

STUDY QUESTIONS

1. Why does Fischer think Frankfurt-type cases are important for moral responsibility? What does he think they show us about the mechanism that actually issues in an agent's behavior?
2. What does Fischer mean by "reasons-responsiveness"? Why does he think it is important to distinguish between strong and weak varieties of reasons-responsiveness?
3. What is causal determinism? Do you think that the lessons we learn from Frankfurt-type cases are applicable to causal determinism?
4. In your own words, explain the view that Fischer calls "semicompatibilism." How does semicompatibilism differ from incompatibilism? Compatibilism?

Freedom of the Will and the Concept of a Person

HARRY G. FRANKFURT

...

In this article Harry Frankfurt offers some intriguing reflections about persons, the will, and freedom. He develops a subtle account of the hierarchical structure of a person's will and relates that account to various problems concerning freedom and moral responsibility. Roughly, Frankfurt suggests that persons are distinct from non-persons, because persons are beings with second-order volitions. Persons, that is, are capable of wanting certain of their motivations to move them. This idea is key to understanding the notions of personhood and freedom of the will, according to Frankfurt. Frankfurt goes on to apply these ideas to various cases, and he draws some interesting conclusions about moral responsibility and determinism.

...

WHAT philosophers have lately come to accept as analysis of the concept of a person is not actually analysis of *that* concept at all. Strawson, whose usage represents the current standard, identifies the concept of a person as "the concept of a type of entity such that *both* predicates ascribing states of consciousness *and* predicates ascribing corporeal characteristics . . . are equally applicable to a single individual of that single type."¹ But there are many entities besides persons that have both mental and physical properties. As it happens—though it

seems extraordinary that this should be so—there is no common English word for the type of entity Strawson has in mind, a type that includes not only human beings but animals of various lesser species as well. Still, this hardly justifies the misappropriation of a valuable philosophical term.

Whether the members of some animal species are persons is surely not to be settled merely by determining whether it is correct to apply to them, in addition to predicates ascribing corporeal characteristics, predicates that ascribe states of consciousness. It does violence to our language to endorse the application of the term "person" to those numerous creatures which do have both psychological and material properties but which are manifestly not persons in any normal sense of the word. This misuse of language is

From *Journal of Philosophy*, LXVIII, No. 1 (Jan. 1971). Copyright © 1971 by Columbia University. Reprinted by permission of the publisher and author.

doubtless innocent of any theoretical error. But although the offence is “merely verbal,” it does significant harm. For it gratuitously diminishes our philosophical vocabulary, and it increases the likelihood that we will overlook the important area of inquiry with which the term “person” is most naturally associated. It might have been expected that no problem would be of more central and persistent concern to philosophers than that of understanding what we ourselves essentially are. Yet this problem is so generally neglected that it has been possible to make off with its very name almost without being noticed and, evidently, without evoking any widespread feeling of loss.

There is a sense in which the word “person” is merely the singular form of “people” and in which both terms connote no more than membership in a certain biological species. In those senses of the word which are of greater philosophical interest, however, the criteria for being a person do not serve primarily to distinguish the members of our own species from the members of other species. Rather, they are designed to capture those attributes which are the subject of our most humane concern with ourselves and the source of what we regard as most important and most problematical in our lives. Now these attributes would be of equal significance to us even if they were not in fact peculiar and common to the members of our own species. What interests us most in the human condition would not interest us less if it were also a feature of the condition of other creatures as well.

Our concept of ourselves as persons is not to be understood, therefore, as a concept of attributes that are necessarily species-specific. It is conceptually possible that members of novel or even of familiar non-human species should be persons; and it is also conceptually possible that some members of the human species are not persons. We do in fact assume, on the other hand, that no member of another species is a person. Accordingly, there is a presumption that what is essential to persons is a set of characteristics that we generally suppose—whether rightly or wrongly—to be uniquely human.

It is my view that one essential difference between persons and other creatures is to be found in the structure of a person’s will. Human beings are not alone in having desires and motives, or in

making choices. They share these things with the members of certain other species, some of whom even appear to engage in deliberation and to make decisions based upon prior thought. It seems to be peculiarly characteristic of humans, however, that they are able to form what I shall call “**second-order desires**” or “desires of the second order.”

