ADDICTION, RESPONSIBILITY, AND NEUROSCIENCE

Michael S. Moore[*]

This Article examines the basic issue of whether addiction is a moral excuse for those otherwise wrongful behaviors done by addicts. Addiction is currently not a legal defense in Anglo-American criminal law, so this moral issue is important because if addiction is a moral excuse then it should provide such a legal defense. Answering the basic issue is pursued in four steps. First, the question is raised as to how addiction should be defined. The definition of addiction used in medicine is taken as a starting point, although the so-called "disease model of addiction" is rejected because it mistakenly attempts to build attributes of excuse into the definition of addiction. Second, a complex mix of psychological explanations of the puzzling behaviors of addicts is examined, the common conclusion being that addicts' acts of use and acquisition of drugs is irrational in a variety of senses of that word. Third, each of the explanations of the behaviors of addicts is probed as to its potential of providing moral excuse for such behaviors. Generally, no such excuse is discovered, although occasionally the irrationalities of addictive behavior does provide partial mitigation from responsibility. In reaching this generally negative conclusion about excuse, no reliance is placed on the responsibility of addicts for becoming addicts in the first place; while addicts have such responsibility, it is no substitute for the responsibility for wrongful acts done as addicts. Fourth, two generations of neuroscience research into addiction is examined for its potential to alter the prior conclusions about how addiction should be conceptualized, explained, or evaluated. Although the first generation of such research—that cast in terms of opioid drugs hijacking the pleasure center of the brain—had the potential to enlarge the excusing potential of addiction,

[*] Charles R. Walgreen University Chair, Center for Advanced Study Professor of Law and Philosophy, Co-Director of the Program in Law and Philosophy, University of Illinois. This Article grew out of the University-wide seminar on addiction, "Addiction and the Law: Perspectives from Philosophy, Economics, and Neuroscience," that I co-taught AY 2016–17 at Yale University. My thanks go to my fellow teachers of the seminar, Hedy Kober, Alan Schwartz, and Gideon Yaffe, for their many insights and suggestions about addiction, as well as for the kind of good colleagueship that gives that otherwise solitary form of activity known as scholarship, the social face that increases its enjoyment. Thanks are also due to Stephen Morse, Douglas Husak, and Richard Holton who lent their energies to helping make the Seminar as educational for me as it was. This Article was presented to the Centre for Agency, Values, and Ethics of Macquarie University, Sydney, Australia, in March 2019, and my thanks go to the participants at the Centre Seminar for their many helpful comments. Lastly, separate thanks to Douglas Husak and to Gideon Yaffe for having read and critiqued on earlier drafts of this Article.

that research was not confirmed in its essential premises. The jury is still out on the potential of the second generation of such research to deepen the excusing potential of addiction.

TABLE OF CONTENTS

I. INTRODUCTION

Addiction to substances generally, and to opioid drugs specifically, is a major problem in the United States. Indeed, it seems to be getting to be a worse problem with every passing year. We have more than two million opioid addicts in our population, of whom approximately 42,000 die of overdosing on opioids

every year. By the U.S. Surgeon General's estimate, substance addiction generally costs us $442 billion annually in health care costs, criminal justice costs, and lost productivity; that figure is $93 billion for drug addiction alone (the rest due to alcohol addiction).[1]

Most of the current attention to the addiction crisis is, rightly enough, focused on methods of prevention and cure. Thus, the National Institute of Health, for example, in April 2018 announced its HEAL ("Helping to End Addiction Long-term") initiative.[2] My interest in this Article, however, is not with these much-discussed issues of prevention and cure. Rather, I focus on the relation of addiction to the criminal justice system. My question is whether addicts who commit crimes deserve to be punished for those crimes, or whether instead they should be fully or partly excused whenever their crimes were the result of their addiction. In short, my question is whether addiction is an excuse, both for moral responsibility and from liability to criminal punishment.

I thus seek to assay the responsibility of addicts for the acts they do because they are addicts. This, too, like the issues of prevention and cure of addiction, is an important contemporary topic. A recent study found that 65% of the jail and prison population of the United States—some 1.5 million inmates—meet the criteria for being diagnosed as suffering from a "substance use disorder" (the American Psychiatric Association's label for substance abuse, severe forms of which are intended to refer to addiction).[3] While most of such incarcerated addicts are not imprisoned for drug-related crimes, surely many are. If addiction should be a legal defense or mitigating factor because it is a moral excuse, then a very large number of those we have imprisoned are being unjustly punished.

Addiction as such is not a defense to crime in any U.S. state or federal jurisdiction.[4] The few jurisdictions that have considered a defense of addiction as a matter of common law have refused to countenance such a defense.[5] At most,

1. Katharine Seelye, *Few Drug Addicts Are Treated, U.S. Finds*, N.Y. TIMES, Nov. 18, 2016, at A14. More recent calculations put these annual figures at 47,600 opioid related deaths and $1 trillion in lost output. Lydialyle Gibson, *The Opioids Emergency: Medicine's Response to America's Largest Public Health Crisis*, HARV. MAG., Mar.–Apr. 2019, 36, at 36–43.

2. *NIH Heal Initiative*, *Research & Training*, NAT'L INSTS. HEALTH, https://www.nih.gov/research-training/medical-research-initiatives/heal-initiative (last visited Jan. 20, 2019).

3. NAT'L CTR. ON ADDICTION AND SUBSTANCE ABUSE AT COLUMBIA UNIV., BEHIND BARS II: SUBSTANCE ABUSE AND AMERICA'S PRISON POPULATION 3 (2010).

4. To my knowledge addiction is not a defense in any Western criminal code. This may be a bit of a surprise in countries such as Portugal, the Netherlands, and Canada, in light of the fact that the legal systems of each have in one context or another regarded addiction as a disease that in its origins and its symptoms is not the fault of the addict. Portugal and the Netherlands have rendered the issue of defense largely moot because by formal decriminalization (Portugal) or informally by nonenforcement (the Netherlands) all drug users, nonaddicts and addicts alike, commit no prosecutable crime when they use drugs, so the issue of defense does not arise for behavior that would be a crime elsewhere. Canada has interpreted its human rights laws against disability-based discrimination so as to prohibit loss of job or housing because one is an addict. Stewart v. Elk Valley Coal Corp., [2017] 1 S.C.R. 591(Can.). Despite this, addiction is not (yet) regarded as a defense to any crime in Canada.

5. The leading case here is *United States v. Moore*, 486 F.2d 1139 (D.C. Cir. 1963), where the District of Columbia Court of Appeals rejected any addiction-based defense to possession of a controlled substance. The case is notable for its two dissents arguing vigorously for the existence of a defense in these circumstances. The very recent decision of the Supreme Judicial Court of Massachusetts in *Commonwealth v. Eldred,* decided July 16, 2018, is in line with *Moore*. *See* 101 N.E.3d 911. The issue in *Eldred* was whether Ms. Eldred's parole could

addiction poses the theoretical possibility of being raised via the insanity defense. Such possibility exists only in those jurisdictions: (1) retaining an insanity defense at all; (2) having a "loss of control" (or "volitional") prong to that defense; and (3) considering addiction to be a legally cognizable mental illness for purposes of the insanity defense. There are almost no instances where jurisdictions in the United States meet these three requirements, particularly the last.[6]

Despite the unavailability of any defense of addiction on a matter of ordinary (statutory and common) law, in 1968 the U.S. Supreme Court came very close to construing the U.S. Constitution to require that addiction be a defense for those accused of using drugs or alcohol to which they were addicted. In *Powell v. United States*, five members of the Court were prepared to hold that those addicted to alcohol could not constitutionally be punished for such addicted use; addiction, it was thought, compels use of that to which one is addicted, and compelled behavior is excused behavior that could not constitutionally be punished.[7] Only Justice White's fine distinction—that although use of alcohol is compelled for alcoholics, *appearing in public* drunk was not compelled—saved the Court from holding addiction is required by the Constitution to be an excuse.[8] Despite these five votes for addiction requiring a constitutional excuse for use, *Powell* has subsequently been interpreted by the lower courts *not* to constitutionally require a compulsion excuse for addicts who use; rather, the only constitutional prohibition is one prohibiting states from conviction of addicts for the *status* of being addicts.[9]

Thus, as a matter of statutory, common, and constitutional law, addiction does not presently serve as any kind of defense in Anglo-American criminal law. But that doesn't answer the normative question of whether the law is not mistaken in this regard. Does doing some wrongful and illegal act in order to satisfy a desperate craving for the drug to which one is addicted, reduce or eliminate one's moral blameworthiness for doing that act? If so, then by standard theories of punishment there should be some legal defense.[10] It is to that moral question that this Article is devoted.

There are two clarifications of this moral question that help to sharpen its focus. One is to stipulate away concerns about whether the acts for which drug

be revoked for use of controlled substances to which she was addicted and thus (she claimed) she was compelled to use; the court held that use of drugs by those addicted to them is not necessarily so compelled (or otherwise not a matter of responsible choice) as to make revocation of parole for such use unfair or impermissible. The *Eldred* court was aided in its decision by an amicus brief signed by myself and many of those thanked in the swordnote to this Article.

6. *See* Stephen Morse, *Addiction, Choice, and Criminal Law*, *in* ADDICTION AND CHOICE: RETHINKING THE RELATIONSHIP 426 (Nick Heather & Gabriel Segal, eds., 2017).

7. Powell v. Texas, 392 U.S. 514, 516 (1968) (plurality opinion).

8. *Id.* at 548–49 (White, J., concurring).

9. *See, e.g.*, *Fisher v. Coleman*, 639 F.2d 191, 192 (4th Cir. 1981); *United States v. Stenson*, 475 F. App's 630, 6531 (7th Cir. 2012); *Joel v. City of Orlando*, 232 F.2d 1353, 1362 (11th Cir. 2000). The only exception to this restrictive interpretation of *Powell* appears to be the Fourth Circuit's recent decision in *Manning v. Caldwell*, 914 F.3d 229 (4th Cir. 2019).

10. *See* Jeanette Kennett, *Why Shouldn't Addiction Be a Defence to Low-Level Crime?*, THE CONVERSATION (June 11, 2014), https://theconversation.com/why-shouldnt-addiction-be-a-defence-to-low-level-crime-27520.

addicts are most frequently prosecuted—acts of possession and use of drugs—should be criminalized in any criminal code properly respecting of liberty. One wants to stipulate away this criminalization question because otherwise one's libertarian intuitions about the lack of any properly prohibited *offense* can color one's judgments about whether addiction should be a *defense*. To raise squarely the issue of defense, it helps to focus less on acts of use and possession and more on those morally wrongful acts of acquisition of drugs that are legally prohibitable by the criminal law, on anyone's view of these matters. In a recent Canadian case, for example, a nurse stole the drugs of the patients in her care in order to feed her own addiction and she claimed in her defense that her addiction compelled her to steal.[11] Her acts of theft were plainly wrong and plainly criminal, raising squarely the issue of whether she should nonetheless have a defense because of her addiction. This is the moral issue raised by this Article that is central to the issue of legal defense.

