



The Importance of Free Will

Author(s): Susan Wolf

Source: *Mind*, New Series, Vol. 90, No. 359, (Jul., 1981), pp. 386-405

Published by: Oxford University Press on behalf of the Mind Association

Stable URL: <http://www.jstor.org/stable/2253093>

Accessed: 07/04/2008 09:43

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=oup>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We enable the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

## The importance of Free Will<sup>1</sup>

SUSAN WOLF

The assumption that we have free will is generally thought to lurk behind the justifications of many of our current practices. That is, it is generally thought that only if we have free will can it be appropriate for us to engage in these practices, and that, if we should conclude that we don't have free will, we would have reason to give these practices up. The importance of the problem of free will in philosophy is often thought to depend on its relation to the justification of these practices. Thus, if an adequate justification of these practices were to be found, it may be thought that either the free will problem would thereby be solved or that, at least, it would thereby cease to be of interest.

In this paper, I shall argue that the justification of these practices need not rest on the assumption that we do have free will, and that the conclusion that we don't have free will gives us no reason at all to abandon these practices. My argument, however, seems to me to leave both the problem and its importance intact. The thought that our wills may not be free is no less disturbing even when all ties to the justification of our practices are completely and irrevocably severed.

Of course, there are some for whom the problem of free will was never disturbing in the first place. This paper is not likely to provide them with any new reason to be disturbed. Moreover, to those who have been and continue to be disturbed by the problem, this paper is not likely to offer much solace. Still, my paper is primarily addressed to this latter group, for, if it cannot provide solace, it may still provide insight into why—and why not—such solace is needed.

I shall begin by outlining a naive attempt at expressing the fears of those who find the problem of free will upsetting, and a naive response by those who think that the problem of free will gives us nothing to worry about. This expression of fear and the response

1 More people have benefited me by their comments and criticisms of drafts of this paper than I am able to acknowledge here. Of the many to whom I am grateful, special thanks are due to Jonathan Bennett and to Martha Nussbaum.

to it constitute a first stage of debate, which focuses on the justification of our practices of overt reward and punishment. The inadequacies at this stage of the argument suggest a way of advancing to a deeper stage, which focuses not on the overt practices themselves, but on the attitudes these practices typically express. Proceeding by way of two analogical cases, I shall argue that these attitudes, too, are safe from the threat of being undermined by reason and metaphysics. Nonetheless, I think that feelings of dissatisfaction may reasonably remain. I shall finally attempt to express what I take to be the appropriate focus of these feelings.

### *The Justification of Reward and Punishment*

I shall hereafter refer to the group who find the problem of free will upsetting as 'the pessimists'. In this, I follow P. F. Strawson in his article 'Freedom and Resentment'.<sup>1</sup> The pessimists include all those who believe, first, that whether or not we have free will depends on which metaphysical hypotheses are true, and, second, that it is not unlikely that the wrong metaphysical hypotheses are true. Perhaps the most common pessimists are those who believe that the thesis of determinism is both incompatible with free will and very likely true. However, they are also pessimists who believe that indeterminism is both incompatible with free will and, at least, very possibly true. And there are some, who may be said to be more pessimistic still, who believe that both determinism and indeterminism are incompatible with free will. For the remainder of this paper, I shall address myself to the first sort of pessimist, but it should be obvious how what I have to say to him can, with minimal adjustments, be addressed as significantly to the concerns of the other sorts of pessimist as well.

The opposing group, the optimists, are likely to doubt that the question of whether or not we have free will can profitably be said to depend on the truth of 'hypotheses', metaphysical or otherwise, at all. But, in any case, they believe that in so far as free will does depend on the truth of hypotheses, the facts already known to us are sufficient to guarantee that the appropriate hypotheses are true.

Of course, in calling the group who believes that we probably lack free will pessimistic, I adopt their accompanying view that

1 P. F. Strawson, 'Freedom and Resentment' in P. F. Strawson (ed.), *Studies in the Philosophy of Thought and Action* (Oxford University Press, 1968), pp. 71-96. I am indebted to this brilliant article for many of the ideas in this paper.

the absence of free will would be a very bad thing. Unfortunately, when it comes to explaining why it would be a very bad thing, the pessimists tend to be distressingly obscure. Of the practices they feel to be potentially undermined by the absence of free will, those associated with attributions of moral responsibility are most often cited. That is, they seem to think that the practices of praising and blaming people, punishing and rewarding them on the basis of the moral quality of their actions would be irrational, inappropriate, and unjustifiable if the thesis of determinism were true.