Besides wanting and choosing and being moved *to do* this or that, men may also want to have (or not to have) certain desires and motives. They are capable of wanting to be different, in their preferences and purposes, from what they are. Many animals appear to have the capacity for what I shall call “**first-order desires**” or “desires of the first order,” which are simply desires to do or not to do one thing or another. No animal other than man, however, appears to have the capacity for reflective self-evaluation that is manifested in the formation of second-order desires.²

I

The concept designated by the verb “to want” is extraordinarily elusive. A statement of the form “*A* wants to *X*”—taken by itself, apart from a context that serves to amplify or to specify its meaning—conveys remarkably little information. Such a statement may be consistent, for example, with each of the following statements: (a) the prospect of doing *X* elicits no sensation or introspectible emotional response in *A*; (b) *A* is unaware that he wants to *X*; (c) *A* believes that he does not want to *X*; (d) *A* wants to refrain from *X*-ing; (e) *A* wants to *Y* and believes that it is impossible for him both to *Y* and to *X*; (f) *A* does not “really” want to *X*; (g) *A would rather die than X*; and so on. It is therefore hardly sufficient to formulate the distinction between first-order and second-order desires, as I have done, by suggesting merely that someone has a first-order desire when he wants to do or not to do such-and-such, and that he has a second-order desire when he wants to have or not to have a certain desire of the first order.

As I shall understand them, statements of the form “*A* wants to *X*” cover a rather broad range of possibilities.³ They may be true even when statements like (a) through (g) are true: when *A* is unaware of any feelings concerning *X*-ing, when he is

unaware that he wants to *X*, when he deceives himself about what he wants and believes falsely that he does not want to *X*, when he also has other desires that conflict with his desire to *X*, or when he is ambivalent. The desires in question may be conscious or unconscious, they need not be univocal, and *A* may be mistaken about them. There is a further source of uncertainty with regard to statements that identify someone's desires, however, and here it is important for my purposes to be less permissive.

Consider first those statements of the form "*A* wants to *X*" which identify first-order desires—that is, statements in which the term "to *X*" refers to an action. A statement of this kind does not, by itself, indicate the relative strength of *A*'s desire to *X*. It does not make it clear whether this desire is at all likely to play a decisive role in what *A* actually does or tries to do. For it may correctly be said that *A* wants to *X* even when his desire to *X* is only one among his desires and when it is far from being paramount among them. Thus, it may be true that *A* wants to *X* when he strongly prefers to do something else instead; and it may be true that he wants to *X* despite the fact that, when he acts, it is not the desire to *X* that motivates him to do what he does. On the other hand, someone who states that *A* wants to *X* may mean to convey that it is this desire that is motivating or moving *A* to do what he is actually doing or that *A* will in fact be moved by this desire (unless he changes his mind) when he acts.

It is only when it is used in the second of these ways that, given the special usage of "will" that I propose to adopt, the statement identifies *A*'s will. To identify an agent's will is either to identify the desire (or desires) by which he is motivated in some action he performs or to identify the desire (or desires) by which he will or would be motivated when or if he acts. An agent's will, then, is identical with one or more of his first-order desires. But the notion of the will, as I am employing it, is not coextensive with the notion of first-order desires. It is not the notion of something that merely inclines an agent in some degree to act in a certain way. Rather, it is the notion of an *effective* desire—one that moves (or will or would move) a person all the way to action. Thus the notion of the will is not coextensive with the notion of what an agent intends to do. For even

though someone may have a settled intention to do *X*, he may nonetheless do something else instead of doing *X* because, despite his intention, his desire to do *X* proves to be weaker or less effective than some conflicting desire.

Now consider those statements of the form "*A* wants to *X*" which identify second-order desires—that is, statements in which the term "to *X*" refers to a desire of the first order. There are also two kinds of situation in which it may be true that *A* wants to want to *X*. In the first place, it might be true of *A* that he wants to have a desire to *X* despite the fact that he has a univocal desire, altogether free of conflict and ambivalence, to refrain from *X*-ing. Someone might want to have a certain desire, in other words, but univocally want that desire to be unsatisfied.

