

PHILOSOPHICAL SUBJECTS

Essays Presented to
P. F. Strawson

EDITED BY

ZAK VAN STRAATEN

CLARENDON PRESS · OXFORD

1980

Free-will and Rationality

A. J. Ayer

In his most interesting and influential essay on 'Freedom and Resentment'¹ Professor Strawson has done us a great service in drawing our attention to the very wide range of attitudes and judgements in which the notion of responsibility plays an essential part. Discussions of the perennial issue of Free-will and Determinism have tended to concentrate far too narrowly on the question of punishment. The result is that those who have sought to argue that we do not need to be free in order to be responsible, or that if we do it is not in any sense that is inconsistent with determinism, have had things made easier for them by the fact that the idea of retributive punishment has fallen rather into disrepute. This has helped to secure a more favourable reception for their contention that nothing of value will be lost if the notion of responsibility is analysed or revised in such a way that agents are held responsible, and therefore liable to be rewarded or punished, for their actions, only in cases where there is good reason to believe that the prospect and bestowal of rewards and punishments will so affect them and others as to increase the likelihood of their behaving in future in ways that are considered socially beneficial and decrease the likelihood of their behaving in ways that are considered socially harmful.

Even in this relatively narrow area, I think that there has been a tendency to underrate the extent to which this purely utilitarian approach is at variance with our ordinary ways of thinking. Even those who are emotionally repelled by the association of punishment with the idea of revenge, or those who regard the notion of expiation as intellectually or morally indefensible, will still be found adhering to two principles which do not fit into the utilitarian pattern. These principles

¹ *Proceedings of the British Academy*, 1962. Reprinted in Sir Peter Strawson's *Freedom and Resentment and Other Essays*.

are first that persons ought not to be punished for offences which they have not committed; and secondly that they ought not to be punished for actions which it was not in their power to avoid. The second principle is subject to some elasticity. For instance, it may be that someone has reached a state in which it is not in his power to avoid acting as he does, but he is still held liable to punishment because his being in that state is thought to be the result of previous actions which it was in his power to avoid: or it may be that someone is exonerated for an action which he was thought capable of avoiding, but only at a price, such as the probable loss of his life, which it was not in the circumstances thought reasonable for him to pay.

I know that utilitarians have made an effort to accommodate at least the first of these principles within their theory. To my mind, this effort has not been successful. I think that their theory will still be found to authorize the taking of punitive action, as a purely preventive or deterrent measure, in circumstances where it would not, in the general view, be morally justified and would, indeed, do violence to their own moral sentiments, if these were not wholly subjugated to their theory. This is not, however, a question that I wish to argue here. I am not concerned to criticize our utilitarians for depriving us of our concept of justice or for their failure, if indeed it is a failure, to make the loss good. I want only to stress the fact that they do deprive us of it and the reasons for which they do so. They deprive us of it by severing the historical connection, which there undoubtedly has been and indeed continues to be for most minds, between the concept of justice and the concept of desert, and the reason why they do so is that, while wishing to preserve as much of the idea of justice as is capable of rescue, they believe that the concept of desert is empty, if not meaningless, in the light of the possibility that all human actions are determined.

The service which Strawson has rendered us is to draw out the full implications of this rejection of the concept of desert. As he rightly points out, it is not only the concept of justice that is affected but almost all the concepts that enter into our estimations of one's own and other people's worth. We do indeed feel admiration or distaste for people's natural endowments, their good or bad looks, their native intelligence or stupidity, the good or bad dispositions which seem to form part of their

genetic inheritance. To a very much greater extent, however, our judgements of their merits and demerits, and the feelings with which these judgements are allied, relate to qualities and dispositions which we imply that the agent need not have developed and to behaviour which it was, or at least once had been, in his power to avoid. Strawson reminds us how often in our relations with one another we display and think that we have good reason to display such feelings as gratitude and resentment, how we take pride in our achievements and feel remorse for our misdeeds, how actions to which we are not a party can evoke our moral approval or indignation. In all these cases there is a tacit or explicit reference to the concept of desert. A person who does me a benefit deserves my gratitude because he could have withheld the favour. I feel remorse for my neglect of my friends, because I could have taken more trouble to please them. In my hatred of tyranny, wherever it is displayed, the idea is implicit that the tyrant could have restrained himself, or even if this is no longer possible, that he himself is at least partly to blame for having become the monster that he is. But how can any of these attitudes be justified if there is even any probability that determinism reigns over all actions, and if one of the consequences of determinism is that no one could ever have acted otherwise than as he did?

