Stop Me Before I Kill Again

There are cases where we think that someone acted, but unfreely, in a way that excuses her of moral responsibility for what she did. I'm thinking of cases where we say things like: "she's not to blame; she acted on an irresistible desire"; "he couldn't help it; he's an addict (kleptomaniac, pyromaniac, etc.)", and "she can't help what she does; she's the victim of a compulsive desire".

When we describe someone in these sorts of ways, we are doing more than just saying that she acted unfreely and that we should sympathize and try to help rather than blame or punish. We are giving a certain kind of explanation of the person's unfreedom and offering a certain kind of reason for not holding her responsible. We think that the person acted unfreely because she was in some sense *psychologically compelled* to do what she did; we think that she was unable to resist the desire which caused her to act. And we excuse her because we think it is unreasonable and unfair to blame someone for what she could not help doing.

This contrasts with another way in which we excuse people from responsibility for what they do -- cases where we say things like: "she didn't realize what she was doing"; "she only meant to help, she didn't know it would make things worse", or, "she's just a child; she didn't understand that it was wrong". Here we excuse, not because the person acted unfreely, but because she lacked some relevant bit of knowledge - she had false beliefs about her action or its consequences, or, in the case of the child, she didn't realize that acts of this kind are wrong.

I don't mean to suggest that it's always easy to distinguish the two kinds of cases, or that there aren't cases where neither excuse seems entirely appropriate. Young children and the clinically mentally ill are an example. Do we excuse them because they lack the appropriate knowledge of what they do or because they are caused to act by desires they cannot resist? Or do we excuse them for some more fundamental reason - because they lack the kind of cognitive skills and capacities required for someone to qualify as a morally responsible being?

The cases I'm interested in, however, are cases where we think that someone acted unfreely because she was psychologically compelled, cases where we think that the problem is, in some broad sense, volitional, rather than cognitive. We may disagree about where to draw the line -- do "workaholics" and "exercise addicts" suffer from psychological compulsion or do they just really like to work or exercise? But we assume

that there is a distinction between those who are psychologically compelled and the rest of us, a distinction which is relevant to moral responsibility.

It is this assumption that I want to question. For it is far from clear how we should understand the sort of metaphors we use when we describe addicts and other compulsives. We use expressions like "victim of her desire" which suggest that we may be helpless with respect to our own desires in the way that we may be helpless with respect to external forces. A prisoner who is dragged through the streets can't help the way his body moves; his body moves regardless of his desires, regardless of how hard he tries to resist. But our desires are not literally forces which push and pull us into action. How, then, can a desire compel someone to act?

First try: When we act freely, what we do depends on our beliefs as well as our desires. But those who are compelled have a desire which causes action regardless of beliefs.

There may be desires which cause behavior regardless of beliefs. The most plausible examples are cases in which a person repeats a simple body movement (washing her hands, for instance) hundreds of times a day, for no apparent reason. But these are not cases of *action*; they are cases of involuntary body movement, like a hiccup or a nervous tic.[1] The cases I'm interested in are cases of action, cases in which what the person does depends on her beliefs as well as her desires. The heroin addict would not stick the needle into his arm if he didn't believe it contained heroin.

It's true that we can distinguish, among involuntary body movements, those which occur regularly from those which occur less often.(But what would be the point of making this distinction? An occasional nervous tic is less troublesome than one which occurs more often, but the former is no more in the person's control than the latter.) And we can draw a similar distinction among actions. Some people drink occasionally, others on a regular basis; the alcoholic drinks nearly all the time. Some people use drugs as recreation; the junkie has a habit. But this difference in behavior is, at best, relevant only as *evidence* of unfreedom. The fact that a person would always (or usually) act on a certain kind of desire doesn't entail that she cannot help doing so. To be driven, single-minded, devoted, fanatic, or even obsessed is not necessarily to act unfreely.

Second try: The difference between someone who acts freely (even if habitually) and someone who is compelled lies in the desire which causes action. Compulsive desires are different from other desires. If the thief suffers from kleptomania, then it's not physically possible, given his desire, his beliefs, and his total psycho-physical state, that he refrains from stealing. He can't help stealing in exactly the way that a rock released from a window can't help falling to the ground. But if his desire to steal is a normal desire,

then it *is* physically possible, given his desire, beliefs, and total psycho-physical state, both that he steals *and* that he refrains from stealing (acting, perhaps for the first time, out of character and on moral principle).

This may be our commonsense way of drawing the distinction, but note what this commits us to. If we say that the kleptomaniac is compelled because there is a deterministic causal explanation of what he does, then we have rejected the claim that free action is compatible with determinism.

At one time, it was widely thought that the problem of free will and determinism is a "pseudo-problem" which comes from confusing causation with compulsion. This, I think, is a mistake. We now have reason to believe that determinism is false, and perhaps also reason to believe that there are some causes which are irreduceably probabilistic. If so, then there is at least one way of distinguishing between causes which compel and those which merely "incline without necessitating", to use Leibniz's quaint phrase.

It might turn out to be true that some of our actions lack deterministic causes. We *might* even learn, at the end of the scientific day, that each of the actions which we thought free had only an indeterministic causal explanation, whereas each of the actions we thought unfree had a deterministic causal explanation. But this co-incidence between our beliefs about freedom and the facts about deterministic causation would not establish the truth of the *philosophical* thesis that free action is incompatible with determinism.

I think, and have argued elsewhere, that free action (and being able to do otherwise) is compatible with the truth of determinism. In this paper, I will simply assume this, and address the question of psychological compulsion only insofar as it presents a problem for someone who believes that an action may be both causally determined and free.[2]

It's one thing to say that free action (or free will) is compatible with causal determinism; another thing to provide a positive account of what free action (or free will) is. Most philosophers are, or think they are, Compatibilists. But there is, at the present time, almost no agreement about what free will is or about what distinguishes those who act freely (in the sense relevant to moral responsibility) from those who act unfreely. Not too long ago, however, there was a widely accepted theory of freedom which I will call Simple Compatibilism. It went something like this:

Even if determinism turned out to be true, there would still be a difference between the falling of a rock and human action. The rock has no choice about whether or not it falls because it has no beliefs and desires and thus is not capable of making *any* choices. But we have beliefs and desires, and we are capable of making choices on the basis of what we believe and want.

And even if determinism turned out to be true, there would still be a difference between me and the prisoner who is dragged through the streets. The prisoner is like me and unlike the rock in having a will - in being capable of making choices on the basis of his beliefs and desires. But he is unlike me in that his will is not causally efficacious. Whether I walk or run depends on what I choose, which in turn depends on what I believe and want. I ran because I wanted to catch the bus and because I believed that I wouldn't catch the bus unless I ran. But if I had wanted exercise more than I wanted to take the bus, I would have chosen to continue walking instead. And if I had chosen to continue walking, I would have done so. But what happens to the prisoner happens to him regardless of his will - regardless of what he chooses, regardless of how much he wants *not* to be dragged. He is the helpless victim of an irresistible force; I am not. My will is free; his is not.

We may summarize Simple Compatibilism with the following slogan: to have free will is to have a causally efficacious will. Or, to put it slightly differently, to have free will is to be capable of acting for reasons which are also causes.

Simple Compatibilism allows us to draw some of the distinctions we intuitively want to draw. Entities which cannot make choices (rocks and other things which lack mental states, the comatose and others suffering from certain kinds of neurological defects or damage) lack free will. We lack free will to the extent that something prevents (or would prevent) us from doing what we want (or might want) to do. This includes cases where we cannot move at all (chains, paralysis) as well as cases where the movements of our bodies are involuntary (falling downstairs, a muscle spasm, nervous tic, and so on).

