Addiction and Causation

MICHAEL CORRADO*

TABLE OF CONTENTS

* Arch Allen Distinguished Professor of Law, University of North Carolina at Chapel Hill. J.D. 1984, University of Chicago; Ph.D. 1970, Brown University; B.S. 1966, B.A. 1965, Pennsylvania State University. An earlier version of this paper was delivered at a University of Pennsylvania conference in honor of the publication of Michael Moore's *Placing Blame*. I thank the participants for their comments. I would especially like to thank Stephen Morse for patient and helpful comments.

I. INTRODUCTION

An interesting thing about addiction is this: When you look into what philosophers have said about it, you find the surprising conviction that addiction is about literally irresistible impulses, and about a complete absence of responsibility.[1] Many philosophers seem to ignore the other possibility, that addiction makes choice *hard* but not *impossible*. The reason for that, I suppose, is that irresistible impulse raises a certain sort of problem for compatibilism, and it is that problem that these writers have been concerned to answer. But *difficult* choices raise another problem of the very same sort.

The question is this: Is it possible for a compatibilist to capture the notion of a choice that is resistible but very, very hard to resist? And, along the same lines, is it possible for the compatibilist to capture the notion of degrees of responsibility, of greater or lesser moral responsibility? Of course, duress may lessen responsibility, and in general the aversiveness of the alternatives facing an agent may lessen her responsibility for an action: The more aversive the alternatives, the less responsible the agent—or at least the less inclined we are to punish the agent. That way of ranking responsibility is clearly intelligible in compatibilist terms. So, for example, the actor is less responsible for something he does under conditions of rational duress, where he is faced with an alternative that involves harm to his children.

But there is another basis for saying that someone is less responsible than he might be otherwise, a basis that has nothing to do with a rational choice involving aversive alternatives. It seems to me that the ordinary notion of addiction involves lessened responsibility in precisely this sense. We conceive of the heavily addicted person as one who, even if he knows that abstaining from the addictive substance is better in every sense, still finds it very difficult to abstain. This difficulty has less to do with better or worse alternatives than with a lessened ability to make the choice that the addict may in fact want to make. I think this is part of our popular conception of addiction, as I say, and it is part of the reason why we are inclined to think that the heavily addicted person is not fully responsible for what he does. My question is whether the compatibilist can account for this notion of difficulty, and for this notion of lessened responsibility.

What I set out to do in this paper was to see if an account of this sort of difficult choice could be found. The account I was looking for would have to have these two properties: (1) it would have to be consistent with

1. *See, e.g.*, JOHN MARTIN FISCHER & MARK RAVIZZA, RESPONSIBILITY AND CONTROL: A THEORY OF MORAL RESPONSIBILITY 48 (1998).

the possibility of universal causation; and (2) it would have to explain the popular sense that the addict is deserving of an excuse or mitigation. I am not interested here in the question whether the addict really deserves an excuse or mitigation, but only in the question whether any compatibilist account can capture that part of the popular sense of what addiction is. Since Michael Moore is the legal philosopher who has most persistently explored the consequences of compatibilism for questions of moral and legal responsibility, it is natural to begin with his approach to this subject.

A. The Causal Theory

Moore has spent a fair amount of time trying to demonstrate that what he and others call the causal theory of excuse should be abandoned.[2] That theory may be formulated as follows:

> **CE: If there is an unbroken causal history for a certain action, extending back to some event over which the agent had no control, then the agent is not (morally) responsible for that action.**[3]

Conversely, according to the causal theory, if an agent is responsible for some action, then the causal history of the action has been broken by some event over which the agent did have control. Moore would like to make this a biconditional, I think.[4] If I read him correctly, he attributes

---

2. *See* Michael S. Moore, *Causation and the Excuses*, 73 CAL. L. REV. 1091, 1091 (1985); Michael S. Moore, *Choice, Character, and Excuse*, 7 SOC. PHIL. & POL'Y 29, 29 (1990). Both are reprinted in chapters 12 and 13 of Moore's *Placing Blame*. *See* MICHAEL MOORE, PLACING BLAME: A GENERAL THEORY OF THE CRIMINAL LAW 481-592 (1997). *See also* Michael Corrado, *Automatism and the Theory of Action*, 39 EMORY L.J. 1191, 1193-1209 (1990).

3. "The principle is that when an agent is caused to act by a factor outside his control, he is excused; only those acts not caused by some factor external to his will are unexcused." MOORE, *supra* note 2, at 487. Notice that in this particular formulation, the "theory" merely claims that causation is a sufficient condition for excuse. Note the similarity of this formulation to what Susan Hurley calls "the regression principle," namely, "to be responsible for something you must be responsible for its causes." Susan L. Hurley, *Responsibility, Reason, and Irrelevant Alternatives*, 28 PHIL. & PUBL. AFF. 205, 206 (1999).

4. For example, he puts the causal theorist in the implausible position of arguing that if mistake-of-fact is an excuse, it must be because the resulting actions are caused by things beyond that actor's control. *See* MOORE, *supra* note 2, at 494-95. That suggests that he sees the causal theory as entailing not only that if something is caused it is

to the causal theorist not only the belief that an agent is not responsible for an action that is caused, but also the belief that an agent *is* responsible for any action that is uncaused. But in one direction the biconditional is utterly implausible. I do not know anyone who believes that in every case in which an action is uncaused the agent is responsible for it. People who are likely to hold something like the causal theory would hold it precisely because they believe that while some actions may be caused, others are uncaused; and among those that are uncaused, common sense tells us, are some that ought to be excused. The pure accident, for example, will result from a choice that is freely made, though without knowing the consequences; and in such a case (the causal theorist might admit) the injurer ought to be excused even though (the causal theorist would insist) his contribution is not caused. Moore has a few quotes that suggest that some people believe that the inference runs in both directions, but we have to lean pretty hard on those quotes to get that interpretation.[5] So I will ignore the biconditional form and consider it only in the form in which I have stated it.

The causal theory certainly has some natural appeal. If we learn that certain behavior is caused by a lesion in the brain, then whether or not the causation was mediated by the agent's choice (the lesion operating *through* the choice), we are inclined to think that the agent is not responsible. Of course this is only true if the agent's *free* choice made no contribution whatever. We are not so quick to excuse an agent who has been made more aggressive, let's say, by a brain lesion, but who still had some control over his aggressive behavior. It is only if the entire action, from beginning to end, choice and all, can be attributed to the physiological defect in the particular circumstances that we are ready to excuse without qualification. If asked why we would excuse, we might give something like the causal theory in response. In addition to its intuitive appeal, the causal theory seems to follow from a principle that many find hard to reject, namely that if an action is caused by forces over which I have no control, then I have no control over the action. Conversely, if I have control over an action, then I have control over at least some of the antecedents of the action. Those who reject the causal theory would see in this argument a *petitio principii,* perhaps, the control principle being nothing other than the causal theory in other clothes. Still, one who rejects the causal theory ought, on his side, to give an argument for rejecting it.

---

excused, but also that if it is excused, it is caused. As I argue in the text, it would be hard to find someone who accepted the second of these conditionals.

5. *See, e.g.*, MOORE, *supra* note 2, at 493.

Moore does give such arguments, and one of the more important, I think, is this: If the causal theory is true, then everything is excused.[6] The missing premise, of course, is this:

> **UC: Everything, including each human action, is in fact caused.**

We may call this the thesis of universal causation.[7]

Now it is true that the causal theory, together with universal causation, entails that no one is morally responsible for anything. But there are three positions we can adopt in the face of that fact. The first is to insist upon the causal theory, and to conclude that the principle of universal causation must be false, and that some human choices are not caused (nor merely random either). This position, which has been called (action-)libertarianism,[8] seems to me to capture pretty well one side of

---

6. *See* MOORE, *supra* note 2, at 504-06.

7. There is another, weaker premise that some consider more plausible than the straightforward thesis of universal causation, and it is this:

UC/R: Every event either has a cause or is a merely random occurrence.

That premise, designed to allow for the possibility that at some level the physical world contains an element of chance, would complicate things a bit, but does not require any change in the statement of the causal theory. For physical movements to constitute an action, they must result from the agent's volition. If the claim that every event either has a cause or is merely random is true, then for any given volition that volition will either be caused by something beyond the agent's control (the backward causal chain will either stop at some random event, or will continue back to the beginning of time), or the volition itself will be a merely random event. Since we can hardly hold someone responsible for an action that originates in a random event, the result is the same: Under the causal theory, if the claim that every event either has a cause or is a merely random event is true, then no one is responsible for what he does. I would like, therefore, to overlook this complication as irrelevant to the discussion, and stick with the simple statement of universal causation. If you happen to believe that causation and randomness logically exhaust all the possibilities—*tertium non datur*—then UC/R will seem to you like a tautology or a logical truth; and the argument of this paragraph may convince you that the causal theory all by itself entails that no one is responsible for anything. The answer is that there may be no other alternative, but that to presume that in this context is to beg one of the questions at issue, whether human freedom is inconsistent with determinism. A third way suggested by some is called "agent causation." Roderick M. Chisholm, *Freedom and Action*, *in* FREEDOM AND DETERMINISM 11, 11-13 (Keith Lehrer ed., 1966); Richard Taylor, *Determinism and the Theory of Agency*, *in* DETERMINISM AND FREEDOM IN THE AGE OF MODERN SCIENCE 224, 227-28 (Sidney Hook ed., 1958). For a discussion of agent causation, see PETER VAN INWAGEN, AN ESSAY ON FREE WILL 106-52 (1983).

8. *See* Keith Lehrer, *Introduction*, *in* FREEDOM AND DETERMINISM 3, 6 (Keith

the ordinary view of things, the view that even philosophers hold when they are not at work. For various reasons, however, not least because of the great advances in neural science, it has come to seem increasingly implausible—which is not to say that it has been shown to be false.

The second possibility is to admit both universal causation and the causal theory, and accordingly to accept the conclusion that no one is morally responsible for anything. This position leads to one or the other of two possible reactions from the law. One is that if no one is morally responsible for anything, then no one is *legally* responsible for anything, either, so that there is no basis for punishment in the law, punishment being necessarily based upon retribution. This resolution is extraordinarily distasteful, as the arguments of Herbert Morris make clear,[9] although the fact that something is distasteful does not all by itself tell us that it is false. The other reaction is to declare that legal responsibility is not based upon moral responsibility, salvaging punishment by giving up any non-consequentialist basis for it. This approach is one way of understanding what some utilitarians are up to. Bentham is an example.[10]

And the third way of handling the relationship between the causal theory, universal causation, and moral responsibility is Moore's: to declare the causal theory false, and look for a substitute. So, for example, we might distinguish among *types* of causes of actions. We might say it is not causation as such that excuses, but causation by certain sorts of events. Causation of an action by education and upbringing does not excuse; causation by unusual brain lesions does. That, of course, requires a theory to distinguish the excusing causes from the nonexcusing causes, and it is just such a theory that Moore tries to provide.

One of the problems for any such theory is to explain why we are inclined to treat certain sorts of conditions—hypnotism, sleepwalking, addiction—as excusing conditions.[11] In post-hypnotic suggestion, as we conceive of it, the hypnotized subject acts upon commands given him

---

Lehrer ed., 1966).

9. *See* Herbert Morris, *A Paternalistic Theory of Punishment*, 18 AM. PHIL. Q. 263 (1981).