Strawson himself does not put the issue quite so sharply, partly no doubt because he begins by allying himself with those who say that they do not know what the thesis of determinism is. He can afford to adopt this standpoint because his ignorance of the exact content of determinism is not such as to preclude his knowing that it is a general thesis about the springs of human conduct, and the main point which his essay is intended to establish is that no thesis which applies to human conduct indiscriminately can pose such a threat as I have been describing. More concretely, if we follow him in speaking of the feelings and judgements, which carry the implication of desert, as entering into 'ordinary inter-personal relationships'² as well as into our assessments of moral worth, and if we follow him also in speaking of someone who judges or responds to human behaviour only from the point of view of its causes and effects as taking an 'objective attitude'³ towards it, then his contention is

² *Op. cit.*, p. 11.

³ *Ibid.*, p. 12.

first that 'it is practically inconceivable'⁴ that the acceptance of the thesis of determinism, whatever it may turn out to be, would lead to a thoroughgoing substitution of 'objective attitudes' for those that currently figure in our moral assessments and our interpersonal relationships, and secondly that even if this substitution could actually be made, it would not be rational to make it.

So far as I can see, the only argument which Strawson advances in favour of the first of these contentions is that whenever we do adopt an objective attitude towards some agent, either in respect of some or all of his actions, it is always for a special reason. Our resentment of what we should otherwise regard as an injury is inhibited, even when it is the work of one whom we take to be a responsible agent, because we have reason to believe that in this particular case he did not intend it or was subject to some pressure which it would not have been reasonable to expect him to resist. In such circumstances his action does not qualify as an injury. In other cases the disqualification may attach to the agent, if we judge that through no fault of his own he wasn't fully in command of himself. These cases are episodic, but there is also an important sub-class of cases in which an agent is persistently absolved from responsibility, because we view him 'as psychologically abnormal – or as morally undeveloped'.⁵ It is only in cases such as these that we tend to treat the person in question not as an appropriate object of our ordinary interpersonal attitudes, but 'as an object of social policy'; someone 'to be managed or handled or cured or trained'.⁶ And the same considerations apply to what Strawson calls our vicarious moral judgements. It is only when for some special reason the agent is seen in an 'abnormal light' that we think it inappropriate to react to his behaviour in the ways that are characteristic of our regarding him as 'a member of the moral community'.⁷

The reason, then, why Strawson thinks it unnecessary for his purpose to explore the general thesis of determinism is that on the comparatively rare occasions on which we do adopt a purely objective attitude towards our fellow human beings, it is always the result of our detecting some special sort of abnormality in their natures or in the conditions under which they have

⁴ *Ibid.*, p. 11.

⁵ *Ibid.*, p. 8.

⁶ *Ibid.*, p. 9.

⁷ *Ibid.*, p. 17.

acted, and that for the rest 'the human commitment to participation in ordinary interpersonal relationships is too thoroughgoing and too deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any such things as interpersonal relationships as we normally understand them'.⁸ Consequently, no matter what the general thesis of determinism turns out to be, and however strong the arguments in its favour, we can rely on the fact that 'a sustained objectivity of interpersonal attitudes, and the human isolation which this would entail, does not seem to be something of which human beings would be capable'.⁹

Now Strawson may very well be right in making this empirical conjecture. Many philosophers and scientists have accepted or believed that they accepted a general thesis of determinism, and no doubt many laymen have followed their lead, but I do not think that any scientific enquiry has yet been made into the effect that this has had upon their social outlook. There is indeed the example of Professor Skinner who has written of the general adoption of the objective attitude not only as a future possibility but as a condition of his picture of Utopia, but I do not know whether he consistently adheres to this position in his own private life. It would cast no deep shadow on his sincerity if he did not, since from his point of view it would prove no more than that his own conditioning had been imperfect. So far as my own observation goes, professing determinists are not the less likely to feel gratitude or resentment, or pride or remorse or moral indignation, or to avoid any of the other commitments which their theory might be expected to deny them. If this is generally true, it is a point in Strawson's favour. Even so, I think that he reaches his conclusion too easily, and my ground for this charge lies in his avoidance of any discussion of the general thesis of determinism. I hope to show that the content which it is reasonable to attach to the thesis is not so irrelevant to his argument as he supposes it to be.

Let me say at once that I do not myself propose to examine the thesis in depth. There are just two points about it that seem to me to have an important bearing upon the present question. The first of them is that the thesis is vacuous if it claims no more

⁸ *Ibid.*, p. 11.