But Simple Compatibilism takes a very simple view about action. According to Simple Compatibilism, anyone who acts intentionally acts freely. This includes those we think victims of compulsive desires. Someone "compelled by her own desire" is, nevertheless, someone who intends or chooses to do something because she believes it will get her what she wants. If Simple Compatibilism is right, we are making a mistake when we say that such a person acts unfreely.

This is a hard view to accept, and most Compatibilists no longer accept it. It's generally agreed that the phenomenon of psychological compulsion shows that Simple Compatibilism is false. The consensus seems to be that if we still want to be Compatibilists we must defend a more sophisticated theory of free will, one which can make sense of the unfreedom of psychological compulsion.

In this paper, I will look at three such accounts and argue that they all fail. I will argue that Simple Compatibilism is right; to act intentionally *is* to act freely in the sense necessary for moral responsibility. Finally, I will argue that it's a mistake to confuse the

question of free will with the question of responsibility; the addict acts freely and with free will, but there may be other reasons for not holding her responsible.

Compulsion as Coercion: Choosing the Lesser Evil

To be physiologically addicted to a drug is to depend on the drug for the normal functioning of one's body. Withdrawal from a drug to which one is addicted is painful. An addict may take a drug, not for the sake of anticipated pleasure, but because she fears the pain of withdrawal.[3]

This suggests that to be addicted is to be unfree in the way that someone who is coerced is unfree. Someone who is coerced ("Your money or your life") is someone who acts in order to avoid a threatened evil. It would be misleading to say that the victim of the gunman hands over her money willingly or that she does so because that's what she wants to do. What she wants is to save her life; she wants to hand over her money only as a means to this end.

On this view, the physiological state of addiction is like a constant threat ("Take the drug or you will suffer terribly"). It would be misleading to say that the addict takes the drug willingly or that she takes it because that's what she wants to do. What she wants is to avoid the pain of withdrawal; she wants to take the drug only as a means to this end.

Not all the cases we call compulsion are cases of physiological addiction. But perhaps what they have in common is that the compelled person acts, not for the sake of some good she hopes to gain by acting, but in order to avoid the pain caused by frustration of the desire.[4]

There is something to this idea. While a desire is not a feeling, there is no denying the fact that desires may be accompanied by feelings and that these feelings may be pleasant or unpleasant. Some desires hurt when they are not satisfied. Perhaps this is the key to understanding our talk about *suffering* from psychological compulsion and being the victim or slave of one's desire.

But there are problems with using the model of coercion to explain why the psychologically compelled are unfree in any sense relevant to responsibility.

First, while it's often said that coercion is a paradigm case of unfreedom, it's not clear why we should think that coerced action is unfree in the sense relevant to responsibility. Granted, someone who does something at gun point is someone who is taken advantage of by the gunman, who exploits his victim's rational fears to get her to do something she would not otherwise have done. Granted, it's bad to be used by the gunman as a mere means to serve *his* ends. But not everything bad involves a loss of freedom. So why should we think that the victim of the gunman acts unfreely?

Is it that the coerced agent is literally unable to do other than what she does, in the way the prisoner being dragged is unable to avoid being dragged? No. The options of the coerced agent are severely limited, but she has a choice. When we say that the coerced agent had "no real choice" or "no reasonable choice" we mean (depending on the circumstances) that she chose the lesser of two evils or that the choice which would have resulted in the greater overall good would have cost her more than it is reasonable to expect.[5]

Is it that the coerced agent doesn't *really* want to do what she does? No. Granted, she doesn't want to have to make a choice in these conditions; she doesn't want to be threatened at gun point. But given the situation she is in fact in, she really wants to do what she does. Given that the only way to save her life is to hand over her money, she wants to hand over the money and doesn't want anything to prevent her from doing so.

But doesn't the fact that we don't blame the coerced agent for doing what would otherwise be blameworthy (e.g., giving away the bank's money) show that we think she is unfree in the sense relevant to responsibility?

No. As J.L. Austin famously pointed out[6], there are two very different ways in which someone might answer an accusation of wrongdoing: She might say: "It was wrong, but I didn't really do it." ("It was a mistake"; "it was an accident"; "I was pushed," and so on.) Or she might say: "I did it, but, given the circumstances, it wasn't wrong." In the first case, the person tries to *excuse* herself from responsibility by claiming that what happened wasn't an action or wasn't the action she intended to perform; in the second case, she accepts responsibility for her action, but claims a justification for what she did. The coerced agent defends her action in the second way, and we absolve her of blame only if we are convinced that the threatened evil justified her action. "The bank's money or your life" is a justification; "The bank's money or I'll sing annoyingly out of tune" is not.

But if the coerced agent acts freely and responsibly, then why do people think that the coerced are unfree, or at least *less* free?

I think the answer is this: There's a difference between acting freely and the extent of one's freedom. We think that we are free to the extent that we have options we value having, and a coercive threat diminishes our options. Before the gunman confronted his victim, she had the option of keeping both her money and her life; now she has to choose between them.

If we are free to the extent that we have options we value having, then the gunman's threat leaves his victim less free than she was before the threat. And if the physiological state of addiction functions like a coercive threat, then the drug addict is less free than she was before she became addicted. But this is a very broad way of thinking

about freedom; on this view, someone may become less free when she has an unwanted pregnancy, becomes poorer, loses a job, or loses her looks. If this is the only sense in which addiction deprives of freedom, then the addict is unfree in the way we all are, at different times, and to different degrees. We should not take the addict literally when she complains that her desire compels her to act, that she is helpless to resist it. She's forced to take the drug in the sense that someone with an unwanted pregnancy is forced to have an abortion or an aging actress is forced to accept a role she would not have taken in her prime.

To lack options we value having is one way in which we may lack freedom worth wanting; so coercion deprives us of freedom worth wanting. Whether there are also other ways of being deprived of freedom worth wanting (and how narrowly or broadly coercion should be understood) is a controversial normative question. My concern here is *not* with this question but with freedom in the sense relevant to moral responsibility. Although it's easy to conflate the freedom worth wanting with the freedom required for moral responsibility, I think it's important to distinguish the two. A prisoner and a resident of a totalitarian state both lack freedom which is worth wanting, but it doesn't follow that they lack the freedom required to be a morally responsible being.

The second problem with the view that the unfreedom of psychological compulsion is the unfreedom of coercion is that the prospect of withdrawal pain does not necessarily count as a coercive threat. The bank clerk's choice may be described as a choice between two evils insofar as both her options involve something bad - handing over the money which it's her job to safeguard or losing her life. And the choice she makes is one that both she and we regard as reasonable, given the circumstances.

But is this true of those we think suffer from psychological compulsion? This depends on the details of the case. Suppose that someone's choice is restricted to these two options:

1. Take drugs for relief from pain of terminal cancer; die sooner.
2. Don't take drugs; live longer in terrible pain.

Here the choice is between two evils, and people may disagree about which is the lesser evil. But most of us would agree that choosing to take drugs is a reasonable choice, given the circumstances.