If this is all that the pessimists are worried about, however, the optimists have a ready reply. For they can provide a justification of the allegedly threatened practices that is in no way invalidated by the truth of determinism. They can argue, in particular, that the way in which we justify—or at least, the way in which we ought to justify—the application of these practices is one that depends on the consequences of engaging in them. We should praise or blame an individual, they may argue, if and only if by doing so we shall improve the moral quality of actions in the future. Or they may argue that we should praise or blame an individual if and only if by doing so we shall be obeying rules the institution of which will improve the moral quality of actions in the future. The hypotheses on which both these justifications of moral praise and blame rest are guaranteed to be true by the facts we already possess. We already know that we can improve the moral quality of actions by maintaining institutions of punishment that serve functions of rehabilitation and deterrence. We already know that we can improve the moral quality of actions by maintaining institutions of reward that provide incentives. Thus, we know that our practices of reward and punishment are justified whether or not the thesis of determinism is true.

I take it that this forward-looking, consequentialist type of justification of the practices of overt moral praise and blame is a good one, and therefore I take it that the intelligent pessimist will think it a good one as well. But in conceding this, the pessimist is likely to withdraw not his fear of determinism, but only his account of it. For the pessimist is likely to feel that the optimist's response somehow misses the pessimist's point. There is a striking difference between the type of justification of moral praise and blame that the optimist offers and a type of justification on which we ordinarily rely. It is in this difference that the pessimist's point, on a revised account, may be said to lie.

The justification of praise and blame the optimist suggests is one that emphasizes the fact that we can view praising and blaming as kinds of action, which, like any other actions, may or may not be sensible conclusions of practical reasoning. Whether or not to engage in these practices is on this view to be decided, like many other practical questions, according to whether engaging in these practices is a good way of achieving other desired ends. But we do not ordinarily praise and blame other persons because, as a result of engaging in practical deliberation, we have reached the conclusion that it would be in our interests to do so. Rather, we praise and blame persons as natural expressions of natural responses to what we see people do. We do not ordinarily *decide* whether a word of praise or a public scolding would be a useful directive to future behaviour. Rather, we find ourselves *reacting* to the actions and characters of others, approving of some, disapproving of others. Unless there is reason to restrain ourselves, we simply express what we feel.

In other words, although moral praising and blaming *can* be considered as kinds of actions, our ordinary experience of these phenomena encourages us to consider them as expressions of a kind of judgment. Accordingly, although one *can* justify these practices in a way that is analogous to justifications of other kinds of action, one can also try to justify these practices in a way that is analogous to justifications of other kinds of judgment. In particular, one can try to justify them by showing how the relevant judgments are fitting for, appropriate to, or, most aptly, *deserved by* the relevant objects of these judgments—in this case, human agents.

To justify praise and blame in the way the optimist suggests is to leave out of account such judgments of individual desert. It is to leave out of account any question of whether it is an individual's fault that he has done something wrong or whether it is to the individual's credit that he has done something right. In short, to justify the praise and blame of persons in the way the optimist suggests is to justify these practices in the same way that we justify the praise and blame of lower animals—in the same way, that is, that we justify the reward and punishment of pets, of pigeons in the laboratory, of monkeys in the circus. It is to justify these practices only as a means of manipulation or training.

The pessimist's fear may now be expressed as the fear that if determinism is true, this consequentialist justification of praise and blame is the only kind of justification that would be available to us.

If determinism is true, the pessimist fears, the type of justification of praise and blame that rests on judgments of individual desert can never be appropriate or valid. He fears that if we discover that determinism is true, we will be rationally obliged to give up making and relying on such judgments—and, more important perhaps, we will be rationally obliged to give up the attitudes which are essentially tied to these judgments.

It is notoriously difficult to give any precise characterization of these attitudes, to do more than be merely suggestive about their range and significance. The attitudes I have in mind include admiration and indignation, pride and shame, respect and contempt, gratitude and resentment. P. F. Strawson, in the article I mentioned earlier, has called this set of attitudes 'the reactive attitudes'. They are attitudes one has toward individuals only in so far as one views these individuals as persons. In contrast to the reactive attitudes, we may take what Strawson calls 'the objective attitude' toward the individuals with whom we interact. This is the attitude we do take—or at least, the attitude we rationally ought to take—toward most animals, present-day machines, and very young children.

### *The Justification of the Reactive Attitudes*

What the pessimist really fears, then, is that if determinism is true, we must give up not the practices of praise and blame themselves, but the attitudes and judgments these practices typically express. We must give up all our reactive attitudes, and adopt the objective attitude toward ourselves and each other, as we do toward everything else. It may be thought that in restating the pessimist's concerns, the scope and importance of his fears have considerably shrunk. The changing of attitudes seems to be such a private and insubstantial affair that it might be thought to make very little difference in the world. On second glance, however, we can see that the abandonment of all the reactive attitudes would make a very great difference indeed. To replace our reactive attitudes with the objective attitude completely is to change drastically—or, as most would say, reduce—the quality of our involvement or participation in all our human relationships.

