing, there would be no difficulty. Some of these words apply to non-human objects, and they never imply that there is a part of the object called the 'self' which adjusts, regulates, or controls the rest, though the object does have a special part without which it would not be self-adjusting, etc. I can point to the self-starter of a car, but not to the self that starts the car; to say that a heating system is 'self-regulating' is to say that it maintains a constant temperature without anyone watching the dials and turning the knobs. Coming to the human scene, to say that a state is 'self-governing' is to say that its inhabitants make their own laws without foreign intervention; and to say that a man is 'self-centred' is to say, not that he is always thinking and talking about something called his 'self', but that he is always thinking and talking about *his* dinner, *his* golf-handicap, the virtues of *his* wife, and the prowess of *his* children. In each case there is a subject and an object; but the 'self' is neither subject nor object.

But if we construe 'self-determined' in this way, it is clear that being self-determined implies only that a man acts freely in the ordinary sense of 'freely' which the libertarian rejects as inadequate in the special case of moral choice. There would be no incompatibility between an action's being 'self-determined' and its being predictable or characteristic of the agent; for 'self-determined' would mean 'determined by *his* motives and character', as opposed to 'forced on him by circumstances or other people'. But the libertarian regards explanation in terms of character as incompatible with genuine freedom and must therefore draw a contrast between 'the self' and 'the character'. But if 'self-determined' is to mean 'determined by the self', it is necessary to give some account of what the 'self' is. And if the question whether an action was determined by the 'self' or not is to be relevant to the ascription of responsibility and the justice of adverse verdicts, we must be able to provide some criterion for deciding whether the self which determined the action is the same self that we are proposing to hold responsible or condemn.

Now the problem of Personal Identity is admittedly a difficult one and the danger of desert-island argument is particularly acute here, since Jekyll-and-Hyde cases that a layman would dismiss as flights of fancy have been known to occur. In fact we decide whether the man I met yesterday is the same that I met last year partly by seeing whether he looks the same, partly by observing an identity of characteristic behaviour, and partly by discovering what he can remember. And if we are to avoid the rather crude course of defining 'same self' in terms of the spatio-temporal continuity of bodily cells, it seems that we must define it in terms of character and memory. But the libertarian's 'self' is neither an empirical object nor displayed in characteristic action.

(3) If it is necessary to decide whether or not a man could have acted otherwise before ascribing responsibility, it is necessary that we should have some criterion for deciding this; and on the libertarian theory such a criterion is quite impossible. For, let us suppose that we know a great deal about his character and also that the temptation which he faced seems to be a fairly easy one for such a man to overcome. On the libertarian hypothesis this information will not be sufficient to enable us to conclude that he could have

acted otherwise. If he in fact does the wrong thing, there are three alternative conclusions that we might draw. (a) The action was not against his moral principles at all, so that no conflict between 'duty' and 'inclination' arose. This is what I have called 'wickedness'; (b) he knew it was wrong and could have resisted the temptation but did not (moral weakness); (c) he knew it was wrong but the temptation was *too* strong for him; he *could* not overcome it (addiction). Now it is essential to be able to distinguish case (b) from case (c), since (b) is a culpable state while (c) is not. By treating 'he could have acted otherwise' in a hypothetical way, the determinist thesis does provide us with a criterion for distinguishing between these cases; but the categorical interpretation cannot provide one, since no one, not even the man himself, could know whether he could have overcome the temptation or not.

(4) The libertarian theory involves putting a very special construction on the principle that 'ought' implies 'can', which it is very doubtful whether it can bear. If we take this principle in a commonsense way it is undoubtedly true. It is no longer my duty to keep a promise, if I literally *cannot* do so. But when we say this we have in mind such possibilities as my being detained by the police or having a railway accident or the death of the promisee; and it is possible to discover empirically whether any of these exonerating conditions obtained. But if 'cannot' is construed in such a way that it covers my being too dishonest a person or not making the necessary effort, it is no longer obvious that 'ought' implies 'can'. These reasons for failure, so far from exonerating, are just what make a man culpable.

(5) Even if it were possible to discover whether or not a man could have acted otherwise by attending to the actual occasion, as the categorical interpretation insists, why should this be held relevant to the question whether or not he is to blame? I shall try to explain this connexion in the next chapter; but on the libertarian hypothesis it will, I think, be necessary to fall back on insight into a relation of fittingness between freedom and culpability.

[4]

The Concept of 'Trying'. It might be thought that the libertarian could discover a criterion for distinguishing culpable weakness of will from non-culpable addiction in the concept of 'trying'. For the addict fails, try as he may, while the weak-willed man fails because he does not try hard enough. The concept of 'trying' is an important one for ethics since, whatever may be the case in a court of law, the question of moral blameworthiness often turns, not on what the agent did, but on what he tried or did not try to do. Morally we blame people, not for failing to live up to a certain standard, but for not trying hard enough to do so; and this is because, while we do not believe that they could always succeed, we do believe that they could always try. We must now see whether the introduction of this concept helps to save the categorical analysis.

We all know what it *feels* like to make an effort. These feelings are phenomena or occurrences that we experience in the same sort of way that we experience aches, pains, qualms, and twinges. And, if we take the introspective language of the libertarian seriously, it would seem that the question 'Did he try?' can be answered only by the man himself and that he answers it by observing whether or not one of these feelings occurred. The logical status of this question will be like that of 'Did it hurt?'. But on this view an effort is not something that a man *makes*; it is something that *happens* to (or inside) him; and it would be highly unplausible to make the question of his responsibility turn on the occurrence or non-occurrence of such a feeling. If 'making an effort' is to be relevant to responsibility, it must be thought of as something which a man can choose to do or not to do. The substitution of the active for the passive voice is an important advance; unfortunately it is fatal to the categorical interpretation of 'he could have acted otherwise'.