The second clarification has to do with the kind of addictions that might raise such a defense. To keep the topic manageable, I have restricted my focus to drug, alcohol, and tobacco addiction, the so-called "substance addictions." And usually I focus on the central one of these, addiction to the use of opioid drugs. The concept of addiction has of course been much more widely employed than that, having been extended to cover what are called "behavioral addictions" such as the addiction to gambling, to sex, to eating, and the like. What is said about the substance addictions can sometimes also be said about the behavioral addictions, but there are also enough differences that in this Article I seek only to deal with what are paradigmatically addictions, leaving to others the degree to which one can analogize the less central cases of addiction to the central case.

II. CONCEPTUALIZING WHAT ADDICTIONS ARE

It is a common injunction to those who write about anything that they should first "define their terms."[12] Heeding such injunction is thought to aid in clarity of one's own thoughts, to aid in successfully communicating those thoughts to others, and to avoid the talking past one another that occurs when communicants fail to talk about the same thing. So I began with some description of what it is we will be talking about.

Given the prevalence of the notion of disease in modern discussions of substance addiction (hereinafter, just referred to as, "addiction"), a place to start in conceptualizing what we will be talking about is with the medical profession's

11. *See* Joseph Brean, *Nurse Who Stole Opioids Wins Her Job Back Because Addiction is a Disease, Arbitrator Rules*, NAT'L POST (Toronto) (Jan. 18, 2019), https://nationalpost.com/news/nurse-who-stole-opioids-wins-her-job-back-because-addiction-is-a-disease-arbitrator-rules.

12. Most memorably, by the Vassar-trained daughter in the film, *Dolores Claiborne*, who admonishes her mother that clarity of thought (about her daughter's earlier "rough patch") demands that her mother first define her terms. DOLORES CLAIBORNE (Castle Rock Entertainment & Columbia Pictures 1995). The injunction is to be taken with a grain of salt. Definitions can aid both speaker and audience secure the reference of words like, "addiction," but such definitions should not themselves be thought to be analytically necessary criteria for the proper use of such terms. It is also and for the same reasons an oversimplification to think that one can entirely separate the definition of "addiction" from either the explanation or the evaluation of addiction.

definitions of "substance use disorders." Consider the latest, fifth edition of the American Psychiatric Association's *Diagnostic and Statistical Manual's* ("DSM-V") definition of the most relevant of substance use disorders, that which is called "Opioid Use Disorder."[13] A "severe" case of opioid use disorder ("severity" being intended to capture the notion of addiction) is when someone exhibits "at least six to seven" of the following eleven symptoms:

1. Taking the opioid in larger amounts and for longer than intended
2. Wanting to cut down or quit but not being able to do it
3. Spending a lot of time obtaining the opioids
4. Craving or a strong desire to use opioids
5. Repeatedly unable to carry out major obligations at work, school, or home due to opioid use
6. Continued use despite persistent or recurring social or interpersonal problems caused or made worse by opioid use
7. Stopping or reducing important social, occupational, or recreational activities due to opioid use
8. Recurrent use of opioids in physically hazardous situations
9. Consistent use of opioids despite acknowledgement of persistent or recurrent physical or psychological difficulties from using opioids
10. Tolerance as defined by either a need for markedly increased amounts to achieve intoxication or desired effect or markedly diminished effect with continued use of the same amount. (Does not apply for diminished effect when used appropriately under medical supervision)
11. Withdrawal manifesting as either characteristic syndrome or the substance is used to avoid withdrawal (Does not apply when used appropriately under medical supervision).[14]

There are a number of observations to be made about this definition. Stephen Morse has extensively reviewed this and similar medical definitions of addiction and has raised a number of relevant considerations.[15] First of all, notice that the definition is intentionally imprecise in the mode of combination of its eleven symptoms. These eleven conditions do not even purport to give *criteria* in the sense of necessary and sufficient conditions of correct application as for example the typical definition of "bachelor" (as (1) unmarried, (2) male, and (3) person) purports to do. Rather, the conditions constitute criteria for addiction only in Wittgenstein's looser, criteriological sense of "criteria": no single condition is necessary for a clump of conditions to be an addiction—in Wittgenstein's famous simile, this is like a piece of rope that is truly one piece of rope even

13. AM. PSYCHIATRIC ASS'N, DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS 541 (5th ed. 2013).

14. *Id.*

15. *See, e.g.*, Stephen Morse, *Hooked on Hype: Addiction and Responsibility,* 10 LAW & PHIL. 3 (2000).

though it is made up of many strands no one of which runs its entire length—and it is unclear whether any conjunctions of conditions are sufficient, short of the entire set.[16]

Second, as Morse notes, several of the conditions considered separately are degree-vague in their specification of the quantities required to satisfy them.[17] Third, Morse notes that such definitions lack a center of gravity, an essence, to addiction; each of these eleven conditions are treated as if they were equally important whereas in truth some conditions seem much more important than others.[18] Morse himself, for example, believes that "craving, a subjectively experienced strong desire, is (or almost always is) a central part of the condition . . . ,"[19] whereas some doctors believe that the essence of addiction lies in "compulsive drug seeking and use, even in the face of negative health and social consequences."[20]

Morse laments these three characteristics of medical definitions of addiction, and he is right to do so—if we were regarding the medical definition as a finished theory as to the nature of addiction.[21] Yet if we regard this definition from the self-consciously tentative and unfinished viewpoint of our current collective understanding of addiction, these three characteristics are but expressions of the provisional nature of our current theory. When we only partially know the nature of something, we would be pretending to know more than we do, and we would freeze inquiry so as to cut off learning more, by stipulating an artificially precise nature to addiction. We might in years past have analogously stipulated, for example, that any person whose heart and lungs have ceased spontaneous functioning, is dead—cutting off the insight that actually some of such persons (if they have been immersed in very cold water) are not really dead because death is *not* a state whose nature is fixed by the long-used heart/lung definition of death.[22]

So it is no defect in the medical conceptualization of addiction that it is provisional and thus somewhat vague about how its criteria are to be combined, about the quantitative variables in those criteria, or about the essential versus the accidental properties of addiction. The scientific nature of the enterprise of describing and explaining addiction cautions patience and thus tolerance of imprecision here.

Apart from these worries about the imprecision of the definition, there is also a worry about whether in law or in ethics we should attend to a definition issued by another profession such as medicine. This is after all a definition of a

16. LUDWIG WITTGENSTEIN, PHILOSOPHICAL INVESTIGATIONS 32e (G. E. M. Anscombe, trans. 1953) (1953). The differences between criteria in these two senses is explored by me in Michael Moore, *The Semantics of Judging*, 54 S. CAL. L. REV. 151, 173–75 (1981).

17. Morse, *Hooked on Hype*, *supra* note 15, at 13.

18. *Id*.

19. *Id*.

20. Alan Leshner, *Addiction Is a Brain Disease, and It Matters,* SCIENCE, Oct. 3, 2005, at 45, 46.

21. Morse, *Hooked on Hype*, *supra* note 15, at 8.

22. Death is a much-discussed example of the point being made in the text, in Michael S. Moore, *A Natural Law Theory of Interpretation*, 58 S. CAL. L. REV. 277, 293–328 (1985).

medical disorder for use by the *medical* profession in order to serve *medical* purposes. The general purpose guiding such a definition is the general aim of the medical profession: to diagnose, cure, and prevent those harmful and unwanted conditions of human beings knows as diseases.[23] Defining particular diseases such as "Opioid Use Disorder" serves these general medical purposes by isolating clinically distinct syndromes, syndromes whose distinctness requires distinct forms of explanation and thus distinct methods of prevention and cure. Whether this definition is a good definition for medical professionals to use depends on how well it serves these medical purposes.

My interests in this Article are not those of medicine. The three tasks undertaken by this Article are: (1) to describe addiction, (2) to explain the behavior of addicts, and (3) to evaluate whether the condition so described explains the wrongful behaviors in addicts in ways tending to excuse them from moral responsibility and legal liability. Those tasks have very little to do with the diagnosis, prevention, and cure guiding medical definitions of addiction, so one might reasonably wonder why I start with a medical definition.

If the medical definition just given were a purely stipulative definition—making a word mean whatever the speaker wants it to mean to serve his purposes in using that word—then we could ignore medical definitions of addiction in this context. Yet treating medical definitions of discrete diseases as stipulative definitions robs them of their ability to serve medical purposes. Medical nosology can serve the diagnostic, preventative, and curative purposes of medicine, only if that taxonomy of diseases describes a natural clumping of disease entities that exist independently of doctors so labelling them. Successful medical definitions capture, but do not create, the clumping together of symptoms into the disease entities that doctors can separately diagnose, explain, prevent, and cure.[24]

This gives moralists and lawyers some interest in medical conceptualizations of natural conditions like addiction that existed long before there were doctors or lawyers. For there is a shared search in all professions for both an accurate description of what addiction is and a true explanation of how addiction explains the drug related behaviors of addicts.[25] This overlapping interest in accurate description and true explanation persists in the face of the fact that medical professionals use such descriptions and explanations for different purposes than do moralists and lawyers. Excuse in morals and law depends upon accurate descriptions of, and true explanations by, what is truly excusing, just as successful prevention and cure in medicine depends on such accurate description and true explanation.

The overlapping concerns of law, medicine, and ethics to discover accurate descriptions and true explanations of conditions like addiction justifies us in

23. *See* Michael Moore, *Discussion of the Spitzer-Endicott and Klein Proposed Definitions of Mental Disorder (Illness)*, *in* CRITICAL ISSUES IN PSYCHIATRIC DIAGNOSIS 85 (Robert Spitzer & Donald Klein eds., 1978) [hereinafter *Spitzer-Endicott*]; Michael Moore, *The Quest for a Responsible Responsibility Test: Norwegian Insanity Law After Breivik*, 9 CRIM. L. & PHIL. 645, 645–47 (2015).

24. Moore*, supra* note 23, at 653.

25. Leshner, *supra* note 20, at 46.

starting with and taking seriously the earlier-quested medical definition of addiction in DSM-V. Such respect does not carry over, however, to the bit of professional aggrandizement done by the medical profession vis-à-vis the legal profession on the question of whether addiction *excuses*.

I refer to what is known as the medical profession's "disease model of addiction." That label has become the slogan for medicine's view that addiction is a brain disease and that because of this fact alone bad acts done by addicts because of their addiction cannot fairly be either blamed or punished.[26]

This would be a much shorter paper if this were true, for addiction is a brain disease—or at least there is nothing improper about the medical profession classifying addiction as a disease in light of the fact that it is an unwanted condition that is at least partially amenable to medical treatment.[27] But it is a serious mistake to infer from that premise, the conclusion that therefore addiction cannot be a moral wrong deserving of punishment, *i.e.*, a crime. As I see it, there are three routes by which this mistaken inference is drawn. The first and least thoughtful is to believe that "sick" and "bad" are exclusive categories, *i.e*., either contradictories or at least contraries: if one is sick, one is not bad; and if one is bad, one is not sick. Yet stated this boldly, this is silly: a murderer who has a cold—or pneumonia, cancer, or spinal meningitis for that matter—is still a murderer, *i.e.*, a culpable killer deserving of blame and punishment. Being sick and being bad are not, on their face, mutually exclusive categories.