⁹ *Ibid.*

than that every human action is subject to law; for any collection of phenomena whatsoever can be made to fit some set of generalizations, if no restrictions are placed on the character of these generalizations or their complexity. The thesis is worth discussing only if it is interpreted as claiming that all human actions are subject to what I call manageable laws; that is to say, the generalizations have to be such that we could actually use them to explain every facet of human behaviour, and that not merely *ex post facto*; they would have to enable us to make consistently successful predictions. In this form, the thesis has surely not been proved to be true. On the other hand, I do not think that we have any *a priori* reason for rejecting it.

The second and more important point is linked to the first. The thesis owes its force in this context to its being taken to entail that no one could ever have acted otherwise than as he did. We need, therefore, to examine what this conclusion comes to. I think it useless to begin with the suggestion that 'he could have acted otherwise' means 'he would have acted otherwise, if he had chosen'. Not only is the equivalence open to doubt, but at best it only leads to the question whether he could have chosen otherwise, on which we gain no foothold with this approach. A more profitable course, as I see it, is to ask under what conditions we actually judge that a man could not have acted otherwise, or conversely that he could. And here the answer does not seem to me far to seek. The cases in which we judge that the man could not have acted otherwise are those in which our beliefs encompass a set of particular facts and a set of well-established hypotheses which together entail or at least make it highly probable that he acted as he did. I ignore for the sake of simplicity the class of cases in which we judge that an action was avoidable but that it would not have been reasonable in the circumstances to expect the agent to avoid it. Conversely, the cases in which we judge that the man could have acted otherwise are those in which we are not in a position to draw the inference that he acted as he did, whether the shortage lies in our command of the requisite facts or the well-established hypotheses, or both. Obviously, to escape making it trivially true that all actions which are known to have occurred were unavoidable, it has to be stipulated that the stock of particular facts to which appeal can be made does not include

the fact that the action in question took place, or any fact which entails it without the assistance of some available hypothesis.

The consequence of this analysis is that the question whether or not a person could on some particular occasion have acted otherwise than as he did is made relative to the state of our knowledge. There might be thought to be a problem as to whose knowledge is at issue, but I think that this can be circumvented. We can regard the claim that a person could have acted otherwise in a given instance as an open challenge to produce the information which would meet the conditions laid down for judging that he could not, and so long as this information is not forthcoming the claim remains acceptable.

Now it is noteworthy that the exceptional cases in which, as Strawson recognizes, we tend to adopt an 'objective' rather than a 'personal' attitude towards a particular action, or towards the over-all behaviour of a particular type of agent, are those in which we do think that we command a set of scientific hypotheses from which, in conjunction with facts which are practically ascertainable if not already ascertained, the conclusion that the behaviour takes place can be derived with at least a high degree of probability and in quite a specific form, even if it does not reach down to every detail. It is indeed true that the actions of those whom we do include in the moral community, with all that Strawson takes this to imply, are not always or even very often such as greatly surprise us. Our friends and acquaintances form habits with which we become familiar. There are social norms with which we can usually trust them to comply. Our treating them as responsible agents is not just dependent on the greater difficulty that we have in predicting what they will say and do. The difference lies rather in the way that we arrive at these conclusions. Whatever justification we may have for attributing to them the personal character, and the particular beliefs and intentions, in terms of which we account for their behaviour, the generalizations involved are not accorded the status of scientific laws. They are statements of tendency which are not thought to be sufficiently strong or far-reaching to make the actions which they govern unavoidable. They do not sustain the verdict that the agent could not have acted otherwise.

This does not mean that the actions which we are accus-

tomed to view in this fashion are thought to be incapable of being explained scientifically. There have indeed been 'libertarians',¹⁰ as Strawson calls them, who have regarded the ascription of free will to human beings as requiring that if the actions which they freely perform are caused, their causes must include as a necessary condition an exercise of the will which is itself spontaneous; and faced with this difficulty have opted for free will. One of the weaknesses of their position, as Strawson remarks, is that this notion of the spontaneous exercise of the will is very obscure. If what is meant is that there is a random factor in what we choose to do, then it does not seem that the libertarians will have gained their objective, even if they are right. For why should the extent to which our actions occur by chance make us any more responsible for them than we would be if they were causally determined?

The more common view nowadays is, however, one that sets no *a priori* limit in this domain to the extension of the empire of science. For instance, many philosophers hold that what we normally classify as mental events are factually identical with events in the subject's brain, and they think it likely that events in the brain are subject to causal laws. Even so, this belief does not appear to inhibit them from making moral judgements, of a not purely utilitarian kind or from entering, as Strawson puts it, into 'ordinary interpersonal'¹¹ relationships. Are we to infer from this that they are inconsistent, or that their materialistic beliefs are not entirely sincere? Is it simply a vindication of Strawson's thesis that the thoroughgoing adoption of an 'objective attitude' is practically inconceivable?