For 'trying' is now thought of as something that a man can choose to do or not to do, and the difficulties encountered in construing 'he could have acted otherwise' will emerge again in construing 'he could have tried to act otherwise'. On the libertarian analysis, if a man fails to act rightly, we must say either that his failure is inexplicable or that it was due to circumstances beyond his control – in which cases he is blameless – or that it was due to his not having tried as hard as he could have tried. For what exonerates is not 'I tried', but 'I tried as hard as I could'; and, in order to distinguish the blameworthy man from the addict who literally couldn't help it because he tried as hard as he could, we must be in a position to answer the question 'Could he have tried harder than he did?'. But how can we answer this question? *Ex hypothesi* he did not try harder than he

did; so that we must say either that his failure to try harder is inexplicable or that it was due to circumstances beyond his control – in which cases he is blameless – or that it was due to his not having tried to try as hard as he could have tried to try.

But this is absurd. In the first place 'try to try' is meaningless; and, if this be doubted, we must push the analysis one stage further. In fact he did not try to try harder than he did. But can he be justly blamed for this? Only if he could have tried to try harder. We must say either that his not having tried to try harder is inexplicable or that it was due to circumstances beyond his control – in which cases he is blameless – or that he failed to try to try harder because he did not try to try to try harder . . . and so on.

Libertarians sometimes speak in terms of our failure to make the best use of our stock of "will-energy"; but this usage gives rise to the same infinite regress. If using will-energy is thought of as something that we do not choose to do, but which just happens to us, it would appear to be irrelevant to responsibility; but if it is something that we can choose to do or not to do, we must be able to distinguish the man whose failure to use sufficient will-energy was due to circumstances beyond his control from the man who failed (culpably) to use it because he did not try hard enough to use it. And this involves answering the question 'Had he sufficient second-order will-energy to enable him to make more use of his first-order will-energy?'.

On these lines there is clearly no way out of the wood. The attempt to discover one is, I think, due to two mistakes. (a) It is noticeable that, on Campbell's analysis, a man's desires and even his character are continually referred to as 'it'; desires are thought of as forces which, sometimes successfully and sometimes unsuccessfully, prod a man into doing what he ought not, and his "character as so far formed" is the sum of these forces. Thus I am said to be able to choose whether or not to "let my desiring nature, my character as so far formed, have *its* way". And this is to treat all cases of 'doing what I want to do' on the model of the opium-addict, as the actions of a man who is a slave to his desires.

And since Campbell uses 'desire' for every motive except the sense of duty, his treatment presupposes that I can choose whether to act from a certain motive or not; and this is not so. If I am both hungry and thirsty I can choose whether to have a meal or a drink; but I cannot choose whether to act from hunger or thirst, unless this strange phrase is used simply as a (very misleading) synonym

for 'choosing whether to eat or to drink'. In the same way, if I have a certain sum of money, I can choose whether to pay a debt or give my aunt a Christmas present. If I choose the former, my motive is conscientiousness; if the latter, it is generosity. And we might, therefore, say that I can choose whether to do the conscientious or the generous thing. But I cannot choose whether to act from conscientiousness or from generosity. What I do will depend on my character; and this 'cannot choose' is not a lamentable restriction on my freedom of action. For to say that my choice depends on my character is not to say that my character compels me to do what I do, but to say that the choice was characteristic of me. The creative 'self' that sits above the battle of motives and chooses between them seems to be a legacy of the theory that a man is not free when he does what he wants to do, since he is then the victim or slave of his desires; and it is postulated to avoid the unplausible doctrine that all action is involuntary.

(b) Campbell takes as a typical and, by implication, the only case of moral choice to which appraisals are relevant, that of a man who knows what he ought to do but is tempted to do something else. Now this, so far from being the only case, is not even the commonest or most important. For in the great majority of cases of moral difficulty what is difficult is not to decide to do what one knows he ought to do, but to decide what one ought to do. This sort of difficulty arises in three main types of case. (i) A humble and unimaginative person who accepts a customary code of morals without much question may find that two rules conflict; the voice of conscience is in this case ambiguous. (ii) A more self-confident, imaginative, and reflective person may wonder whether he ought, in the case before him, to do what the customary rule enjoins. He knows very well what the rule enjoins; but what prompts him to depart from it is not "part of his desiring nature", but a suspicion that the rule is one that, in this particular case, he ought not to follow. (iii) A man of fixed moral principles (whether or not they are those customarily adopted) may find himself in a radically new situation that is not catered for in his code. What is he to do? It is here, if anywhere, that the idea of an unpredictable 'creative' choice seems to make sense. He takes a leap in the dark, but just because it is a leap in the dark I doubt if we should be inclined to blame him if he leapt in what turned out to be the wrong direction.

Men who belong to a generation for whom the questioning of

accepted principles has been no mere academic exercise and who have found themselves faced with momentous choices in situations not covered by their traditional rules will be less likely than their fathers perhaps were to suppose that the only sort of moral difficulty is that of resisting temptation.

If, in the first two of these three cases, a man decided that he ought to do something and did it, he might still be held to blame. For reasons given in chapter 17 conscientiousness is so valuable a motive that we should be chary of blaming a man who did what he honestly thought he ought to do, however misguided we thought him. But we should not necessarily excuse him, which we should have to do if all wrong-doing were failure to resist temptation. Integrity is not the only moral virtue, any more than it is the only virtue in an artist; and the belief that it is is one of the more regrettable consequences of the Romantic Movement. We blame people, not only for failing to live up to their moral principles, but also for having bad moral principles; and I shall examine the logic of this type of blame in the next chapter.

Perhaps the most crucial objection to the libertarian thesis lies in the sharp discontinuity which it presupposes between moral and non-moral choice and between moral and non-moral appraisal. It is not enough to admit that we can, within broad limits, predict what a man of known habits, tastes, and interests will do and to insist that our powers of prediction only break down in the small, but important area of moral choice. For it is not the extent of the area open to prediction that is at issue.