Suppose that a physician engaged in psychotherapy with narcotics addicts believes that his ability to help his patients would be enhanced if he understood better what it is like for them to desire the drug to which they are addicted. Suppose that he is led in this way to want to have a desire for the drug. If it is a genuine desire that he wants, then what he wants is not merely to feel the sensations that addicts characteristically feel when they are gripped by their desires for the drug. What the physician wants, in so far as he wants to have a desire, is to be inclined or moved to some extent to take the drug.

It is entirely possible, however, that, although he wants to be moved by a desire to take the drug, he does not want this desire to be effective. He may not want it to move him all the way to action. He need not be interested in finding out what it is like to take the drug. And in so far as he now wants only to *want* to take it, and not to *take* it, there is nothing in what he now wants that would be satisfied by the drug itself. He may now have, in fact, an altogether univocal desire *not* to take the drug; and he may prudently arrange to make it impossible for him to satisfy the desire he would have if his desire to want the drug should in time be satisfied.

It would thus be incorrect to infer, from the fact that the physician now wants to desire to take the drug, that he already does desire to take it. His second-order desire to be moved to take the drug does not entail that he has a first-order desire to take it.

If the drug were now to be administered to him, this might satisfy no desire that is implicit in his desire to want to take it. While he wants to take the drug, he may have *no* desire to take it; it may be that *all* he wants is to taste the desire for it. That is, his desire to have a certain desire that he does not have may not be a desire that his will should be at all different than it is.

Someone who wants only in this truncated way to want to *X* stands at the margin of preciosity, and the fact that he wants to want to *X* is not pertinent to the identification of his will. There is, however, a second kind of situation that may be described by "*A* wants to want to *X*"; and when the statement is used to describe a situation of this second kind, then it does pertain to what *A* wants his will to be. In such cases the statement means that *A* wants the desire to *X* to be the desire that moves him effectively to act. It is not merely that he wants the desire to *X* to be among the desires by which, to one degree or another, he is moved or inclined to act. He wants this desire to be effective—that is, to provide the motive in what he actually does. Now when the statement that *A* wants to want to *X* is used in this way, it does entail that *A* already has a desire to *X*. It could not be true both that *A* wants the desire to *X* to move him into action and that he does not want to *X*. It is only if he does want to *X* that he can coherently want the desire to *X* not merely to be one of his desires but, more decisively, to be his will.⁴

Suppose a man wants to be motivated in what he does by the desire to concentrate on his work. It is necessarily true, if this supposition is correct, that he already wants to concentrate on his work. This desire is now among his desires. But the question of whether or not his second-order desire is fulfilled does not turn merely on whether the desire he wants is one of his desires. It turns on whether this desire is, as he wants it to be, his effective desire or will. If, when the chips are down, it is his desire to concentrate on his work that moves him to do what he does, then what he wants at that time is indeed (in the relevant sense) what he wants to want. If it is some other desire that actually moves him when he acts, on the other hand, then what he wants at that time is not (in the relevant sense) what he wants to want. This will be so despite the fact that the desire to concentrate on his work continues to be among his desires.

II

Someone has a desire of the second order either when he wants simply to have a certain desire or when he wants a certain desire to be his will. In situations of the latter kind, I shall call his second-order desires "**second-order volitions**" or "volitions of the second order." Now it is having second-order volitions, and not having second-order desires generally, that I regard as essential to being a person. It is logically possible, however unlikely, that there should be an agent with second-order desires but with no volitions of the second order. Such a creature, in my view, would not be a person. I shall use the term "**wanton**" to refer to agents who have first-order desires but who are not persons because, whether or not they have desires of the second order, they have no second-order volitions.⁵

The essential characteristic of a wanton is that he does not care about his will. His desires move him to do certain things, without its being true of him either that he wants to be moved by those desires or that he prefers to be moved by other desires. The class of wantons includes all non-human animals that have desires and all very young children. Perhaps it also includes some adult human beings as well. In any case, adult humans may be more or less wanton; they may act wantonly, in response to first-order desires concerning which they have no volitions of the second order, or less frequently.