The second route attempts to add some precision to the first route. Of course murderers can have colds while they kill and still remain bad people; but the point is that the conditions that make them bad cannot be the same conditions as those that make them sick, and it is in this sense that the categories of the sick and the bad are exclusive. The argument for this more precise conclusion begins with the observation that all illnesses involve incapacitation of some kind. This is made explicit in the overall definition of disease ("medical disorder") proposed by the then Chairman of the American Psychiatric Association's Committee on Nomenclature and Statistics for inclusion in the Diagnostic and Statistical Manual, Third Edition: being diseased is to suffer dysfunction, distress, disability, and disadvantage.[28] The argument then proceeds by observing that all forms of volitional excuse are based on the offender being in some sense *incapacitated* from doing better than he in fact did. Therefore, the argument concludes, to be properly classified as being diseased is to be excused.

26. *Id.*; *see also* Nora D. Volkow et al., *Neurobiologic Advances from the Brain Disease Model of Addiction*, 374 NEW ENG. J. MED. 363, 364 (2016).

27. Amenability to medical treatment is what I called the "jurisdictional" justification for classifying a condition as a "disease." Lawyers use the same justification for classifying a problem as a "legal problem," *i.e.*, a problem only those with professional legal training can resolve without being guilty of the unauthorized practice of law. *See Spitzer-Endicott*, *supra* note 23, at 87–89.

28. *See id.* at 15. As a consultant to Spitzer's Committee, I urged a narrowing of this overall definition of medical disorder. *Spitzer-Endicott*, *supra* note 23, at 89. Some of my suggested narrowings found their way into the overall definitions of medical disorder to be found in the third, fourth, and fifth editions of the Diagnostic and Statistical Manual. *See* Dan Stein et al., *What is a Mental/Psychiatric Disorder? From DSM-IV to DSM-V*, 40 PSYCHOL. MED., 1759, 1759–65 (2010).

The problem for this version of the inference (from disease to excuse) lies in the middle terms of that inference, disability and incapacitation. Put simply, the disability that makes for being diseased need not be the same as the incapacity that makes for excuse.[29] This obvious enough mistake is obscured by the fact that there *is* a large overlap between the disability that makes for disease and the incapacity that makes for excuse. It is this overlap that makes it absurd to blame and punish *all* conditions properly classified as diseases, for most diseases are not our fault. Samuel Butler caricatured this absurdity in his invention of a country, Erewhon (close to "Nowhere" spelled backwards), where all diseases are punished as offences against the state.[30] Butler describes an Erehwonian sentencing hearing for the offense of having pulmonary consumption:

> Prisoner at the bar, you have been accused of the great crime of laboring under pulmonary consumption, and . . . you have been found guilty . . . yours is no case for compassion: this is not your first offence . . . . You were convicted of aggravated bronchitis last year: and I find that though you are now only twenty-three years old you have been imprisoned on no less than fourteen occasions for illnesses of a more or less hateful character . . . . It is all very well for you to say that you came of unhealthy parents, and had a severe accident in your childhood which permanently undermined your constitution; excuses such as these are the ordinary refuge of the criminal . . . . There is no question of how you came to be wicked, but only this—namely, are you wicked or not? . . . You may say that it is not your fault . . . . I answer that whether your being in a consumption is your fault or no, it is a fault in you . . . . You may say that it is your misfortune to be criminal; I answer that it is your crime to be unfortunate.[31]

Butler is of course correct: it would be absurdly unjust to punish all diseases because most diseases are things that happen to us and are not conditions constituted by things we do. In the ancient etymology of "patient," most sufferers of diseases are passive not active, in both the bringing about of their condition and of the symptoms manifesting that condition.

But overlap is not co-extensiveness. Some diseases we do culpably cause to exist in ourselves (cigarette-caused lung cancer, *e.g.*), and the symptoms of some diseases are actions we do and not conditions we suffer under (using drugs, *e.g.*). We can thus be at fault for having such a disease and for manifesting its symptomatic behaviors. It would be absurdly unjust to punish people *because* their condition and behaviors are rightly classified as being a disease; but such absurdity does not infect the idea that we may punish people for their conditions and behaviors *despite* those conditions and behaviors being rightly considered to

29. The incapacities that make for excuse are two: first, that the actor could not have *acted* other than he did. And second, that the actor could not have *chosen* other than he did. *See infra* notes 36–45 and accompanying text. Not needed is a third incapacity, namely, that the actor could not have *desired* differently than he did. *See infra* notes 113–117 and accompanying. None of these incapacities is what is meant by the "disability" that defines disease.

30. SAMUEL BUTLER, EREWHON AND EREWHON REVISITED 88 (2d ed. 1927).

31. *Id.* at 106–10.

be diseases. Being diseased is not a reason to punish people but it is also not a reason not to punish them.

Unfortunately, it is easy to confuse these two relations between being diseased and justified punishment such that the absurdity of the one (punish because diseased) is thought to infect the other (punish despite being diseased). Consider these two statements from the opinions of the U.S. Supreme Court in *Robinson v. California*.[32] Justice Stewart wrote for the majority of the Court, holding that no one can constitutionally be punished for being a drug addict, on the rationale that being a drug addict was to be in a particular kind of status, that of being diseased:

> It is unlikely that any State at this moment in history would attempt to make it a criminal offense for a person to be mentally ill, or a leper, or to be afflicted with a venereal disease . . . a law which made a criminal offense of such a disease would doubtless be universally thought to be an infliction of cruel and unusual punishment . . . . Even one day in prison would be a cruel and unusual punishment for the 'crime' of having a common cold.[33]

Justice Douglas echoed Stewart's disease rationale in his concurrence in the same case:

> [T]he prosecution [of addiction] is aimed at penalizing an illness . . . . We would forget the teachings of the Eighth Amendment if we allowed sickness to be made a crime and permitted sick people to be punished for being sick. This age of enlightenment cannot tolerate such barbarous action.[34]

Stewart and Douglas get it wrong: while it would be unjust to punish someone just because doctors had properly classified his condition as a diseased one—Butler's point—it would not necessarily be wrong (and certainly not "barbarous") to punish someone for morally blameworthy acts even if those acts were causative of or symptomatic of a condition properly classified as a disease. The disease classification cannot rule out moral blameworthiness and legal punishment even though it is not itself a basis for such blameworthiness and punishment.

I come then to the third route by which the medical profession and its acolytes have sought to show how addiction being a disease *ipso facto* means that addicts are to be excused from responsibility. This route depends on a feature of the "disease model of addiction" that we have not yet addressed. This feature is the deeper, physical nature to addiction that scientists believe they have discovered in the human brain. Although we will detail these discoveries (and what they betoken about the folk psychological states that explain addictive behavior) later on in Part V of this Article, consider this early summary of the brain pathology that is thought to underlie the behavior and phenomenology of addiction to drugs:

> Although each drug that has been studied has some idiosyncratic mechanisms of action [within the brain], virtually all drugs of abuse have common effects, either directly or indirectly, on a single pathway deep within

32. 370 U.S. 660 (1962).

33. *Id.* at 666–67.

34. *Id.* at 678 (Douglas, J., concurring).

> the brain. This pathway, the mesolimbic reward system, extends from the ventral tegmentum to the nucleus accumbens, with projections to areas such as the limbic system and the orbitofrontal cortex. Activation of this system appears to be a common element in what keeps drug users taking drugs. This activity is not unique to any one drug; all addictive substances affect this circuit.[35]

Such seeking of some deeper, unifying nature to the natural kinds that disease entities have long been supposed to be, is an expected and legitimate part of science.[36] In our daily life we brush into surface indicators that things like water, gold, polio, or addiction, each might be a natural kind, and we expect science to investigate and reveal to us whether such surface indicators are or are not underlain by some deeper, unifying nature that marks a kind as a natural kind. Such natures can themselves be of different kinds, but for behavioral/phenomenological surface indicators like the symptoms of addiction, underlying states of the brain is a plausible place to look.

So there is nothing suspect about the disease model of addiction seeking a unifying nature for addiction in some pathological states of the brain. This is good science doing the work it is supposed to do. It is the second step of the inference from disease to excuse that is the mistake here. That step is taken when one assumes that any behavior that is physically caused is, by virtue of that causation, to be excused.[37]

There are two stunning problems for this bald assertion of an "incompatibilism" existing between physical causation of behavior and responsibility for that behavior. The first and main one is that the assertion is demonstrably false. For several hundred years—roughly since David Hume—philosophers have worked through a series of "compatibilisms."[38] Recently I surveyed ten of these compatibilisms for neuroscientists, who like other scientists seem woefully ignorant that such a literature even exists, let alone what its content might be.[39]

All forms of compatibilism share a denial of the supposed incompatibility of being caused to do what we do and being responsible for what we do. Of

35. Leshner, *supra* note 20, at 46. For an update of the same view, see Nora D. Volkow, George F. Koob & A. Thomas McLellan, *Neurobiologic Advances From the Brain Disease Model of Addiction*, 374 NEW ENG. J. MED. 363, 365 (2016).

36. Hilary Putnam used diseases as examples of natural kinds in his early papers on the famous "Kripke-Putnam" theory of direct reference to such natural kinds. *See* Hilary Putnam, *Dreaming and 'Depth Grammar'*, *in* ANALYTICAL PHILOSOPHY 211–35 (R.J. Butler ed., 1st series, photo reprt. 1966) (1962); Hilary Putnam, *Brains and Behavior*, *in* ANALYTICAL PHILOSOPHY 1, 19 (R.J. Butler ed., 2d series, 1965). For some doubts about whether the diseases of DSM-V really are natural kinds, *see* GEORGE GRAHAM, THE DISORDERED MIND: AN INTRODUCTION TO PHILOSOPHY OF MIND AND MENTAL ILLNESS 58–59 (2d ed. 2013).

37. This is what I dubbed long ago, "the causal theory of excuse" (Michael S. Moore, *Causation and the Excuses*, 73 CALIF. L. REV. 1091, 1091 (1985), *reprinted in* MICHAEL S. MOORE, PLACING BLAME: A GENERAL THEORY OF THE CRIMINAL LAW 486 (Oxford, 1997)), and that Stephen Morse has long called "the Fundamental Psycho-Legal Error" committed by scientists and others. *See* Stephen J. Morse, *Culpability and Control*, 142 U. PA. L. REV. 1587, 1660 (1994).

38. *See* Morse, *supra* note 37, at 1588–89.

39. Michael S. Moore, *Compatibilism(s) for Neuroscientists*, *in* 10 LAW AND THE PHILOSOPHY OF ACTION 1, 25 (Enrique Villanueva ed., 2014), *expanded and reprinted in* Michael S. Moore, *Stephen Morse on the Fundamental Psycho-Legal Error*, 10 CRIM. L. & PHIL. 43, 56 (2016).

course, such compatibilists owe us an account of what it means to say that some action is excused because "he couldn't have done other than he did."[40] If ability in this principle (usually called the "Principle of Alternative Possibilities") does not mean "uncaused," compatibilists need to tell us what it does mean. On my own version of compatibilism—known as "classical compatibilism," or, in its revised form, "new conditionalist compatibilism"—one *could* not have done other than they did on some occasion, if and only if that person *would* not have done other than they did even if he were presented with very strong reasons to do so.[41] The relevant question of excuse is whether a wrongdoer lacked the capacity to do better, in this counterfactual sense of "capacity"—not whether his wrongdoing was caused by events in his brain.[42] Although Leshner's "activation of the mesolimbic reward system" may figure in an account of why addicts are rightly to be excused (see Part V below), that will not be because such reward system activation is a physical cause of addicts' behavior. Rather, such activation of the reward system will have to be shown to cause an incapacitation in the relevant sense.