It is true that we can, within broad limits, predict what a man will choose from a menu, whether he will make a century to-day, or finish his cross-word puzzle; but we can also predict, again within broad limits only, whether or not he will resist the temptation to run away or to cheat at cards. Our reliance on the integrity of a bank clerk is not different from our reliance on his accuracy. In neither case do we believe that he 'must' or 'is compelled to' be honest or accurate; and what is paradoxical is not so much the libertarian's defence of moral freedom as his willingness to accept mechanical determinism as an explanation of non-moral action. For the rigid distinction between 'formed character' (where determinism reigns) and 'creative choice' (which is in principle unpredictable) it would be better to substitute a conception of continual modification of character in both its moral and its non-moral aspects. This not

only does justice to the fact that we use both choosing and appraising language in the same way in moral and non-moral contexts, but it is closer to the facts. A man can grow more or less conscientious as time goes on, just as he can become better at tennis or more fond of Mozart.

CHAPTER 20

Freedom and Responsibility (2)

[1]

IN the last chapter I tried to show that 'could have' sentences in non-moral contexts can be analysed in terms of 'would have . . . if . . .'; and we must now see whether the application of this analysis to moral cases is consistent with our ordinary use of moral language.

The first question to be considered is the question: what sorts of if-clauses are in fact allowed to excuse a man from blame. Clearly 'I could not have kept my promise because I was kidnapped' will exculpate me while 'I could not have kept my promise because I am by nature a person who takes promises very lightly' will not. Translated into the hypothetical form, these become respectively 'I would have kept my promise if I had not been kidnapped' and 'I would have kept my promise if I had been a more conscientious person'. Again it is clear that the first exculpates while the second does not. The philosophical difficulties, however, are to decide just why some 'would . . . ifs' excuse while others do not and to provide a criterion for distinguishing the exculpating from the non-exculpating cases. Forcible seizure exculpates; but do threats or psychological compulsion? And if, as some suggest, desires are internal forces which operate on the will, do they exculpate in the way in which external forces do? The problem of freewill is puzzling just because it seems impossible, without indulging in sheer dogmatism, to know just where to stop treating desires as 'compelling forces'.

Now before tackling this difficulty it will be prudent to examine what goes on in a place where questions of responsibility are settled every day and have been settled daily for hundreds of years, namely a court of law. Lawyers have evolved a terminology of remarkable flexibility, refinement, and precision and, although there may be a difference between moral and legal verdicts, it would be strange if the logic of lawyers' talk about responsibility were very different from our ordinary moral talk.

To establish a verdict of 'guilty' in a criminal case it is necessary

to establish that the accused did that which is forbidden by the law or, in technical language, committed the *actus reus*, and also that he had what is called *mens rea*. This last phrase is sometimes translated 'guilty mind' and in many modern textbooks of jurisprudence it is supposed to consist of two elements, (a) foresight of the consequences and (b) voluntariness. But, whatever the textbooks may say, in actual practice lawyers never look for a positive ingredient called volition or voluntariness. A man is held to have *mens rea*, and therefore to be guilty, if the *actus reus* is proved, *unless* there are certain specific conditions which preclude a verdict of guilty. "What is meant by the mental element in criminal liability (*mens rea*) is only to be understood by considering certain defences or exceptions, such as Mistake of Fact, Accident, Coercion, Duress, Provocation, Insanity, Infancy."[1] The list of pleas that can be put up to rebut criminal liability is different in different cases; but in the case of any given offence there is a restricted list of definite pleas which will preclude a verdict of guilty.

This is not to say that the burden of proof passes to the defence. In some cases, such as murder, it is necessary for the prosecution to show that certain circumstances were not present which would, if present, defeat the accusation. The essential point is that the concept of a 'voluntary action' is a negative, not a positive one. To say that a man acted voluntarily is in effect to say that he did something when he was not in one of the conditions specified in the list of conditions which preclude responsibility. The list of pleas is not exhaustive; we could, if we wished, add to it; and in making moral judgements we do so. For example we sometimes allow the fact that a man acted impulsively to exonerate him morally or at least to mitigate his offence in a case in which the law would not allow this. But it remains true that, in deciding whether an action was voluntary or not, we do not look for a positive ingredient but rather for considerations that would preclude its being voluntary and thereby exonerate the agent. In moral cases the most important types of plea that a man can put forward are (a) that he was the victim of certain sorts of ignorance, and (b) that he was the victim of certain sorts of compulsion.

[1] Professor H. L. A. Hart: *Proceedings of the Aristotelian Society*, 1948–9. Aristotle in effect defines 'the voluntary' in the same negative way as what is done not under compulsion and not through ignorance.

[2]

Ignorance. A man may be ignorant of many elements in the situation in which he acts. For example he may not know that it was a policeman who told him to stop, that the stuff he put in the soup was arsenic, that the money he took was not his own. In such cases he would be blamed only if it was thought that he ought to have known or taken the trouble to find out. And his vicious trait of character was not contumacy or callousness or greed or disregard for any moral principle, but carelessness; and carelessness can amount to a vice. Fire-arms are so notoriously dangerous that the excuse 'I didn't know it was loaded' will not do. The reason why he is blamed for carelessness and not for the specific vice for which he would have been blamed if he had done any of these things intentionally is that, although he intended to do what he did, he did not intend to break a moral rule. He intended to take the money, but not to steal. His action was not, therefore, a manifestation of the particular vice that the actions of thieves manifest. Ignorance of fact excuses or reduces the seriousness of an offence; but there is one type of ignorance that never excuses; and that is, in legal contexts, ignorance of the law and, in moral contexts, ignorance of right and wrong.

Now why should ignorance of fact excuse while ignorance of rules does not? Why should a man who takes someone else's money, thinking it to be his own, be guiltless of anything (except possibly carelessness), while a man who takes it, knowing it not to be his own but because he sees nothing wrong in taking other people's money, be held guilty and therefore blameworthy? We are not here concerned with the question why some types of action should be stigmatized as 'wrong', but solely with the question why ignorance of what is wrong should not be held to exculpate.

The reason is that while the man who thought the money was his own did not intend to act on the maxim 'It is permitted to take other people's money', the thief does act on this maxim. If a man does something because he does not think it wrong he cannot plead that he did not choose to do it, and it is for choosing to do what is *in fact* wrong, whether he knows it or not, that a man is blamed. The situation is exactly analogous to that in which some non-moral capacity is concerned. 'I would have solved the problem, if I had known all the data' would, if substantiated, allow me to get full

marks. But 'I would have solved the problem if I had known more mathematics' would not. Since competence at mathematics is not a moral trait of character, men are not blamed for lack of it; but they are given low marks and denied prizes.