Imagine for a moment what a world would be like in which we all regarded each other solely with the objective attitude. We would still imprison murderers and thieves, presumably, and we would still sing praises for acts of courage and charity. We would applaud

and criticize, say 'thank you' and 'for shame' according to whether our neighbours' behaviour was or was not to our liking. But these actions and words would have a different, shallower meaning than they have for us now. Our praises would not be expressions of admiration or esteem; our criticisms would not be expressions of indignation or resentment. Rather, they would be bits of positive and negative reinforcement meted out in the hopes of altering the character of others in ways best suited to our needs.

An act of heroism or of saintly virtue would not inspire us to aim for higher and nobler ideals, nor would it evoke in us a reverence or even admiration for its agent. At best we would think it a piece of good fortune that people occasionally do perform acts like this. We would consider how nice it must be for the beneficiaries and decide to encourage this kind of behaviour. We would not recoil from acts of injustice or cruelty as insults to human dignity, nor be moved by such acts to reflect with sorrow or puzzlement on the tide of events that can bring persons to stoop so low. Rather, we would recognize that the human tendency to perform acts like this is undesirable, a problem to be dealt with, like any other, as scientifically and efficiently as possible.

The most gruesome difference between this world and ours would be reflected in our closest human relationships—in the relations between siblings, parents and children, and especially spouses and companions. We would still be able to form some sorts of association that could be described as relationships of friendship and love. One person could find another amusing or useful. One could notice that the presence of a certain person was, like the sound of a favourite song, particularly soothing or invigorating. We could choose friends as we now choose clothing or home furnishings or hobbies, according to whether they offer, to a sufficient degree, the proper combination of pleasure and practicality. Attachments of considerable strength can develop on such limited bases. People do, after all, form strong attachments to their cars, their pianos, not to mention their pets. Nonetheless, I hope it is obvious why the words 'friendship' and 'love' applied to relationships in which admiration, respect, and gratitude have no part, might be said to take on a hollow ring. A world in which human relationships are restricted to those that can be formed and supported in the absence of the reactive attitudes is a world of human isolation so cold and dreary that any but the most cynical must shudder at the idea of it.

It is such a world in which the pessimist fears we would be rationally obliged to live if we came, once and for all, to the conclusion that the thesis of determinism was true. It is such a world, so much bleaker and more barren than our present world, to which the pessimist fears the truth of determinism would rationally force us. Once the optimist recognizes just what it is that the pessimist fears is at stake, however, the optimist once again has a ready reply. One thing he can point out is that even if the truth of determinism would give us some *reason* to regard ourselves differently, we would be psychologically incapable of changing our attitudes in the appropriate way. Another is that even if the truth of determinism would give us *some* reason to regard ourselves differently, we would have an overriding reason to keep the attitudes we currently hold. The overriding reason, of course, is that were we to give up our reactive attitudes, we would drastically reduce our sense of the meaning and value of our lives.

In light of the magnitude of this potential loss, it seems to me not irrational for the pessimist once again to concede the optimist's point. Once again, however, in conceding this, the pessimist is likely to withdraw not his fear of determinism but only his account of it. In other words, the pessimist might accept the optimist's argument—but he will accept it with despair. For with the first of his arguments, the optimist does not even attempt to allay the pessimist's fear that we will be forced to the conclusion that our attitudes towards ourselves are unjustified. Rather, he only seeks to show the pessimist how impotent this conclusion, if reached, would be. With the second of his arguments, the optimist suggests a way to avoid the feared conclusion. However, in so far as the optimist's justification takes the form of providing reasons that *override* other reasons, the justification can be only as satisfying as the acceptance of the lesser of two evils can be. How satisfying that is depends, in turn, on how evil is the evil we are forced to accept. Thus, it is worthwhile to get clear about the evil with which the pessimist now thinks we are left—the reason, in other words, for giving up the reactive attitudes which the optimist's argument must override.

Recall, then, that for the pessimist, whether or not we have free will is a matter of metaphysical fact. If determinism is true, then we do not have free will—that is, we are not free and responsible beings. In so far as we take reactive attitudes towards ourselves and each other, however, we regard ourselves as free and re-



sponsible beings. If determinism is true, then by continuing to take these attitudes, we live in a way that is discordant with the facts.<sup>1</sup>

The reason for giving up our reactive attitudes, then, is that by doing so we will be living in accordance with the facts. We will be accepting our status as creatures who are no more responsible for their lives and characters than are animals and machines. We will be accepting our status as agents to whom notions of personal credit, discredit, and desert fail to apply. If, despite the knowledge that this is our status, we choose to retain our reactive attitudes, we choose to live as if we were a kind of being that we know we are not. In doing this, we choose something akin to self-deception.<sup>2</sup>

As I said earlier, I believe such a choice may be rational. With Strawson, I think that it may be rational to choose *not* 'to be more purely rational than we are'.<sup>3</sup> It may be rational for a man to choose not to face the fact that he has a terminal illness or for a woman to try to avoid discovering that her husband is having an affair. If the costs would be high enough, it may be rational to override the reason for a course of action that is given by the acknowledgment that only that course of action would constitute living in accordance with the facts. To override this reason, however, is not just to choose to leave a desire unsatisfied. It is to choose to leave a value unrealized, a value, moreover, which is arguably one of considerable depth and importance.<sup>4</sup> To choose to act against, or contradict, a value as deep as this one, is inevitably to suffer a significant loss. It should not be surprising if the conviction that such a choice may be rational fails to bring the pessimist peace of mind.