But suppose that an addict thinks that his choice is restricted to these options:

1. Steal from my children's college fund and take drugs.
2. Don't take drugs and feel very sick for a few days.

If these are the addict's only options, then his choice is a choice between two evils. But this doesn't mean that the threat of withdrawal pain counts as a *coercive* threat. Coercion is a normative concept. Whether a threat counts as coercive depends on whether compliance with the threat would be justified; that is, on whether compliance would be the lesser of two evils ("The bank's money or your life") or on whether the cost of noncompliance would be too great to expect the agent to bear. ("Give us the name of your comrades in the Resistance or we'll torture you for three more days.")

In the situation we are imagining, the addict's withdrawal pain is a considerably lesser evil than the loss of his children's college fund. And it's doubtful whether it's unreasonable to expect someone to suffer a few days of pain for the sake of his children's future. So it seems at best doubtful whether the prospect of withdrawal pain counts as a coercive threat.

The last point is this: Even if some of those we call compelled are also coerced (and thus lack freedom, though not responsibility, for this reason), psychological compulsion is not the same as coercion. Not all compulsives try to justify their behavior by claiming that the pain of unsatisfied desire is so great that it's not reasonable to expect them to resist. On the contrary, an addict might say something like this:

> A few days of feeling very sick are not so bad, taking the long view, as a drug free life. I agree I ought to quit. But I just can't help myself. I need the drug; I've got to have it.

Such a person is not claiming that his action is the rational response to pain; he is confessing weakness and irrationality.

This suggests another way of accounting for the unfreedom of psychological compulsion.

Compulsion as Irrationality: Acting against one's Better Judgment

Consider the addict who agrees that the life of addiction is not a good life, and that stealing from his children's college fund is a terrible thing to do, but who nevertheless goes ahead and does it. Unlike the addict who tries to justify his action by claiming that it's reasonable given his physiological dependence on the drug, this addict agrees there is *no* justification for what he does, and feels terrible and hates himself. This person has done what he most wants in one sense, yet we are inclined to say that there is another sense in which he has failed to do what he most wants. He has acted on his strongest desire, but he's failed to act according to what he most values and believes he ought to do.

This suggests another way of understanding the addict's plea that he is enslaved by his own desire. Perhaps the addict's helplessness lies in his failure to make his will conform to his reason, and, in particular, to his beliefs about what's good for him, his beliefs about right and wrong, and his beliefs about what he ought to do. Whereas the rest of us are able to conform our choices and hence our actions to what we think we ought to do, the addict's desire overcomes his reason.

This view, which dates back to Plato, has been revived by Gary Watson.[7] In an important and well-known paper, Watson argues that we cannot make sense of the unfreedom of the addict unless we reject Hume's account of value and his view of the role of reason in deliberation and motivation. According to Hume, to value something is just to care about it, to want it, to be motivated (in the appropriate circumstances) to try to get it. On the Humean view, we may use reason to figure out what to believe and how to most effectively get what we want. But reason has no role to play in determining what we *should* want, or what is *worth wanting.*

Watson argues that if Hume is right, then it makes no sense to say that someone acts on his strongest desire but fails to do what he most values or thinks best. For if values are just desires, then there is no difference between what we most value and what we most strongly desire. Since our strongest desire is the one that actually moves us to act (or try to act), it follows that whenever we act intentionally we act according to what we most value or think best. Despite his protests to the contrary, the addict who steals from his children does what he thinks best; his action shows that he values his short-term pleasure more than he values the future of his children.

Plato's view of value, reason, and motivation was very different. He thought that to value something is to believe it to be good and worth wanting, and he thought that reason *can* tell us what's good and worth wanting. That is, reason can tell us, not just what we ought to do given what we already want, but also what we *ought* to want. Since what we ought to want is not necessarily what we *actually* want, we may want something without thinking it good or worth wanting.

Watson suggests that the key to understanding the unfreedom of addicts lies in Plato's view that we may be motivated either by what reason tells us we ought to do or by mere desire. If we can be motivated either by reason or by desire, there is a possibility of conflict. We may do what we most strongly desire without doing what we think is best or most worth doing.

Watson argues that if we reject Hume's account of value, then we can retain a Compatibilist account of freedom, while being able to draw distinctions Hume was unable to draw. By saying that to do what one most wants is to act freely, the Simple

Compatibilist view in effect says that an agent's will is what she most wants to do. If there is no distinction between someone's causally strongest desire and what she most values, as Hume thought, then Simple Compatibilism reduces to the view that an agent's will is her strongest desire. But if there is a distinction between wanting most in the causal sense and valuing most, then it makes sense to identify the agent's will, not with her causally strongest desire, but with her value judgment about what she ought to do. Given this, we can say that the addict who steals from his children acts, not just against his better judgment, but also against his will, and thus unfreely.

Watson's suggestion is appealing. While we may disagree about how to best interpret the behavior of the addict, it seems intelligible that he steals from his children despite his judgment that this is not what he ought to do, even given his need for the drug. If Hume's account of value cannot make sense of this, then we should reject Hume's account.

But is Watson also right in proposing that an agent's will simply *is* her judgment about what she ought to do? And is he right in claiming that someone who acts against her better judgment thereby acts unfreely?

Intuitively, there is a difference between someone who acts against her better judgment despite the fact that she *wants* to do what she believes she ought to do, and someone who doesn't want to do what she agrees she ought to do.

Bob is a utilitarian and thus believes that he ought to do what will maximize the greatest good of the greatest number. But sometimes when he's forced to choose between the greatest good and his good, or between the greatest good and the good of those near and dear to him, he does what's best for him or those he loves. It's far from obvious that whenever Bob fails to live up to his demanding moral theory this is because he is unable to conform his conduct to his better judgment. A more plausible explanation is that he cares more about his good and the good of those he loves than he cares about doing what he believes to be best. Or, to put it more bluntly, there are occasions on which Bob does not want to do what he believes he ought to do.

X lies to Customs officials. When we question him later, he agrees that breaking the law is wrong, but candidly explains that he did it because he didn't want to spend hundreds of dollars on import duties. X acted contrary to his judgment about what he ought to do. But we've got no reason to suppose he acted unfreely.

Suzy has a weakness for trashy novels. She is sitting on her front steps, basking in the early morning sun, reading a Harlequin romance. She'd be embarrassed if any of her friends caught her doing it, and she knows she should really be grading papers. But she indulges herself and continues to read. Suzy may be weak willed; she herself calls it

procrastination and self-indulgence. But most of us would say that she acts freely in the sense relevant to moral responsibility.

These three cases differ from each other in some ways and from the case of our unhappy addict in other ways, but they are alike with respect to what Watson has told us is relevant to unfree action - they are all cases in which someone acts contrary to her judgment about what she ought to do.

Cases like Bob, X, and Suzy pose a problem for Watson. There are only two ways he can respond.

He might bite the bullet and admit that Bob, X, and Suzy act against their better judgment, hence unfreely in the sense relevant to moral responsibility. But this is wildly implausible. If these agents act unfreely, then it's hard to see how anyone can ever be responsible for doing something wrong. For either she did not know that it was wrong (in which case she is not responsible because she was ignorant) or she knew it was wrong (in which case she is not responsible because she acted against her judgment about what she ought to do).

Or Watson might deny that Bob, X, and Suzy really act against their better judgment. He might argue that X is more plausibly described as someone who thinks that breaking the law is only prima facie wrong. Perhaps X thinks that the Customs laws are absurd and should not be obeyed. If so, then X does not act contrary to his judgment about what he ought to do, *all things considered.* And he might make a similar claim about Bob and Suzy. Bob acts contrary to his beliefs about what *morality* requires, but he doesn't act contrary to his judgment about what he ought to do, all things considered. Bob weighs the demands of morality against the cost to himself (or those he loves) and concludes that on this occasion morality is not worth the cost. Suzy acts contrary to her beliefs about she ought to do, qua conscientious teacher; but her all things considered judgment is that she's been working very hard recently and deserves a break.