[3]

Compulsion. So long as 'compulsion' is used in the literal sense it is not difficult to see why it should be held to exonerate. If a man is compelled to do something, he does not choose to do it and his action is not a manifestation of his moral character or principles. Now, since the purpose of blame and punishment is to change a man's character and principles, neither blame nor punishment is called for in such a case. It would be unjust to punish him since the rules for punishing lay down that a man who acts under compulsion is not to be punished; and the rules lay this down because, with due allowance for superstition and stupidity, we do not have pointless rules. Once more we must be careful to avoid the mistake of saying that the justice of a sentence turns on the question whether the accused is likely to be reformed by it. What is at issue here is not our reason for exonerating this accused, but our reason for making a *general* exception in the case of men whose actions are not expressions of their moral character. Physical compulsion is an obvious case where this is so.

But what if the source of compulsion is within the man himself? It is not an accident that we use 'compulsion' in a psychological way and exonerate compulsives. There are two questions that are relevant here. In the first place we ask whether the man could have resisted the 'compulsion'; and we decide this in the way that we decide all 'could have' questions. We look for evidence of his past behaviour in this, and also in related matters; for the behaviour of the compulsive is usually odd in matters unconnected with his special compulsion; and we compare his case with other known cases. Once the capacity to resist the compulsion is established beyond reasonable doubt we do not allow unsupported sceptical doubts about his capacity to resist it in a particular case to rebut the conclusion that he could have helped it. And we do not allow this because there is no way of establishing or refuting the existence of a capacity except by appeal to general evidence. If the capacity has been established and

all the necessary conditions were present, we would not say that, in this case, he was the victim of a compulsion. Indeed a 'compulsion' is not something that could be said to operate in a particular case only; for to say that a man has a psychological compulsion is to say something about his behaviour over a long period. A compulsion is more like a chronic disorder than like a cold; and it is still less like a sneeze.

It is also relevant to raise the question whether he had any motive for doing what he did. Part of the difference between a kleptomaniac and a thief lies in the fact that the former has no motive for what he does; and he escapes blame because the point of blame is to strengthen some motives and weaken others. We are sometimes inclined to take the psychologists' talk about compulsions too seriously. We think that a man is excused because he has a 'compulsion', as if the compulsion could be pointed to in the way that an external object which pushed him could be pointed to. But compulsions are not objects inside us; and we use the word 'compulsion', not because we have isolated and identified the object which caused him to do what he did, but because we want to excuse him in the same sort of way that we excuse someone who is literally pushed; and we want to excuse him for the same sort of reason. We know that it will do no good to punish him.

Desires. A man might plead that he would have acted otherwise if he had not had a strong desire to do what he did; but the desire was so strong that, as things were, he could not have acted otherwise. Would this plea be allowed to exonerate him? In some cases it would; for there are, as we have seen, cases of addiction in which we allow that a man is not to blame since his craving was too strong for him. But in most cases it would be considered frivolous to say 'I would have done the right thing if I hadn't wanted to do the wrong thing'; for it is just for this that men are blamed.

To distinguish an overwhelming desire from one that the agent could have resisted is not always easy; but the criterion that we in fact use for making the distinction is not difficult to understand. We know from experience that most men can be trained to curb some desires, but not others; and we assume that what is true in most cases is true in a given case unless special reasons are given for doubting this. Now it might seem that, although this evidence enables us to predict that we shall be able to train the man to curb his desire in future, it sheds no light on the question whether he could have curbed it on the occasion in question. I shall say more about this question of moral training later; here I only wish to point out that we have no criterion for deciding whether a man could have resisted a desire on a given occasion other than general evidence of his capacity and the capacity of others like him. We do not, because we cannot, try to answer this question as if it referred solely to the given occasion; we treat it as a question about a capacity.

Character. Finally a man might plead that he could not help doing what he did because that's the sort of man he is. He would not have done it if he had been more honest or less cowardly or less mean and so on. This sort of plea is paradoxical in the same sort of way that the plea of ignorance of moral rules and the plea that he did it because he wanted to are paradoxical. And all three paradoxes stem from the same source, the uncritical extension of 'ought implies can' and of the exculpatory force of 'he could not have acted otherwise' to cases which they will not cover. We know that these pleas are not in fact accepted; the puzzle is to see why.

The plea 'I could not help it because I am that sort of person' might be backed up by an explanation of how I came to be that sort of person. Just as the discovery of a compelling cause exonerates, so, it might be argued, to reveal the causes of my character being what it is is to show that I could not help being what I am and thus to exonerate me. But this argument is fallacious. In the first place to discover the cause of something is not to prove that it is inevitable. On the contrary the discovery of the cause of a disease is often the first step towards preventing it.

Now it is logically impossible to prevent something happening if we know the cause of it, since it could not have a cause unless it occurred and therefore it was not prevented. So when we talk of preventing diseases or accidents we are not talking about preventing cases which have occurred but about ensuring that there are no future cases. Similarly, if I know how Jones came to be a dishonest man I cannot prevent him from being dishonest now; but it may be possible to prevent others from becoming dishonest and to cure Jones of his dishonesty.

Secondly, the discovery of a cause of something has no necessary bearing on a verdict about that thing. We know that a man has come to be what he is because of three main types of cause, heredity, education, and his own past actions. These three factors are not independent of each other and it is not the business of a philosopher

to say exactly what is the effect of each or which is the most important for moral training. The question 'Granted that we want people to be better and that we have fairly clear ideas about what "being better" means, should we try to breed a superior race or pay more attention to education?' is not a philosophical question. But it is the business of a philosopher to show in what ways these 'causes' are related to responsibility.