The second problem for the asserted incompatibilism of cause and responsibility, is that even if the asserted incompatibilism were true, it could not provide a theory of why addicts are excused from a responsibility that nonaddicts bear. For given the plausibility of there being physical causes in the brain for all that we choose and do, physical causation of addictive behavior can hardly be the basis for excusing addicts; rather it could only be the basis for excusing everybody from any responsibility for anything. Such a theory of *universal* excuse is not a theory of excuse at all; it is a theory why no one needs excuses because no one is responsible for anything anyway.

The only way of avoiding this unwelcome extension to universal excuse is by hedging one's bets about the physical causation of nonaddicts' behaviors. Thus, one might say that addicts' behaviors are "fully caused" by events deep in the brain, but that nonaddicts are only "partly caused" to do what they do, or that addicts are "mechanically caused" to take drugs, whereas nonaddicts are only "predisposingly caused" to do wrongful acts; or that brain-event causation forms a larger part of the set of factors sufficient to explain addicts' behaviors but that such brain-event causation plays a more minor role in the set of factors sufficient to explain nonaddicts' behaviors; etc.[43] Such strategies aim to dilute the kind, degree, or strength of the physical causes of nonaddicts' behaviors so that such nonaddicts can be held responsible, while the more strongly caused behaviors of

40. The exception is the so-called "source compatibilist," who deny that one must have had the ability to have done other than one did to be responsible for doing that thing.

41. For a much more nuanced account of the counterfactuals involved here, see Moore, *Compatibilism(s) for Neuroscientists*, *supra* note 39, at 28. Classical compatibilism stems from G.E. MOORE, ETHICS (1912), and much of what classical compatibilism now consists in are the ten or so amendments one must make to Moore to accommodate the century of criticism that has intervened.

42. Notice that one may have the capacity to do other than he did (in this counterfactual sense of "capacity") even though his action and the choice behind it were sufficiently caused by factors not under the actor's control.

43. I explore five such "partial libertarianisms" in Moore, *Causation and the Excuses*, *supra* note 37, at 1114–21; and in Moore, *The Quest for a Responsible Responsibility Test*, *supra* note 24, at 666–69.

addicts can then be excused without leading to universal excuse. One should say of all these maneuvers what Peter Strawson once said about one of them: the metaphysics such maneuvers require is "grotesque," and "obscure and panicky" in its motivation.[44] No respectable science of human behavior should find these maneuvers even slightly tempting.

The upshot is that the disease model of addiction's thesis—that the behavior of addicts is physically caused by events deep in the brain—is a valid piece of scientific theorizing that is nonetheless not determinative of responsibility. If addiction excuses (about which, much more in Part IV of this Article), it will not be because addiction is a physical cause of the behavior of addicts.

This allows us to put aside the shortcut offered up by the disease model of addiction to answering the concerns of this Article. We can also more generally put aside all three forms of the thought that medical classification of a condition as a disease carries any moral freight with regard to moral excuse. This frees us to look at medical definitions such as that with which we began in their proper light: as an attempt to pick out a natural phenomenon, addiction, by a profession with as much interest as that of the law in getting its descriptions and explanations of that phenomenon right. The stamp of the medical profession upon such a definition adds nothing to the authority that it has for us by virtue of its accuracy. Such a definition can nonetheless aid us in the task of providing some conceptualization of addiction with which to work. The definition can use considerable simplification and compression, however; Morse is right, there is no center of gravity to the definition.[45] Here I think that we can put both Morse and Leshner together, combining their views into the following idea about the essential features of addiction: addiction is indeed a state characterized by the phenomenology of craving for the thing to which one is addicted, as Morse holds; and such cravings characteristically bring with them an experience of being compelled by them, as Leshner holds, inasmuch as yielding to such cravings goes against other things the addict needs, desires, and values.[46]

Very little turns on whether this is a complete (or even accurate as far as it goes) account of the essence of addiction. Given the provisional nature of any definition of addiction (provisional before the insights of an advancing science), all we need are indicators that succeed in referring to the natural kind of phenomenon that is an addiction. We more completely explore the nature of that thing as we explain why it exists and how it works to produce its symptomatic behavior. It is to that explanatory task to which we now turn.

44. Peter Strawson, *Freedom and Resentment*, *in* 48 PROCEEDINGS OF THE BRITISH ACADEMY, 1, 25 (1962).

45. At least not explicitly; one might discern an implicit core to the first nine conditions as being compulsion to use drugs.

46. *See* Leshner, *supra* note 20, at 46; Morse, *Hooked on Hype*, *supra* note 15, at 19.

III. THE FOLK PSYCHOLOGICAL EXPLANATION(S) OF ADDICTION

With this conceptualization of addiction under our belt, we can now turn to the explanation of addiction. This I intend to explore in two parts: in this Part III, I shall outline what I take to be the folk psychological explanation(s) of addiction. These are explanations couched in terms of the familiar concepts of action, belief, desire, intention, emotion, character, and the like; the evidence for such explanations being true is drawn from phenomenology and behavior. In Part V, I will explore what recent academic psychology and neuroscience have learned that promises deeper explanations of addiction.

A. A Schema of Practical Rationality

The first folk psychological explanation of addiction that we will explore will be the rational choice model. On this view behavior by addicts differs in no interesting way from the behavior of other rational agents: addicts become addicts, and continue to engage in addictive behaviors as addicts, because this is what they see as their best option and they choose to act accordingly.[47] Alcoholics, Herbert Fingarette wrote years ago, are just people who like to drink a lot and do so for that reason.[48]

To flesh out the rational choice model of addiction requires a background understanding of the nature of rational action generally. Such a general understanding of practical rationality will in any event be necessary to understand other folk psychological explanations of addiction; for to pinpoint where addictive choices and behaviors differ from ordinary rational choice requires a prior understanding of rational choice itself.

A schematic way of presenting the ingredients of practical rationality is via the following charts of the aetiology of actions and their consequences, an aetiology that is divided into two charts only because of limitations of space.

47. Gary S. Becker & Kevin M. Murphy, *A Theory of Rational Addiction*, 96 J. POL. ECON. 675, 675 (1988).

48. HERBERT FINGARETTE, HEAVY DRINKING (1988). I reviewed Fingarette in Michael S. Moore, *Review of Fingarette's Heavy Drinking: The Myth of Alcoholism as a Disease*, 99 ETHICS 660 (1989).

CHART 1: THE PRE-HISTORY OF ACTIONS IN TERMS OF THEIR LONG TERM CAUSES

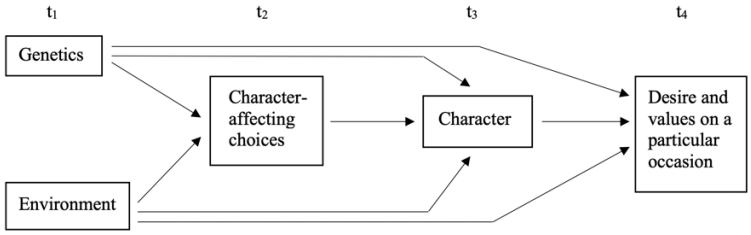
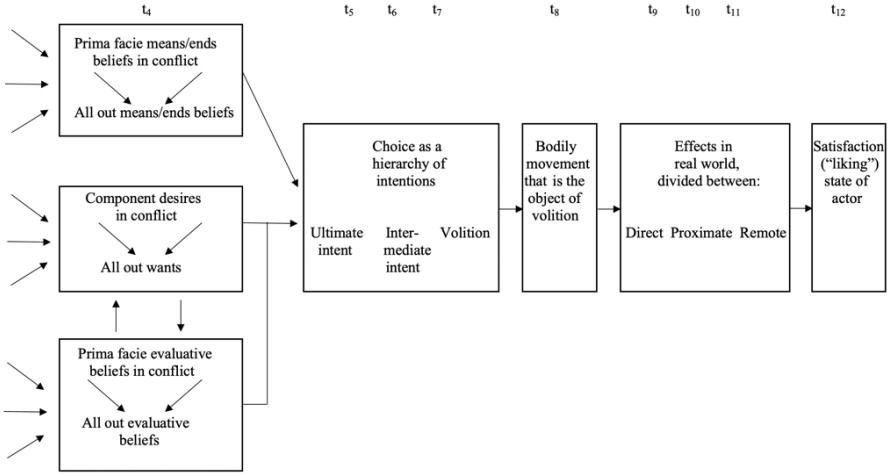


CHART 2: THE MORE IMMEDIATE ETIOLOGY OF RATIONAL ACTION



The "prehistory" of action in terms of its more remote causes is less relevant to the present purpose than is the immediate history, but because such prehistory does bear on some accounts of addictive behavior, I shall start there. A romantic, unrealistic view of such prehistory is that of the metaphysical libertarian. On this view, we choose our characters, either by large, existential choices where what kind of character we have is the object of such choices; and/or we form our characters through ordinary, first order choices about doing particular actions on particular occasions, and character accretes from such choices although it is not chosen as such.[49] In either case, the libertarian regards such character-forming choices as uncaused.

Such a romantic view is wrong on two counts. First, it is implausible in the extreme that such character-forming choices of either kind are uncaused by the genetic and environmental factors that precede them. On any plausible metaphysics, choices ultimately rest on (environmental and genetic) factors that are

49. The first is the view of Robert Kane (ROBERT KANE, THE SIGNIFICANCE OF FREE WILL 125 (1996)), who hopes that what he calls "self-forming willings" are the prima causa in his account of human actions. Arguably existentialists like Jean Paul Sartre had a somewhat similar view, thinking that we could "just choose" who we would be on some occasions. The second is closer to Aristotle's view of how our choices affect our characters.

themselves not the objects or products of choice. Thus, behind the character forming choices at t_2, Chart 1 shows at t_1 the environmental and genetic factors productive of such choices.

Secondly, it is also implausible that character-forming choices of the first kind have much causal power (if they have any). Self-made persons, in a noneconomic sense of that phrase, are a rarity if they exist at all. Even choices of the second kind are not plausibly viewed as the *exclusive* causes of character. That is why Chart 1 shows genetic and environmental factors at t_1 causing character at t_3 directly, without the mediation of character-forming choices at t_2.

Character itself at t_3 consists of those long-term, relatively stable traits that make us the kind of person that each of us is. Such traits include dispositions to act in certain ways, to feel certain emotions in certain situations, to be more or less gullible in our beliefs, to be subject to certain moods, to react to various situations with certain emotions, and as having a general preference ordering of our desires—along with much else.

When we believe, desire, or value things on particular occasions (t_4), such beliefs, desires, and evaluations are usually "in character" for us, in which case, our general character traits may plausibly be regarded as causing (in a certain sense) our more particular mental states on a given occasion. This is depicted in the causal arrow between character at t_3 and our desires/evaluations on a particular occasion at t_4. Yet sometimes we "act out of character" because of desires and evaluations that are themselves out of character for us; in addition, the so-called "situationists" in contemporary social psychology make some case for thinking that the occasions in which character is bypassed in the causation of intentional actions are more numerous than just occasional acts out of character.[50] In either event, our characters are not among the causes of the desires and evaluations that motivate such actions. Yet such deviant desires and evaluations cannot be utterly uncaused; presumably the total package of unchosen causes of behavior (which I see environmental and genetic factors as exhausting) cause such out-of-character desires and evaluations directly, without character serving as a causal intermediary. This is depicted in Chart 1 by the direct causal arrows between t_1 and t_4, bypassing character at t_3.