10. *See, e.g.*, JEREMY BENTHAM, AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION 156-73 (J.H. Burns & H.L.A. Hart eds., 1970). This willingness to disconnect legal and moral responsibility also helps explain the willingness of contemporary consequentialists to punish *more* severely in cases of, for example, provocation. If moral responsibility, and therefore desert, do not act as a limit on punishment, what becomes relevant is that the provoked actor is harder to deter than one who is not provoked, and hence the punishment must be heavier. *See, e.g.*, Richard A. Posner, *An Economic Theory of the Criminal Law*, 85 COLUM. L. REV. 1193, 1223 (1985).

11. *See* Corrado, *supra* note 2, at 1213-21.

earlier by the hypnotist. Somehow or other the hypnotist succeeds in planting certain of his own preferences or desires in the subject, who then acts upon them as if they were his own. But why should that excuse? The fact that we are not in control of the preferences the hypnotist places in us is not enough to excuse. Most of the time we are not responsible for our preferences; they come to us through education or upbringing. What we are responsible for is acting upon our preferences. So if our inclination is to think that the hypnotic subject should not be responsible for his actions, it can't be because he is not responsible for the preferences he is acting on. There must be some other explanation.

One suggestion might be this, that in the case of hypnotism the subject has no choice. He *makes* a choice, but he has no alternative but to make the particular choice he makes. His choice is caused by the preferences the hypnotist has planted in him. But if universal causation is true then *all* our behavior, when we act upon our preferences, is caused by those preferences together with the surrounding circumstances. On this view of things, the only difference between post-hypnotic action and ordinary action is the causal origin of the preferences upon which we act, and it is difficult to see why that should make a difference in responsibility. I will not labor the point. It is the burden of a theory like Moore's to distinguish between education and hypnotism and to locate things like brainwashing, which lie on the continuum between education and hypnotism, on one side or the other of the divide—a project to which Moore himself has devoted considerable effort. Of course if, in the case of hypnosis, the connection between preference and action were causal and in the case of ordinary actions the connection between preference and action were not causal—as the popular conception would seem to have it—there would be no problem. But that would require the thesis of universal causation to be false. And of course there would be no problem (there would not be *this* problem, anyway) if, rather than abandoning universal causation, we were to concede that no one is responsible for any action, hypnotic suggestion or not. That, of course, would require some reshuffling of our views about punishment.

Sleepwalking raises different sorts of problems.[12] The sleepwalker apparently acts on her own volition, aware of the world around her but in a drastically diminished way; and she may be acting out a part in an

12. *See* NORVAL MORRIS, THE BROTHEL BOY AND OTHER PARABLES OF THE LAW 128-57 (1992).

unfolding dream. Automatism, of the sort caused by some kinds of epilepsy for example, appears to present a similar case. Sleepwalking and automatism are different from hypnotism in that the preferences and beliefs the agent acts upon arise within her, and are not caused by some other agent. Of course, the origin of the desires that lead to the sleepwalker's actions are deviant, and the sleepwalker is not responsible for them. But we are not responsible in general for our desires or preferences; we are responsible for acting upon them. Why should the sleepwalker not be responsible for acting upon her own preferences? It sometimes happens that we hear of a sleepwalker who injured another in mistaken self-defense or defense of another: the sleepwalker believed that she (or a family member) was being attacked and was in mortal danger, and in that deluded state attacked someone. That sort of mistake warrants a familiar kind of excuse, the excuse of mistaken self-defense, and goes some way toward explaining our inclination to excuse the sleepwalker. The fact is, however, that there is much more to it than that. For imagine someone who, acting out a dream in which she is robbing a bank, kills someone she imagines to be a policeman preventing her escape. I assume that we would all agree that she should be excused, though I may be mistaken about that. But if you do agree, it cannot be because she would be excused if the facts were as she imagined them to be. If the facts were as she imagined them to be, she would be guilty of second-degree murder (or possibly first-degree felony-murder). What would explain the inclination to excuse her, then? It might be that the connection between desire and action was not in her control, and that in the context her desire caused her action. But if universal causation is right, preferences or desires always cause our actions.

The general problem that Moore's answer faces is the problem of making a case for compatibilism, that is, making a case for the general position that responsibility is compatible with universal causation. The history of the compatibilist struggle is well known. Whereas compatibilists for many years believed that a conditional sort of freedom was all that responsibility required, that analysis proved in the end inadequate for defining the ability to do otherwise.[13] Then Frankfurt showed that the ability to do otherwise is not a necessary condition of responsibility, and compatibilists started off down another road: showing that responsibility depends not on what might be true in other worlds, but only on what is true of the agent in this world.[14] The results are not

13. Surprisingly, Moore seems to accept the conditional analysis. *See* MOORE, *supra* note 2, at 553.

14. *See* Harry G. Frankfurt, *Alternate Possibilities and Moral Responsibility, in*

yet in, but the prospects do not look good: the long and short of it is that no one has been able to make an entirely plausible analysis of responsibility that is consistent with a rejection of the causal theory.[15]

B. Addiction and the Difficulty of Doing Otherwise

But, you may ask, what has all this to do with addiction? For addiction does not seem to raise the same problems as hypnotism and sleepwalking. In those cases it is plausible to argue that the actions are caused, and thus Moore's position requires him to explain why such actions should be excused since, on his theory, causation by outside forces does not, by itself, excuse. No one would suggest that addictive behavior is caused—but I must be careful not to give away the store here. Since Moore seems ready to concede that all action may be caused, addictive behavior may be caused as well. Here is the point: whereas some (not Moore) would argue that the reason for excusing sleepwalking or post-hypnotic behavior is that it is caused, no one would argue that the reason for excusing or mitigating addictive behavior is that it is caused. Whatever it is that makes us treat sleepwalkers and hypnotic subjects as automatistic (whether the fact that their behavior is caused, according to action-libertarians, or the way in which it is caused, according to Moore and others), no one believes that addicts are automatons.

So the causal theory would seem to be irrelevant to the question of the responsibility of addicts. The addict does not have a diminished capacity to perceive the world around him, as the sleepwalker does; and he is not subject to someone else's suggestion. His preferences have been changed, but he is able to perceive the world and to carry out ordinary calculations. He is driven to consume a certain drug but he seems to have a choice; he can forgo it for a time, if necessary, or even

---

HARRY G. FRANKFURT, THE IMPORTANCE OF WHAT WE CARE ABOUT 1, *passim* (1988). Although Frankfurt showed that alternative possibilities are not a necessary condition for responsibility, the examples that undermined the conditional analysis still work. All that is necessary is the possibility that an agent be in a position that deprives him of the ability to choose otherwise *and* deprives him of responsibility: for example, that he be in an automatistic state. It might still be true that there are alternatives which are such that if he had chosen them, he would have done otherwise. Thus the conditional is true, but in the circumstances the agent is not responsible.

15. There may be a developing consensus to this effect among philosophers. *See* Peter van Inwagen, *Logic and the Free Will Problem*, 16 SOC. THEORY & PROC. 277, 289 n.3 (1990).

forever. His preferences, as I say, have changed, in particular his preference for the substance he is addicted to. But preferences, even very strong preferences, don't excuse.[16] A person might, because of some hormonal change, say, experience a greatly increased desire for money. But as long as we believe he is in control of the connection between his desire and his action, we will hold him responsible for stealing money. If we are inclined to excuse the addict or mitigate his responsibility, then, it is not because of his increased preference for a certain substance, but because we believe that the connection between that preference and his actions is to some extent out of his control, or is very difficult for him to control.

And, in fact, Moore appears to accept some version of the following principle concerning difficult choices:

> **DIFF: If some event beyond an agent's control makes an action unreasonably difficult to avoid, then the agent is not responsible for that action.[17]**

Like CE above, this is a principle with a good deal of intuitive appeal. There are, of course, all sorts of qualifications we might want to acknowledge if our interest were in perfecting the principle. The unreasonableness of the difficulty is a normative notion that needs to be filled out, and even assuming an answer to that question there is the further question whether and at what level difficulty will excuse rather than merely mitigating responsibility. But these are questions that do not have to be answered for the purposes of this paper.

The question that does concern me here is this: Can Moore accept DIFF while rejecting CE? At least at first glance accepting DIFF would seem to entail accepting CE (and rejecting CE to entail rejecting DIFF). For it would seem that any event that causes an action would certainly make it unreasonably difficult to avoid that action; indeed, that seems to be something of an understatement. If a brain lesion causes Brian's arm to twitch, then it would seem that the twitch is (in the circumstances) out of Brian's control and unavoidable. And by the same reasoning if a brain lesion causes a volition in Brian resulting in an action, the volition being a sort of twitch, then the volition and the action would seem to be out of Brian's control and unavoidable. That is to say, it would seem to

16. *See* Stephen J. Morse, *Hooked on Hype: Addiction and Responsibility*, 19 LAW & PHIL. 3, 3-4 (2000).

17. *See* MICHAEL S. MOORE, LAW AND PSYCHIATRY: RETHINKING THE RELATIONSHIP 87 (1984) ("Each [of a list of enumerated circumstances] at least mitigates the actor's responsibility because of the difficulty of refraining from doing what he ought not to do."). In some cases where the difficulty is great enough, of course, it eliminates responsibility entirely.

be *extremely* difficult to avoid, in the same sense in which zero is an extremely small number.

Indeed, we can paint a sort of picture of the intuitive connection between the two notions. To say that an action would have been easy to avoid appears to mean something like this: There were alternatives to the action which (1) the actor was free to choose, and (2) were not especially onerous. To say that an action would have been difficult to avoid, then, would be to say that there were no such alternatives. If an action is caused, there are no alternatives at all that the agent was free to choose. And so if an action is caused, it would have been (unreasonably) difficult to avoid, and the actor would not be responsible for it. And if causation of an action implies the unreasonable difficulty of avoiding the accident, then DIFF implies CE. To put it another way, if one reason for rejecting CE is that it leads to the conclusion that no one is responsible, then that should be a reason to reject DIFF.

But at this point the Frankfurt examples become relevant. Frankfurt has shown that alternative possibilities are not a condition of responsibility, so that whatever "easy to avoid" may mean, it does not imply the existence of alternative possibilities.[18] To recap a familiar argument briefly, the examples involve situations like the following: Black wants Jones to perform a certain act. Black has the technology to electronically manipulate Jones's brain, to cause him to choose to perform the action. But it would be best for Black's purposes if Jones did it on his own, without intervention. And, as it happens, Black has some way of knowing whether or not Jones will choose to perform the desired action. Accordingly Black rigs things this way: If Jones will choose to perform the action on his own, then there will be no interference; but if Jones will not choose to do it on his own, the machinery will step in and cause him to choose it. Either way Jones will perform the action, and, more importantly, he will choose to perform it. But things work out for the best, and Jones chooses to perform the action on his own, and no intervention is necessary. Most who have thought about the example agree that Jones should be held responsible, because what he did he did on his own. There was no intervention in fact. And yet Jones could not have done otherwise; he had no *alternative possibilities* open to him. Indeed, he could not have chosen to do otherwise. Hence the existence of alternative possibilities, which had

18. *See* Frankfurt, *supra* note 14.

proved a stumbling block to compatibilism, is not a necessary condition for responsibility. The fact that someone could not in any way avoid making a certain choice does not always mean that he is not responsible for making it.[19]

If you accept this argument, as you should, you should be willing to accept this further consequence: We must be able to say of someone (counterintuitively) that certain choices would have been easy for him to avoid and others difficult for him to avoid, even in conditions in which no other alternatives were open to him: In other words, we must be able to say that his action would have been easy or difficult to avoid in conditions in which his action would have been *impossible* to avoid. Here is what I mean: We might imagine two people about to try heroin, one of them already addicted, the other not. It would be difficult for the first to avoid taking the drug, but easy for the second, and so we might hold the second responsible but not the first—or at least we would distinguish between their levels of responsibility. Now imagine that each of them is in a set of Frankfurt conditions, so that if either of them would in fact choose not to take it a mechanism would kick in that would cause her to choose to take it. Nevertheless, each of them does choose to take it on her own, without the aid of the mechanism. While they are not free to choose to do otherwise (that option being blocked by the mechanism), we would still be inclined to hold the second more responsible than the first. It was harder, *in some sense or other*, for the first to avoid doing it than for the second to avoid doing it. Consequently it is not precise to put the difference between them in terms of the difficulty of doing otherwise, because the difference between them is consistent with neither of them being able to do otherwise at all.