There is always room to argue about particular cases. Despite what he says, perhaps X really believes that lying to Customs officials is justified. Despite what he says, perhaps Bob believes that we are sometimes justified in doing what will not bring about the greatest good. Despite what she says, perhaps Suzy really believes that reading the Harlequin romance isn't just what she *wants* to do; it's what she *ought* to do (all things considered). But Watson needs to make a stronger claim, if he is to argue that, despite appearances, people like Bob, X, and Suzy do not act contrary to their all things considered value judgment. He has to argue that we must *always* disregard what people say and what they appear to believe, and look instead to their actions to find out what they think best, all things considered.

But now I think that Watson's account is in danger of collapsing back into Hume's. On Hume's view, there is no difference between what someone thinks best and what she wants most. Watson criticizes Hume on the grounds that his view makes it impossible to make sense of someone acting against her better judgment. But if thinking best is not the same as wanting most, it must be possible to say, of someone, that she thinks it best (all things considered) to do x, even though her action shows that she wants something else more. Watson *seems* prepared to say this, but only in certain sorts of cases -- cases like that of the drug addict and others we think victims of psychological compulsion. But if wanting most doesn't entail thinking best, then this should be true for *any* kind of desire, including X's desire to save money, Suzy's desire to bask in the sun reading trashy novels, and Bob's desire concerning the well-being of those he loves. If Watson nevertheless insists that people like Bob, Suzy, and X do what they think is best, all things considered, then he should apply the same standard to the addict who steals from his children. His actions speak louder than his words; despite his protests to the contrary, he does what he thinks best, all things considered.

Watson makes two independent points. He says that Hume went wrong in failing to distinguish between wanting most and thinking best, and he says that Simple Compatibilism went wrong in thinking that someone's will is what she wants most. I am persuaded by Watson's criticism of Hume's account of value; I think we need an account of value on which it's possible for someone to act contrary to her all things considered judgment about what she ought to do. But the cases of Bob, X, and Suzy show that it's a mistake to identify the agent's will with her judgment about what she ought to do. Bob, X, and Suzy all fail to do what they think they ought to do, but they don't act "against" or "despite" their will, nor do they suffer from any psychological compulsion, impairment or unfreedom of will. Hume and Simple Compatibilism get the right verdict here; Bob, X, and Suzy act freely because they do what they most want to do.

I think that Watson's mistake was in rejecting only half of Hume's account of value. On the Humean account, to want something is to think it good, and to think something good is to want it. Watson rejects the first half of Hume's account - that to want something is to think it good. But he follows Plato in retaining the second half - that to think something good is to want it.[8] Given this view, it's natural to identify a person's judgment about what she ought to do, all things considered, with her will. And it's natural to think that someone who acted contrary to her "rational desire" - her desire to do what she thinks she ought to do - is someone who did so *because* she had a nonrational desire (appetite, passion, etc.) which "overwhelmed" or "overpowered" her rational desire.

If you hold this Platonic view of the connection between reason, value, and motivation, then you will tend to overlook the possibility of people like Bob, X, and Suzy, people who seem not to have any desire to do the act they claim to think best. You will be forced either to call these people unfree or to cast doubt on their sincerity. (Or, perhaps, you might argue that they are somehow momentarily confused or mistaken about what they ought to do.)[9]

The moral I would draw is this: Hume was mistaken on both counts: It's false that to want something is to think it good and it's also false that to think something good is to want it. To think that something is good or that one ought to do a certain kind of action is to have a belief, and no belief necessarily causes any desire. (Hume *was* right about this.)

What kind of belief? This depends on the person's normative views. In Bob's case, it's the belief that the right action is the one that has the best overall consequences, taking into account everyone affected by the action. Someone else might have the belief that she ought to do what will maximize the satisfaction of her preferences, provided that in so doing she isn't breaking any laws or causing harm to others. Someone else might have the belief that she ought to do what will maximize the satisfaction of her preferences, regardless of the cost to others. And so on. On any of these views, it is possible for someone to believe that she ought to do something without wanting (at that time) to do it. And in all these cases I think that Simple Compatibilism gets it right; the person acts freely in the sense relevant to moral responsibility.

But you don't need to agree with me about value to accept my criticism of Watson's account. Watson's account fails because it cannot distinguish people like Bob, X, and Suzy from those we think victims of psychological compulsion. If there is a relevant difference between those who act freely and those who are compelled, it lies elsewhere.

But what is this difference? Intuitively, it has something to do with the fact that Bob, X, and Suzy have no desire to do what they think best (or to do what they claim to think best), whereas the unhappy addict *wants* to do what he thinks best(or to do what he claims to think best). We may disagree about whether this difference in desire constitutes a difference in value judgment. But I think we can agree that the relevant difference lies in desire.

This brings us to Frankfurt's second order desire account.

Compulsion as Unfreedom of Will: Acting against Second order Desire

Let's begin by considering another addict.

Mary is trying to quit smoking. We have excellent evidence that she's sincere not only in her claim that she thinks she ought not to smoke, but also in her claim that she

wants to stop smoking. She's gotten rid of all the cigarettes in the house, avoids places where smokers hang out, has instructed her friends not to give her cigarettes even if she begs for them, and engages in all the displacement activity she can think of. She succeeds in not smoking for three days. But on the evening of the fourth day, she says "I can't stand it any longer; I need some nicotine". She searches the house until she finds a cigarette in a coat pocket. She smokes it, cursing herself for her weakness.

Contrast Mary to Suzy, who is a self-confessed trashy novel junkie. Suzy says things like "When I start reading a good one; I can't stop until I get to the end." But she doesn't mean this literally; she means that once she starts she usually doesn't *want* to stop before she gets to the end. Nor does Suzy (despite her slight embarrassment about having such lowbrow tastes) want to get rid of her desire to read trashy novels. She enjoys reading them and thinks this is a good way to relax. Today, however, she was planning to work. But somehow she found herself on the front steps and now she is halfway through the first chapter. She stops, thinks about it, and says to herself "I really ought to stop reading". She gives this just a moment's thought, then happily continues to turn the pages.

When Mary smokes the cigarette, she acts against her better judgment. But what distinguishes her from Suzy (and Bob and X) is the fact that she acts on a desire she wants not to be moved by, a desire she's been trying to eliminate. It is for this reason that it seems true of Mary, unlike the others, that she acts on her desire *against* or *despite* her will. It also seems true of her, if it is ever true of anyone, that Mary suffers from the kind of unfreedom we've been calling psychological compulsion.

This brings us to Frankfurt.[10] In a paper which is arguably the most influential article on free will written in the last quarter century, Frankfurt pointed out that young children and animals are never conflicted in the way that Mary is. That's because they lack the capacity to have second order desires. To have a second order desire is to have a desire about a desire. For instance, someone may want to have a desire she doesn't yet have, she may want to get rid of a desire she has, she may want to be motivated more often by one of her desires, and so on.

It seems plausible to suppose that young children and animals lack free will. It also seems plausible to suppose that Mary lacks free will, but in a different kind of way. She is someone of whom it intuitively makes sense to say that she has free will in some respects and lacks it in others. She lacks free will with respect to her desire to smoke, but enjoys free will with respect to her desire to drink wine, to eat ice cream, to read the newspaper, and so on.