Now these three factors also play a part in situations in which non-moral verdicts are given. Leopold Mozart was a competent musician; his son Wolfgang was given a good musical education and practised his art assiduously. Each of these facts helps to explain how he was able to compose and play so well. There is plenty of evidence that musical ability runs in families and still more of the effects of teaching and practice. But, having learnt these facts, we do not have the slightest tendency to say that, because Mozart's abilities were 'due' to heredity, teaching, and practice, his compositions were not 'really' his own, or to abate one jot of our admiration. In the same way, however a man came by his moral principles, they are still *his* moral principles and he is praised or blamed for them. The plea that, being what he is he cannot help doing what he does, will no more save the wicked man than it will save the bad pianist or actor who has the rashness to expose his incompetence in public. Nor is he saved by being able to explain how he has come to be what he is.

Hereditary tendencies are not causes and do not compel, although a man may inherit a tendency to some form of psychological compulsion. In general to say that a man has a tendency to do something is to say that he usually does it; and to add that the tendency is hereditary is to say that his father also used to do the same sort of thing; and neither of these facts has any tendency to exculpate.

The belief that heredity or a bad upbringing excuse a man's present character is partly due to the false belief that to explain something is to assign an antecedent cause to it and that, to be voluntary, an action must be uncaused. But there is also a good reason for this belief. In fact we do sometimes allow these factors to exculpate; and if the question of explanation was as irrelevant to the question of responsibility as I have suggested it would be hard to understand why we do this. Why do we tend to deal less harshly with juvenile delinquents who come from bad homes than with those who have had every chance? The question is not one of justice, since it is not a question whether Jones ought to be punished, but

whether the law should lay down that people whose bad characters are due to certain causes should be punished. We must therefore ask what is our reason for differentiating between two boys whose characters and actions are the same but who come respectively from bad and good homes. And the reason is that in the first case we have not had a chance to see what kindness and a good education could do, while in the second we know that they have failed. Since punishment involves the infliction of pain and since it is a moral rule that unnecessary pain should not be inflicted, there is a general presumption that people should not be punished if the same end could be achieved without the infliction of pain. This consideration is, of course, irrelevant to the question whether Jones should be punished; but it is highly relevant to the question whether a distinction should be made between those whose characters have come to be what they are because of a bad education and those whose characters are bad in spite of a good one.

But suppose a man should plead that he cannot now help doing what he does because his character was formed by his own earlier actions? This also will not excuse him. The logic of this plea is that he did X because he was, at the time, the sort of man to do X and that he became this sort of man because he did Y and Z in the past. But if he cannot be blamed for doing X now, can he be blamed for having done Y and Z in the past? It would seem that he cannot, for he will exculpate himself in exactly the same way.

Once again the argument presupposes that if his present character can be explained in terms of what happened in the past he necessarily escapes blame. The assumption is that a man's actions form a causal chain in which each necessitates the next. Now, if we suppose that, to be free, an action must be uncaused, either we shall find a genuinely uncaused action at the beginning of the chain or we shall not. If we do not, then no action is culpable; and if we do, then we must suppose that, while most of our actions are caused and therefore blameless, there was in the past some one uncaused action for which alone a man can be held responsible. This theory has in fact been held, although even in the history of philosophy it would be hard to find another so bizarre. The objections to it are clear. In the first place we praise and blame people for what they do now, not for what they might have done as babies; and secondly this hypothetical infantile action could hardly be said to be an action of the agent at all, since it is *ex hypothesi* inexplicable in terms of his character.

The conclusion of the foregoing argument is that 'He could not have acted otherwise' does not always exculpate and, in particular, that it does not exculpate if the reason which is adduced to explain just why he could not have acted otherwise is that he was a man of a certain moral character. We have seen that 'He could have acted otherwise' is to be construed as 'He would have acted otherwise, if . . .' and we have seen which types of 'if' are not allowed to exculpate. We must now see why they are not.

[4]

What is moral character? The key to the logical relationships between the five types of judgement seems to lie in the judgement of moral character (5). For (2) is thought to be a necessary condition of (1), (3), and (4) only because we exclude those cases of incapacity to act otherwise in which the incapacity lies in the moral character. If it is due to an external force or to a 'compulsion' (which we talk of as if it were an external force), or to some non-moral defect, the incapacity to act otherwise is allowed to excuse; but not if it is due to a moral defect. And it is now necessary to provide some criterion for deciding what a moral defect is.

Moral traits of character are tendencies or dispositions to behave in certain ways. How are they to be distinguished from other tendencies? If any tendency were to count as 'moral' we should have to say that conformity to physical laws was a universal trait of human character and that susceptibility to colds was part of the moral character of a particular man.

The first and most obvious limitation lies in the fact that the names of virtues and vices are not purely descriptive words. They are terms of praise and blame used to express approval and disapproval and to influence the conduct of the person whose character is appraised and also of others. These three functions are tied together in a way that should by now be familiar. Appraising, praising, and blaming are things that men *do* and can only be understood on the assumption that they do them for a purpose and use means adapted to their purpose. The logic of virtue- and vice-words is tailor-made to fit the purposes and conditions of their use.

Men would not employ a special form of speech for changing the character and conduct of others unless they had a pro-attitude

towards those changes; so that the first limitation that can be put on 'moral character' is that traits of character are tendencies to do things that arouse approval or disapproval. But moral verdicts do not just express the attitudes of the speaker; they are couched in impersonal language and imply accepted standards because the traits of character that a given man wants to strengthen or inhibit in others are usually those that other men also want to strengthen and inhibit. The impersonal language of morals implies a rough community of pro- and con-attitudes. Moreover men would not have adopted the moral language they have unless it was likely to achieve its purpose; and its purpose is achieved because most men dislike disapprobation. The power of moral language is greatly enhanced by the very facts which make impersonal moral language possible. No one likes to be universally condemned and most men are willing to take considerable pains to avoid it.

But this limitation is not enough. There are many things for which men are applauded and condemned which do not count as parts of their moral character. A great musician, mathematician, actor, or athlete is applauded and rewarded for what he does and his ability may be called a 'virtue', but not a moral virtue. Conversely, if a man fails to save a life because he cannot swim, we may regret his incapacity and urge him to learn, but his incapacity is not called a vice.