Even to this last account of the pessimist's fear, I believe that the optimist has a reply which should make the pessimist withdraw his attempt to express and explain the threat of determinism yet another time. For I believe that even if determinism is true, and

1 Discordance with the facts is weaker than logical inconsistency. It is logically consistent to take attitudes that are essentially subject to certain standards of justification while at the same time believing that no such justifications are possible.

2 On this I part company with Strawson. For an excellent alternative account of these matters more faithful to Strawson's own views, see Jonathan Bennett, 'Accountability', in Zak Van Straaten (ed.), *Philosophical Subjects: Essays presented to P. F. Strawson* (the Clarendon Press, 1980).

3 Strawson, *op. cit.* p. 84.

4 For a good account of the distinction between values and other desires, see Gary Watson, 'Free Agency', *The Journal of Philosophy*, lxxii (24 April 1975), pp. 205-220.

even if this implies that as a matter of metaphysical fact we are not free and responsible beings, this gives us *no reason at all* to regard ourselves as unfree, irresponsible beings. That is, we have no reason at all to abandon our reactive attitudes and to adopt the objective attitude in their place. If we have no reason at all to abandon these attitudes, then we have nothing we need to override, no value we need to contradict in choosing to keep these attitudes. Our retention of the reactive attitudes need not be viewed as a choice between the lesser of two evils.

At first glance, it may appear that *this* conclusion must finally put the pessimist's mind at ease. We shall return to this claim later. First we must understand why this conclusion is warranted.

*The case of the addict.* Let us consider a hypothetical but not unrealistic situation: the situation of a drug addict who cannot help but take the drug to which he is addicted regardless of the attitude or value or second-order desire he has concerning his addiction. Let us further assume that in other respects our addict is a normally functioning, intelligent human being. Then the degree to which we hold this individual responsible for his drug-taking actions will vary in proportion to the degree to which we think he approves of—or, at least, doesn't disapprove of—the fact that he takes these actions. If the addict, with apparent sincerity, says and shows that he is relatively content to be an addict, that he sees no sufficient reason for trying to resist his addiction, then he is, in effect, accepting responsibility for taking the drug. He is affirming the fact that his efforts to obtain and to take the drug are *his actions*, that they effect and contribute to his character and his life in a way that may fairly enter into an assessment of what kind of person he is. It is therefore rational for us to regard him as responsible for taking the drug. If, on the other hand, the addict says and shows that he repudiates his addiction, that he makes all possible efforts to resist taking the drug, then he effectively removes himself from responsibility for taking the drug. He shows that he takes the drug only because he is addicted and that he would not take the drug if he could help it. It is therefore rational for us to regard him as not responsible for taking the drug. In other words, the addict's own attitude toward taking the drug gives us a reason (perhaps *the* reason) by which to establish ours—that is, it gives us the means by which to decide whether we ought to regard him as responsible for taking the drug.

The addict's actions are not free because whether or not he

chooses to take the drug, he will take the drug because he is compelled to do so. However, the addict's responsibility for his actions turns on the truth or falsity of an independent claim: namely, that whether or not the addict is compelled to take the drug, he will take the drug because he chooses to do so.<sup>1</sup> The addict, then, in taking an attitude toward his unfree actions, can thereby claim or disclaim responsibility for them. But whichever attitude the addict does take, the addict, in taking *an* attitude, asserts himself as a free and responsible being. By this I mean that if the addict accepts responsibility for taking the drug, he claims in effect that as a free and responsible being he chooses to take it, and if the addict rejects responsibility for taking the drug, he claims in effect that as a free and responsible being he does not choose to take it. The fact that we take the addict's own attitudes to his drug-taking actions seriously—that is, the fact that his attitudes count as a reason for us to hold him more or less responsible for these actions—rests on our belief that the addict, qua attitude-taker, *is* a free and responsible being. If we believed that the addict's approval or disapproval of his actions were itself determined by the influence of the drug, we would not regard his attitude as giving us a reason by which to establish ours.

*The case of the robot.* Let us turn now to a second case which takes us into the realm of science fiction. Let us imagine an individual who has been and continues to be very completely and elaborately programmed. He is programmed not only to make various choices and perform various actions, but also to engage in various thought processes, to form various second-order volitions and so forth, in coming to perform these actions. Indeed, this individual is programmed in such a way as to appear to be an ordinary human being in every respect. If no one were informed that this individual was programmed, he would appear both to us and to himself to be 'one of us'. I shall hereafter refer to this individual as a robot, but I believe that whether he is a member of the human species or not is irrelevant to the case.