If we put these two thoughts together, we are led to Frankfurt's basic idea, which is this: Simple Compatibilism is correct as a theory of freedom of *action*; freedom of action is, roughly, being free to do what one wants to do. But there is more to freedom than freedom of action. Animals and young children may enjoy freedom of action, but they aren't motivationally complex enough to have freedom of will. Freedom of *will*, is, roughly, being free to have the will one wants to have. More precisely, someone has freedom of will to the extent that she is free to be motivated by the first order desires she wants to be motivated by.

Note what Frankfurt is offering us. He is offering us a Compatibilist way of accounting for our commonsense way of thinking about ourselves. We think that we are (ordinarily) free to act on the desires we want to act on and to refrain from acting on desires we don't want to act on. This commonsense view is often thought to be committed to a metaphysically and empirically dubious view of the self as an uncaused entity which desirelessly chooses which of its desires to act on. But if Frankfurt is right, then this is not so. Frankfurt's claim is that we can use second order desires to give a satisfactory account of the freedom we have (and which the addict lacks). It is no part of his story that second order desires are uncaused or enjoy any special status other than the fact that they are desires about first order desires. His account is, or seems to be, strictly parallel to the Simple Compatibilist account of freedom of action. The prisoner dragged through the streets lacks freedom of action because his body moves *regardless* of whether or not he *wants* his body to move this way. His first order desires concerning the movement of his body are (or would be) inefficacious, impotent. On Frankfurt's view, the addict lacks free will (with respect to his drug-taking desire) because "his desire to take the drug will be effective *regardless* of whether or not he *wants* this desire to constitute his will". (My italics.)[11] His second order desires concerning his drug-taking desire are (or would be) inefficacious, impotent.

This looks like the kind of account we've been looking for. We've been looking for a way of making sense of the idea that the psychologically compelled suffer from a specific kind of unfreedom, an unfreedom which would make it unreasonable and unfair to hold the person responsible for what she does. And we were looking for a way of making sense of the idea that the psychologically compelled are helpless with respect to the desires on which they act in the way that we are helpless with respect to some external forces. Frankfurt seems to offer us both.

I think, however, that appearances are deceptive. To see why, we need to take a closer look at what Frankfurt's account is. Consider the question of what distinguishes Mary from Suzy. Why is it that Mary lacks free will (with respect to her desire to smoke)

whereas Suzy has free will (with respect to her desire to read trashy novels)? Here are two very different answers Frankfurt might give:

He might say that the relevant difference lies in the second order desires which Mary and Suzy have. When Mary smokes the cigarette, she lacks free will because she acts on a desire she wants *not* to be motivated by. When Suzy reads the novel, she has free will because she acts on a desire she wants to be motivated by.

On this view, to have free will (with respect to the desire on which one acts) *is* to have the will (motivating desire) one wants to have and to lack free will *is* to fail to have the will one wants to have. Can't someone be indifferent about whether or not she acts on a certain desire? Yes, but to the extent that this occurs, Frankfurt suggests that she is no longer a person, but what he calls a "wanton"; someone who, like a child or animal, lacks free will "by default".[12] If moral responsibility requires free will, then this has the consequence that those who are indifferent about their motivation lack responsibility for what they do. Let's call this the "Endorsement" view, for it says that someone has free will with respect to a desire if and only if this desire is appropriately endorsed by a second order desire.

Or Frankfurt's claim might be that the difference between Mary and Suzy lies in the fact that Suzy has an *ability* which Mary lacks - the ability to refrain from acting on the first order desire which moves her to act. On this view, the fact that Mary smokes despite her second order desire not to yield to this desire is relevant only insofar as it gives us reason to believe that she *cannot* refrain from smoking. Let's call this the "Ability" account.

The Ability account disagrees with the Endorsement account about some cases. Consider Mary's younger sister Kate, who is an unrepentant and enthusiastic smoker. Kate doesn't want to quit smoking; she approves of her desire to smoke and she wants this desire to be her will. According to the Endorsement account, Kate has free will with respect to her desire to smoke. Kate *thinks* she's got free will; she thinks she can quit any time she wants. Mary thinks Kate is just kidding herself. According to the Ability account, Mary might be right.

Insofar as Frankfurt claims to be offering a Compatibilist way of accounting for our commonsense (Libertarian) beliefs about free will, you might expect his account to be the Ability account. And up to a point, it is. It's clear that he thinks that addicts lack the ability to refrain from acting on their desire for the drug. He describes an addict who:

> ..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 desire 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.

And he says that a willing addict (someone who approves of his desire for the drug and who would continue to take it even if he weren't addicted) also lacks free will because "his will is outside his control"; his desire for the drug will cause him to act "regardless of whether or not he wants this desire to constitute his will".[13]

But Frankfurt says that the willing addict is morally responsible for taking the drug. This is a strange view. On the one hand, he believes that the addict is as helpless with respect to his desire as the prisoner is with respect to the external force which moves his body. On the other hand, he seems to be saying that the addict's second order Endorsement of this desire is sufficient to make him morally responsible for what he cannot avoid doing. And he says that the willing addict's "will is not free" but he nevertheless takes the drug "of his own free will".[14] It appears, then, that Frankfurt holds *both* the Ability and the Endorsement accounts of free will and that he thinks that someone who is free according to the Endorsement account is morally responsible even if she is unfree according to the Ability account.

This seems implausible. Someone pushed off the roof may have wanted to be pushed - she was standing at the edge, trying to work up the nerve to jump - but this first order endorsement of what happens to her doesn't make her responsible for her descent to the ground. (She fell, after all; she didn't jump. She would have ended up on the ground even if her first order desires had been different; even if she had tried as hard as she could to resist the push.) Frankfurt takes seriously the analogy between compulsive first order desires and forces too powerful to resist. He thinks that the willing addict would have ended up taking the drug even if his second order desires had been different, even if he had struggled desperately to resist his desire. On his view it seems that there is no relevant difference between the willing addict and the suicidal murder victim. Neither can prevent what they end up doing; both nevertheless get what they want. Yet Frankfurt insists that the willing addict's second order endorsement makes her responsible for what she is powerless to prevent.[15]

What Frankfurt says about the willing addict is not plausible, but what he says about the unwilling addict may nevertheless seem plausible. The unwilling addict is our paradigm case of someone who is compelled by her desire, who acts unfreely, and without moral responsibility for what she does. She is someone who acts on her desire for the drug even though she wanted *not* to act on this desire. It also seems reasonable to think that she *could not* help acting on her desire and that she is for that reason not responsible for what she does. (If someone falls off the roof despite wanting not to fall,

the reasonable inference is that she could not help falling and that she is therefore not responsible for her descent to the ground.) Given this, we may think that Frankfurt is right insofar as he says that to act on a desire despite one's second order desire not to do so *is* to lack the ability to refrain from acting on the desire.[16] And we may think that anyone who acts against second order desire (and thus without ability to refrain from so acting) is therefore not responsible for what she does.

But I think that this is a mistake. Consider the following cases:

Julia decides that she's been drinking way too much coffee and resolves to quit drinking it altogether. She says to herself: "I want to get rid of my desire to drink coffee, and in the meantime, I want to stop acting on it". She quits for three days. But on the fourth day, she has a deadline. Coffee helps her work better, but she doesn't think that this is a good enough reason to have a cup. She says: "It would be best for me not to have a cup of coffee; if I get that surge of adrenaline, I'll just want more." She reminds herself that she wants to quit. She gets up and makes herself a cup of coffee.