A man may fail to achieve some worthy object because he is physically or intellectually incompetent, too weak or too stupid. But he may also fail because he is too cowardly or too dishonest or has too little regard for the welfare of others. Why do we call the first set of traits 'non-moral' and never condemn them, while the second are called 'moral' and condemned? It is clear that it will not help to say that we intuit a non-natural relation of fittingness which holds between blameworthiness and dishonesty or meanness but not between blameworthiness and physical weakness or stupidity. For this is only to say that the former traits deserve blame while the latter do not and that we cannot understand why.

To discover why we draw the line in the way that we do we must first ask exactly where we draw it; and all that is necessary for this purpose is to construct two lists, the one of moral traits, the other of non-moral. Cowardice, avarice, cruelty, selfishness, idleness would go into the first list; clumsiness, physical weakness, stupidity, and anaemia into the second. The second list will, of course, contain

items of many different sorts, since we are interested, not in the way in which non-moral characteristics differ from each other, but in the distinction between moral and non-moral.

If we construct these lists we shall find that the items in list 1 have two properties in common which the items in list 2 do not have. (a) We believe that if a man's action can be explained by reference to a list 1 characteristic, he could have acted otherwise. And it would appear at first sight that this is the crucial feature which distinguishes moral from non-moral characteristics. Why does a schoolmaster punish a lazy boy but not a stupid one for equally bad work if not because he believes that the lazy boy could have done better while the stupid boy could not? But why does the schoolmaster believe this? In fact he appeals to the evidence of past performance. On the libertarian view this would scarcely be relevant, since the boy might not have been lazy in the past but was lazy at just that moment. And perhaps his momentary laziness was no more under his control than the stupid boy's stupidity? An analysis on these lines could hardly fail to lead to the paradoxical conclusion that no one has any reason whatever for ascribing responsibility. And even if it were possible to answer the question whether he could have acted otherwise, we should be left with the question why this is considered relevant to the propriety of holding him responsible.

Moreover it would be circular to make the phrase 'he could have acted otherwise' the distinguishing criterion of moral characteristics; since, as we have seen, it is necessary to make use of the distinction between actions explained by reference to moral, and actions explained by reference to non-moral characteristics in order to elucidate the phrase 'he could have acted otherwise'.

(b) There is, however, another element which all the characteristics in list 1 have and those in list 2 do not. It is an empirical fact that list 1 characteristics can be strengthened or weakened by the fear of punishment or of an adverse verdict or the hope of a favourable verdict. And when we remember that the purpose of moral verdicts and of punishment is to strengthen or weaken certain traits of character it is not difficult to see that this feature, so far from being synthetically connected with the notion of a 'moral' characteristic, a virtue or a vice, is just what constitutes it. What traits of character can be strengthened or weakened in this way is a matter of empirical fact. Knives can be sharpened, engines decarbonized, fields fertilized, and dogs trained to do tricks. And men also can be

trained, within certain limits, to behave in some ways and not in others. Pleasure and pain, reward and punishment are the rudders by which human conduct is steered, the means by which moral character is moulded; and 'moral' character is just that set of dispositions that can be moulded by these means. Moral approval and disapproval play the same role. It is not just an accident that they please and hurt and that they are used only in cases in which something is to be gained by pleasing or hurting.

We might therefore say that moral traits of character are just those traits that are known to be amenable to praise or blame; and this would explain why we punish idle boys but not stupid ones, thieves but not kleptomaniacs, the sane but not the insane. This is not to say that amenability to praise and blame is what justifies either of these in a particular case; that, as we have seen, is a question to be decided by reference to the rules. But a breach of a moral rule is only considered to be culpable when it is attributable to the agent's character, his vice or moral weakness; and our theory is intended to explain just what is included in and what excluded from 'moral character' and to explain why this distinction should be considered relevant to responsibility.

According to this explanation there is no need to postulate any special insight into necessary connexions between the five moral judgements with which we started; for the whole weight of the analysis is now seen to rest on the proposition that people only do those things which are either objects of a direct pro-attitude (i.e. that they want to do or enjoy doing for their own sake) or are believed to produce results towards which they have pro-attitudes. It is absurd to ask why a man who thinks that praise and blame will alter certain dispositions which he wishes to alter should praise and blame them. For this is a special case of the question Why do people adopt means that they believe to be the best means of achieving their ends?'; and this is an absurd question in a way in which 'Why does a man deserve blame only if he acted voluntarily and has broken a moral rule?' is not.

Nevertheless this way of tracing the connexions between pro-attitudes, moral rules, verdicts on character, and ascriptions of responsibility is obviously too simple and schematic. It is more like an account of the way in which moral language would be used by people who knew all the facts and thoroughly understood what they were doing than like a description of the way in which moral

language is actually used. In practice these connexions are much looser than the theory suggests; and there are two reasons for this. In the first place there is the inveterate conservatism of moral language. Even when it is known that a certain type of conduct, for example homosexuality, is not amenable to penal sanctions or moral disapproval, it is difficult to persuade people that it is not morally wrong.

The second reason is more respectable. We are still very ignorant of the empirical facts of human nature, and this ignorance both makes it wise for us to make moral judgements in accordance with a more or less rigid system of rules and also infects the logic of moral language. Our moral verdicts do not, therefore, always imply that the person condemned has in fact done something 'bad' or 'undesirable' in a non-moral sense. An act of cowardice or dishonesty might, by chance, be attended with the happiest consequences; but it would still be blamed. But this fact does not involve any major modification in the theory that bad traits of character are those which (a) tend to bring about undesirable results in most cases and (b) are alterable by praise and blame. For, in deciding whether a trait of character is vicious or not, we consider its effects in the majority of cases. We do not want to reinforce a tendency to behave in a certain way just because it turns out, on rare occasions, to be beneficial. And, in making a moral judgement, we do not consider the actual consequences of the action concerned. Nor do we even need to consider the consequences that such actions usually have. A man has broken faith or been cowardly or mean; we condemn him forthwith without considering why such actions are condemned. The fact that deceitful, cowardly, and mean actions are, by and large, harmful is relevant, not to the questions: 'Has Jones done wrong? Is he a bad man? Does he deserve to be blamed?', but to the question 'Why are deceitfulness, cowardice, and meanness called "vices" and condemned?'.