Let us further assume that the robot's programming is not of any normal or familiar kind. In particular, let us assume it is not

1 I take it that I am agreeing here with Harry Frankfurt in 'Freedom of the Will and the Concept of a Person', *The Journal of Philosophy*, lxxviii (14 January 1971), pp. 5–20. My description of the addict was suggested by his discussion of three kinds of addict—the willing addict, the unwilling addict, and the wanton, who has no attitude or second-order desire concerning his addiction. I am concerned only with the first two kinds.

the case that any complete program is installed in the robot before he is, as it were, released into the world. Rather, let us assume that the robot is programmed on a day-to-day or moment-to-moment basis: the programmer implants the robot's responses to situations as these very situations arise. One might imagine the relation between robot and programmer to be very much like a possible relation between author and character; or, perhaps even better, one might imagine the relation to be like the relation between a magician and a human being over whose thoughts and bodily movements the magician has complete control.

In light of the nature of the robot's programming, I believe that the only way of living in accordance with the facts would be by regarding the robot solely with the objective attitude.<sup>1</sup> That is, I believe that the robot is not a free and responsible being in whatever sense of 'free and responsible' the objects of our reactive attitudes are ordinarily assumed to be. Were we to be purely rational, we would allow ourselves to feel some emotions toward the robot, but we would not feel those emotions or sentiments constitutive of our reactive attitudes. For though the robot might choose to perform the actions he performs, he chooses to perform them only because he is programmed to so choose. Though his decisions and judgments may be preceded by thoughts which look or sound like reasons, he cannot be said to reason to these conclusions in the way we do. He is not in ultimate control of his values, his personality, or his actions. He is, properly speaking, only a vehicle for carrying out the plans (if plans there be) of his programmer.

Were such a robot to live within our society, it may well be that we would not ultimately regard him in the way that I have suggested it would be purely rational for us to regard him. The sheer difficulty of keeping in mind the fact that the robot is programmed (along with all its implications) may make it psychologically impossible for us to take the objective attitude towards him consistently. Moreover, we might decide that, though there is some reason for us to treat the robot objectively, there are overriding reasons to treat him as a normal member of the community. Or perhaps we would take some sort of middle ground. (For example, we might treat him as if he had some of the rights of a normal person, but

1 Some philosophers will resist this conclusion, and the few remarks I add (directly below) by way of support may be insufficient to convince them. However, I believe the pessimist would think that this conclusion is correct, and it is with the pessimist's position that I am primarily concerned.

we would shrink from allowing one of our daughters to marry him.)<sup>1</sup> All that is important for my purposes is that we take it to be purely rational to regard the robot objectively. For this should carry at least some weight in determining how we ought ultimately to regard him.

Let us now assume that, after years of thinking himself to be like other human beings, the robot comes to believe that he is completely programmed. If at this point we were to adopt the robot's interests as our own, would it be rational for us to urge (to the programmer, presumably) that the robot take the objective attitude toward himself?

By this question I mean to approach the question 'Would it be rational for the robot to adopt the objective attitude toward himself?' as closely as my standards of conceptual coherence allow. I am not sure that we can make sense of the question 'What would it be rational for the robot to do?', because the attempt to answer it seems to require that we imaginatively endow the robot with the powers of a free and responsible being while, at the same time, remaining convinced that the robot lacks these same powers. However, I see nothing to prevent us from reasoning *on the robot's behalf*. Thus, to repeat my question, I ask, if we were to adopt the robot's interests as our own, would it be rational for us to urge that the robot take the objective attitude toward himself?

Well, I can imagine some situations in which it might be. If, for example, the robot were an individual unusually tormented by an awareness of his limitations, the belief that he was not responsible for the meagreness of his abilities might be a source of some comfort to him. Or, if the members of the robot's community did take the purely rational attitude toward the robot, the robot's own adoption of the objective attitude toward himself might allow him to take this treatment less personally. (Of course, here as always, it will ultimately be up to the programmer whether the robot's adoption of the objective attitude would serve the purposes that I have suggested they might serve. But let us assume at this point that the programmer is cooperative.)

In so far as we argue that the robot should adopt, or try to adopt,

1 This recalls a scene from a play by Woody Allen: Two characters in the play appeal to help in their dialogue from the members of the audience. An attractive woman comes to their aid, with whom one of the characters begins to fall in love. The other character, trying to discourage the romance, asks his friend, 'What kind of children would you have? She's Jewish, you're fictional!'

the objective attitude toward himself for reasons such as these, however, we are not arguing for the adoption of this attitude simply on the grounds that the attitude is appropriate. That is, we are not arguing that the robot should take this attitude simply because of his (and our) value in living in accordance with the facts. Our reasons for urging that the robot should take the objective attitude are, rather, utilitarian ones: his life will be less painful if he takes the objective attitude toward himself.