Patrick is in love with Sofia. He knows she doesn't love him, and knows that she knows of his feelings, is embarrassed by them, and that she wishes that he would get over it, or go away, or both. He knows that he should do this, that there is no chance whatsoever that she will ever love him. He wants to stop wanting her, and, failing that, he wants not to be motivated by his desire to spend time in her presence. Yet he continues to seek out her company, embarrassing both himself and her.

Helen is angry at Michael. She wants to hurt him, and she figures that the way to do it is by flirting with John. She succeeds. But Helen is a basically good person; among her standing desires is the desire not to use anyone merely as a means to serve her ends, especially when the end is to hurt someone she loves.

These are all cases in which someone acts on a desire despite his or her second order desire not to do so. But Julia, Patrick, and Helen don't think that they are the helpless victims of their desires; they think that they have free will and that they are responsible for what they do. They think of themselves as weak willed, or (in Helen's case) as willfully self-indulgent. And most of us would agree.

Frankfurt faces a dilemma similar to the one I posed for Watson. He has two options. He could bite the bullet and argue that Julia, Patrick, and Helen are, like the unwilling addict, neither free nor responsible for what they do. Or he could argue that there is some relevant difference between the addict and these three.

Biting the bullet is not an attractive option. It would be a disaster for our conception of ourselves as morally responsible beings if acting contrary to our second order desires were sufficient for not being responsible for what we do. Here's the

problem. It's part of being a full-fledged moral person that one not only has beliefs about what's good and bad, right and wrong, but that one wants, for the most part, to be motivated by what one thinks good and right. And since we are all far from perfect, we are all motivated, to a lesser or great extent, by desires by which we would rather not be motivated.

The fact that we don't like the results of an account is not a good reason for rejecting the account. *If* we had good reason to believe that Frankfurt's account of freedom and unfreedom of will is correct, then we should accept the consequences of this account even when they don't correspond to our pre-theoretic beliefs.

But Frankfurt can't defend his bullet-biting this way. He doesn't have an account of free will; we've seen that he equivocates between two, and possibly three, different accounts. I've argued that his second order desire account is plausible only insofar as it's understood as an analysis of the ability to refrain from acting on one's first order desire. And whether it succeeds as an analysis of this ability is precisely what is in question. The fact that it gets the intuitively right verdict in the unwilling addict case is an argument in its favor, but the fact that it gets intuitively wrong verdicts in other cases counts against it.

We should guard against a possible source of confusion here. You might be tempted to think that there is some sense in which Julia, Patrick, and Helen are unfree. Doesn't Julia suffer from a physical craving for coffee, a craving she can't get rid of simply by willing it to go away? And isn't Patrick helpless with respect to his feelings for Sofia? They will eventually fade, but there is nothing he can do right now which will make them go away. And isn't Helen's anger at Michael also something that she cannot help feeling?

Let's grant that Julia, Patrick, and Helen may lack control over the feelings which accompany the desires on which they act. (They need not, given the stories I told, but they might.) And let's also grant that having control over one's cravings, yearnings, and other feelings might be a kind of freedom worth having. Then Julia, Patrick, and Helen may be less free than, say, a Buddhist monk. But Frankfurt's account is an account of freedom of *will*, and the ability to control one's feelings is relevant only insofar as it impairs the ability to control one's will. There are cases in which we think that people are "swept away" or "overpowered" by their feelings; someone might be paralyzed with fear, or go berserk with rage. But the cases I've described are not, intuitively, of this sort.

Frankfurt's other option is to find some relevant distinction between the unwilling addict and people like Julia, Patrick, and Helen.

Intuitively, there is this difference: What the unwilling addict *wants most*, at the second order level, is that she stop acting on her desire for the drug. But while the other three want not to act on the desire which motivates them, this is not what they want most.

If Julia had *really* wanted not to give in to her desire for coffee, she would not have had a cup. If Patrick's desire to resist his desire to see Sofia had been *stronger*, he would have stayed home. If Helen had wanted to act on her desire to be a good person *more* than she wanted to act on her desire to hurt Michael, she would not have flirted with John.

To evaluate this reply we need to consider how to evaluate claims about the strength of someone's second order desires.

The measure of strength of desire is preference, and the test for preference is actual or counterfactual choice. Is my desire to eat an apple stronger than my desire to eat cheesecake? That depends on what I would choose, given a choice between the two. If I'd choose the cheesecake, then I prefer eating cheesecake to eating an apple and my desire for cheesecake is stronger than my desire for an apple.

If it makes sense to talk of strength of second order desires, then a similar test applies. Is my desire to be motivated by my apple-eating desire stronger than my desire to be motivated by my cheesecake-eating desire? That depends on which *desire* I would choose to be motivated by, given a choice between the two. If I would chose to be motivated by my apple-eating desire, then I have a second order preference for my apple-eating desire over my cheesecake-eating desire and my second order desire to act on my apple-eating desire is stronger than my second order desire to act on my cheesecake-eating desire.

There is a problem lurking here. It might seem that this counterfactual test for strength of second order desires always yields the same result as the counterfactual test for first order desires. We evaluate counterfactuals by considering the closest (most similar) worlds at which the antecedent is true. And it seems that the *closest* worlds where I have a choice between being motivated by my apple-eating desire and between being motivated by my cheesecake-eating desire are worlds very much like our own, where I have the same fondness for cheesecake I in fact have. At these worlds, I choose the cheesecake, revealing that my first order desire for cheesecake is stronger than my first order desire for an apple. But my action also reveals that my *second order* pro-cheesecake desire is stronger than my second order pro-apple desire.

But I don't think we have to say this. We can distinguish between someone's first and second order preferences by making the counterfactual test for second order preferences *different* from the counterfactual test for first order preferences.[17] Consider the unwilling heroin addict. His actions reveal that his first order desire for heroin is stronger than his first order desire not to take heroin, but we want to say that his second order anti-heroin desire is stronger than his second order pro-heroin desire. We can make sense of this by using the following counterfactual as the test of his second order

preference: If he were offered the Heroin Cure (a pill which would get rid of his desire for heroin), would he take it? If the answer is "yes", then his anti-heroin second order desire is stronger than his pro-heroin second order desire.

But now let's apply this counterfactual test to our three cases. If Julia were offered the Caffeine Cure, would she take it? Yes. If Patrick were offered the Sofia Love Cure, would he take it? Yes. (There are people who would decline, preferring the pain of unrequited love, but Patrick is not one of them.) If Helen were offered the Desire-to-Get Revenge-by-Flirting Cure, would she take it? Yes; she would prefer to be the kind of person who expresses her anger openly and directly, rather than in devious ways. Julia, Patrick, and Helen act, not just contrary to the desire they want to act on, but also contrary to the desire they *most want* not to act on.

I conclude that the appeal to the strength of second order desires will not help Frankfurt to distinguish the unwilling addict from people who intuitively have free will and are responsible for what they do.

Frankfurt assumes (thinking of his Unwilling Addict) that if someone is motivated by a desire she wants not to be motivated by, then this is because she has an abnormal first order desire, a desire which she is unable to resist even though that's what she most wants to do. But the cases we've been looking at suggest another possibility. Someone who most wants to get rid of her first order desire (she prefers taking the Desire Cure to not taking it) may nevertheless act on it because, given that she still has the desire (and the feelings and cravings which accompany it) *that's* what she most wants.