Turning to Chart 2, the chart on the immediate causes of action, the states depicted at t_4 need considerable elaboration from the black box used on Chart 1 at t_4 ("Desires and Values"). Let me do such elaboration via four complications of the simple "Desires and Values" of Chart 1. First complication: we need to distinguish desires from evaluations. To be sure, "desire" (and "wish," "want," etc.) can be used idiomatically to describe both desires (in a narrower sense I

50. *See, e.g.*, Gilbert Harman, *Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error*, *in* 99 PROCEEDINGS OF THE ARISTOTELIAN SOCIETY 315, 329-30 (1999) (reviewing this debate between virtue ethicists and situationist social psychologists). For the more recent summary of the situationism vs. character debate, see Hagop Sarkissian, *Minor Tweaks, Major Payoffs: The Problems and Promise of Situationism in Moral Philosophy*, 10 PHILOSOPHERS IMPRINT 1, 2 (2010).

shall henceforth employ) and evaluations. Indeed, Donald Davidson's "pro attitude" label was intended precisely to achieve this merger.[51] Yet we experience brute desires as being different than judgments of desirability. To see something as desirable is to judge it as worth attaining or promoting; to desire it simplicatur is to desire that state of affairs to obtain without judging it to be worthwhile. True enough, most of the desires we experience ourselves as having have desirability characteristics that we readily affirm if asked.[52] Yet (importantly for present purposes) that is not always the case, and even when it is the case, we experience the desiring itself separately from the judgment of desirability.

When we distinguish "values" or "evaluations" from desires in this narrow sense, we do not mean by the former terms to designate objective (as opposed to subjective) reasons for action. In explaining human behavior rather than justifying it, we operate entirely in the realm of subjective reasons. By "values," we thus do not here mean the moral properties of actions or states of affairs that give rational actors objective reasons for action. Rather, what is meant is subjective: within the psychology of the actor, what does he/she regard as objectively valuable? Let us henceforth call these "evaluative beliefs."

The second complication is necessitated by the fact that the relationships between desires and evaluative beliefs are complicated. There is, again, a romantic view of this that is so implausible that it needs to be mentioned only to put it aside. This is the view that truly rational agents make their evaluations of some state of affairs S, and that this evaluation then causes the desire for S to obtain.[53] Although this may be true sometimes, surely many of our desires are not the products of our evaluation (if anything, it is the other way around). Such desires may spring directly from our characters and not from any prior evaluations. (The romantic here worries that such a more complicated view betrays rational control—yet responsible agency (on the kind of compatibilist view of it that I have long defended) requires that we be able to choose to act in light of our desires, not that we be able to choose those desires themselves.)

The third complication at t_4 takes into account the fact that rational action requires more than that that action satisfy one's desires and one's evaluative beliefs, rational action also requires the agent to have certain factual and causal beliefs. More specifically, a rational agent chooses to do action A not only because she desires (and values) state S, but also because that agent believes the world to be such that doing A now will (or at least might) produce S.

Notice that unlike the relationships between what one desires and what one values, there is no causal relationship between what a rational actor believes, on the one hand, and what he desires or values, on the other. (There are thus no causal arrows between these two boxes on Chart 2.) A rational actor does not believe something because he wants something to be the case or because he values that it become the case—that is irrational, wishful thinking. Conversely, a

51. Donald Davidson, *Actions, Reasons, and Causes*, 60 J. PHIL. 685, 685 (1963).

52. *See* A.J. Watt, *The Intelligibility of Wants*, 81 MIND 553, 559 (1972).

53. Katie Steele & H. Orri Stefansson, *Decision Theory*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY (Dec. 16, 2015), https://plato.stanford.edu/entries/decision-theory.

rational actor does not desire something, or judge it to be desirable, because he believes that thing is attainable through her action—a rational actor does not regard possibility as a criterion of desirability. In short, factual beliefs are independent of our desires and our values in their rational validation.

The fourth complication at t_4 is that often the mental states in play here—desires, factual beliefs, evaluative beliefs—are themselves the products of componential versions of themselves. There are componential desires ("wishes"), prima facie ("tentative") factual beliefs, and prima facie evaluative beliefs (alternatively but equivalently, all-out beliefs in "prima facie goodness or rightness"). On many occasions we face conflicts in our componential desires, our prima facie beliefs, and our prima facie evaluations, and to act rationally is to resolve these conflicts and to act in light of what we "all-out" want, believe, or value.

I have depicted each of these four complications on Chart 2.

When we all-out value and all-out want some state of affairs S, we have not yet decided (chosen, intended) to bring S about through our action. That is the role of choice at t_5–t_7. Choice/intention is a different modality under which S is represented in the actor's psychology. S may be predictively believed as likely to occur, all-out valued and all-out desired that it occur, yet choosing or intending that S occur is a distinct and different state of mind.[54] In choosing to bring S about, the actor has decided that S will occur. A rational actor's desires, factual beliefs, and evaluative beliefs all have a hand in the production of that distinct mental state we think of as a *choice* or *intention*, but as Hume taught us, such causes cannot be identical to the thing they cause.

Unlike desires and beliefs, intentions cannot conflict in the mind of a rational agent. This is not to say that intentions cannot conflict as a matter of psychological fact; only that, if an actor has intentions that conflict, that is criticizably irrational even though psychologically possible. Chart 2 thus does not depict conflicts of intentions as it did depict conflicts of desire, belief, and evaluation. Still, as Michael Bratman has shown in detail, intentions do come in packages even if not in the conflict-ridden packages characteristic of beliefs, desires, and evaluations.[55] Rather, intentions come in hierarchically organized packages of means to ends fitting into an overall plan. This is because rational actors typically do not just intend, for example, to move their finger in order to inherit their uncle's wealth. Rather, they (all-out) want their uncle's wealth, and intend to acquire it; they believe that (because of the uncle's will) if they kill him they will inherit his wealth; they further believe: that if they put a bullet in his heart, the uncle will die, that if they shoot the gun pointing at him, that will put a bullet in the heart; that if they pull the trigger of the gun, that will shoot the gun; that if they move their finger which is on the trigger, they will pull the trigger. Bratman's point: they intend each of the stages they see as necessary to attain their

54. A central thesis in MICHAEL E. BRATMAN, INTENTION, PLANS, AND PRACTICAL REASON 111 (1987). The older view, identifying intention as a species of all-out want, is represented in DONALD DAVIDSON, *Intending, in* ESSAYS ON ACTIONS AND EVENTS 83, 83 (1980).

55. BRATMAN, *supra* note 54, at 127.

ultimate intention, which is to achieve what they most want.[56] As Kant said, to intend the end is to intend the means—all of them.[57]

In Chart 1, I have somewhat artificially divided this plan-constituting hierarchy of intentions into three stages: the ultimate intention at t_5 (which is to attain the object of what one most wants), intermediate means intentions at t_6, and the last intention at t_7 that executes one's plan, the intention to move one's body. This last intention I call a volition or a willing, although Bratman calls it an "endeavoring" to mark the fact that this is where tryings begin.[58]

Such a choice (*i.e.*, a hierarchy of intentions) causes the bodily movement that is the object of the most discrete of those intentions, the volition. Such volitionally-caused-bodily movements are *actions*, as I have defined the concept.[59]

Such bodily movements as are produced by volitions themselves produce real world effects. Such effects also have a structure, being organized into chains or cones of causal links. Some effects are closer than others (within such chain/cone structures) to the bodily movements of the actor which initiates the chain, and some are less proximate. I accordingly have divided such real world effects between direct effects at t_9, indirect but still "proximate" effects at t_{10}, and remote and freakish ("unproximate") effects at t_{11}.[60] When rational action is successful, among those effects will be state S, the most wanted, most valued, and intended state that motivated the actor to act on this occasion.

It did not take Freud to convince us that getting what we want—state S, say—does not always give us happiness, pleasure, or enjoyment. Being psychologically *satisfied* that S was achieved by one's action thus is a separate thing from the combination of wanting S, valuing S, and causing S to exist by one's action. With some abuse of the language, the future fact of whether or not some state S will satisfy the actor who intentionally produced it has come to be called a presence or absence of *liking* state S at the time the actor chooses S.[61]

Notice that there are two importantly different ways of not being satisfied at the occurrence of S because of one's action: most common is to find that S is accompanied by unwanted and negatively evaluated consequences that one did not anticipate when one acted. This kind of disappointment may involve some cognitive failure of the kind ordinarily called negligence (although it need not), but it does not draw into question one's desires and values in light of one's likes and dislikes. More problematic is the less common case where achieving state S by itself, unaccompanied by unforeseen consequences, gives no pleasure to the

56. *Id.* at 134.

57. *An end-in-itself,* BBC: ETHICS GUIDE, http://www.bbc.co.uk/ethics/introduction/endinitself.shtml (last visited, Jan. 20, 2019).

58. This is not to say that Bratman means by "endeavoring" exactly what I mean by "volition." For exploration of the differences, compare Michael E. Bratman, *Moore on Intention and Volition*, 142 U. PA. L. REV. 1705, 1708 (1994), *with* Michael S. Moore, *More on Act and Crime,* 142 U. PA. L. REV. 1749, 1821–22 (1994).

59. MICHAEL S. MOORE, ACT AND CRIME: THE PHILOSOPHY OF ACTION AND ITS IMPLICATION FOR CRIMINAL LAW 158–59 (1993).

60. *See generally* MICHAEL S. MOORE, CAUSATION AND RESPONSIBILITY: AN ESSAY IN LAW, MORALS, AND METAPHYSICS 391–425 (2009).

61. *See* Terry E. Robinson & Kent C. Berridge, *The Neural Basis of Drug Craving: An Incentive-Sensitization Theory of Addiction*, 18 BRAIN RES. REVS. 247, 247–91 (1993).

actor who wanted and valued S. If such divorce of liking from both wanting and valuing occurred very much with respect to states like S, a rational actor would have reason to re-examine his positive evaluations and his desiring of S.

B. Applying the Schema to Explain the Behavior of Addicts: The Rational Choice Model of Addiction

Recall that we seek to explain two kinds of behaviors by addicts: (1) at t_1, the using of drugs, alcohol, etc., while not addicted but which eventually cause one to become addicted; and (2) behavior at t_2 while addicted that is motivated by the addict's need/desire for drugs, alcohol, etc. (the latter category includes both possession and use of such items and behaviors such as theft done in order to fund acquisition of such items).