This theory enables us to understand why it is not only moral weakness that is blamed, but also wickedness; and it also enables us to distinguish between moral weakness and addiction in a way that the libertarian theory could not. A wicked character can be improved by moral censure and punishment; and if we really thought that a man was so bad as to be irremediable we should, I think, cease to blame him, though we might impose restraints on him as we would on a mad dog. Moral weakness is considered to be a less culpable

state, since the morally weak man has moral principles which are good enough, but fails to live up to them. He is therefore more likely to be improved by encouragement than the wicked man is. What he needs is the confidence which comes from knowing that others are on the side of his principles. But both he and the wicked man differ from the addict or compulsive in that the latter will respond neither to threats nor to encouragement.

[5]

Moral Principles. Traits of character, then, are dispositions to do things of which a spectator (including the agent himself) approves or disapproves and which can be, if not implanted or wholly eradicated, at least strengthened or weakened by favourable and adverse verdicts. But they are dispositions to *do* things, in the active sense of 'do', dispositions to choose certain courses of action. It is not, therefore, an accident that the names of virtues and vices, such as 'generosity' and 'avarice', are motive-words which necessarily imply a pro-attitude towards doing the things called 'generous' or 'greedy' for their own sake. And since moral principles are also dispositions to choose, they also must be classed as 'pro-attitudes'. How do they differ from other pro-attitudes?

(a) In the first place a pro-attitude does not count as a moral principle unless it is a relatively dominant one and concerned with an important matter. However regularly I choose to drink coffee for breakfast no one would call this disposition to choose one of my moral principles. To act on principle is consistently to pursue a policy of doing certain sorts of things for their own sake; and for this reason 'acting on principle' must be sharply distinguished from 'acting from a sense of duty', although we shall see later that the two are connected. The reason for distinguishing them is that to act from a sense of duty is consistently to pursue a policy of obeying certain rules for the sake of obeying those rules; it is therefore a special case of acting on principle. 'Acting on principle' cannot, therefore, be identified with either the 'sense of duty' or the 'impulses' which, according to some philosophers, are the only types of motive. It is distinguished from 'acting on impulse' by regularity and consistency and from 'acting from the sense of duty' by the fact that the man who acts on principle does what he does for its own sake.

Now since a moral principle is a disposition to choose, a man cannot be said to have a certain moral principle if he regularly breaks it, and we discover what a man's moral principles are mainly by seeing how he in fact conducts himself. But this is not the only test. A man's moral principles are 'dominant' in the sense that he would not allow them to be over-ridden by any pro-attitude other than another moral principle. Thus a man may belong to many organizations and be allowed by the laws of his country to do something that he is not allowed to do by the rules of his trade union, profession, or church. When a conflict of principles or loyalties arises he may wonder what he ought to do; but it is part of the force of the phrase 'moral principles' that he cannot (logically) wonder what he ought to do if there is a moral principle on one side and not on the other. If I regard something as immoral, then, however trivial it may be and however great may be the non-moral advantages of doing it, I cannot debate with myself whether I ought to do it; and we discover what our own moral principles are very often by putting just this sort of question to ourselves.

A similar limitation in the use of the phrase 'moral principle' comes out in our attitude to compensation. A man will not lightly give up a moral principle; nor will he lightly give up anything else that he regards as valuable. But our attitude towards giving up a moral principle differs from all other cases. If a man has a picture that he values very highly he may reject a low price and be more inclined to part with it if the bid is raised. But if a man refuses a bribe of ten pounds and you offer him a hundred, he might say: "You don't understand; it is not a question of how much; doing that sort of thing is against my moral principles". Indeed he must say this, if it is really a matter of moral principle, unless he can manage to bring the acceptance of the offer under some other moral principle. It is for this reason that Napoleon's dictum that every man has his price sounds so cynical; it implies that no man has any moral principles.

(b) But consistency in action is not the only test of a man's moral principles. Although a man cannot claim that it is against his moral principles to be cowardly or mean if he regularly does cowardly or mean things, he can do such things occasionally and still justify this claim. His claim is justified if he is prepared to condemn his own actions and if he feels remorse. His moral principles are not those on which he always acts, but those which he acknowledges or avows

and those about which he feels remorseful when he breaks them. His moral principles are those on which, in his more reflective moments, he honestly says that he would like to act; they are the moral principles of the person he is striving to become. I shall return to this point in the last section of this chapter.

(c) A principle is not usually called a moral one unless the person who adopts it is prepared to apply it universally. If a man says that he does something as a matter of principle, he cannot (logically) make exceptions unless another moral principle is involved. However narrow in scope it may be, a moral principle must be applied to all cases that are alike in all relevant respects. If there are two people of roughly similar character, tastes, and habits, it may well be that a man likes one of them better than the other. If asked why, he may be unable to give a reason; he just happens to like Jones, although he concedes that Smith is just as virtuous, charming, and amusing. And, although there is an oddity about his taste that might interest a psychologist, there is nothing logically odd about it. But he is abusing language if he says that it is a matter of moral principle with him to pay his debts and he pays Jones, while refusing to pay Smith, without being able to give any reason for the discrepancy.

The logical fact that a pro-attitude is not called a 'moral principle' unless a man is prepared to universalize it has led some philosophers to suppose that it can be proved that we ought to be impartial. But this is to commit the fallacy of deducing a moral injunction from a feature of moral language. A man who has no principles that he is prepared to apply impartially has no moral principles; but we cannot prove that he ought to have any moral principles by pointing out how the phrase 'moral principles' is used.

(d) The fact that a man's moral principles are those which he acknowledges in his more reflective moments throws some light on the connexion between moral principles and rules. A man's moral principles are those on which he thinks he ought to act and the word 'ought', like all deontological words, is only used in connexion with rules and therefore in connexion with relatively long-range principles and policies that we avow and adopt in our more reflective moments.