But isn't there a difference between the unwilling addict and people who merely act on desires they would prefer not to act on? Frankfurt describes his Unwilling Addict as someone who "struggles desperately" to resist his desire. Doesn't this suggest that this addict's desire is not *just* a desire he prefers not to act on, but a desire which is in some sense external to or even alien to his "true" or "real" self? And if this is so, then it's a mistake to say that the addict most wants, all things considered, to act on this desire. Rather, what the addict most wants is to act according according to his second order preference. He tried to do this, but failed. The only reasonable inference is that he failed because he could not resist the alien desire.

We *might* think this, but if this is our account of what distinguishes the unwilling addict from Julia and the others, then we are appealing to the very intuitions we were trying to explain. We were looking for an account which would make sense of the idea that the addict's desire compels her to act in the way an external force might compel someone's body to move. It's no help to be told that this is so in virtue of the fact that the desire for the drug is external to the addict's "real" self and that it causes action even

though this real self tries its best to resist. We need to be given *reasons* for believing that the addict is identical to this inner "real" self rather than to the ordinary person, and we need to be told why some desires, but not others, are alien to the person's real self. Insofar as Frankfurt has an account, it is the Endorsement account.[18] External or alien desires are desires we most want not to be motivated by. But this doesn't distinguish the addict's desire from the desires of Julia, Patrick, Helen and anyone who acts on a desire she would prefer not to act on.

I conclude that Frankfurt's second order desire account fails as an analysis of the unfreedom of psychological compulsion.

Simple Compatibilism and Moral Responsibility

To someone unfamiliar with the free will literature of the last twenty-odd years, it may seem that I've been laboring the obvious. But anyone with even a passing acquaintance with the literature will think differently. While there has been a great deal of criticism of Frankfurt's account of free will, almost all of it takes for granted that addicts and other "obsessive-compulsives" lack free will. (The criticism most often made is that Frankfurt's account fails because it ignores *other* ways in which people may lack free will.) And nearly everyone agrees with Frankfurt's main point - that Simple Compatibilism fails as a theory of free will because its conditions for freedom are too easily satisfied. The general consensus in the literature (at least among Compatibilists) is that we need a more sophisticated theory of free will, one which places more stringent constraints on what counts as free will.

I think that this search for a more sophisticated Compatibilist theory is a mistake. I think that Simple Compatibilism is basically right, not just as an account of freedom of action, but also as an account of *free will*. It's not, however, a theory of moral responsibility, and much confusion will be avoided once we realize this.

Simple Compatibilism says that free will is the capacity to make choices on the basis of one's beliefs and desires. (Roughly, it's the capacity to choose to do what one thinks will bring about what one wants.) According to Simple Compatibilism, to act intentionally is to exercise this capacity. That is, to act intentionally is not just to act freely, but also to act with free will. Finally, Simple Compatibilism says that someone who acts intentionally is able, ceteris paribus, to choose and to do otherwise. Why? Because the following counterfactuals are ordinarily true of someone who acts intentionally: If she had wanted something else more, she would have chosen differently; if she had chosen differently, she would not have done what she did.

This is a simple account, though not quite as simple as you might think. Consider, for instance, someone who claims that she is unable to do something because of a phobia or pathological fear. Depending on the details of the case, Simple Compatibilism may agree. For while some failures are actions, not all are. And if the claustraphobiac's failure to enter the room is not an action, then it may be true that what she most wanted was to enter the room. If so, then Simple Compatibilism can agree with our intuitive verdict about this case -- that she could not do what she most wanted to do.

There is another way in which Simple Compatibilism is not as simple as some people think. Simple Compatibilism says that facts about free will are a certain kind of psychological fact about a person at a particular time. (Did she have the capacity to make a choice? Or did extreme fear, panic, pain, etc. render her either incapable of choosing at all, or incapable of choosing to do an act of a certain kind?) But Simple Compatibilism is true only if Compatibilism is true, and Compatibilism is a metaphysical and modal thesis -- the thesis that this capacity to choose is consistent with the existence of earlier deterministic causes of our choices. There are sophisticated incompatibilist arguments for the conclusion that if determinism is true, then we are *never* capable of making any choices (or never capable of making any choices other than the ones we in fact make). These arguments can, I think, be answered, but not without engaging in substantive metaphysical and modal discussion of counterfactuals, causation, and laws of nature.[19]

I think that Compatibilism is true, and I also think that Simple Compatibilism is true. The question of moral responsibility, however, is a larger and more complex question. To see why, let's begin by considering the question of why young children are not morally responsible.

Young children are capable of making choices on the basis of their beliefs and desires; I think that they have free will. But there are a number of *cognitive* differences between them and us, differences which are relevant to responsibility. Young children live in the present; they lack any significant capacity to remember the past or to anticipate the future. They lack awareness of themselves as creatures enduring through time and thus are unable to understand what's going on when we blame them for something they did in the past. While they can make some choices, their inability to think beyond the immediate future severely restricts the kinds of choices they can make. Their limited understanding of the world and of themselves further restricts their capacity for intelligent, prudent, or moral choice-making. And since they lack any significant capacity to have general beliefs or to think abstractly, they are incapable of understanding what counts as a moral justification of an action. And so on.

Which cognitive capacities are required for moral responsibility is a controversial and difficult normative question. What we say here depends on our views about rationality and morality as well as our views about metaethics. We should not confuse these normative and metaethical questions with the metaphysical and modal question of free will. Free will is the capacity to choose, not the capacity to choose wisely, prudently, or morally. Free will is a necessary but not sufficient condition of moral responsibility.

Some philosophers think that children lack free will as well as responsibility. I suspect that this is because they think that children lack *all* the freedom worth wanting. But it's a mistake to conflate having all the freedom worth wanting with having free will. Recall the victim of coercion. She lacks all the freedom worth wanting - she chooses among sadly diminished options. But she has free will.

I've been arguing that moral responsibility requires a bundle of cognitive capacities in addition to free will. But there is also another way in which questions of free will are distinct from questions of responsibility. Questions about free will are questions about an *agent* - a person at a particular time. But questions about moral responsibility are questions about a *person* - someone who endures through time. Someone is responsible for a past action only if she is the same person as the person who did the action. So a theory of personal identity through time must be part of a theory of moral responsibility. We need an account which connects this up with the question of why we hold persons, but not children and other sorts of creatures, responsible for their past acts. Memory and continuity of character are not enough; a significant degree of causal control over one's future self seems crucial. We are, to some significant degree, who we are today because of what we did in the past. Here questions about what's required in order to be a morally responsible person and questions about what's required in order to have all the freedom worth wanting are not always easy to distinguish. The literature on autonomy is, I think, engaged in addressing both these questions, as is some of the literature on character and virtue.

My concern here is not with a general theory of moral responsibility, but with the moral responsibility of addicts and others who claim to be victims of psychological compulsion. On my view, they have free will and act freely. Whether they are responsible is a separate and more difficult question, and one which depends on the details of the particular case, as well as on our theory of responsibility. But on my view the fact that someone is a drug addict (kleptomaniac, alcoholic, sex addict, etc.) does not, by itself, excuse from responsibility. The addict who hates himself for stealing from his children may lack all the freedom worth wanting, and we may pity him for that. But I think that he is responsible for what he does.

On my view, the fact that someone acts on a desire she neither values nor wants to act on is not relevant to the question of her responsibility. But what about those who ask for help? I'm thinking of the addict who checks herself into a drug clinic, the alcoholic who is a member of AA, the person who wrote the legendary graffiti on the New York subway wall - "Stop me before I kill again", Ulysses who had himself bound to the mast so that he would not follow the seductive song of the sirens.