The Frankfurt examples generate a powerful argument, but I would like to observe what the argument does *not* prove. It shows that the distinction in levels of responsibility is consistent with the actors being unable to choose otherwise. It does *not* show that the distinction is consistent with each of them being caused to choose as they do. Suppose, for example, it had gone otherwise, and Frankfurt's evil intervenor had actually activated the mechanism that caused both of the subjects to choose to take heroin. Would we still be inclined to distinguish the responsibility of the one from the responsibility of the other? I doubt that we would. We are more likely, I believe, to think that neither would be responsible in those circumstances. The inconclusiveness of the argument results from the fact that although, in the thought-experiment as originally described, the actors could not have

19. *See id.* at 8-9.

chosen otherwise, it was left open whether the choices they actually made were *caused.* And although we know, thanks to Frankfurt, that alternative possibilities are not a necessary condition for responsibility, we do not know that a lack of causation is not a necessary condition. And thus we cannot conclude, just because of the Frankfurt argument, that the distinction in responsibility between the actor who, as we would have said, found it easy to do otherwise and the actor who, as we would have said, found it difficult to do otherwise is consistent with universal causation.

To summarize the argument so far: The difficulty of doing otherwise seems, at first glance, to suggest the need for alternative possibilities. That is, to say that you find it difficult to do otherwise implies that doing otherwise is a genuine alternative possibility for you. And we know that alternative possibilities are something that compatibilism cannot deal with.[20] But if compatibilism is false, then causation is inconsistent with responsibility, so that it appears that anyone who accepts DIFF should accept CE as well: If DIFF, then alternative possibilities; if alternative possibilities, compatibilism is false; if compatibilism is false, then CE is true. But the Frankfurt examples show that alternative possibilities are not really needed to understand the difficulty of doing otherwise, without really resolving the question whether the diminished responsibility of the addict can be shown to be compatible with universal causation. How are we to resolve that question? What I will do is this: I will survey action-theoretic explanations of addictive behavior with the following question in mind: Can talk about the difficulty of avoiding such behavior be shown to be consistent with universal causation? That is, can some interpretation of "it would have been difficult for him to do otherwise" be found that will be consistent with determinism? My hope is that by examining the various ways of conceptualizing addiction we can get closer to an answer to the question of the true role of causation in the criminal law.

20. That is to say, the idea that a person does one thing in a certain set of circumstances, but could have done something different (the alternative possibility) in the very same set of circumstances, is inconsistent with determinism. Modern compatibilists argue that all that is required for responsibility is that the actor would have acted differently in some *other* set of circumstances, somehow related to the actual circumstances. We'll take up that argument in the last part of the paper. *See infra* Section II.C.2, *Defects of Will.*

II. THEORIES OF ADDICTION

A. Overview

Let's turn then to theories of addiction and see what they imply about the relationship of addiction to responsibility. According to some theories addictive behavior is rational, and according to others it is not. Theories according to which addiction is rational are of two sorts: there are theories that attempt to explain the facts of addiction as a rational reaction to the circumstances of the addict's life together with the addict's pursuit of utility (rational addiction theories), and there are theories that try to explain the facts of addiction as a rational reaction to the threat of withdrawal pains (withdrawal theories). Theories according to which addiction is irrational are also of two sorts; there are theories that explain the facts of addiction as the consequence of a distorted or uninformed or conflicted rationality, and theories that explain the facts of addiction as an inability to choose (or an unreasonably great difficulty in choosing) in accordance with the instructions of rationality. These last are sometimes called "disease" theories. I should add that these four types of explanation of addictive behavior comprise all the theories of addiction that I am aware of. As far as I can see, every explanation of addiction, including biomedical explanations, will find a place in the following schema:

THEORIES OF ADDICTION

RATIONAL	IRRATIONAL
1. rational addiction	3. distortion theories
2. withdrawal theories	4. "disease" theories

A quick preliminary survey of what the various theories would say about responsibility, and of their implications for policy, would show this: According to the rational addiction theories, addictive behavior is rational and responsive to the (generally positive) consequences of such behavior.[21] Although it doesn't follow directly from the fact that addictive behavior is rational (if it is) that the actor is responsible for what he does, proponents of these theories generally argue for that further conclusion, and for the legal implications that follow from that: addiction is not an excuse; the sale of legal addictive substances does not impose a duty to warn of addictive properties; and the addict's own welfare is not a sufficient ground for regulating addictive substances. (The good of others, of course, may be.) The problem with such

21. *See infra* Section II.B.1, *Theories of Rational Addiction.*

explanations (for our purposes) is that though they are consistent with determinism and suit the compatibilist's purposes in that way, they do not explain why it is difficult for the addict to do otherwise, or why the addict's responsibility is diminished in any way.

According to what I have labeled "withdrawal" theories, the effect that addiction has upon responsibility is explained by the addict's rational reaction to the pain of withdrawal symptoms.[22] The analogy here is to the excuse of duress: if the withdrawal pains are sufficiently intense, the rational agent will choose to continue to consume the addictive substance, and the community should not punish that agent for not doing something that a person of reasonable firmness would not do. If, on the other hand, withdrawal symptoms are weak or nonexistent, the addict should not be excused—at least not on *that* ground. In the former case, it would seem appropriate also to impose a duty to warn on the seller of legal addictive substances, since the fact of serious withdrawal pains makes of addiction a trap for the ignorant. And in that case it may be appropriate also to regulate addictive substances in order to protect the addict from himself, among other reasons.[23] The difference between the rational addiction and the withdrawal theories is the difference between attraction and avoidance, the difference between bribes and threats. On the first sort of theory the addict acts to secure a good for herself. On the second the addict acts to avoid a bad. These explanations give an interpretation of the difficulty of doing otherwise, and of diminished responsibility, that is consistent with determinism. Unfortunately they do not fit the facts of addiction.

There are theories according to which addiction is a result of a defect in the addict's ability to reason to appropriate practical conclusions—perhaps an inability to calculate consequences correctly, or an excessive discounting of future costs and benefits. We will call these defect of reason explanations of addiction.[24] Some of these theories find the defect in ordinary human rationality: *everyone* distorts in this or that irrational way, everyone discounts in this or that irrational way. Others treat the defect as something caused specifically by consumption of addictive substances. While at least some of those who hold either sort

22. *See infra* Section II.B.2, *Addiction as Duress.*

23. Husak has pointed out the possibility that withdrawal symptoms might be severe enough to justify regulation, and not severe enough to excuse. Douglas N. Husak, *Addiction and Criminal Liability*, 18 LAW & PHIL. 655, 682-83 (1999).

24. *See infra* Section II.C.1, *Defects of Reason.*

of theory believe that such irrationality excuses or at least mitigates criminal behavior, most do not. And when it comes to regulation of addictive substances, the conclusion again is not clear. If the defect is one that all human reasoning suffers from, then it would seem to be correctable with sufficient information and so no regulation would be called for.[25] On the other hand, if the defect is caused by the addictive substance itself, it would probably not respond to ordinary education, and it may be that state regulation would also be called for.[26] The difficulty here is in understanding just what is meant by a defect of reason. When we pursue this question we will find that such theories either do not fit the facts of addiction, or if they do fit the facts of addiction they do not explain the sense that the addict's responsibility for his behavior is diminished, or they collapse into disease theories, the discussion of which constitutes the final section of the paper.

"Disease" theories are what are sometimes described as defect of will theories.[27] They include both the *automaton* theories dismissed above, according to which the addict has no choice in and no control over his behavior whatever, and theories according to which the addict finds it not impossible but extremely difficult to make the choice to give up his addictive behavior, in spite of the fact that he understands that it is rational to do so. Since according to these latter explanations addiction makes it difficult, and perhaps unreasonably difficult, to carry out the conclusions of practical reason, these may also be seen as irrationality theories. Such theories would justify excusing the addict, by way of an extension of the dictum that

> *ought* implies *can*
>
> to
>
> *ought* implies *can without unreasonable difficulty.*

That is, if addiction makes it unreasonably difficult for the addict to conform his behavior to the law, then the addict may not be required to conform. On this explanation, too, addiction becomes a serious trap, and both warnings and regulation may be required.[28] The problem with this

---

25. *See infra* Section II.C.1.a, *Universal Irrationality.*
26. *See infra* Section II.C.1.b, *Specific Irrationality.*
27. *See infra* Section II.C.2, *Defects of Will.*
28. In spite of the general reluctance to create an excuse for addiction, the law seems to assume generally that the addict may not be in control of his behavior. On addiction generally, see the definition of "addict" in the U.S. Code, 21 U.S.C. § 802(1) (1994): "The term 'addict' means any individual who . . . is so far addicted to the use of narcotic drugs as to have lost the power of self-control with reference to his addiction." *Id.* In connection with alcoholism, see D.C. Code § 24-522 (1981), which defines a

sort of theory is to find a compatibilistic explanation of it.

There may of course be some overlap among the four types of theories. It may be that one sort of theory accounts for one sort of addiction, and another sort of theory for other sorts of addiction. It may also be that some forms of addiction are overdetermined, explainable by more than one theory. If, as I believe, defect of will explanations are required to explain at least some of our beliefs about some kinds of addiction, and defect of will explanations cannot be given a compatibilist interpretation, then the popular idea that the responsibility of addicts may be diminished by the difficulty of doing otherwise presupposes the causal theory of excuse.

B. Rational Choice Theories

1. Theories of Rational Addiction

One possible explanation of addiction is that the experience of consumption is so pleasant that the "addict" simply chooses it in spite of all the possible risks. If this is so, then addiction is not a trap at all, except in the sense in which truffles are a trap—once you taste them you want very much to taste them again. The difference between addictive drugs (or some addictive drugs) and truffles is that the pleasure the drugs give is much, much more intense than the pleasure that truffles give. If this is the explanation of addiction, or of some forms of addiction, then the implication for the law is clear: The addict is responsible for what he does. The law does not take into account, in judging a rapist, that the experience gave him an intense pleasure, and neither should it take into account, in judging a user of illegal drugs, that the drugs gave him such and such a degree of pleasure. Conversely, if the drug is legal the seller of the drug need not warn of its addictive properties, because consumption of the drug is not really a trap for the unwary. To force cigarette manufacturers to warn about the addictive properties of tobacco would be no different from forcing the manufacturers of chocolate to put warnings on their product:

"chronic alcoholic" as "any person who chronically and habitually uses alcoholic beverages to the extent that . . . he has lost the power of self-control with respect to . . . such beverages." *See also* Easter v. District of Columbia, 361 F.2d 50, 52 (D.C. Cir. 1966).