Moreover these deontological words contextually imply a background of general agreement; so that, in deliberating about what to do, we tend to use the language of 'ought' only in connexion with principles of action that we know to be generally approved. Now,

for reasons given in chapter 14, moral codes never contain injunctions to people to pursue their own pleasure; and most moral rules are concerned with the welfare of others. These pervasive features of moral codes infect the logic of deontological words. It is odd to describe a man as a 'conscientious egoist' or to say that pleasure-seeking is his highest moral principle, because people do not in fact use the language of 'ought' when they are being deliberately and consistently selfish. And the reason for this is that it is hard to dissociate this word from its moorings in the language of advice, exhortation, and command. Nevertheless, if a man regularly decides that he ought (in the verdict-giving sense of 'ought') to do whatever brings him pleasure or profit, his dominant pro-attitude is towards his own pleasure or profit. Whether or not we choose to call selfishness a moral principle with him, depends on the criterion we are using for the phrase 'moral principle'. If he behaves selfishly without acknowledging his wickedness and without feeling remorse, we could say that selfishness was one of his moral principles; and we hesitate to say this partly because he almost certainly does not address himself in the language of 'ought' (in the self-hortatory sense) and partly because we are reluctant to believe that he really is what he makes himself out to be.

[6]

Can a man choose to act against his own moral principles or choose to change them? Some moral principles are fundamental in the sense that we can give no reasons for adopting them; they do not follow from any higher principles. And it follows that a man cannot, at the moment of choosing, question the validity of the principle on which he chooses to act. For to do this would be to criticize the principle in the light of a higher principle; and in that case the principle in question is not a fundamental one. A man cannot condemn the principle on which he acts unless he has a con-attitude towards it; and in that case it is not a fundamental pro-attitude.

Now this seems to entail that a man cannot choose to act against his own moral principles, that he cannot choose to do what he knows to be wrong. But this is not so. Self-criticism is possible because, in criticizing my own character or conduct, I apply, not the principles on which I act, but the principles that I acknowledge on those

occasions when there is no question of their being manifested or not manifested. I can, for example, think that I ought to be less greedy, vindictive, or sanctimonious than I am, and this implies a con-attitude towards these particular traits in my character. But I cannot (logically) condemn any of these vices in myself while at the same time exercising them. For if I behave vindictively while at the same time condemning myself for doing so, I am a weak-willed but not a vindictive person. If, on the other hand, I deliberately choose to do something vindictive, then I am a vindictive person; and I can still claim that to be vindictive is against my principles only in the sense that, in my more reflective moments, I am prepared to condemn what I did.

The answer to the question whether a man can choose to change his moral principles is partly logical, partly empirical. In the case of principles that are not fundamental there is no logical difficulty, since we adopt these for reasons and both can and should abandon them if we find that the reasons are bad reasons, although it may be in practice difficult to do so. Traditionally a large part of moral philosophy has consisted in the attempt to show that many moral principles are subordinate in this way to one or a few very general principles, such as the Golden Rule or the Greatest Happiness Principle.

But, although there is no logical difficulty in the notion of trying to change a subordinate principle, there must, at any given moment, be some principles that are, here and now, fundamental moral principles for me. If this were not so, we could not talk about *choosing* or *trying* to change a principle, since this implies having a pro-attitude towards making the change. And it is here that the logical difficulty arises.

To try to change a principle implies having a pro-attitude towards making the change, and this implies that the principle is not a fundamental one. But it does not follow from this that there are any moral principles that are unchangeable. The fact that it makes no sense for me to ask whether I ought to act on a certain principle that is for me a fundamental one has often been cited as a proof that there are self-evident principles. For is not to say that it is senseless to question the principle to say that it is self-evident? But this argument confuses the practical impossibility of asking a certain question at a certain time with the logical impossibility of asking it at any time; and it also confuses the role of the advocate with that of the judge.

So long as a man is considering whether or not to act in a certain way, he addresses himself in the split-personality language of 'you ought'. But sooner or later he must make up his mind; he must decide. No doubt perpetual indecision is logically possible; but in many cases not to decide is to take a momentous decision, since the situation alters and the opportunity for choosing has passed. Moreover the logic of practical language is adapted to the practice of ordinary men, not to that of mental paralytics.

Sooner or later, then, he must proceed to a verdict "This is what I ought to do; this is the principle on which I shall act". And it is logically impossible for him to question this decision only in the sense that, if he questioned it, he would be returning to the standpoint of the advocate and it would not be a decision. It does not follow that at some future time he might not reconsider the decision and wonder whether he had been right. But to question the morality of a decision or principle is to criticize or appraise it in the light of a higher principle. Could this principle be questioned in its turn? Unless it were tautologous (in which case it could not serve as a moral principle at all, since it would be compatible with every course of action), it could be. Self-guaranteeing moral principles are impossible; and the demand for them rests on the failure to notice that 'there must always be some moral principle that I cannot now question' does not entail 'There must be some moral principle that I cannot ever question'. Every sentence must (logically) end with a full stop; but there is no point in any sentence at which a full stop must (logically) be put.

A man can, therefore, question the morality of his own principles and try to change them; but he cannot do so while applying them or if he has no pro-attitude towards making the change. Whether or not he can change them if these logical conditions are satisfied is an empirical question, to which the only answer is: "Sometimes. He may not always succeed; but he can always try". And since no one, not even the man himself, knows the limits of what he can do if he tries, it is a question to which no more precise answer can be given. There are moral principles which it is difficult to imagine any man wanting to change, because it is difficult to imagine what it would be like to adopt the contrary principle or to have a pro-attitude towards adopting it. But we must not confuse the difficulty of imagining something with its logical impossibility.

What sort of principles a man adopts will, in the end, depend on

his vision of the Good Life, his conception of the sort of world that he desires, so far as it rests with him, to create. Indeed his moral principles just *are* this conception. The conception can be altered; perhaps he meets someone whose character, conduct, or arguments reveal to him new virtues that he has never even contemplated; or he may do something uncharacteristic and against his principles without choosing to do it and, in doing it, discover how good it is. Moral values, like other values, are sometimes discovered accidentally. But the one thing he cannot do is to *try* to alter his conception of the Good Life; for it is ultimately by reference to this conception that all his choices are made. And the fact that he cannot choose to alter this conception neither shields him from blame nor disqualifies him from admiration.