The fact that a wanton has no second-order volitions does not mean that each of his first-order desires is translated heedlessly and at once into action. He may have no opportunity to act in accordance with some of his desires. Moreover, the translation of his desires into action may be delayed or precluded either by conflicting desires of the first order or by the intervention of deliberation. For a wanton may possess and employ rational faculties of a high order. Nothing in the concept of a wanton implies that he cannot reason or that he cannot deliberate concerning how to do what he wants to do. What distinguishes the rational wanton from other rational agents is that he is not concerned with the desirability of his desires themselves. He ignores the question of what his will is to be. Not only does he pursue whatever course of action he is most strongly

inclined to pursue, but he does not care which of his inclinations is the strongest.

Thus a rational creature, who reflects upon the suitability to his desires of one course of action or another, may nonetheless be a wanton. In maintaining that the essence of being a person lies not in reason but in will, I am far from suggesting that a creature without reason may be a person. For it is only in virtue of his rational capabilities that a person is capable of becoming critically aware of his own will and of forming volitions of the second order. The structure of a person's will presupposes, accordingly, that he is a rational being.

The distinction between a person and a wanton may be illustrated by the difference between two narcotics addicts. Let us suppose that the physiological condition accounting for the addiction is the same in both men, and that both succumb inevitably to their periodic desires for the drug to which they are addicted. One of the addicts hates his addiction and always struggles desperately, although to no avail, against its thrust. He tries everything that he thinks might enable him to overcome his desires for the drug. But these desires are too powerful for him to withstand, and invariably, in the end, they conquer him. He is an unwilling addict, helplessly violated by his own desires.

The unwilling addict has conflicting first-order desires: he wants to take the drug, and he also wants to refrain from taking it. In addition to these first-order desires, however, he has a volition of the second order. He is not a neutral with regard to the conflict between his desire to take the drug and his desire to refrain from taking it. It is the latter desire, and not the former, that he wants to constitute his will; it is the latter desire, rather than the former, that he wants to be effective and to provide the purpose that he will seek to realize in what he actually does.

The other addict is a wanton. His actions reflect the economy of his first-order desires, without his being concerned whether the desires that move him to act are desires by which he wants to be moved to act. If he encounters problems in obtaining the drug or in administering it to himself, his responses to his urges to take it may involve deliberation. But it never occurs to him to consider whether he wants the relation among his desires to result in his having

the will he has. The wanton addict may be an animal, and thus incapable of being concerned about his will. In any event he is, in respect of his wanton lack of concern, no different from an animal.

The second of these addicts may suffer a first-order conflict similar to the first-order conflict suffered by the first. Whether he is human or not, the wanton may (perhaps due to conditioning) both want to take the drug and want to refrain from taking it. Unlike the unwilling addict, however, he does not prefer that one of his conflicting desires should be paramount over the other; he does not prefer that one first-order desire rather than the other should constitute his will. It would be misleading to say that he is neutral as to the conflict between his desires, since this would suggest that he regards them as equally acceptable. Since he has no identity apart from his first-order desires, it is true neither that he prefers one to the other nor that he prefers not to take sides.

It makes a difference to the unwilling addict, who is a person, which of his conflicting first-order desires wins out. Both desires are his, to be sure; and whether he finally takes the drug or finally succeeds in refraining from taking it, he acts to satisfy what is in a literal sense his own desire. In either case he does something he himself wants to do, and he does it not because of some external influence whose aim happens to coincide with his own but because of his desire to do it. The unwilling addict identifies himself, however, through the formation of a second-order volition, with one rather than with the other of his conflicting first-order desires. He makes one of them more truly his own and, in so doing, he withdraws himself from the other. It is in virtue of this identification and withdrawal, accomplished through the formation of a second-order volition, that the unwilling addict may meaningfully make the analytically puzzling statements that the force moving him to take the drug is a force other than his own, and that it is not of his own free will but rather against his will that this force moves him to take it.

The wanton addict cannot or does not care which of his conflicting first-order desires wins out. His lack of concern is not due to his inability to find a convincing basis for preference. It is due either to his lack of the capacity for reflection or to his mindless indifference to the enterprise of evaluating his own

desires and motives.⁶ There is only one issue in the struggle to which his first-order conflict may lead: whether the one or the other of his conflicting desires is the stronger. Since he is moved by both desires, he will not be altogether satisfied by what he does no matter which of them is effective. But it makes no difference to *him* whether his craving or his aversion gets the upper hand. He has no stake in the conflict between them and so, unlike the unwilling addict, he can neither win nor lose the struggle in which he is engaged. When a *person* acts, the desire by which he is moved is either the will he wants or a will he wants to be without. When a *wanton* acts, it is neither.