To explain addiction in this way would be to make it a rational activity, with the addict simply choosing the most utility for himself. There are, however, a number of problems for this rather naive theory. First of all, it has to explain why the addict continues to consume the addictive substance when the phenomenon of tolerance sets in; that is, when she begins to derive less and less pleasure from the drug. It also has to explain why the addict continues to use the drug when it begins to affect the rest of her life in a negative way. The search for pleasure, all by itself, doesn't seem to explain that. Nor does it explain why quitting an addictive substance seems to require complete abstinence—cold turkey withdrawal—or why the addict cannot maintain a low level of consumption of the drug but sometimes, at least, tends to binge. Finally, it does not explain the sense that addiction is a trap into which it is easy to fall but difficult to get out. The theories I am going to consider here are more or less sophisticated versions of this sort of theory. They are sometimes called "rational addiction" theories.

According to rational addiction theories, addiction should *not* excuse addictive behavior. This is a point of view according to which the addict is fully responsible for his behavior. We see this point of view in the rational choice theory of economists like Gary Becker, who maintains that addicts are maximizing their own utility,[29] and in the work of legal theorists like Alan Schwartz, who argues that purveyors of (legal) addictive substances do not have a duty to warn consumers of the addictive nature of their goods because addicts do not lose the ability to control their behavior and are never trapped by their habit.[30] We see it in the work of philosophers like Herbert Fingarette, who argues that addiction is largely a myth.[31] A common thread here is the combination of the notion that addicts act upon their own preferences with the claim that anyone who acts upon his own preferences acts freely and willingly.

On the face of it, rational choice theories are badly suited to dealing with addiction. Rational choice theories attempt to explain human behavior as the maximization of utility, where utility consists in the

29. *See* Gary S. Becker & Kevin M. Murphy, *A Theory of Rational Addiction*, 96 J. POL. ECON. 675, 675 (1988), *reprinted in* GARY S. BECKER, ACCOUNTING FOR TASTES 50 (1996) [hereinafter BECKER, ACCOUNTING FOR TASTES].

30. *See* Alan Schwartz, *Views of Addiction and the Duty to Warn*, 75 VA. L. REV. 509, 514, 545, 559 (1989).

31. *See* HERBERT FINGARETTE, HEAVY DRINKING: THE MYTH OF ALCOHOLISM AS A DISEASE, *passim* (1988) [hereinafter FINGARETTE, HEAVY DRINKING]; HERBERT FINGARETTE & ANN FINGARETTE HASSE, MENTAL DISABILITIES AND CRIMINAL RESPONSIBILITY, *passim* (1979).

satisfaction of preferences. The utility we derive from a good depends upon our preference for that good. One problem for such theories is that they generally assume that preferences are fixed; if we allow preferences to change, the theory will lead us into contradictions. But addiction seems to involve changing preferences: the addict's desire for the substance and his consumption of it increase over time until he is abusing the substance. In such circumstances it is impossible to favor one course of action over another as the more rational. If I am faced with two courses of action, one of which involves consuming an addictive substance and the other of which does not, how am I to decide between them? I may dislike the addictive substance now, and I may continue to dislike it if I refuse to consume it. But if I consume it I may come to like it a great deal. From which point of view should I appraise the two courses of action—with the preferences I have now? With the preferences I will have if I become an addict?[32] Once I am addicted the problem does not disappear; each choice will affect my future preferences.

It isn't only addiction that raises the problem of changing preferences; indeed, the problem is commonplace. Alan Gibbard describes the case of the young man trying to decide whether to enter the seminary or to go to the university for a secular career.[33] If we are trying to work out the utility maximizing alternative for him, the problem is whether to judge both the seminary alternative and the university alternative from the point of view of the preferences he would have if he went to the seminary, or from the point of view of the preferences he would have if he went to the university, or from the point of view of the preferences he has right now before making the decision. Obviously none of those alternatives is exactly right: he will develop one set of tastes if he goes to the university, and a different set if he goes to the seminary. Is it fair to judge either alternative by the tastes of the other? Or of neither?

32. *See* Cass R. Sunstein, *Legal Interference with Private Preferences*, 53 U. CHI. L. REV. 1129, 1159 (1986); Menahem Yaari, *Endogenous Changes in Tastes: A Philosophical Discussion*, *in* DECISION THEORY AND SOCIAL ETHICS 59, 64-66 (Hans W. Gottinger & Werner Leinfellner eds., 1978).

33. *See* Allan Gibbard, *Interpersonal Comparisons: Preference, Good, and the Intrinsic Reward of a Life*, *in* FOUNDATIONS OF SOCIAL CHOICE THEORY 165, 176 (Jon Elster & Aanund Hylland eds., 1986).

> Perhaps we should judge his going to the university by the . . . preferences he has by virtue of going to the university, but judge his going to the seminary by the . . . preferences he would have if he had gone to the seminary. But to try to do so would be incoherent . . . .[34]

The problem has been tackled in a number of ways, the best known of which may be Becker's. According to the standard rational choice theory, how much of a good an agent consumes is a function of his preference for that good—something that remains unchanged throughout the calculation. For Becker, consumption is a function of a kind of super-preference the agent has. Instead of just applying a preference to a good to get an amount of consumption, we must apply the super-preference to both the good and the effects of a prior history of consumption. In other words, our preference for a good, and consequently how much we consume, is influenced not only by what the good is, but by the prior history of our experience with the good. These super-preferences are themselves unchanging, making rational choice theorizing possible, and yet how much of a good we consume at any given time and how much utility we derive from it will vary with our history of consumption.[35]

Everyone has personal capital, according to Becker, which consists in a stock of prior consumption which depreciates over time. This stock is the combined effect of past experience and consumption and present social influences on behavior. The utility you get from a good depends not only upon your consumption of the good, but also upon your stock. Different people, with different backgrounds, bring different stocks to the consumption of any good and get different amounts of utility out of it. Thus we do not need to assume that one person has different preferences at different times, or even that different people have different preferences. Differences in utility arise from differences in stock. The preferences themselves remain constant.[36]

Applied to addiction, the picture that Becker has drawn for us[37] is one

---

34. *Id.* at 177.
35. *See* BECKER, ACCOUNTING FOR TASTES, *supra* note 29, at 7-10, 20-23.
36. *See id.* at 57.
37. Here is how the theory applies to the explanation of addiction: Part of personal capital may be an addiction stock, which is the effect of past consumption of some addictive good. The addictive stock tends to decay or depreciate over time, so that if you wait long enough between episodes of consumption, you will lose all your addictive stock and will not be an addict. This is what happens when an addict breaks the habit. Consumption of the addictive good increases the addictive stock an individual has, which tends to increase future consumption. That happens because an increase in the stock of past consumption means an increase in the present marginal utility of consumption. At the same time, for most addictive substances, an increase in the stock of past consumption means a decrease in future utility. This is the phenomenon known as "tolerance." Consumption will thus be reinforcing if the present marginal increase in

in which an addict's rate of consumption will vary with the amount and timing of his past consumption, which Becker calls the addict's addictive stock, and which we may think of as the monkey on his back. If the addict fails to consume his addictive substance, the monkey diminishes in size; as he consumes, the monkey grows. If his consumption is exactly enough to offset the effect of depreciation, then he will remain in a steady state of consumption. If it more than offsets the depreciation, the monkey will grow and consumption will increase. If his consumption is not great enough to offset the loss, his consumption will decrease. An actor whose stock is large enough to dictate an amount of consumption that will more than make up for the stock lost through depreciation is on the road to ever increasing consumption and addiction. Under the right conditions the addict will quit "cold turkey"; under other conditions he will binge.[38]

What do such theories say about the relationship of addiction to responsibility? They start with the observation that addictive behavior is the consequence of the addict's own preferences,[39] and that the addict alters his addictive behavior in response to changing circumstances. This latter point is buttressed by Becker with the empirical studies of cigarette smokers who change their habits in response to changing prices, and by Schwartz and Fingarette with references to the decline in the use of drugs among soldiers returning from Vietnam, where drugs were freely available, to the United States, where they were not.[40] Upon

---

utility exceeds the future decrease, discounted to present value. Much depends, therefore, on the addict's "time preference:" how he discounts future utility. *See id.* at 56-60, 68-70.

38. Becker assumes that the addict knows all this, and that that is what accounts for his responsiveness to price changes. For example, cigarette smokers reduce their consumption in response to future increases in the price of cigarettes (instead of smoking as much as they can at today's lower price). That is because they know they will be able to afford fewer cigarettes in the future, and by reducing the consumption today they reduce the amount they will consume in the future. When a future drop in prices is announced, similarly, smokers will increase their consumption, even at today's higher prices, because they know that they will want to consume more in the future. Becker has attempted to support all of these claims with empirical studies. According to Becker, this same explanation of addictive behavior makes clear why withdrawal must often be "cold turkey," and why there may be binges in response to price reductions. *See id.* at 70-71.

39. For some writers the fact that addictive behavior results from the addict's own preferences is enough to demonstrate that the addict is responsible. For if the behavior results from preferences it is volitional, and if it is volitional it is voluntary. *See, e.g.*, FINGARETTE & HASSE, *supra* note 31, at 160-63; Schwartz, *supra* note 30, at 542.

40. *See* BECKER, *supra* note 29. *See also* FINGARETTE & HASSE, *supra* note 31, at

the reasonable assumption that the addict knows what he is doing, the fact that his behavior is the result of his own volition and the fact that he had a choice in the matter are taken to show that he is responsible for what he does.[41]

Now if the rational addiction theorist is right, then the difference between addictive behavior and ordinary behavior can easily be squared with universal causation: the difference between the addict and everyone else lies in the stock of addictive capital that the addict brings to his choices—not unlike a difference in software. The problem is that the rational addiction theorist cannot account for the supposition that the addict ought to be excused or his responsibility mitigated. Remember that the common wisdom about addiction (accepted by Michael Moore) is that addicts may find it extremely hard, perhaps unreasonably hard, to avoid their addictive behavior. But the rational addiction theorist, when it comes down to it, has got to deny that that is so. The theory predicts that when the substance stops being beneficial, the addict will stop taking it. It would make addicts satisfied, rational actors, and give the lie to the common wisdom. But addicts often are not happy, and often would like to overcome their addictions. Becker's version of the approach does not explain the regret addicts suffer when they consider their condition, and their sense that they have harmed themselves:

> The rational approach . . . has been widely criticized as inconsistent with observed regret among addicts and incompatible with any role for information and public policy. . . . Winston (1980) points out that addicts in the Stigler and Becker model are 'happy addicts,' choosing their addiction[42] after careful consideration of the alternatives and never doubting their actions.

Nor is this something that Becker would deny. He acknowledges that addicts may not be happy; but he maintains that in the circumstances they would be even more unhappy if they were prevented from consuming the addictive goods. He is not talking about withdrawal pains here; he means that the addict's addictive behavior helps him deal

---

157-58; Schwartz, *supra* note 30, at 542.