Furthermore, if the robot were programmed to take this attitude, he would not really be taking an attitude that would be in accordance with the facts. In believing that reactive attitudes toward himself are inappropriate, he would not be accepting all the implications of the fact that he is programmed; he would not be denying his freedom and responsibility completely. For the robot, in taking an objective view of himself, necessarily leaves a part of himself out of this view—specifically, he leaves out that part of himself which is taking the objective attitude. The robot perhaps takes comfort in the fact that *he*—as it were, ‘his self’—is not responsible for his meagre abilities; or, the robot takes comfort in thinking that he is not responsible for the fact that he is merely a robot. But the robot’s alienation from his abilities on the one hand, or from his robotness on the other, presupposes a self from whom these things are alienated, a self whose fault these things are not. And of course, in this example, the robot’s self is itself the result of his programming. To the extent that being programmed justifies a denial of responsibility for *any* feature of the robot’s existence, it justifies—indeed, demands—a denial of responsibility for *every* feature of his existence—including, in particular, his denial of responsibility for every other feature.

We can bring this out by comparing the case of the robot to that of the addict. For, recall that the addict, in taking any attitude toward his addiction, asserts himself as a free and responsible being. If he accepts responsibility for taking the drug, he claims that as a free and responsible being he chooses to take it. If he rejects responsibility for taking the drug, he claims that as a free and responsible being he does not choose to take it. Similarly, the robot in taking an attitude toward himself, asserts himself as a free and responsible being. But unlike the addict, the robot is not a free and responsible being in any respect whatsoever. He is in a position analogous to that of the addict whose attitude toward his addiction is itself determined by the influence of the drug. Thus, the robot’s

own attitude toward himself cannot have any weight for us. If the robot, as a matter of metaphysical fact, is an unfree, irresponsible being, then *his* acceptance of this fact gives us no extra reason to regard him as such.

*The case of our (determined) selves.* We may finally turn to the question of what it would be purely rational for us to do if we came to believe that the thesis of determinism was true and that this implied that, as a matter of metaphysical fact, we were not free and responsible beings. Perhaps the pessimist thinks that if determinism is true, then the whole world is in a position analogous to that of the robot—that, in other words, the thesis of determinism is not different in any relevant respect from the thesis that the whole world is, like the robot, completely programmed. In this case, would it be rational for us to take the objective attitude toward ourselves? Again, we might answer, as we did when reasoning on behalf of the robot, that we can imagine some situations in which it might be. In particular, it would be rational for us to take this attitude if by doing so we would become, on the whole, better off. But considerations of the sort I suggested earlier make this possibility seem very unlikely. It is hard to believe that more of our desires (all orders inclusive)<sup>1</sup> would be satisfied if we ceased to take the reactive attitudes and adopted the objective attitude in their place. Still, among our desires, we must include the often very strong desire to live in accordance with the facts. Indeed, as I said earlier, this is not just a desire, but a value of considerable depth and importance. If, by taking the objective attitude toward ourselves, we would better realize this value, then, regardless of our ultimate decision, we would have at least some reason to adopt the objective attitude.

It should now be clear, however, that we would *not* be realizing this value by adopting the objective attitude. If we were to view ourselves objectively, we would, like the robot, necessarily leave a part of ourselves out of this view. In taking any attitude toward ourselves, including the attitude that we are not free or responsible beings, we would be asserting ourselves *as* free and responsible beings.<sup>2</sup> Any attitude we take, then, would involve a false step—any attitude would be unjustified. Thus, it seems that the only way we could live in accordance with the facts would be by ceasing to

1 The idea of higher-order desires is taken from Harry Frankfurt, *ibid.*

2 Perhaps this claim can be taken as a reformulation of the liar paradox: 'Do not take me seriously.'

have any attitudes at all—by ceasing, that is, to make or rely on any judgments about an individual's responsibility or lack of it at all.

The truth of determinism, then, gives us no reason at all to replace our present reactive attitudes with the objective attitude. Some might think, however, that it gives us reason to do something even more drastic—namely, to give up the taking of attitudes altogether. For I have said that we place a considerable value in living in accordance with the facts. And I have also said that if determinism were true, and if this implied that, as a matter of metaphysical fact, we were not free and responsible beings, then the only way in which we would be living in accordance with the facts would be by giving up the taking of attitudes altogether.

Of course, in answer to the question 'Would it be rational for us to give up all our attitudes?', pragmatic replies of the sort I suggested earlier will be all the more poignant. That is, even if determinism gave us some *reason* to give up all our attitudes, we would be psychologically incapable of meeting this demand. And even if determinism gave us *some* reason to give up our attitudes, we would have overriding reason to retain them.