It is tempting just to say 'Yes' to all three of these questions, but I think that it would be wrong. I suggest that what saves these philosophers from inconsistency is that their theory of the physical determination of mental events is not a working theory. They may sincerely believe that some specific theory of this sort is true, but they do not in fact know how it operates. They are not in a position to put any such theory to any widespread practical use. Physiology does not yet provide us with a complete explanation of the transition from one particular brain-state to another, or with the means of identifying or even uniquely correlating particular mental with cerebral

¹⁰ Ibid., p. 25.

¹¹ Ibid., p. 17.

events. Possibly these theoretical advances will be made, but even if they are, there will still be some way to go before the theory can be put to much practical use. To make only one simple point, it is not easy to see how we are going to be furnished with the requisite information about the current states of people's brains. Consequently, if my analysis is correct, a philosopher who believes that a given action was physically determined can still truly say that the agent could have acted otherwise. He can say so because we do not have the detailed scientific knowledge from which to derive the conclusion that the agent acted as he did.

But now suppose that this were altered. Suppose that the requisite physiological or psycho-physical theories were developed and that we could use them in everyday life to make mainly accurate predictions. Or if this be thought too fanciful, let us suppose that the theory of conditioning were developed to a point where it became possible to implant desires and beliefs and traits of character in human beings, to an extent that it could be deduced, at least in fairly general terms, how any person who had been treated in this way would most probably behave in a given situation, and that we lived under a regime in which these powers were exercised upon us, let us say from early childhood. It may be remarked that there have been, and are, regimes in which a very large measure of such control has been seriously attempted; and that the historical evidence tends to show that its effects are not so overwhelming as some behavioural psychologists would have us believe. This is not, however, the point at issue. We are once again postulating a scientific advance to which, so far as I can see, there is no objection of logical principle. We are supposing that the methods of conditioning, to whatever ends they may be directed, are both universally employed and almost totally effective, and the question is whether our interpersonal attitudes would be likely to withstand them.

My own view is that they most probably would not, unless, what is also conceivable, the conditioning itself were directed towards their preservation. Otherwise, I think that if it were a matter of common knowledge that these methods were practised, and if we understood the ways in which they operated, or even just the nature of their effects, sufficiently well to be able to

account by their means for almost everything that anybody said or did, we should be most strongly disposed to assume an objective attitude not only towards our fellow human beings but even towards ourselves. It would be as if we were spectators of a play in which we also participated with no other option than to enact the roles allotted to us. Even those who were charged with writing and directing the play would see themselves as guided to frame it along certain specific lines. We could still attach moral and aesthetic values to the ways in which the various characters performed, but the judgements and attitudes which depend on our seeing people as responsible agents would be missing.

Would such a point of view be rational? It might be argued that the question does not arise. If we have come to a point where we see everything as decided for us, we shall hardly be troubled with the question what decisions we ought to take. 'Things and actions are what they are and the consequences of them will be what they will be.'¹² We should no longer have the inclination to 'desire to be deceived'.¹³ Once again, however, the position may not be quite so simple. For instance, if the general belief that we cannot act otherwise than as we do were derived not from the prevalence of conditioning but from the general acceptance of a working physiological theory, then the fact of our believing that a causal explanation of each of our choices was readily available might not preclude our seeing the choices as needing to be made. I think that it would make some difference here whether we knew that the result of our deliberation had actually been predicted, and one need not assume that this would always be the case. It is anyhow worth noting that our ability to explain such things as the drawing of inferences does not dispense them from their subjection to normative standards. However detailed our knowledge of the way in which a calculating machine works, we can still coherently raise the question whether its calculations are correct.

An easier course of argument, which will have the advantage of giving the question of rationality a keener point, is to suppose that we are at a stage where the development of efficient processes of over-all conditioning is admitted to be practically feasible, and the question which we have to decide is whether we

¹² Bishop Butler, *Fifteen Sermons*, No. 7, para. 16.