III

There is a very close relationship between the capacity for forming second-order volitions and another capacity that is essential to persons—one that has often been considered a distinguishing mark of the human condition. It is only because a person has volitions of the second order that he is capable both of enjoying and of lacking **freedom of will**. The concept of a person is not only, then, the concept of a type of entity that has both first-order desires and volitions of the second order. It can also be construed as the concept of a type of entity for whom the freedom of its will may be a problem. This concept excludes all wantons, both infrahuman and human, since they fail to satisfy an essential condition for the enjoyment of freedom of the will. And it excludes those suprahuman beings, if any, whose wills are necessarily free.

Just what kind of freedom is the freedom of the will? This question calls for an identification of the special area of human experience to which the concept of freedom of the will, as distinct from the concepts of other sorts of freedom, is particularly germane. In dealing with it, my aim will be primarily to locate the problem with which a person is most immediately concerned when he is concerned with the freedom of his will.

According to one familiar philosophical tradition, being free is fundamentally a matter of doing what one wants to do. Now the notion of an agent who does what he wants to do is by no means an altogether

clear one: both the doing and the wanting, and the appropriate relation between them as well, require elucidation. But although its focus needs to be sharpened and its formulation refined, I believe that this notion does capture at least part of what is implicit in the idea of an agent who *acts* freely. It misses entirely, however, the peculiar content of the quite different idea of an agent whose *will* is free.

We do not suppose that animals enjoy freedom of the will, although we recognize that an animal may be free to run in whatever direction it wants. Thus, having the freedom to do what one wants to do is not a sufficient condition of having a free will. It is not a necessary condition either. For to deprive someone of his **freedom of action** is not necessarily to undermine the freedom of his will. When an agent is aware that there are certain things he is not free to do, this doubtless affects his desires and limits the range of choices he can make. But suppose that someone, without being aware of it, has in fact lost or been deprived of his freedom of action. Even though he is no longer free to do what he wants to do, his will may remain as free as it was before. Despite the fact that he is not free to translate his desires into actions or to act according to the determinations of his will, he may still form those desires and make those determinations as freely as if his freedom of action had not been impaired.

When we ask whether a person's will is free we are not asking whether he is in a position to translate his first-order desires into actions. That is the question of whether he is free to do as he pleases. The question of the freedom of his will does not concern the relation between what he does and what he wants to do. Rather, it concerns his desires themselves. But what question about them is it?

It seems to me both natural and useful to construe the question of whether a person's will is free in close analogy to the question of whether an agent enjoys freedom of action. Now freedom of action is (roughly, at least) the freedom to do what one wants to do. Analogously, then, the statement that a person enjoys freedom of the will means (also roughly) that he is free to want what he wants to want. More precisely, it means that he is free to will what he wants to will, or to have the will he wants. Just as the question about the freedom of an agent's action has to do

with whether it is the action he wants to perform, so the question about the freedom of his will has to do with whether it is the will he wants to have.

It is in securing the conformity of his will to his second-order volitions, then, that a person exercises freedom of the will. And it is in the discrepancy between his will and his second-order volitions, or in his awareness that their coincidence is not his own doing but only a happy chance, that a person who does not have this freedom feels its lack. The unwilling addict's will is not free. This is shown by the fact that it is not the will he wants. It is also true, though in a different way, that the will of the wanton addict is not free. The wanton addict neither has the will he wants nor has a will that differs from the will he wants. Since he has no volitions of the second order, the freedom of his will cannot be a problem for him. He lacks it, so to speak, by default.

People are generally far more complicated than my sketchy account of the structure of a person's will may suggest. There is as much opportunity for ambivalence, conflict, and self-deception with regard to desires of the second order, for example, as there is with regard to first-order desires. If there is an unresolved conflict among someone's second-order desires, then he is in danger of having no second-order volition; for unless this conflict is resolved, he has no preference concerning which of his first-order desires is to be his will. This condition, if it is so severe that it prevents him from identifying himself in a sufficiently decisive way with *any* of his conflicting first-order desires, destroys him as a person. For it either tends to paralyse his will and to keep him from acting at all, or it tends to remove him from his will so that his will operates without his participation. In both cases he becomes, like the unwilling addict though in a different way, a helpless bystander to the forces that move him.