The most straightforward folk-psychological explanation of such behaviors is that of a fully rational action. That is, the potential or actual addict, although perhaps conflicted about it, most wants to take drugs over all the other things he also desires at a given time; values the experience of intoxication (perhaps as Aldous Huxley's "voyage of discovery for those with the courage to take it");[62] believes correctly that satisfying such a desire will rule out satisfying the other, conflicting desires that he also has; believes that if he steals the money in front of him, he can use drugs immediately; therefore chooses to use drugs as soon as he gets his hands on them, to steal the money now, and to move his arm now to reach for the cash; does move his arm now in response to that last, most discrete intention (volition); causes himself to be in possession of the money, to be in possession of the drugs, to use the drugs, and to upset those who care about him; despite the costs that he acknowledges his drug using causes, nonetheless feels as satisfied by his theft, possession, and use as he predicted he would feel when he decided upon this course of action. One might add that such an action, choice, desire, evaluation, and enjoyment ("liking") are all in character for this addict, who might also have what is called an "addictive personality." And we might further add that he has such a personality in part because that is the kind of person he has chosen to be.

Notice that there is nothing in the rational choice explanation of addictive behaviors that evaluates the addict's chosen action, choice, etc., as actually being desirable, morally permissible, or prudentially wise choices, actions, etc. What I mean by the rational choice explanation of addiction is thus not what economists such as Gary Becker appear to mean when they refer to the rational choice model of addiction.[63] My rational choice explanation takes no position on the normative correctness of an addict's choices and actions—these may well be the best a person in the addict's situation can get, or they may not. Rather, what is *rational* in the rational choice explanation is the way the addict's character forming choices,

62. Huxley's description to me when I was an undergraduate when Huxley had just returned from experiencing the hallucinogenic effects of Mexican mushrooms (LSD) for the first time.

63. Gary Becker's classic article is Gary S. Becker & Kevin M. Murphy, *A Theory of Rational Addiction*, 96 J. POL. ECON. 675, 675–76 (1988). Most of what I understand about Becker's theory I have learned from Alan Schwartz. *See* Alan Schwartz, *Views of Addiction and the Duty to Warn*, 75 VA. L. REV. 509, 530–31 n.35 (1989).

character, beliefs, desires, evaluations, choices, actions, and enjoyments line up together. The addict described above is rational because these items for him line up in the way that constitutes practical rationality. No position is taken whether this fully practically rational addict just depicted is actually choosing and doing the right or even the sensible thing.[64]

It is interesting to speculate how we would regard addiction if, contrary to fact, all people we now call "addicts" conformed to the rational choice model and were thus not excused. Would addiction still exist, or would the rational choice explanation have explained addiction away? There are, after all, what Saul Kripke once called "skeptical explanations."[65] Such explanations show us that the thing we are explaining has a nature that differs dramatically from what we thought it had. To use Kripke's example: did Hume try to show us that causation exists but has a quite surprising nature (no necessitation of one event by another, only regularity of co-occurrence)? Or did Hume try to show us that causation does not exist? That depends on whether we require anything properly called "causation" to share the pre-Hume understanding of what causation must be, viz., necessitation.

Likewise with addiction, if at least some of the failures of practical rationality described in Section III.C below are necessary for any condition to properly be called an addiction, then the rational choice model, if true universally of those now called "addicts," quite literally explains away the phenomenon of addiction. Behaviors we call "addictive" would simply be one kind of practically rational action, with there being no interesting criterion of differentiation from all other instances of practically rational action.

In any case, even if the concept of addiction survived the discovery that all addicts were rational in the sense just depicted, what would not survive such a discovery is any claim that addiction excuses the drug-related behaviors of addicts. For on the rational choice model, the choices of addicts would be just like the choices of nonaddicts. The pertinent question is thus how many addicts choose to take drugs, steal, etc., with the psychology of rational choice. Perhaps some do. Years ago, Harry Frankfurt distinguished what he called the "willing addict" from the "unwilling addict."[66] Willing addicts are rational choice addicts while unwilling addicts are not. And perhaps some addicts are indeed willing addicts. But surely most are not. Most are unwilling addicts whose psychology does not match that of the rational choice model, and because of this fact, are at least within the ballpark of possessing a plausible excuse.

64. Of course, if one's ethics is that of a monistic utilitarian—where the only intrinsically good state of affairs is the satisfaction of human preference—that will blur this distinction between objective versus subjective rationality. For on such an ethics, satisfying subjective desire will necessarily also be objectively valuable.

65. Kripke's actual phrase was, "skeptical solutions." SAUL A. KRIPKE, WITTGENSTEIN ON RULES AND PRIVATE LANGUAGE 66–68 (1982).

66. Harry Frankfurt, *Freedom of the Will and the Concept of a Person*, 68 J. PHIL. 5 (1971).

C. Indirectly Applying the Schema to Explain the Behavior of Addicts in Terms of Less than Full Practical Rationality

The remainder of the folk psychological explanations that we will consider depart from the rational choice model in a variety of ways. None of them fully abandon that model; all, that is, employ the concepts of the folk psychology to explain why addicts do what they do, even though the items doing the explaining do not line up in the way needed to completely satisfy the above schema for fully rational action.

Because of the 2000-year-old prominence of akrasia, or "weakness of will," I divide these less-than-fully rational kinds of explanations into two main camps: those where there is a failure between t_4 and t_5, that is a failure of the actor's choices to reflect what he most wants, most values, most enjoys, or most believes to be true; and those where the actor's actions fail to reflect what he has chosen (intended) to do.[67] (I shall restrict the notion of weakness of will to failures of this second kind only, recognizing that others employ the phrase more broadly.) Both kinds of these failures result in actors doing things that are not in accordance with what they most want and/or most value—but how and where such failures occur differs between the two camps.

I also consider in less detail two other forms of failure in practical rationality that are here relevant: first, where there is no intention to match or mismatch with what one most wants or most values because no such intention is formed; and second, where what one most wants, most values, and intends, brings the actor no psychological satisfaction. I will consider each class of failure seriatim in what follows, starting with several species of the failure to form an intention that could match up with what one most wants and/or most values.

1. Failures to Form an Intention—"Automaticity" Models of Addiction

Surprisingly prominent in the literature describing the phenomenology of addiction is one or other of various models according to which addicts use drugs and steal to get drugs on "automatic pilot." These are presented as cases where choice/intention is (largely) absent so that one's desires to use drugs never gets integrated with one's other desires or with one's evaluative beliefs in the way that they do when we deliberately choose to do something. Such failures to form an intention are one way that addicts may use drugs despite that act not being what one most wants or most values. I see three or possibly four variations to be teased out of the literature on addictive by-passing of intentions as the addict acts. I consider each below.

67. A division I employed in Michael S. Moore, *The Neuroscience of Volitional Excuse*, *in* PHILOSOPHICAL FOUNDATIONS OF LAW AND NEUROSCIENCE (D. Patterson & M. Pardo eds., 2016).

a. The Seeming Automaticity of Preconscious Actions

It is a familiar feature of daily life that we engage in intelligently executed, goal-directed activities without having to direct our conscious attention to the matter at hand. With skills that we have mastered, or habits that we have acquired—such as, for example, going to lunch at our regular time, driving a curvy mountain road, stacking lumber, and playing the piano—we can do such things without consciously directing the routines by which we do them. As both James and Freud noted, we experience these behaviors as being within our control because of two dispositions we have that accompany them: one, we can direct our attention to the guidance or cessation of such behaviors if, because of environmental surprises, we need to do so; and second, when we do direct our conscious attention to such behaviors we are not surprised at finding out that we have been doing them, for all along we had the ability to describe what we were doing if asked to do so.[68]

In such ordinary cases of "preconscious actions" we might think that the actor has formed the requisite intention, has made the requisite choice, just that such intention is preconscious.[69] Contrast such ordinary cases with those studied more recently by the self-proclaimed psychologists of the "new unconscious" such as John Bargh and Daniel Wegner: despite being on a diet and resolving not to eat fattening foods, we find ourselves nibbling on the cookies that sat next to us as we were reading.[70] In such cases, we feel less in control because the routines in which we engage are not in the service of our conscious goals, as is the case in playing the piano, etc. Rather, we are surprised and, in some cases, even irritated to discover that we have been acting without having consciously chosen to so act and in a way that is against what we most strongly valued and desired. As Daniel Wegner concluded, "these actions seem to roll off in a way that skips intention."[71]

Some think that this account explains why at least some addicts take the drugs to which they are addicted. As Jeanette Kennett sees it, "drug-related stimuli . . . cue action *automatically*," which she believes results in a "subsequent loss of self-control."[72]

68. WILLIAM JAMES, 1 THE PRINCIPLES OF PSYCHOLOGY 516–17 (1890); Sigmund Freud, *The Unconscious*, (1915), *translated and reprinted* in GENERAL PSYCHOLOGICAL THEORY, PAPER ON METAPSYCHOLOGY 116, 126–27 (Philip Rieff ed., 1963).

69. I distinguish such preconscious actions from truly unconscious actions in Michael S. Moore, *Responsibility and the Unconscious*, 53 S. CAL. L. REV. 1563, 1576 (1980).

70. John Bargh, *The Automaticity of Everyday Life*, *in* THE AUTOMATICITY OF EVERYDAY LIFE 3 (John Bargh & Robert Wyer, eds., 1997); DANIEL WEGNER, THE ILLUSION OF CONSCIOUS WILL 20 (2002).

71. WEGNER, s*upra* note 70, at 130.

72. Jeanette Kennett, *Addiction, Choice, and Disease*, *in* NEUROSCIENCE AND LEGAL RESPONSIBILITY 271 (Nicole A. Vincent ed., 2013) (emphasis in original). Richard Holton and Kent Berridge call this the "Habit Account" of addiction. Holton & Berridge, *Addiction Between Compulsion and Choice*, *in* ADDICTION AND SELF CONTROL 239, 244–45 (Neil Levy, ed., 2013).

b. Too Many Rather than Too Few Intentions

In other work I distinguished a second way in which this direct causing of behavior by desires such as an addict's cravings could occur.[73] The one just explored is where the component desire (or "craving"), although not being what the actor most wants or most values, nonetheless causes the behavior satisfying it because the actor's component desires have not issued in an intention whose content corresponds to the content of what they most want. An alternative possibility is one where, although intention as such is not bypassed (because the actor has formed some intention) the actor has formed *two* intentions whose contents merely duplicate the contents of his conflicting desires and conflicting evaluative judgments; in this second kind of case there is also no overall intention or choice that is resolving of this conflict. In either the first or the second kind of case, the defect in practical rationality is that there is no *conflict-resolving* intention, and it is this absence that allows a weaker and less positively evaluated desire to cause behavior satisfying it, *i.e.*, the addict takes drugs and does what needs doing to obtain drugs.