If our sense of the value and meaning of our lives would be sharply reduced in a world without reactive attitudes, it would be altogether eliminated in a world in which no attitudes were taken at all. For the only way we could give up taking either the attitudes that regard others as responsible for their actions or the attitude that regards others as not responsible for their actions would be by giving up thinking in terms of the notions of responsibility and desert at all. Giving this up, I believe, would require in turn that we give up a great deal more. We would have to stop thinking in terms of what ought and ought not to be. We would have to stop thinking in terms that would allow the possibility that some lives and projects are better than others.<sup>1</sup> Were we to make ourselves into the kind of creatures that ceased to think in these terms, we would lose the distinction between desires and values and, therefore, our distinction as valuing creatures. We would lose our ideals, our senses of self, and, I think, our status as persons. A world without reactive attitudes would be a tragic world of human

1 Here I assume that 'ought' and 'better' have the force of objective reason. Once the thought that some things ought to be is allowed, so is the thought that some things ought to be done (by oneself, for instance). And this, I think, leads inevitably to the thought that one is, *ceteris paribus*, responsible for doing them.



isolation; a world without reactive attitudes or the objective attitude would be a bleak, blank world of human brutes.

Thus, as I said, in answer to the question 'Would it be rational for us to give up all our attitudes?', pragmatic replies will be all the more poignant. If the optimist's arguments stopped here, however, the pessimist could still sigh and point out once again that pragmatic replies are merely pragmatic. As such, they can be only as satisfying as the acceptance of the lesser of two evils can be. However, I believe that the optimist's arguments need not stop here, with the merely pragmatic. If we *had* some reason to give up all our attitudes, we would have overriding reason to retain them. But, in fact, I believe, we have no reason at all to take this very drastic step. We have no reason at all to fulfil our desire to live in accordance with the facts, when the facts in question are facts such as these. In other words, the desire is itself irrational in relation to facts such as these. If the facts are that we are, in all relevant respects, like the robot, there is no point to living in accordance with them.

Unfortunately, it is not at all clear what the point of this desire normally is. The desire to live in accordance with the facts is more easily felt than explained. It is this desire, I take it, that people sometimes express when they say that they want to live in the Real World. It is this desire that makes people shudder at the thought of passing their days hooked up to a pleasure machine. This desire shows up in more realistic situations when we consider how important it is to us that we not only feel ourselves to be loved, but that we truly be loved, or when we see how important it is to us that we not only believe that our efforts to achieve something in the world have succeeded, but that they really have succeeded.

Why is it so important to us that our conception of our lives correspond to some more objective fact? Why does it matter so much that the world we live in is the Real World? I can think of two possible answers.

First, I think it plausible that we place a primitive, unanalysable value on 'getting things right'. Perhaps, that is, we value being right for its own sake. From this value, the value of living in accordance with the facts would follow as a direct corollary. If so, it should at least give us pause to notice that living in accordance with the fact that we are not free and responsible beings would require us to give up our value in being right. For living in accordance with the fact that we are not free and responsible beings

would require us to give up all our values. More important, if we were to live in accordance with the fact that we are not free and responsible beings by giving up the taking of attitudes altogether, we would not even realize our (past) value in getting things right. We would admittedly cease to be getting some things wrong, for we would cease to regard ourselves as free and responsible beings. But we would do this at the cost of ceasing to regard ourselves as anything at all.

On the other hand, our desire to live in accordance with the facts—our desire, that is, to live in the Real World—may rest essentially on the belief that it is the Real World, and the beings within it, that matter. In other words, we may want to live in accordance with the facts because we want ourselves to matter in the right sort of way, to make the right sort of difference to the world and the beings who do matter and to whom we might matter. But all the beings that could possibly be encompassed by these concerns must themselves be within the grasp of the same determinism as ourselves.<sup>1</sup> If the point of living in accordance with the facts is to make the right kind of difference to the right kind of beings, then it cannot possibly be an achievement to eliminate the right kind of beings *en masse*.

It might be rational for the robot to commit a kind of suicide of self as a result of the realization that he is, unlike the rest of us, a robot. For it seems plausible that the realization that you cannot, and/or rationally ought not, matter to the people or to the world that matters to you—indeed, to the people or to the world that matter independently of you—might give you a reason to commit suicide.<sup>2</sup> But the realization that you are determined because your whole world is determined cannot generate such a reason. For us, either this world matters or none at all. If this world matters, then it would be irrational to destroy it. And if this world does not matter, then it certainly doesn't matter that we do or do not choose to destroy it.

Thus, we reach the conclusion that the truth of determinism

1 As perhaps all the people a person in a dream can concern himself with are themselves dream-people. Perhaps, one might think, we can also concern ourselves with the programmer (or God), if there is one, and this individual would not be in the grasp of the same determinism as ourselves. Even if this were correct, however, we could in no way improve our status with such a being by living in accordance with the facts.

2 Douglas MacLean once suggested to me that Kafka's *Metamorphosis* might be interpreted as an illustration of just this point.

gives us no reason at all to give up our reactive attitudes. Let me briefly review the argument.