Another complexity is that a person may have, especially if his second-order desires are in conflict, desires and volitions of a higher order than the second. There is no theoretical limit to the length of the series of desires of higher and higher orders; nothing except common sense and, perhaps, a saving fatigue prevents an individual from obsessively refusing to identify himself with any of his desires until he forms a desire of the next higher order. The tendency to generate such a series of acts of forming desires, which

would be a case of humanization run wild, also leads toward the destruction of a person.

It is possible, however, to terminate such a series of acts without cutting it off arbitrarily. When a person identifies himself *decisively* with one of his first-order desires, this commitment "resounds" throughout the potentially endless array of higher orders. Consider a person who, without reservation or conflict, wants to be motivated by the desire to concentrate on his work. The fact that his second-order volition to be moved by this desire is a decisive one means that there is no room for questions concerning the pertinence of desires or volitions of higher orders. Suppose the person is asked whether he wants to want to concentrate on his work. He can properly insist that this question concerning a third-order desire does not arise. It would be a mistake to claim that, because he has not considered whether he wants the second-order volition he has formed, he is indifferent to the question of whether it is with this volition or with some other that he wants his will to accord. The decisiveness of the commitment he has made means that he has decided that no further question about his second-order volition, or any higher order, remains to be asked. It is relatively unimportant whether we explain this by saying that this commitment implicitly generates an endless series of confirming desires of higher orders, or by saying that the commitment is tantamount to a dissolution of the pointedness of all questions concerning higher orders of desire.

Examples such as the one concerning the unwilling addict may suggest that volitions of the second order, or of higher orders, must be formed deliberately and that a person characteristically struggles to ensure that they are satisfied. But the conformity of a person's will to his higher-order volitions may be far more thoughtless and spontaneous than this. Some people are naturally moved by kindness when they want to be kind, and by nastiness when they want to be nasty, without any explicit forethought and without any need for energetic self-control. Others are moved by nastiness when they want to be kind and by kindness when they intend to be nasty, equally without forethought and without active resistance to these violations of their higher-order desires. The enjoyment of freedom comes easily to some. Others must struggle to achieve it.

IV

My theory concerning the freedom of the will accounts easily for our disinclination to allow that this freedom is enjoyed by the members of any species inferior to our own. It also satisfies another condition that must be met by any such theory, by making it apparent why the freedom of the will should be regarded as desirable. The enjoyment of a free will means the satisfaction of certain desires—desires of the second or of higher orders—whereas its absence means their frustration. The satisfactions at stake are those which accrue to a person of whom it may be said that his will is his own. The corresponding frustrations are those suffered by a person of whom it may be said that he is estranged from himself, or that he finds himself a helpless or a passive bystander to the forces that move him.

A person who is free to do what he wants to do may yet not be in a position to have the will he wants. Suppose, however, that he enjoys both freedom of action and freedom of the will. Then he is not only free to do what he wants to do; he is also free to want what he wants to want. It seems to me that he has, in that case, all the freedom it is possible to desire or to conceive. There are other good things in life, and he may not possess some of them. But there is nothing in the way of freedom that he lacks.

It is far from clear that certain other theories of the freedom of the will meet these elementary but essential conditions: that it be understandable why we desire this freedom and why we refuse to ascribe it to animals. Consider, for example, Roderick Chisholm's quaint version of the doctrine that human freedom entails an absence of causal determination.⁷ Whenever a person performs a free action, according to Chisholm, it's a miracle. The motion of a person's hand, when the person moves it, is the outcome of a series of physical causes; but some event in this series, "and presumably one of those that took place within the brain, was caused by the agent and not by any other events" (18). A free agent has, therefore, "a prerogative which some would attribute only to God: each of us, when we act, is a prime mover unmoved" (23).