Intentions, unlike desires, cannot rationally conflict; it is criticizably irrational to both intend some action A and to intend to omit doing A. But such conflict is not psychologically impossible. So it is possible that some addicts have their cravings for drugs cause drug-seeking behavior through one of a pair of conflicting intentions; such actions are still automatic in the sense that the component intention simply mirrors the context of the craving but does not integrate that intention/desire pair with other, conflicting intention/desire pairs.

c. Addictive Cravings as a Kind of Emotion-Driven Bypassing of Intention

Undeniably one of the salient markers of addiction is the experience by addicts of their desires (to do that to which they are addicted) as cravings. Cravings are rightly seen as a kind of emotion. It is commonly thought that strong emotions can "unhinge the will" in the sense that intention and choice are bypassed or sidelined by the frenzy of such emotional storms. The thought is not unique to cravings. Some such explanation is given for actors who are provoked to do things they would not otherwise do by the outrageous behavior of their victims, behavior that makes such actors so angry that their anger is said to unhinge their reason. Likewise to explain why some actors again do acts they otherwise would not do, and know that they should not do, when driven by the fear aroused by a threat of another. Nor is the thought, when it is confined to cravings and not other emotions, limited to the cravings of addicts; I take the depiction of the craving for high social position in Theodore Dreiser's *An American Tragedy* to be an accurate portrayal of how strong emotions of craving that are not those of an addict can cause an actor to do what he cannot bring himself to do through intention and choice—in Dreiser's example, strike the blow, capsize the boat,

73. Moore, *The Neuroscience of Volitional Excuse*, *supra* note 67, at 194–98.

and fail to rescue the victim he so wants to be rid of in order to satisfy his craving for high social position.[74]

I have long been suspicious of whether in truth emotions generally have the capacity to bypass the will in the way this model of addiction supposes. As I said on an earlier occasion:

> Are any emotions truly free of corresponding judgments that justify them to the agent whose emotions they are? Is any rage truly blind, or any anxiety without its object? Do the emotions that allegedly cause action by 'short-circuiting' choice ever proceed except by a chosen letting go, a chosen self-indulgence?[75]

But watching up close and personal the voluptuous pleasure taken in letting go of all restraint of emotion by one I was at one time close to, has perhaps colored my judgment here. I thus include this explanation of addictive behavior as a psychological possibility, despite my doubts. What it would explain, if true, is why an addict would go "on automatic pilot" in bypassing his will, on those occasions that addicted behavior is such a bypassing of will. It differs thus from the distracted, nonemotion driven automaticity studied by John Bargh, Daniel Wegner, and others mentioned in the earlier Subsection.[76]

d. "Half-way to Intention" Models?

Although critical of applying the habit model of automaticity to addicted behaviors, Berridge and Holton themselves propose a kind of automaticity explanation of addicted behavior.[77] They liken the addict's craving-desires for drugs to a mental state half-way between a desire and an intention, one that has "a tendency to lead directly to action."[78] As they more completely describe their conceptualization:

> Desires are the inputs to deliberation . . . . Intentions are the outputs of deliberation . . . and lead directly to action . . . . Cravings seem to come somewhere between the two. While they have many of the features of standard desires, they are not easily thought of as inputs to deliberation. Rather, they lead directly to action unless something [such as an exercise of self-control] stops them.[79]

I mention this model of automatic addicted behavior because it seemingly occupies a logical space: just as there can be *no* intention mediating between a desire and the action done to fulfill that desire, and just as there can be two intentions in conflict mediating between conflicting desires and behavior, so it might seem that there can be "half an intention." Yet in truth I think we can dispense with this model of automaticity because it adds nothing to the forgetful

74. *See* THEODORE DREISER, AN AMERICAN TRAGEDY 431 (1925).

75. Michael S. Moore, *Choice, Character, and Excuse*, 7 SOC. PHIL. & POL'Y 29, 39 (1990) *reprinted in* MICHAEL S. MOORE, PLACING BLAME: A THEORY OF THE CRIMINAL LAW 560 (1996).

76. *See supra* notes 70–72 and accompanying text.

77. Holton & Berridge, *supra* note 72, at 240–41.

78. *Id*. at 241–42.

79. *Id*. at 261.

habit, emotional storm, and duel intention accounts earlier explored.[80] The last three are all accounts of how addicts' craving desires can lead directly to action; and there is no need (or justification) for conceptualizing such directly causing desires as "half-way" to being an intention. In truth, a desire need not be "an input to deliberation" in order to remain a standard issue, not deviant, desire. My desire to beat a famous chess player in chess can cause my heart rate to go up; it can cause my fatally stupid move in chess; it can even cause me (van Gough-like) to cut off my ear; and in none of such cases need the desire in question be seen as unusual or nonstandard. Such "mental causes" of behavior (as the ordinary language philosophers used to call them) are a familiar feature of our everyday phenomenology of desire. The direct causing of behavior by desire is worth remarking on in the context of explaining addictive behavior, but that feature seems adequately explained in terms of the habit, emotion, and dual intention models described before, leaving no room for the "hallway-to-intention" conceptualization.

2. Failures in the Intention that is Formed and on Which One Acts, to Match What One Most Wants or Most Values

Unlike the automaticity models where there is no choice or intention (or at least no *conflict-resolving* intention), here the addict does form the intention to use drugs or to engage in drug acquiring behavior. Yet the choice fails to match what the addict most wants or most values. There are three possible reasons for this, corresponding to the three inputs (depicted on Chart 2) of choice, namely, the inputs of factual beliefs, wants, and values. I consider each kind of failure below.

a. Cognitive Failure: Not Keeping Degrees of Belief Responsive to the Evidence Available to Support It

Part of the phenomenology of addicted behaviors seems to be the erosion of belief. Dieters, for example, may come to believe momentarily that they can eat the dessert in front of them and that yet such consumption is not inconsistent with losing weight. Or a person who has not in the past found enjoyment in taking drugs may at a certain point in time believe that this time the taking of drugs will produce satisfaction, pleasure, and enjoyment. Or yet again, an addict who has tried to quit before and failed may come to believe that failure is inevitable (despite the social science showing that it is not) and, being thus resigned to such failure, does not try to resist his temptations.[81]

Such erosions of belief—erosions that result in the actor temporarily believing things contrary to what the evidence available to him would support—seem to be due to wishful thinking, itself a kind of self-deception. One craves the

80. *See supra* Sections III.C.1.a–b and accompanying text.

81. Jeanette Kennett dubs addicts suffering from this kind of erosion of beliefs, "resigned addicts." Jeannette Kennett, *Just Say No? Addiction and the Elements of Self-Control*, *in* ADDICTION AND SELF-CONTROL 144, 160 (Neil Levy ed., 2013).

drug so powerfully that this causes one to temporarily "forget" what one knows, and instead believes that one can "have his cake and eat it too," that pleasure is just an injection away, that success in quitting is impossible.

b. Motivational Failure: Not Integrating One's Desires into What One Most Wants

Just as an actor's decisional processes may skip the forming of a choice or intention, as described in Subsection 1 above, so an actor may skip the forming of an all-things-considered desire, what I have called what he *most wants*.[82] The beliefs of some schizophrenics give us a model for how this can work. A delusional belief that one is being persecuted, for example, can be "frozen" in the sense that it is immune to correction or outweighing by other, contrary beliefs. Frozen beliefs don't "play nice" with their fellow beliefs, that is, don't combine with those other beliefs the way that ordinary beliefs do.

The cravings of addicts can operate vis-à-vis competing desires the way frozen beliefs operate against competing beliefs in the minds of schizophrenics. Such cravings are not just strong desires, although they are that too; they are not just emotion-laden desires, although they are that too. In addition, cravings are "asocial" vis-à-vis their fellow desires in the way that frozen beliefs are asocial vis-à-vis their fellow beliefs. Being not combinable with their fellows, they are not amenable to correction by desires that are, in some sense at least, stronger.[83] The result is an unresolved conflict of desires whereby the craving may directly cause choice in accordance with it, without there being any overall want operating as a causal intermediary.

This "noncombinability" characteristic of the cravings of addicts is not so much a feature of the cravings themselves, as if they possessed some intrinsic "reverse magnetism" vis-à-vis other desires. Rather, it is that addicts are robbed of one of our main tools for integrating conflicting desires into an overall want, namely, *attention*. In resolving conflicts of desires, we need to be able to put aside attention to one desire while we attend to the other items that we also desire. Whereas this is just what addicts have a difficult time doing: the craving monopolizes attention on itself, precluding attention to the other desires that may be of

82. *See supra* Section III.C.1.

83. It is no small matter to specify the sense in which a desire that loses out to another desire in a head to head competition between them to motivate a given action, is nonetheless *stronger*. As Donald Davidson once remarked, "it is unclear how a want is shown to be overriding except by the fact that it overrides." Donald Davidson, *Freedom to Act, in* ESSAYS ON FREEDOM OF ACTION 137, 151 (Ted Honderich, ed., 1973). The most obvious senses to be given to strength of desire, other than Davidson's behavioral test, are: (1) the intensity of the craving experienced by the desirer; (2) the degree of satisfaction experienced by the desirer if the object of his desire is attained; (3) the degree to which the desirer judges the object of his desire to be desirable, *i.e.*, how much he values it and identifies it as being part of his self; (4) how temporally dominant a desire is over a stretch of time, *i.e.*, although not determining action on this occasion it would determine action at most other times because its *long term* strength is greater. Yet each of these senses of "strength of desire" form an independent model of addiction, as is explored in other subsections of this part of the Article. Still, what is left for giving sense to "strength of desire" is this: (5) one desire is stronger than another if it would win out in open competition with the other desire, "open competition" meaning a comparison and integration between two desires when neither of them is frozen (or asocial) in the manner described in the text.

greater strength.[84] In this too addictive cravings resemble the frozen beliefs of paranoid schizophrenics: in both cases the craving or belief is an obsessional mental state, a state the actor cannot easily not focus on and a state that drives out focus on other states (of desire or belief) and thus prevents them from being considered and compared.

c. Normative Failure: Acting and Wanting Against One's Better Judgment

Perhaps most common to the phenomenology of addiction is the failure of the content of what one most wants, to match the content of what one values, *i.e.*, judges one should want. In such cases one's evaluative beliefs fail to match one's desires and wants. Although choice and action lines up with what one most wants in such cases, they do not line up with one's evaluative judgments of desirability. The addict chooses against his own best judgment.

As noted earlier, the relationship between desires and judgments of desirability ("values") can easily be missed; in Donald Davidson's terminology, both are kinds of "pro attitudes" that motivate action.[85] Even those who, unlike Davidson, wish to mark this distinction between two kinds of pro-attitudes, often have done so in terms of desire. Early Harry Frankfurt, for example, spoke of there being "second-order desires," which are second order in the sense that their content contains other, first order desires.[86] I desire, for example, not to desire to eat cake. Such second-order desires thus function like evaluative beliefs, evaluating first order desires as being or not being desirable. Frankfurt raised such second-order desires, not because he thought that they were necessarily stronger than first-order desires,[87] but because he thought that persons identified themselves more with such second-order desires (and with the desires secondarily desired) than with brute first-order desires. Later Frankfurt sought to capture this greater centrality to self-identity with his notion of wholeheartedness: some desires we wholeheartedly endorse, whereas others (such as the cravings of an addict) we do not, or perhaps we even disavow them.[88] Second order desires for later Frankfurt thus reveal themselves to be in actuality judgments about desirability. As another example, Michael Smith speaks of desires that match the desires the actor *believes* he should have, translated (for Smith) into the strongest

84. The attention monopolization is a much remarked-upon feature of addictive cravings. *See, e.g.*, R. Jay Wallace, *Addiction as Defect of the Will*: *Some Philosophical Reflections*, 18 LAW & PHIL., 621, 645–47 (1999); Walter Sinnott-Armstrong, *Are Addicts Responsible?*, *in* ADDICTION AND SELF-CONTROL 128 (Neil Levy, ed., 2013).