¹³ *Ibid.*

should favour or oppose it. The possibility of there being a causal explanation for our coming to whatever decision we do is not a factor in this argument since we are still in a position in which, if I am right in my analysis, either decision is open to us. Let us suppose also that we are taking for granted the truth of the very dubious proposition that the policy to which the conditioning was adapted would be and remain beneficent, and that we see it as leading to the prevalence of objective attitudes. Would it be rational for us in this case to favour 'the march of mind'?¹⁴

Strawson's answer is that it would not. He views rational choice 'in the light of an assessment of the gains and losses to human life, its enrichment or impoverishment'¹⁵ and considers that from this point of view a state of affairs in which objective attitudes were generally prevalent would be of less value than one in which our present commitments to interpersonal relations were maintained. So far as I can see, he offers no argument in favour of this conclusion, and indeed I think that we should not expect him to. He had taken the issue to a point where it becomes a matter of moral sentiment, and we have only, as William James once put it, 'to confess to each other the motives for our several faiths'.¹⁶

For my part, I am strongly inclined to side with Strawson on this matter, but I have a qualm which he does not seem to share. As he hints in a footnote,¹⁷ there is another sense of 'rational' in which the rationality of an attitude is measured not by the probable consequences of adopting it but by the standing of the beliefs which enter into it. In this sense, an attitude is irrational if it rests on a belief which we have no good reason for accepting; more seriously so, if it rests on a belief which we have good reason to reject. Such reasons may be various: but one of them surely is that the belief requires the satisfaction of a concept which dissolves on analysis, where what I mean by its dissolving is that a dissection of it reveals no possible circumstances in which its application would be justified.

Now my trouble is that I believe this to be true of the concept of desert. I do not deny that the extension with which this concept is unwarrantably credited could be roughly covered by

¹⁴ See T. L. Peacock, *Crotchet Castle*. ¹⁵ *Op. cit.*, p. 13.

¹⁶ William James, *Essays in Radical Empiricism*, p. 276. ¹⁷ *Op. cit.*, p. 13.

a different concept of a utilitarian character; but that is beside the point. I am speaking of the concept that we actually employ, the one that sustains our actual attributions of moral responsibility and the whole range of attitudes into which they enter. Even in our present state of knowledge, without any forecast of further scientific developments and their probable consequences, I cannot discover any circumstances in which the application of *this* concept would be justified. Whether we conceive of human actions as subject to causal or statistical laws, or whether we explain them in terms of motives the emergence of which leaves room for the play of chance, there is, as Lord Melbourne said of the award of the Garter 'no damned merit about it'; and no demerit either. Our actual attributes of merit and demerit depend upon the use of a metaphysical idea of self-determination, which Strawson himself dismisses as inane.¹⁸

But if the concept of desert is empty, so is the concept of responsibility which is founded on it. The moral judgements, in which persons are credited with such responsibility, are consequently irrational, in the sense of the term with which we are now concerned, and so are all the attitudes that either involve such moral judgements or commit us to the same ascriptions of responsibility; in effect, all the attitudes that Strawson prizes.

I find myself, therefore, in a dilemma. I attach a strong value to being rational in this second sense. I am dismayed when any belief that I hold appears to me irrational, in the way that I have just defined irrational belief, and in such a case I think that I ought to discard the belief, so far as this seems to be within my power. Correspondingly, I see it as an objection to an attitude that it rests on any such irrational belief, and again I think that I ought to try to refrain from adhering to it, so far as I can. From this it appears to follow that I should set myself to cultivate an objective attitude towards myself and others, and to welcome an ordering of society in which it was generally prevalent. What should concern me morally would be just the beneficence of the conditioning. At the same time I have to confess that the prospect of any such Brave New World repels me. Why it should do so is not clear to me. I see no harm, but

¹⁸ *Ibid.*, p. 24.

¹⁹ *Op. cit.*, p. 13 fn.

rather goodness, in the attempt to influence people by argument and precept. This too is a form of conditioning, though not in many cases as effective as one would like it to be. Why then should I recoil from the idea of our all being subject to forms of conditioning that really would be effective and beneficial? If I were pressed for an answer, I should probably be reduced to saying something to the effect that it seemed to me an infringement of liberty or an affront to the dignity of man. But what would this mean? It looks as if I half-consciously hold the metaphysical belief in self-determination which my reason repudiates.

This shows, it may be said, that I just am not able to be rational, in the way I think I ought to be, which should not prevent my looking forward hopefully to scientific advances that will cause such disabilities to be overcome. But there is more to it than that. The fact is that I hold strong moral convictions, and strong personal attachments and antipathies, in which the unredeemed concepts of merit and demerit play an essential part; and that the idea of replacing these feelings and convictions by objective attitudes is not, on balance, attractive to me, not only as applying to myself, but also as a matter of general policy. Like Strawson, I am disposed to see the outcome as an impoverishment of 'human life'.

I am inclined therefore to endorse Strawson's epigram that 'if such a choice were possible', as unlike him I think it quite easily might be, 'it would not necessarily be rational to choose to be more purely rational than we are'.¹⁹ I wish only that I could accept it as blithely as he does; without my present feeling of intellectual discomfort.