This account fails to provide any basis for doubting that animals of subhuman species enjoy the

freedom it defines. Chisholm says nothing that makes it seem less likely that a rabbit performs a miracle when it moves its leg than that a man does so when he moves his hand. But why, in any case, should anyone *care* whether he can interrupt the natural order of causes in the way Chisholm describes? Chisholm offers no reason for believing that there is a discernible difference between the experience of a man who miraculously initiates a series of causes when he moves his hand and a man who moves his hand without any such breach of the normal causal sequence. There appears to be no concrete basis for preferring to be involved in the one state of affairs rather than in the other.⁸

It is generally supposed that, in addition to satisfying the two conditions I have mentioned, a satisfactory theory of the freedom of the will necessarily provides an analysis of one of the conditions of moral responsibility. The most common recent approach to the problem of understanding the freedom of the will has been, indeed, to inquire what is entailed by the assumption that someone is morally responsible for what he has done. In my view, however, the relation between moral responsibility and the freedom of the will has been very widely misunderstood. It is not true that a person is morally responsible for what he has done only if his will was free when he did it. He may be morally responsible for having done it even though his will was not free at all.

A person's will is free only if he is free to have the will he wants. This means that, with regard to any of his first-order desires, he is free either to make that desire his will or to make some other first-order desire his will instead. Whatever his will, then, the will of the person whose will is free could have been otherwise; he could have done otherwise than to constitute his will as he did. It is a vexed question just how "he could have done otherwise" is to be understood in contexts such as this one. But although this question is important to the theory of freedom, it has no bearing on the theory of moral responsibility. For the assumption that a person is morally responsible for what he has done does not entail that the person was in a position to have whatever will he wanted.

This assumption *does* entail that the person did what he did freely, or that he did it of his own free

will. It is a mistake, however, to believe that someone acts freely only when he is free to do whatever he wants or that he acts of his own free will only if his will is free. Suppose that a person has done what he wanted to do, that he did it because he wanted to do it, and that the will by which he was moved when he did it was his will because it was the will he wanted. Then he did it freely and of his own free will. Even supposing that he could have done otherwise, he would not have done otherwise; and even supposing that he could have had a different will, he would not have wanted his will to differ from what it was. Moreover, since the will that moved him when he acted was his will because he wanted it to be, he cannot claim that his will was forced upon him or that he was a passive bystander to its constitution. Under these conditions, it is quite irrelevant to the evaluation of his moral responsibility to inquire whether the alternatives that he opted against were actually available to him.⁹

In illustration, consider a third kind of addict. Suppose that his addiction has the same physiological basis and the same irresistible thrust as the addictions of the unwilling and wanton addicts, but that he is altogether delighted with his condition. He is a willing addict, who would not have things any other way. If the grip of his addiction should somehow weaken, he would do whatever he could to reinstate it; if his desire for the drug should begin to fade, he would take steps to renew its intensity.

The willing addict's will is not free, for his desire to take the drug will be effective regardless of whether or not he wants this desire to constitute his will. But when he takes the drug, he takes it freely and of his own free will. I am inclined to understand his situation as involving the overtermination of his first-order desire to take the drug. This desire is his effective desire because he is physiologically addicted. But it is his effective desire also because he wants it to be. His will is outside his control, but, by his second-order desire that his desire for the drug should be effective, he has made this will his own. Given that it is therefore not only because of his addiction that his desire for the drug is effective, he may be morally responsible for taking the drug.

My conception of the freedom of the will appears to be neutral with regard to the problem of

determinism. It seems conceivable that it should be causally determined that a person is free to want what he wants to want. If this is conceivable, then it might be causally determined that a person enjoys a free will. There is no more than an innocuous appearance of paradox in the proposition that it is determined, ineluctably and by forces beyond their control, that certain people have free wills and that others do not. There is no incoherence in the proposition that some agency other than a person's own is responsible (even *morally* responsible) for the fact that he enjoys or fails to enjoy freedom of the will. It is possible that a person should be morally responsible for what he does of his own free will and that some other person should also be morally responsible for his having done it.¹⁰

On the other hand, it seems conceivable that it should come about by chance that a person is free to have the will he wants. If this is conceivable, then it might be a matter of chance that certain people enjoy freedom of the will and that certain others do not. Perhaps it is also conceivable, as a number of philosophers believe, for states of affairs to come about in a way other than by chance or as the outcome of a sequence of natural causes. If it is indeed conceivable for the relevant states of affairs to come about in some third way, then it is also possible that a person should in that third way come to enjoy the freedom of the will.