41. This conclusion is drawn explicitly by Schwartz, *supra* note 30, at 531. It is attributed to Becker in a roundabout way by Richard Herrnstein and Drazen Prelec: "[T]he logic of this approach argues against criminalization of addictive substances, or any other form of government action that raises prices or otherwise impedes their use. It also argues against attempts to change demand by education, restrictions on the advertising of addictive substances, and the like." Richard J. Herrnstein & Drazen Prelec, *A Theory of Addiction*, *in* CHOICE OVER TIME 331, 357 (George Loewenstein & Jon Elster eds., 1992) (internal citations omitted). "If there is an overriding social welfare argument against drug abuse even though it is individually optimal, the 'rationalists' [like George Stigler, Gary Becker, and Kevin Murphy] have, to our knowledge, not yet made it." *Id.* at 356.

42. Athanasios Orphanides & David Zervos, *Rational Addiction with Learning and Regret*, 103 J. POL. ECON. 739, 740 (1995).

with his unhappiness. Addicts, after all, "are rational and maximize utility."[43] And the same would be said, I think, by Schwartz and others who hold similar views.

I conclude that although a rational addiction view shows how addictive behavior may be distinguished from normal behavior in a deterministic world, it does not explain our sense that the addict is entitled to special consideration when it comes to breaking the law.

2. *Addiction as Duress*

There is another explanation of addictive behavior that sees it as rational. But instead of pointing to the utility the addict derives from consumption, it emphasizes the disutility of withdrawal, the fear the addict has of the consequences of quitting. The analogy that is sometimes drawn is with the excuse of duress. The criminal law allows that a person who commits a criminal act because of threats against him or his family may be excused from criminal liability. Suppose that Jones threatens to harm Smith's wife unless Smith drives the getaway car for a bank holdup. If the threat is credible, and Smith has no alternatives (like going to the police), and if the harm threatened against his wife is so serious that a person of reasonable firmness would have given in and driven the getaway car, then Smith will not be held responsible; he will be excused for participating in the holdup. The argument is that the reasoning behind the defense of duress should apply to other situations as well, including the case of someone who is threatened with withdrawal pains if he does not secure and consume a prohibited substance, and the pains are so serious that a person of reasonable firmness would have done the same thing under the circumstances.[44]

Now it may be that the person who acts under duress and the person who acts under fear of withdrawal pains are not acting rationally. The fear itself may have disturbed their ability to process information; they may lose the ability to draw reasoned conclusions, and may in panic act irrationally. There is no need for us to deal with this possibility here, since later on I will examine all the explanations from irrationality together. What I am interested in here is the rational version of the defense: a person who acts under duress will be excused even if he has

43. Becker & Murphy, *supra* note 29, at 691.

44. *See* Husak, *supra* note 23, at 656-57. *See also* Gary Watson, *Excusing Addiction*, 18 LAW & PHIL. 589, 590 (1999).

coolly appraised the alternatives and has chosen to break the law, so long as the threat was of the appropriate sort and sufficiently serious.

The implications of this explanation are relatively clear. Addiction is indeed a trap, if the pains are severe enough and the explanation is correct, for the person who is not aware of the withdrawal pains, so that the consumer would be entitled, even in the case of legal addictive substances, to a warning about those pains. Furthermore, the addict whose violation of the law is attributable to his cool appraisal of withdrawal pains, where those pains are serious enough, is entitled to an excuse. The question is, then, are withdrawal pains serious enough to excuse the addict for his addictive behavior? Doug Husak's considered opinion, after a careful review of the empirical literature, is that there simply is no evidence, for any drug, of overwhelming withdrawal pains. His conclusion is amply supported by the evidence. For some drugs, like cocaine, there is no withdrawal pain at all; for others, like nicotine, there is mild discomfort; for others like heroin the most serious cases have been compared to the discomfort of having a bad cold for several days. None of these things would excuse breaking the law.[45]

What we have in the duress theory, then, is another sort of theory that would make the difference between addictive behavior and normal behavior consistent with universal causation. The addict is a person faced with threatening possibilities the normal person is not threatened with: he calculates his options, and chooses to consume.[46] The normal person, not faced with the same threats, may choose another option. Both will choose the option that satisfies their search for utility, however. Nevertheless, the theory fails on empirical grounds; the evidence simply does not support the conclusion that in those cases where withdrawal pains do exist, they are severe enough to explain addiction. And the theory does not begin to explain our willingness to believe that even the addict not faced with withdrawal pains may find it difficult to control his choices and may deserve to be excused for his behavior.

---

45. *See* Husak, *supra* note 23, at 680-83.

46. Notice that while the idea of alternative possibilities all accessible from one and the same set of circumstances is not consistent with determinism, the idea that one might *consider* alternatives and choose one *is* consistent with determinism. The background and makeup of the chooser might determine the outcome, but consideration and weighing of alternatives might be part of the process by which the choice is implemented. A machine, for example, might be programmed to perceive and choose among alternatives, for example, and the explanation of that is surely mechanistic.

C. Irrationality Theories

1. Defects of Reason

One of the things we believe about addicts is that they have a distorted view of what is important. If in fact the addict's world view is badly distorted, then addiction may be a kind of insanity: the addict, like the legally insane person, is not really in a position to appreciate the consequences of his actions. If so, then perhaps he should not be held responsible for what he does. This point of view, addiction as irrationality, finds sophisticated expression in the work of social scientists like Herrnstein[47] and Ainslie,[48] and in the work of legal commentators like Morse.[49] Moore's own approach to addiction would seem to fit here. In *Placing Blame* he says:

> Addiction is a cause of behaviour, but it is not because of this fact that addictions operate as compulsions and therefore provide an excuse. Rather, addictions can be compulsions because they make it difficult to do what the law demands .... Both the duress example and the addiction example show that what furnishes an excuse is *the disturbance of practical reasoning*, not the fact that the disturbance was caused.[50]

We must distinguish at the outset between two different claims that such theories might make. They might, on the one hand, claim that the defect in rationality that the addict suffers from is universal, found in all human beings, with especially unfortunate consequences for the addict; or they might, on the other hand, claim that the defect of rationality is one that is *caused* by addiction. After considering these two sorts of irrationality explanation we will consider a third sort, one that depends upon levels of preference or desire. Let's begin with theories that talk about a universal defect of reason.

a. Universal Irrationality

Herrnstein's work on this subject is found in two papers by Herrnstein

47. *See* Herrnstein & Prelec, *supra* note 41, at 356-57.
48. *See generally* George Ainslie, *A Research-Based Theory of Addictive Motivation*, 19 LAW & PHIL. 77 (2000).
49. *See* Morse, *supra* note 16, at 49. Fingarette, mentioned above in connection with rational addiction, makes some of the same points. *See* FINGARETTE & HASSE, *supra* note 31, at 6, 191-95.
50. MOORE, *supra* note 2, at 526 (emphasis added).

and Prelec, *Melioration*,[51] and *A Theory of Addiction*.[52] Like the work of rational addiction theorists, Herrnstein's explanation of addiction is based upon the addict's desire to maximize utility. Unlike the rational addiction theorists, however, Herrnstein does not assume that addicts are perfect reasoning machines. Indeed, it is because of his defective reasoning that the addict lands in a less than optimal situation, a kind of internal market failure. The defect in the addict's reasoning is not one caused by drugs, according to Herrnstein and Prelec; it is a standard defect that we all share, which is activated in a particularly unfortunate way to produce the condition of addiction.[53]

Herrnstein posits that human beings have two different approaches to choice. If the choice is a one of a kind choice, involving objects of larger dimensions, like buying a house, then human beings tend to make rational decisions; they tend to maximize their utility. But if it involves repeated small choices over time, then the human reasoning apparatus is simply inadequate. In the case of repeated small choices—like the choice of what to have for dinner—human beings tend to pursue those goods that promise the most utility, *on the average*.[54]

Let me give a herrnsteinian example to show why this approach to decision-making does not maximize utility. Suppose your choice is between eating carrots with dinner or eating green beans with dinner. You could eat carrots every day of the week, or you could eat green beans, or you could vary them in any combination. Now suppose that there is a declining marginal utility from the repeated consumption of vegetables. If you eat carrots once, that gives you a certain utility; but if you follow that at the next meal with another portion of carrots, you will get a bit less utility. Suppose that carrots give the higher utility to begin with, but that the utility declines rapidly with repetition; green beans start out lower on the utility scale, but do not decline so fast. Then it is entirely possible that you will fail to vary your choices in a way that would maximize utility. I give a numerical example in the footnote.[55]

---

51. Richard J. Herrnstein & Drazen Prelec, *Melioration*, *in* CHOICE OVER TIME 235 (George Loewenstein & Jon Elster eds., 1992).

52. Herrnstein & Prelec, *supra* note 41.

53. *See* Herrnstein & Prelec, *supra* note 41, at 339-42; Herrnstein & Prelec, *supra* note 51, at 235-37.

54. *See* Herrnstein & Prelec, *supra* note 51, at 240.

55. Suppose that the declining utility of eating carrots and green beans can be represented by the following numbers:

$$C = 8, 7, 6, 5, 4, 3, 2$$
$$GB = 5.5, 5, 4.5, 4, 3.5, 3, 2.5$$

That is, on the first day of eating carrots you will derive 8 units of utility, on the second day 7 units, and so on; on the first day of eating green beans you will derive 5.5 units of utility and so on. Our subject has tried both carrots and green beans: he knows the utility

The defect in reasoning is this: where repeated small choices are made, we substitute average utility for marginal utility.

This error, this distortion, is supposed to explain addiction. The addict will move in the direction of the addictive substance instead of in the direction of alternatives, in spite of the fact that the next unit of the addictive substance will produce less utility than the alternative. She will do so because the addictive substance has, over the past units of consumption, yielded the higher *average* utility. This explains why addiction seems like a trap. The addict is unlikely to avoid consumption so long as the average utility of consumption remains high enough. And it explains why addicts sometimes refrain from consuming the addictive substance for a period of time, fully intending to return to the habit: they are trying to raise the average utility.

The theory of Herrnstein and Prelec is a good example of the sort of theory that attributes addiction not to rational action but to a form of irrationality that afflicts every human being.[56] The implications of any such theory are clear enough. The choice to engage in addictive behavior is at no point irrevocable, and therefore, as soon as the average utility of consumption falls below that of an alternative course of action, the addict will stop his addictive behavior. He has no special disability that would require the state to excuse him for violations of the law. His problem is a lack of information, or a lack of appreciation of the information. But a lack of information about how much utility he would derive does not excuse anyone's violation of the law. The addict may be punished for possession and consumption of prohibited materials, therefore. Although the addict is responsible for what he does, a warning may be helpful. Thus the state might require cigarette manufacturers to warn smokers that they might be led to consume beyond the point at which smoking maximizes utility.

Again, although such theories depend upon a disability, this is not the sort of disability that would explain our sense that the addict is not fully

---

produced. If he were a maximizer, he would consume carrots four days a week and green beans for three days, for a total of 41 units of utility. Instead Herrnstein and Prelec's work shows, according to them, that a person will choose the vegetable with the highest average utility. As long as the *average* utility of eating carrots exceeds the average utility of eating beans, carrots will be chosen, and eventually the person will find himself eating carrots six times a week and green beans only once, for a total utility of 38.5. *See* Herrnstein & Prelec, *supra* note 51, at 237-42.