On my view, these facts are relevant, not to the question of free will, but to the question of responsibility. The question of moral responsibility, unlike the question of free will, is not settled by facts at the time of action. We hold the drunken driver responsible for causing the death of the child, not because of what she chose or could have chosen at the time of the accident, but because of her earlier decision to drink and drive. But there are other ways in which earlier and later facts are relevant to the question of responsibility. Someone may mitigate her responsibility for an action by what she does afterwards - by trying to repair the damage, by apologizing, and so on. And I think someone may mitigate her responsibility for an action by what she does *before* the time of action.

Why are we less inclined to blame Mary (the unwilling smoker) than her sister Kate (the unrepentant and enthusiastic smoker)? Not, I think, because of Mary's mental state at the time she smokes the cigarette, but because of what she does earlier. She *was* trying to quit, she had managed not to smoke for three days, she had asked her friends not to give her cigarettes even if she begged for them, and so on. I think that this is why we blame her less, and *should* blame her less.

I think that this is the key to understanding why we excuse (some) unwilling compulsives. Not because their desperate struggles are evidence of their lack of free will, but for the more straightforward reason that we should not blame those who are trying to help themselves.

The question of how much weight we should give to the earlier (and later) mitigating actions of the person who pleads psychological compulsion is a difficult normative question, one which may involve considerations of charity as well as justice. Remember Patrick. Suppose that he decides that he is the helpless victim of his love for Sofia and begins following her around, turning up on her doorstep, phoning her every night, and so on. She would blame him for pestering her, and rightly so. He might decide he needs professional help, and he may be right. He starts seeing a therapist, but continues to hound Sofia. Should Sofia change her verdict of responsibility and blame to one of compassion and pity? I think that she would be charitable in doing so, but that she would not act unjustly if she continued to regard him as responsible and blameworthy.

Sartre said that we are condemned to be free; that we are free with respect to everything we do and that we cannot avoid being so. Many people have rejected this view because they think it implies that we are responsible for everything we do (which is what Sartre thought). But if I'm right, then this is not so. Insofar as we act for reasons, we act freely and with free will. We cannot escape our freedom by identifying ourselves with our higher or better or rational self or by disowning the desire which causes us to act. But this doesn't mean that we are condemned to be responsible, for two different sorts of reasons. We *could* become convinced that our practice of holding people responsible does more harm than good or that there is some fundamental unfairness in it, and that we should blame less, or differently, or not at all. But suppose we think that something like our current practice is justified. We may still have good reason not to blame (or not blame as much) some of those who plead that they are the helpless victims of their desires. For they have asked for help and it would be uncharitable to respond with blame. There are complex and difficult questions concerning responsibility, questions of justice as well as questions of charity. But these are normative questions, not metaphysical questions about free will.[20]

NOTES

[1] Oliver Sacks describes a Parkinsonian patient who managed to come up with a purpose for the compulsive movement of her hand to her face - she used it to adjust the position of her glasses. But this rationalization did not convert her tic into an action; she would have moved her hand in the same way even if she had not believed that doing so would adjust her glasses. (*Awakenings*, Harper Perennial, 1990, p.136.)

[2] See my "Freedom, Causation, and Counterfactuals", *Philosophical Studies* 64 (1991), 161-184, and "Freedom, Necessity, and Laws of Nature as Relations between Universals", *Australasian Journal of Philosophy* 68 (1990), 371-381. For the best recent defense of incompatibilism, see Peter van Inwagen, *An Essay on Free Will*, Clarendon Press, Oxford, 1983.

[3] The argument given in this section, as well as most of the argument of the next section, is substantially reproduced from my "Are Drug Addicts Unfree?", in *Drugs, Morality, and the Law*, ed. Curtis Brown and Steven Luper-Foy, Garland, New York, forthcoming (February, 1994). I thank the editors for permission to use this material. In that paper I offered an account of the unfreedom of addiction, an account I now think mistaken.

[4] Patricia Greenspan suggests something like this as an account of the unfreedom of addiction and certain extreme cases of aversive conditioning in "Behavior Control and Freedom of Action", *The Philosophical Review* 87 (1978), 225-240. But she doesn't draw any conclusions about moral responsibility, and none of my remarks are intended as criticisms of her account.

[5] And of course, not all coercive threats succeed. Some people refuse to make what we call their only "real choice"; depending on the circumstances, we call them heroes or fools.

[6] J.L. Austin, "A Plea for Excuses", *Philosophical Papers*, Clarendon Press, Oxford, 1961.

[7] Watson, "Free Agency", *Journal of Philosophy* 72 (1975), 205-220. Also in *Free Will*, ed. Gary Watson, Oxford University Press, 1982. All references are to the Watson book.

[8] He says: "...it is one thing to value (think good) a state of affairs and another to desire that it obtain. However, to think a thing good is at the same time to desire it (or its promotion). Reason is thus an original spring to action." (p.106)

[9] Although Watson doesn't consider the question of the unfreedom of people like Bob, X, and Suzy, he realizes that his account is committed to the claim that there is no distinction, so far as freedom at the time of action is concerned, between the weakwilled and the compelled. In "Skepticism About Weakness of Will", *Philosophical Review* 86 (1977), 316-338, he argues that we should accept the conclusion that the weakwilled act unfreely. I think that his argument is convincing only if you've already accepted his Platonic view of value and the will.

10 Frankfurt, "Freedom of the Will and the Concept of a Person", *Journal of Philosophy* 68 (1971), 5-20. Also in *Free Will*, ed. Gary Watson, ibid. Page references are to the Watson book.

11 p.94.

12 p.91.

13 p.87 and pp. 94-95

14 p.94

15 I think that Frankfurt's views about the moral responsibility of the willing addict are due chiefly to his tendency to equivocate between the Endorsement and Ability accounts of free will. But some of his remarks about the willing addict suggest that Frankfurt is relying on yet another account of free will, one considerably less plausible than the other two; he says that the willing addict acts of his own free will because his second order desire *causes* his desire for the drug to be his will. Whereas the Endorsement and Ability accounts are motivated by the idea that we have free will when we have (or are able to have) the will we want to have, this account is motivated by the more dubious idea that we have free will if and only if our actions are the outcome of a certain kind of causal process - one which includes a second order decision or choice or act of will.

16 That is, even though the willing addict case shows that acting contrary to the relevant second order desire is not a necessary condition for lacking the ability to refrain, it may nevertheless be true that it is a sufficient condition for lacking the ability to refrain.

17 I owe this suggestion to Terrance Tomkow.

18 Perhaps supplemented by the somewhat obscure idea that the addict "identifies himself" with his desire not to take the drug and "withdraws himself" from his desire for the drug. (p.88) Insofar as this suggests that one may disown one's desires (and responsibility for their causal upshots) by an act of will, this is not an attractive suggestion.

Watson criticizes Frankfurt on this very point and draws the moral that the significance of second order desires lies in the fact that they are desires to do what the person thinks best. But this doesn't help with the problem cases we've been discussing, for they are all cases in which someone acts against both second order desire and value judgment, yet intuitively acts freely and without alienation.

19 See note 2.

20 I'm grateful to Steven Burns, Curtis Brown, Richmond Campbell, Sue Campbell, Steven Luper-Foy, Phillip Scribner, Milton Wachsberg, David Zimmerman, and especially Terrance Tomkow for helpful discussion and for comments on earlier drafts and ancestors of this paper.