NOTES

1. P. F. Strawson, *Individuals* (London: Methuen, 1959), pp. 101–2. Ayer's usage of "person" is similar: "it is characteristic of persons in this sense that besides having various physical properties . . . they are also credited with various forms of consciousness" (A. J. Ayer, *The Concept of a Person* [New York: St. Martin's, 1963], p. 82). What concerns Strawson and Ayer is the problem of understanding the relation between mind and body, rather than the quite different problem of understanding what it is to be a creature that not only has a mind and a body but is also a person.
2. For the sake of simplicity, I shall deal only with what someone wants or desires, neglecting related phenomena such as choices and decisions. I propose to use the verbs "to want" and "to desire" interchangeably, although they are by no means

perfect synonyms. My motive in forsaking the established nuances of these words arises from the fact that the verb "to want," which suits my purposes better so far as its meaning is concerned, does not lend itself so readily to the formation of nouns as does the verb "to desire." It is perhaps acceptable, albeit graceless, to speak in the plural of someone's "wants." But to speak in the singular of someone's "want" would be an abomination.

3. What I say in this paragraph applies not only to cases in which "to *X*" refers to a possible action or inaction. It also applies to cases in which "to *X*" refers to a first-order desire and in which the statement that "*A* wants to *X*" is therefore a shortened version of a statement—"A wants to want *X*"—that identifies a desire of the second order.
4. It is not so clear that the entailment relation described here holds in certain kinds of cases, which I think may fairly be regarded as non-standard, where the essential difference between the standard and the non-standard cases lies in the kind of description by which the first-order desire in question is identified. Thus, suppose that *A* admires *B* so fulsomely that, even though he does not know what *B* wants to do, he wants to be effectively moved by whatever desire effectively moves *B*; without knowing what *B*'s will is, in other words, *A* wants his own will to be the same. It certainly does not follow that *A* already has, among his desires, a desire like the one that constitutes *B*'s will. I shall not pursue here the questions of whether there are genuine counterexamples to the claim made in the text or of how, if there are, that claim should be altered.
5. Creatures with second-order desires but no second-order volitions differ significantly from brute animals, and, for some purposes, it would be desirable to regard them as persons. My usage, which withholds the designation "person" from them, is thus somewhat arbitrary. I adopt it largely because it facilitates the formulation of some of the points I wish to make. Hereafter, whenever I consider statements of the form "*A* wants to want to *X*," I shall have in mind statements identifying second-order volitions and not statements identifying second-order desires that are not second-order volitions.
6. In speaking of the evaluation of his own desires and motives as being characteristic of a person, I do not mean to suggest that a person's second-order volitions necessarily manifest a *moral* stance on his

part toward his first-order desires. It may not be from the point of view of morality that the person evaluates his first-order desires. Moreover, a person may be capricious and irresponsible in forming his second-order volitions and give no serious consideration to what is at stake. Second-order volitions express evaluations only in the sense that they are preferences. There is no essential restriction on the kind of basis, if any, upon which they are formed.

7. "Freedom and Action," in *Freedom and Determinism*, ed. Keith Lehrer (New York: Random House, 1966), pp. 11–44.
8. I am not suggesting that the alleged difference between these two states of affairs is unverifiable. On the contrary, physiologists might well be able to show that Chisholm's conditions for a free action are not satisfied, by establishing that there is no relevant brain event for which a sufficient physical cause cannot be found.
9. For another discussion of the considerations that cast doubt on the principle that a person is morally responsible for what he has done only if he could have done otherwise, see my "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* (1969): 829–39.
10. There is a difference between being *fully* responsible and being *solely* responsible. Suppose that the willing addict has been made an addict by the deliberate and calculated work of another. Then it may be that both the addict and this other person are fully responsible for the addict's taking the drug, while neither of them is solely responsible for it. That there is a distinction between full moral responsibility and sole moral responsibility is apparent in the following example. A certain light can be turned on or off by flicking either of two switches, and each of these switches is simultaneously flicked to the "on" position by a different person, neither of whom is aware of the other. Neither person is solely responsible for the light's going on, nor do they share the responsibility in the sense that each is partially responsible; rather, each of them is fully responsible.

KEY TERMS

Second-order desires	Freedom of will
First-order desires	Freedom of action
Second-order volitions	Determinism
Wanton	