56. Another closely related example is Ainslie's theory of hyperbolic discounting. *See* George Ainslie & Nick Haslam, *Hyperbolic Discounting*, *in* CHOICE OVER TIME 57 (George Loewenstein & Jon Elster eds., 1992).

in control of his behavior. The disability is one that is shared by every human being, and does not just fall upon addicts. And such theories do not explain the sense that abandoning an addiction is difficult, however much utility the addict is deriving at the time. The cigarette addict does not simply abandon his habit when the average utility of consumption falls way below the average utility of non-consumption. Let's turn now to theories that claim that addicts suffer from a special defect of reason, one caused by their addiction.

b. Specific Irrationality

According to Stephen Morse, addictive cravings interfere with rationality by clouding the addict's understanding of what he is doing.[57] The addict might, because of his addiction, severely discount the value of alternatives to addictive behavior, and that is the sort of mistake that someone not addicted would not be likely to make. He might not be able to see, or to reason, clearly about the alternatives because of the addiction, and may reach the wrong conclusion about what he should do. Since responsibility presupposes rationality[58] and addicts are not fully rational, Morse concludes that addicts are not generally fully responsible for their addiction. For that reason they should have at least a partial excuse in the criminal law, an excuse for possession and use, and perhaps for some other crimes as well.[59] Here, then, is an explanation that might account for the sense that the addict is different from the ordinary person and not fully in control of his behavior.

I see two different sorts of irrationality that may be involved here. The first is a kind of *objective* irrationality. Though the addict's behavior follows from his preferences—in this case, his time/value-preferences—we judge his preferences themselves, and in particular his preferences relating to the value of time, to be disordered. There is nothing wrong with the way the addict's reason functions; it gets him to the appropriate conclusion, and the appropriate action, given the premises he starts from. The other sort of irrationality is *subjective* or internal. Here there is nothing wrong with the addict's assessment of the alternatives, but his defective reason prevents him from drawing the appropriate conclusion. It prevents him from drawing the appropriate conclusion about the circumstances he is in from the evidence he has, or it prevents him from drawing the appropriate conclusion about the action

57. *See* Morse, *supra* note 16, at 38-43.
58. *See id.* at 24.
59. *See id.* at 45-49.

he ought to take from his preferences and his beliefs about his circumstances.

According to the first of these approaches, the addict tends to value the next consumption of his substance much more highly than he values more distant rewards that would be precluded by that consumption. He discounts those distant rewards severely, and that sort of discounting is caused by his addiction. He chooses, of course, in line with his valuation, and thus will choose consumption of the addictive substance over the more rational course of action. This approach is similar to the one taken by George Ainslie in *A Research-Based Theory of Addictive Motivation.*[60] The problem with this approach is, I think, that it does not give any reason to want to excuse the addict, and so cannot explain the sense that the addict does not control his behavior. The addict may discount differently from the non-addict, but all that means is that he prefers nearer rewards over future rewards much more than the non-addict does. Why should that excuse his behavior? Discounting is a kind of preference, and we are not normally excused for acting upon our preferences; we are held responsible precisely for that. Furthermore, how could such distortion excuse? Suppose that addiction makes the addict value the next instance of consumption over the possibility of staying out of jail, and chooses accordingly. Why in the world should that excuse him? We do not inquire of our criminals whether they did not discount the importance of going to jail too much; why would the answer make any difference to us?

But what about the other sort of defect of reason, the subjective sort? This refers to the possibility that the addict's reasoning is so disturbed that his behavior cannot be said to be reasonably related to his beliefs and desires. To keep this suggestion distinct from other explanations of addiction, we must have the addict acting in accordance with the conclusions he draws from the premises; but the conclusions must not be those that properly follow from those premises. His behavior must be rationally inexplicable, but at the same time there must be nothing wrong with his ability to choose otherwise. His reasoning is simply haywire. Such irrationality, I am willing to concede, should excuse behavior. But I see no reason to think that anything of that sort is wrong with the addict. Indeed, that is what makes addiction such an interesting case. The sleepwalker might draw the wrong practical conclusions,

60. *See* Ainslie, *supra* note 48, at 77.

conclusions that have no relationship to his actual preferences. But the addict, granted that he may have distorted desires, seems perfectly capable of drawing the conclusions that follow from those desires.

There are two sorts of thing to be said, then, about the subjective irrationality explanation of addiction. The first is that if the addict did systematically draw the wrong practical conclusions from his beliefs and desires, he might be entitled to an excuse. But it would not be because of the difficulty of doing otherwise, which after all is what we are trying to capture here. It would be because of a kind of insanity or incompetence. The second response is that that condition does not fit very well with our picture of addiction; the addict is not a bumbling, helpless wanderer who can't fit his actions to his desires. Addictive behavior, however destructive, is purposeful and often efficient in satisfying the addict's immediate desires.

c. Irrationality as Inconsistent Desires

There is a third sort of explanation that best fits in the "defect of reason" classification, but is very different from the above two sorts. In this sort of explanation, advanced primarily by Harry Frankfurt, the addict *is* responsible for his addictive behavior, *if* his addictive behavior accords not only with the immediate desires that lead to it, but also with the addict's higher level desires, his desires about what sort of person he wants to be and so on.[61] The *willing* addict, the addict who has no regrets, the addict whose consumption of the addictive substance is consistent with his deepest desires, does what he does willingly, and can be held responsible. On the other hand, the unwilling addict, the addict who is unhappy about what he is doing, the addict who does not want to be someone who is dependent upon an addictive substance, is not responsible for his addictive behavior. In light of his concerns about the sort of person he is or will become, it is irrational for him to continue; and yet he does, though "against his will."[62]

To understand the motivation for this explanation of the connection between responsibility and addiction, we must look at another part of Frankfurt's work. In the attempt to show that freedom and responsibility have nothing to do with an actor's ability to do otherwise Frankfurt constructed the example, mentioned above,[63] of Jones and Black.[64] Either way, Jones will end up doing what Black wants him to do, and doing it

61. *See* Harry G. Frankfurt, *Freedom of the Will and the Concept of a Person, in* HARRY FRANKFURT, THE IMPORTANCE OF WHAT WE CARE ABOUT 11, 24-25 (1988).
62. *Id.*
63. *See supra* notes 18, 19 and accompanying text.
64. *See* Frankfurt, *supra* note 14, at 4-10.

intentionally. In one case he does it on his own; in the other case he is caused to do it by Black's interference. Now assume that Jones does it on his own, without interference. Is he responsible for what he has done?

If the ability to do otherwise, or the ability to choose to do otherwise, were a necessary condition of responsibility, then Jones would not be responsible. For Jones could neither do otherwise nor choose to do otherwise. But why should Jones not be responsible for what he has done? He acted without interference; he did what he wanted to do, and he did it willingly. Black had nothing to do with it, and everything about Jones's state of mind was just what it would have been had Black never entered upon the scene. The conclusion Frankfurt draws from the example is that the ability to do or choose to do otherwise is not a necessary condition of freedom or responsibility.[65] Since incompatibilist theories of excuse are built upon the assumption that the ability to do otherwise *is* a necessary condition of moral responsibility, this argument was a step in the attack upon the incompatibilist and libertarian positions.[66]

It is a short step from here to the conclusion that the person who is truly responsible and free is the one whose actions are in accord with his desires, including what we can call his higher level desires, about what sort of person he wants to be and so on. That is the state that Jones is in when he acts without interference. And the person whose desires are in conflict, who chooses to consume the drug on this occasion in spite of his overriding desire not to be an addict, that person acts *against* his will. That person acts unwillingly. We can, if we want to, say that he is constrained to do what he does not what to do, but that is not the important point; the important point is that he acts against his will.[67] The willing addict is responsible for what he does; the unwilling addict is not.[68]

This explanation of how addiction is related to responsibility is unwieldy. How would the law make use of it? Would unwilling addicts be let go, and willing addicts imprisoned? Indeed, in what sense is the willing addict an addict at all? Isn't he just a pleasure seeker? What if a willing addict finds some reason to quit, and learns that he can't, and so

65. *See id.* at 9.
66. *See* Frankfurt, *supra* note 61, at 23.
67. *See id.* at 18.
68. *See id.* at 25.

continues unwillingly; was he responsible before he tried to quit, but no longer responsible now? But beyond its awkwardness, it simply does not jibe with our notion of human freedom. Consider the case in which Jones is going to decide not to do the thing Black wants him to do. Black must interfere. Now he may interfere simply by causing Jones to intentionally do the thing, something that would perhaps conflict with the deeper desires that would have left him, absent interference, to refuse to do the thing. In that case, Jones is not acting freely, and is not responsible for his behavior, which is as it should be. On the other hand, Black might interfere by not only causing Jones to intentionally do the thing, but to have the higher level desire to do it as well. In other words, Black might cause Jones to do the thing "willingly." And yet Jones is no more free or responsible here than he was in the first case. The fact that his desires are in line does nothing in that situation to change matters, and so the story about desires of different levels does not succeed in explaining for us addiction and responsibility.

To summarize: The three types of explanation we have looked at—rational addiction, duress, and defect of reason—may or may not help us to understand why addicts behave as they do. None of them, however, is a very plausible explanation of the general sense that there is something about addiction that calls for compassion, that the addict is not fully responsible for what he does, and that the severely addicted person is entitled to something like an excuse or a mitigation. Neither is any of them a very plausible explanation of the general sense that addiction is something of a trap warranting state regulation. The rehabilitated "disease" theory that I want to discuss here, on the other hand, is a relatively successful explanation of both of those things.

2. Defects of Will

a. "Disease" Theories and Compatibilism

What exactly is a disease theory of addiction? The *automaton* theories that I mentioned at the outset would seem to qualify as disease theories. So would any explanation that had addictive behavior following from some dysfunctional physiological state, without any input from the addict himself: the movements would have to occur without the addict intending or willing them to do so. Any explanation of the first sort is easily discredited; it took a few sentences in the introductory section of this paper to do just that. And the addict of course acts intentionally, which defeats the second sort of explanation. Furthermore, it seems that the addict can avoid his addictive behavior, both in the particular instance, when the policeman is nearby, for example, and in general, by

licking his addiction. So the addict is not a helpless automaton, like the sleepwalker may be.

But automatistic theories do not exhaust the possibilities. A theory according to which the addict *could* carry out his deep desire to break his habit, but only with unreasonable difficulty, might be a disease theory. What distinguishes disease theories from other theories is that according to the disease theory the addict does not behave rationally, nor can his irrationality be described as an inability to reason correctly. According to these theories what is involved in addiction is not a difficulty in reaching a rational conclusion, but rather a difficulty in acting in accordance with that conclusion. Such theories may also be described as defect of will theories. According to Jay Wallace, "[a] defect of the will . . . should be understood as a condition that impairs our ability to act well, without necessarily depriving us of the capacity to think clearly and rationally about what we are to do."[69] It is not that every addict wants to avoid his addictive behavior and experiences difficulty in doing so; but for anyone to be addicted it must be the case that *were he* to want to avoid the addictive behavior, were he to reach the conclusion that all things considered it would be better for him if he refrained from consuming the drug, he would have difficulty in conforming his behavior to that conclusion. It is for this reason that such theories may be classified as irrationality theories: though the addict reasons to a certain conclusion, he may experience a great deal of difficulty in acting in accordance with that conclusion. His behavior, therefore, will be irrational.

Disease theories appear to presuppose a difference between the ordinary person who, given a certain structure of preferences, is free to act upon them or not as she chooses, and the addict who, given a certain structure of preferences, will suffer from a diminished capacity to act upon certain of them. It requires us to suppose that there is a difference between ordinary preferences which, however strong, need not be acted upon, and addictive preferences, which, like bodily needs, insist on being acted upon. This suggests a familiar model of human choice: a person with a given desire in a given setting may choose to act upon it and, conversely, may choose not to act upon it, and his choice is not determined by the setting (including his own makeup). The difference then between ordinary preferences and addictive preferences lies in the

69. R. Jay Wallace, *Addiction as Defect of the Will: Some Philosophical Reflections*, 18 LAW & PHIL. 621, 629 (1999).

ease with which the actor can make either of those choices.