We first considered the suggestion that the recognition that, as a matter of metaphysical fact, we were not free or responsible beings would give us a reason to regard ourselves as unfree, irresponsible beings—it would give us a reason, that is, to replace our present reactive attitudes with the objective attitude toward ourselves and each other. But we saw that this change would fail to achieve its purpose; it would not satisfy the desire to live in accordance with the facts. For it is only rational to take some particular attitude toward ourselves in the context of the belief that we are, at least in our capacity as attitude-takers, free and responsible beings. Therefore, we would be no less irrational if we chose to take the objective attitude than if we chose to take the other alternative.

Second, we considered the suggestion that the desire to live in accordance with the facts might give us a reason to cease to take attitudes altogether. But we saw that if living in accordance with the facts required *this* change, we would have no reason to live in accordance with them. For there seemed to be two possible sources of our desire to live in accordance with the facts. According to the first, this desire rests on the belief that by living in accordance with the facts we will promote our ability to get things right. According to the second, this desire rests on the belief that living in accordance with the facts would put us in the world that is most worth living in—the world, that is, with valuable and valuing selves. But if determinism is true, and if this implies that we are not free and responsible beings, then neither of these beliefs are justified. On neither of these accounts would it be rational to live in accordance with the fact that determinism is true.

Thus, the truth of determinism gives us no reason at all to choose to take one attitude rather than another. And the truth of determinism gives us no reason at all to choose to take no attitude rather than some.

Since the truth of determinism gives us no reason at all, we must look elsewhere for reasons by which to decide which attitudes, if any, it would be best for us to take. Presumably, we would have to look at the consequences of these various decisions—and, looking at these, we would, presumably, choose to keep our present reactive attitudes. This brings us to the apparently optimistic conclusion that, even if determinism were true, and even if this implied that, as a matter of metaphysical fact, we were not free or responsible

beings, it would still be rational—and without impurification—to retain our present reactive attitudes.

*The importance of free will*

Some might think that this conclusion must finally silence the pessimist. For it should convince him that no answer to the problem of free will can have any practical, pessimistic consequences whatsoever. But here again, I believe the pessimist might withdraw not his fear of determinism, but only his account of it. If the argument I have presented as the justification of our attitudes is the only justification we can have, the pessimist again, might accept this justification—but he will accept it with despair. For the position I have outlined might be said to reduce, in effect, to something like this: ‘Even if we are puppets on the strings of the hands of God, there is nothing at all we can do about it. It would therefore not be rational to try to do anything about it, nor would it be rational, because of *this*, to commit suicide. Since there are no rational options by which to respond to this possibility, the option we do take cannot be irrational. So we are rationally permitted—perhaps, even obliged—to go on living our (possibly puppet—) lives.’

This argument, unfortunately, takes nothing away from the fact that we don’t *want* to be puppets. We don’t want to be, or be no better than, objects of someone else’s manipulation. Of course, it is nice to know that, whatever the facts, the rationality of our practices is not open to criticism. It is nice to know that, whatever the facts, we are not making fools of ourselves. It is also nice to know that, as new facts come to light, nothing can happen that will generate, or that even ought to generate, a practical crisis. We will not have to choose between the lesser of two evils; we will not have to choose self-deception. But the guarantee that we are not behaving irrationally or serving as the unwitting agents of our own humiliation and error—the guarantee, in other words, that *we* cannot be faulted for taking an inappropriate attitude towards ourselves and our place in the world—is not the only guarantee that one can reasonably wish for. And the onset of a practical crisis, of the recognition of the need to confront an inconsistency in ourselves and to change our personalities and practices in undesirable ways, is not the only state of affairs that one can reasonably fear.

The pessimist fears that if determinism is true, then we are no

better off than puppets. And the lives of puppets, the pessimist thinks, are meaningless and absurd. No one would dream of faulting the puppets for this—the thought that puppets are blameworthy for not recognizing their puppethood and integrating their recognition into the way they live their lives is at worst incoherent and at best simply false. Nonetheless, the puppets' lives are meaningless, and, from the puppets' point of view, that would be too bad. The pessimist fears that if determinism is true, then we are no better off than puppets. Naturally, from the pessimist's point of view, if determinism were true, that would be too bad. The fact that we don't have to *change* our values is of little solace if it may be the case that we are, now and forever, incapable of *realizing* our values. The fact that we don't have to think that our lives are meaningless is of little comfort if, for all that, our lives may actually *be* meaningless.

Thus, the apparently optimistic conclusion that it is completely rational for us to regard ourselves as free and responsible beings must, in order to silence the pessimist, be reached in a more optimistic way. No position which allows that as a matter of metaphysical fact we might not be free and responsible beings—even if this gives us no reason at all to regard ourselves as such—can properly be called optimistic. The pessimist will only give up his pessimism if the possibility of this state of affairs is directly refuted. In other words, for the pessimist, who asks for a justification of the fact that we treat ourselves as free and responsible beings, only one kind of justification will do—a justification, in particular, which relies on the fact that we *are* free and responsible beings.