Since the disease theories depend upon the notion of a defective ability to choose, and since that seems to implicate alternative possibilities for human actions, I want to retrace at somewhat more length my earlier discussion that led us from addiction through alternative possibilities to the Frankfurt examples. Then I want to look at what might be considered the leading attempt to resolve the post-Frankfurt question of compatibilism without alternative possibilities, the solution proposed by John Martin Fischer and Mark Ravizza,[70] and to look particularly at their suggested way of dealing with addiction. My line of reasoning will be this: the disease theories give us one way of understanding the diminished responsibility of addicts; none of the other sorts of theory does an adequate job of explaining the diminished responsibility of addicts; therefore rejection of the causal theory of excuses will depend upon showing that the disease theories are consistent with universal causation.

The first three types of theory got around the problem of finding a substitute for alternative possibilities by taking a different route to the explanation of addictive behavior. Both rational addiction theories and withdrawal theories were consistent with the actor having no real freedom of choice, so long as the choice he made was explainable in terms of the projected consequences, whether attractive or aversive. And the defective reason theories would explain responsibility or the lack of it in terms of the reasoning mechanism that resulted in action. But the defect of will theories raise the problem of alternative possibilities directly. Talk about a defective ability to choose requires us to believe that there is an action that rationally recommends itself to the agent, an action that he could choose to perform, which is yet difficult (though not impossible) for him to choose. And that seems to require that at some point in time both performing the action and not performing it are causally consistent with the existing state of affairs, and that it all depends upon the agent's choice, and yet that there is some sense in which it is easy to perform the action, and difficult not to perform it.

Now if this is really what an explanation of diminished responsibility for addicts requires, then it is inconsistent with universal causation, because the idea that at some point in time both an action and its omission are causally consistent with the existing state of affairs is itself inconsistent with universal causation. What might save us from this conclusion, as we saw earlier, is the Frankfurt example. The implication

70. *See* FISCHER & RAVIZZA, *supra* note 1. Fischer's earlier, slightly less complex, theory is found in JOHN MARTIN FISCHER, THE METAPHYSICS OF FREE WILL: AN ESSAY ON CONTROL (1994).

of Frankfurt's work is that alternative possibilities are not a necessary condition for responsibility, leading compatibilists to search for a "same-world" explanation of responsibility. And we saw that these examples could be adapted to our problem: we would want to hold the nonaddict fully responsible for the consumption of a drug and the addict less responsible, even in a Frankfurt situation where neither of them could have done otherwise at all. But this left us without an explanation of the difference between the two. The Frankfurt examples gave new hope to compatibilism, and what we are looking for is a compatibilist explanation of the difficulty of doing otherwise.

b. Responsibility as Guidance Control

Perhaps the most sophisticated contemporary attempt to formulate a compatibilist theory is that of Fischer and Ravizza. They propose a theory of moral responsibility which says roughly the following:[71]

1. To be responsible for our actions we must be in control of them.
2. The control we need is not regulative control (genuine alternatives and the ability to choose among them) but guidance control.
3. An agent has guidance control over an action if the mechanism (for example, practical reasoning) that results in the action (1) belongs to her, and (2) responds to reasons in an appropriate way.
4. A mechanism belongs to an agent if the agent has taken responsibility for it in the past.
5. A way of responding to reasons is appropriate if the agent can appreciate and rank reasons for acting (this is the issue of the mechanism's *receptivity*), and if the mechanism that produces her action would respond differently to at least one reason, under some set of circumstances (this is the question of *reactivity*).

An agent is in guidance control of her behavior B only if in some other possible set of circumstances she would avoid doing B. But it is not required that she be able to refrain from doing B under exactly the

71. *See* FISCHER & RAVIZZA, *supra* note 1, at 240-44.

same set of circumstances as the one in which she did B, because if that were what it took then guidance control would be inconsistent with causation: the set of conditions would not be sufficient conditions for the occurrence of B, since not-B would also be possible given those conditions. Hence that set of conditions could not be a causal condition of B. Talking about the choice the agent would make in different circumstances (rather than in the same circumstances) retains as much of the "ability to do otherwise" as is consistent with causation, and at the same time avoids the problem that the Frankfurt examples create for alternative possibilities.[72]

But we don't want to be talking about just any other set of circumstances; to appraise the control that a hypnotized subject has over her behavior, we don't want to bring in the set of conditions in which she is not hypnotized. The question is, given that she has been hypnotized, how much control does she have over her behavior? So the list of sets of conditions that are relevant must be limited to conditions which include the fact that she has been hypnotized. This is the point of talking about the "mechanism" through which the agent operates. The mechanism by which the hypnotized agent operates is different from the mechanism that operates when she is not hypnotized. The question is not whether there is any set of conditions whatever, including a change of mechanisms, under which the agent would do otherwise, but rather whether there is any set of conditions under which the very mechanism that operated in this case would produce a different result. (Fischer and Ravizza acknowledge that they have no general way of individuating "mechanisms;" but they suggest that a certain indeterminacy in mechanisms may reflect a certain indeterminacy about responsibility.[73]) If universal causation is true, then the conditions under which the mechanism actually operates *cause* the action that occurs; but that is consistent with supposing that other conditions might elicit a different action from the same mechanism.

To return to the case of addiction: To say that the addict, in taking the drug he is addicted to, is not an automaton we must be able to say that under some set of conditions he would refrain from taking that drug. Of course he might refrain from taking it if he were not addicted; that isn't of any interest to us, and so the question must be whether there is any set of conditions under which the addict, with relevant parts of his operating machinery—i.e., his addiction—intact, would refrain from taking the

72. The Frankfurt examples show that there can be responsibility even though in the given set of conditions the agent could not have done otherwise. *See* Frankfurt, *supra* note 14, at 8-10.

73. *See* FISCHER & RAVIZZA, *supra* note 1, at 40.

drug. If he would not, then his addiction is irresistible, he is indeed an automaton of a sort, and he is not morally responsible for his drug-taking.

Of course, a "crazy" mechanism might produce different results under different sets of conditions, without rhyme or reason. Someone who takes a drug and would refrain but only when he had stood upon a blue carpet within the last twelve hours, and under no other circumstances, would hardly be an example of someone who was morally responsible for his addictive behavior. Thus the requirement that the pattern of responses be appropriate: it must in some way or other be an intelligible pattern.

And finally what about the possibility of a Frankfurt-style "evil manipulator" who actually electronically manipulates our decision processes from without so that we perform a desired piece of behavior? If that is all that he does, then the agent is not responsible for the behavior for a number of reasons. For one thing, the operating mechanism—reasoning manipulated from without—is not responsive to reasons: it would give the same result under all conditions. For that reason alone the agent is not morally responsible. But it is also the case that the mechanism producing the behavior does not belong to the agent; it is external. This fact also rules out the case in which the evil manipulator implants a mechanism that gives the right result under current conditions but is capable of giving different results under different conditions: although the mechanism is properly responsive to reasons, it is not the agent's own. What would it take to make it the agent's own? It must be a mechanism that he has taken responsibility for in the past; and he has taken no responsibility for this implanted mechanism. (All of this presupposes, of course, that we can distinguish two mechanisms. The idea here is that the two may be identical in present features; what distinguishes them is that the agent took responsibility for one in the past, but not the other. But to make that work we must be able to maintain that they are two mechanisms and not one.)

This hardly does justice to the theory, but it is enough for our purposes. The great beauty of this effort, it seems to me, is that it reflects all of the difficulties that face compatibilism; its structure mirrors the history of the debate. There is no simpler theory, at least not at this point in time, that will settle the question of moral responsibility.

c. *Guidance Control and Possible Worlds*

I want now to carve out of this theory as much as will be useful for our discussion of addiction. One helpful way of talking about mechanisms and the different sets of conditions under which they operate is to talk about possible worlds. Remember that Frankfurt argued that alternative possibilities were not necessary conditions for responsibility. Alternative possibilities, in possible worlds talk, are possible worlds *accessible* from the actual world. To be accessible from this world, a possible world must be identical with it up to some point in time, and then follow a different history. If a person is faced with a genuine choice, so that in the circumstances it is up to him which of two things happens, we might describe that as saying that there are two possible worlds, both accessible from this one, and which one becomes actual depends upon that person's choice. Now if determinism is true, then there are no other possible worlds accessible from this one, because at each point in time what happens at the next moment is determined; there are no genuine choices of the sort that I have described. So any possible worlds model that hopes to make responsibility consistent with causation must do without accessibility.

But alternative possibilities, the sort of alternatives that Frankfurt showed were not necessary for responsibility, do require that sort of identity of worlds. The notion of freedom and responsibility that Frankfurt was out to attack required that the agent be capable of doing otherwise under precisely the same conditions, a requirement that is inconsistent with causation. And the point of the Frankfurt examples was to show that responsibility does not require alternatives of that sort. But not every action is free; there must be something that distinguishes free, responsible actions from others.[74] And what Fischer and Ravizza have attempted to give us is a stand-in for alternative possibilities.

The Fischer-Ravizza possible worlds are logically possible worlds that have the same natural laws that exist in this world. But they need not be *accessible* from this world in the sense described: It need not be the case that they are identical with this world up to the point in history at which the agent whose responsibility we are questioning chooses to perform the action. If our evaluation of responsibility required us to look only at

74. One possibility, not examined here, is that one of the conditions of free choice and responsibility is that the choice be uncaused. In the Frankfurt examples there is a striking difference between our perception of the agent's choice when he acted on his own, and the agent's choice had the manipulator stepped in to cause it, even though in both cases the agent could not have done otherwise. Some writers think the intuitive difference is the persistent belief that the first choice is uncaused, while the second is caused.

worlds identical to this one up to the moment of choice, then the assumption of universal causation would entail that there was no other world in which the agent did otherwise: an identical history entails an identical result, if determinism is true.

Let me labor this a bit longer, because I think that it is important to the discussion that follows. There are two ways to picture control in terms of possible worlds. The first is to imagine worlds branching off from this one at each point at which an individual is faced with a genuine choice. Each branching world contains one of the alternatives he is choosing among, and which of those worlds becomes actual is up to him. It is a necessary part of this picture that all the branching worlds are identical with this one up until the moment of choice. Under those conditions the actor has a kind of control—regulative control, in the Fischer-Ravizza lexicon. It is this picture of control that compatibilism cannot deal with: it involves the idea of a choice that might go either way in one and the same set of circumstances, and that is inconsistent with universal causation. If this is what control is, then the causal theory is true.

The second way to picture control in terms of possible worlds is to allow worlds that are not identical to the actual one for *any* period of history—to put it another way, worlds that are never accessible from this one—to count in the evaluation of control. If we could do that, we could account for control in the Frankfurt examples. In those examples there is no world *accessible from this one* in which the agent does otherwise. That rules out regulative control, as described above. But if it is enough that the agent does otherwise in worlds *not* accessible from this one, then that might explain why the agents do have control in the Frankfurt examples—what Fischer and Ravizza call guidance control. Of course those other possible worlds cannot simply be any logically possible worlds at all; and that is the function of the notion of a mechanism. The only worlds that come into play in evaluating responsibility are those in which the same mechanism is operating—though under other sets of conditions in those other worlds, with different reasons for acting present. This second picture is compatible with universal causation. (Of course, to generate other possible worlds with the same causal laws, we must assume that each world has a different starting set of conditions.)

Now might we not describe two different senses of "the difficulty of choosing otherwise," making use of these two pictures of control? In one sense (call it the regulative sense) what it means to say that it would

have been difficult for someone to choose otherwise means, first of all, that she could have chosen otherwise in the given circumstances; her choice was not determined by the circumstances (including her own makeup). Second, it means that to choose otherwise than she did choose would take a great deal, perhaps an unreasonable amount, of effort. It is the choosing itself that is hard, and not the action chosen. The action might be easy to perform if she could but choose it; and the actions and its consequences might even be what she would prefer to make happen. But she finds it difficult to initiate that action.

To capture the second sense (call it the guidance sense) we must picture an array of possible worlds, in some of which the agent makes the choice she makes in this world, and in some of which she makes a different choice. In none of those worlds is any of the other possible worlds accessible: in none of them could the agent have done otherwise under the circumstances in which she finds herself in that possible world. To capture the difference between ease of doing otherwise and difficulty of doing otherwise we must count the worlds in which the agent in fact does do otherwise. If she does otherwise in many of the other possible worlds—that is to say, she responds otherwise when faced with any of a number of different reasons for doing otherwise—we will say that it would have been easy for her to do otherwise. But if she does otherwise in few of the other possible worlds, we will say that it would have been difficult for her to do otherwise. (That is exactly the same outcome we should expect in the regulative picture as well, except that in that picture all of the relevant worlds are accessible from the actual world.)

d. Guidance Control and Addiction

Here, then, is a sketch of a compatibilist notion of difficulty of doing otherwise that might make sense of accepting DIFF, the principle distinguishing levels of responsibility and excuse according to the difficulty of doing otherwise, while rejecting CE, the causal theory of excuse. To bring all this back to the example of addiction: on this understanding, the addict who is less responsive to reasons for refusing to take his drug of choice is the more seriously addicted, and the less responsible for what he does. The less seriously addicted person is more responsive to reasons, and accordingly more responsible for what he does. I want now to explore this proposal further. But before I go on I want to point out that the proposal simply builds upon the theory of Fischer and Ravizza, and should not be attributed to them. To the extent that Fischer and Ravizza talk about addiction at all, they talk only about cases involving literally irresistible urges, which I have argued is not

true of normal addictions.[75] To the extent they talk about difficulty of doing otherwise (as far as I can see) they talk about it in terms of duress; and by duress they appear to mean calm reasoned choice between aversive alternatives, choice for which the agent is responsible but which might be found not to be blameworthy on grounds, for example, of justification.[76] I have argued that addiction is not like that either.

The first question is whether a theory of this sort, supplemented as I have suggested, will permit us to say that any addict, at any level of addiction, bears any degree of responsibility for what he does. The problem is this: the theory requires us to distinguish mechanisms that belong to the agent from mechanisms that do not, as a way of explaining the difference between the person who acts on his own and the person whose action is directed by an outside manipulator. In both cases, according to the theory, the action and the choice that leads to it may be caused; but in one case it is caused through the agent's own mechanism, and in that case the agent is responsible. In the other case, the case in which the agent's choice is actually manipulated from without, it is caused by a mechanism not the agent's own, and so the agent is not responsible. But what should we say about addictive behavior? If the mechanism responsible for addictive behavior is not the agent's own, then the agent is not responsible for it, no matter how light the addiction might be. And that simply cannot be the case. Addiction is generally supposed to proceed through stages, and in the early stages it is relatively easy to defeat. And if we want to hold the addict responsible at least in those cases, we must say that the mechanism resulting in his drug-taking behavior is the agent's own, and is not like the mechanism created by the evil manipulator who deprives the agent of control.

But we want a different result when the agent is more heavily addicted. Perhaps we could draw a line between the lighter cases and the more serious cases, and argue that in the former the mechanism of addiction is the agent's own, but in the latter cases not. But that would mean that there were two possibilities only: either the agent is responsible, when the addiction is light, or he is entirely deprived of responsibility, when the addiction is heavier. That would make the heavily addicted person an automaton, a sleepwalker, a victim of the evil manipulator. And that is a possibility that we have rejected from the outset. In order to make our proposal work, we have to assume that the

75. *See, e.g.*, FISCHER & RAVIZZA, *supra* note 1, at 35, 48.
76. *See, e.g.*, *id.* at 83.

mechanism of addiction, light or serious, is the addict's own, and that gradations occur in the responsiveness of the mechanism to reasons.

But that resolution is not entirely satisfactory, either; for it is hard to see why the mechanism of addiction should be seen as belonging to the addict. The picture of addiction we have drawn requires a mechanism that yields a certain result under one set of conditions and another under others; that has to do with its responsiveness. That mechanism is created through the influence of addictive substances. If our evil manipulator were to create the same sort of mechanism in the agent, one that caused him to react one way under one set of conditions and another way under other conditions, it would be irrelevant that the mechanism is responsive to reasons: the agent would not be responsible because the mechanism would not be the agent's own.[77] Why is addiction any different? It is important to understand that this problem does not arise only for the Fischer-Ravizza model, but is a problem generally for compatibilist theories. Everyone would, or should, concede that an agent is not responsible for action caused by an evil manipulator. And that remains true if the way in which the manipulator causes the action is by creating in the agent a mechanism that will produce the desired result under these conditions, but would produce different actions under different conditions, where the reasons for acting were different. That is, nothing would change if the manipulator created a reasons-responsive mechanism, which caused the agent to behave differently in different possible worlds. We would still not hold the agent responsible for what he did. But addiction acts in very much that way, on the assumption that universal causation is true. Addiction creates a mechanism that causes the addict to take the drug under these circumstances, and not to take the drug under other circumstances in which the reasons to avoid the drug might be stronger. And that is true of mild addiction and strong addiction alike, even if the numbers of worlds in which the agent behaves differently are different. So if the agent is not responsible for behavior caused by the manipulator, why should he be responsible for behavior caused by an addiction, mild or strong? I do not know of a satisfactory answer.

Let's assume, however, that there is an answer, and let us suppose that the addictive mechanism, unlike the mechanism created by the evil manipulator, is the addict's own. That will permit us to talk about degrees of responsibility based upon the degree of responsiveness to reasons. I want now to suppose that there is a person who is addicted to a certain drug and who would, in the face of certain reasons for refusing to take the drug, take the drug anyway, while in the face of certain other

77. *See, e.g., id.* at 232-35.

reasons he would in fact refuse to take the drug. Suppose that there are just ten reasons in all, and that they can be labeled according to their strength from one to ten, with ten the strongest reason to avoid the drug. Our agent consumes the drug unless he is confronted with reason number ten, in which case he abstains. Suppose, finally, that when he is confronted with reasons two through nine his choosing to abstain would carry differing degrees of difficulty, but that if he were confronted only with reason number one, the weakest reason for abstaining, it would be impossible for him to abstain. Now suppose that he is in fact only confronted with reason one, and consequently cannot do otherwise but to consume the drug. If it is impossible for him in those circumstances to choose otherwise, then he is not responsible for what he does, even though the mechanism is his and is reasons-responsive.

Now it is true that this presumes that it makes sense to talk of the difficulty, in each individual possible world, of doing otherwise, and it also presumes that the difficulty might be different in different possible worlds. But why should that not be possible? Doesn't it make perfectly good sense to argue that the difficulty the addict has when confronted with one scenario and set of reasons might be vastly different from the difficulty he has when confronted with another scenario? And setting to one side the case in which the addict finds it impossible to do otherwise, consider that under one set of conditions the addict might find it terribly hard to do otherwise (so that we would be inclined to hold him less responsible under those conditions), while under another set of conditions he might find it quite easy to do otherwise. The regulative version of the possible worlds picture of the difficulty of doing otherwise makes sense of all this: What matters in judging responsibility for an action is the difficulty of avoiding it under the actual circumstances in which the agent operated, and not under other circumstances and other reasons. But the regulative picture, as we know, is not compatible with causation. Can we salvage the guidance picture of difficult choice?

We could argue, I suppose, that the mechanisms involved in the different worlds were not identical: the mechanism which acted upon reason one, in the absence of other reasons, was one mechanism, while the mechanism that acted upon the other reasons was a different mechanism. Then we can say that the agent was not responsible in the case in which he was confronted only by reason one, and responsible in the other situations. Similarly, where the choice was difficult there

might be one mechanism, and a different one where the choice was easy. But we have already considered that answer and rejected it; it pushes the notion of mechanism identity pretty hard, and I doubt whether it could be made to work.

The other answer would be to accept the conclusion that the guidance picture entails: Even in world one, we might conclude, the agent is responsible for taking the drug; just how responsible he is would depend upon how many other worlds he succumbed to temptation in. And he would bear exactly that level of responsibility in all the possible worlds, because in all of them the total of possible worlds in which he consumed the drug and of possible worlds in which he abstained would be the same. We would reject the idea of relativizing levels of difficulty to particular sets of conditions and particular reasons; the only thing it would mean to say that an action was more or less difficult to avoid in a certain situation would be to say something about how the agent would act under other circumstances. Here again the outcome is not entirely satisfactory. It is hard to avoid the conviction that there are some circumstances that make choice difficult and others that make choice easy. It may be that when an addict confronts a certain complex of reasons he might find it so difficult to avoid addictive behavior that his responsibility ought to be mitigated, while confronted with a different complex of reasons the choice to do otherwise might be easy. And the difference between them might have nothing to do with the number of reasons that would cause him to do otherwise or the number of possible worlds in which he does otherwise.

So it seems to me that the possible worlds interpretation of the difficulty of doing otherwise does not do a very good job of capturing that notion. I don't say that it can't be done, but for now it hasn't been done, and it is not easy to see how it might be done. What we are after is a notion of degrees of control, and if that notion cannot be constructed out of responsiveness to reasons it is hard to imagine what materials it might be constructed out of. So far, then, there is no reason to believe that a compatibilist explanation for the defect of will interpretation of addiction is available. Clearly this is not the end of the story; such an explanation may be forthcoming. But pending the development of such a theory we must acknowledge that this fourth class of theories, though it does in fact seem to capture our sense of the addict's difficulty, does not satisfy the second requirement of our quest, that it be shown to be compatible with universal causation.

III. CONCLUSION

Thus, in the end, it may be impossible to square our sense that the addict is not fully responsible with compatibilist theories of excuse or with a rejection of the causal theory. If that is so, we have, I think, only three choices. First, we may reject the effort to construct a compatibilist theory of excuses and accept instead a libertarian view of responsibility and addiction. That is, we would try to save such excuses as addiction by denying the truth of universal causation. Or, second (and this seems to me the most inviting possibility), we may disengage legal responsibility from moral responsibility, constructing our theory of excuses upon a more clearly consequentialist basis. That is my own preference. But I admit that from the argument I have given a third, less disruptive possibility follows, that of denying that addiction should be an excuse at all. For one thing, addiction is the only condition that raises the precise problems discussed here; the consequences of rejecting it would be narrow. For another, the law does not presently grant an excuse for addiction, and seems unlikely to change; the reasoning here can undergird the law's position. The only problem with this outcome is that it conflicts with the abiding sense that addiction should excuse or mitigate, and the feeling that general propositions about responsibility should be based upon such abiding convictions, rather than the other way around.