85. Davidson, *supra* note 51.

86. Frankfurt, *supra* note 66.

87. As Philip Pettit so construes him. Philip Pettit, *The Capacity to Have Done Otherwise, in* RELATING TO RESPONSIBILITY: ESSAYS FOR TONY HONORÉ ON HIS EIGHTIETH BIRTHDAY 25 (Peter Cane & John Gardner eds., 2001).

88. Harry Frankfurt, *Identification and Wholeheartedness*, *in* RESPONSIBILITY, CHARACTER, AND THE EMOTIONS 27, 43–44 (Ferdinand Schoeman ed., 1987). On avowal and disavowal of desires, and emotions, see HERBERT FINGARETTE, SELF-DECEPTION (1969).

desires the actor believes he *would* have if he were fully rational.[89] Desires like those of the addict are not, for Smith, the objects of such hypothetical, evaluative beliefs.[90] They may even be acknowledged by the agent to be defective desires in that they conflict with the desires he believes he would have in greater strength if he were fully rational.

Painting with a somewhat broad brush, I see these various formulations all referring to roughly the same thing: some desires are tightly woven into an agent's view of who he is and should be and others are not. Victor Tadros more recently speaks of desires that are not "accepted [by the agent] in light of the agent's values."[91] This way of putting things finds resonance in the contemporary literature on addiction: addicts are said to want what they do not value.

3. Failures of One's Actions to Match One's Intentions

I turn now from failings that occur before choice (failings either to make a choice at all or failings to match one's choice to what one most wants or most values) to failings that occur after choice is made. I turn, that is, from failures to choose in light of what one most wants and most values, to cases where one fails to do what one chooses. (Notice that in either case, one ultimately *acts* in ways not fulfilling of what one most wants or most values.)

Ordinary failures to achieve what we intended to achieve are not our concern here, for these are a dime a dozen. None of us succeed at everything we undertake to do. But such failures are due to the world not cooperating with our plans. These are failures of our bodily movements at t_8 on Chart 2 causing such of the results at t_9–t_{11} as we intend. Our concern here is rather with failures between t_7 and t_8; we do not even try to do what we intended to do. We do not execute our intentions at t_5–t_7 into those bodily movements at t_8 that would give our plan a chance of success in the external world. I choose, for example, not to eat the dessert in front of me, and such choice is fully in line with what I most want and most value; but I intentionally eat the cake anyway. I am classically considered to be akratic, that is, to suffer from weakness of will.

a. Synchronic Weakness of Will

Many would deny that anyone actually intends at the time he is acting not to do A, and then despite or perhaps even because of that intention does A anyway. Yet we should separate criticizable irrationality from psychological impossibility. The akratic who acts against his present intention is indeed highly irrational; but that does not mean there are no such cases.

One suspicion about such cases stems from the thought that an intention that produces the opposite of the action intended—when nothing intervenes to induce

89. *See generally* Michael Smith, *Responsibility and Self-Control*, *in* RELATING TO RESPONSIBILITY: ESSAYS FOR TONY HONORÉ ON HIS EIGHTIETH BIRTHDAY (Peter Cane & John Gardner eds., 2001).

90. *Id*.

91. VICTOR TADROS, CRIMINAL RESPONSIBILITY 343 (2005).

a change of mind, and no mistakes in beliefs about means of execution are present—goes against what intentions are. On a dispositional view of intention, it is part of the "logic of the concept" (as the ordinary language philosophers used to say) that one does the act intended when the occasion to do so arrives and nothing relevant has changed in the actor's mental states. Yet this old view of intentions is too behavioral to be credible; "intention" refers to a natural kind whose deeper nature is functional and physical, not phenomenological and not behavioral. Van Gogh can intend to be a great artist, and this can cause him to cut off one of his ears. This is wildly irrational, but not psychologically impossible.

A second worry about this picture of weakness of will is that such weakness occurs only in automatic actions. This is because of the absence of any intention mediating between all out wants and action in such cases. True, this worry could concede, there is in such cases an intention such as an intention to refrain from eating some desirable slice of cake, and true, that intention can serve as a causal intermediary between the strongest desire (to remain on the diet) and the action (of eating the cake) despite the mismatch of act done both to act both most wanted and intended. Yet will not most such cases where this array of mental states is present be cases that tempt us to say that the weaker desire (to eat cake) directly caused the action desired? And if this is true, such actions will be part of the "automaticity of everyday life" earlier discussed.[92]

The worry is that many cases of plausibly compelled and even obsessive behavior are not sudden yieldings to temptation. As J.L. Austin observed, one can take the second dessert at High Table (which one knows one should not have) with delicacy, deliberation, and graceful slowness; yielding to temptation need not always be, and often is not, the wolfing down of such dessert.[93] The latter is the exaggerated depiction of yielding that is the stuff of grade B movie scripts.

Neither of these two problems to my mind rule out the possibility of weakness of the will as I have depicted it, where the strongest desire to do the right thing is realized in a choice (intention) to do that very thing, and yet (with no other mental state intervening), the actor does the opposite. Even though deeply irrational such behavior is psychologically possible. Even so, surely such deeply irrational behavior is comparatively rare, indeed, too rare to capture the range of cases we intuitively think of as weakness of will.

b. Diachronic Weakness of Will

There is a better conceptualization of the phenomenon, one that can be seen by attending to these examples of Thomas Schelling, a noted game theorist and economist (and my some-time correspondent in the 1980s). Schelling was looking for a rational consumer whose preferences would be worth maximizing in a utilitarian calculus:

> How should we conceptualize this rational consumer whom all of us know and who some of us are, who in self-disgust grinds his cigarettes

92. *See supra* note 70 and accompanying text.

93. J. L. Austin, *A Plea for Excuses*, 57 PROC. ARISTOTELIAN SOC'Y 1, 24 n.13 (1956).

> down the disposal swearing this time he means never again to risk orphaning his children with lung cancer and is on the street three hours later looking for a store that is still open to buy some cigarettes; who eats a high calorie lunch knowing that he will regret it, does regret it, cannot understand how he lost control, resolves to compensate with a low calorie dinner, eats a high calorie dinner knowing he will regret it, and does regret it; who sits glued to the TV knowing that again tomorrow he'll wake early in a cold sweat unprepared for that morning meeting on which so much of his career depends; who spoils a trip to Disneyland by losing his temper when his children do what he knew they were going to do when he resolved not to lose his temper when they did it?[94]

What Schelling's familiar examples from daily life suggest to me is not actions that go against both strongest desire and intention, as modeled above. Nor is it what seems to tempt Schelling himself (which seems to be the simultaneous dual intention model I mentioned briefly earlier in Subsection 1). Rather, Schelling's examples suggest that we go diachronic: keep the match between object of intention and object of all-out wants, keep the match between action done and action intended, and thus keep the overall match of strongest desire to action done.[95] But see Schelling's agents as oscillating over time between sets of mental states and actions, each of which obey these requirements. So at t_1, the smoker intends not to smoke in line with what he most wants and values. Yet at t_2, the mental constellation of mental states change, resulting in the opposite action. The smoker reverses field on his wants, values, and intentions, and now chooses to smoke in line with his changed constellation of wants and values. And then at t_3, the period of immediate regret, he oscillates back to the first constellation of mental states.

Notice that neither of the objections raised earlier to synchronic weakness of will apply to this diachronic conceptualization of weakness of will. There is no need to qualify the view that ties dispositions to behave to intentions (for the agent is disposed to behave in accordance with the objects of his intentions at each time). There is no need for the act of smoking to be sudden or automatic, because it is not directly caused by desire but is rather guided by an appropriate intention. Another objection does apply, however. This agent's intentions are decidedly "nonsticky."[96] Unlike ordinary, sticky intentions, nonsticky intentions

94. Thomas Schelling, *The Intimate Contest for Self-Command*, 60 THE PUB. INTEREST, 94, 96 (1980).

95. A suggestion one also finds in Thomas E. Hill, Jr., *Weakness of Will and Character*, *in* 14 PHIL. TOPICS, 130 (1986) (reprinted in THOMAS E. HILL, JR., AUTONOMY AND SELF-RESPECT (1991)):

> [W]e cannot identify weakness of will simply by looking to see whether at each moment the agent's acts correspond to his deliberative conclusions at that moment; we need to survey several aspects of the agent's history over time, including . . . the frequency and reasons for 'changes of mind.'

Michael Smith is also atune to the comparative ease of conceptualizing weakness of will, "diachronically" rather than "synchronically." *See supra* note 89, at 5–9.

96. "Stickiness" is my nontechnical term for the rational commitments having an intention commits us to. MICHAEL BRATMAN, INTENTION, PLANS, AND PRACTICAL REASON (1987). Of particular relevance is the rational commitment to nonreconsideration of the predecision desires that incline one in different directions. *See* GIDEON YAFFE, ATTEMPTS 148–56 (2010). Joseph Raz has long conceptualized such commitments to nonreconsideration in terms of negative second-order reasons (Raz calls them "exclusionary reasons," so called because they exclude what were formerly good reasons pro or con some past decision). Exclusionary reasons are second-order reasons

do not preclude constant re-evaluation of what the agent most wants to do (or thinks, all things considered, that he should do). Such nonstickiness is criticizably irrational because intentions are not performing as they should to give us some stability and continuity in our daily life. But it is not only psychologically possible, but it is just as common as Schelling plainly thinks it is.

Moreover, is not this a good match to the idea of a will that is weak on a given occasion? The decisions (choices, intentions) of such a weak-willed person do not control his behavior much into the future because they themselves are so constantly subject to being changed. Such lack of much if any *psychological* commitment to the nonreconsideration that having an intention rationally commits us to, well unpacks the idea of a will that is weak.[97]

4. Failures of Experiential Satisfaction to Match What One Wants and Chooses

Even if all else is in perfect working order, one's action may still represent a failure in practical rationality because the state of affairs desired, valued, intended, and caused by one's acts, may give the actor little or no satisfaction. Such lack of satisfaction can occur in two ways, only one of which is relevant here. The irrelevant way is that the action chosen has unforeseen and unwanted consequences, and the badness of these (together with the badness of foreseen consequences) outweighs the goodness and desirability of the intended consequences of that action. One can regret both the choice and the action because of such unforeseen side consequences. Sometimes this might represent a kind of cognitive failure—not having predictive beliefs proportionate to the evidence available to the actor at the time of acting—but it need not represent such kind of cognitive failure. When it does not, the disappointment of the actor does not bear on the rationality with which he acted.

More relevant is the second kind of disappointment, one where the attainment of the actor's chosen objective itself brings no joy or happiness. Freud's example of this was the all-too-common pursuit of wealth: although wanted, valued, and acted upon by many, money gives no happiness, Freud thought, because

not to act for certain reasons. JOSEPH RAZ, PRACTICAL REASON AND NORMS (Oxford University Press, 1999); JOSEPH RAZ, THE MORALITY OF FREEDOM 23–109 (1986). For several interpretations of Raz's exclusionary reasons, see MICHAEL S. MOORE, *Authority, Law, and Razian Reasons*, 62 S. CAL. L. REV. 827 (1989), *reprinted in* EDUCATING ONESELF IN PUBLIC: CRITICAL ESSAYS IN JURISPRUDENCE 128–89 (2000).

97. The diachronic conceptualization of weakness of will suggests that there also should be a diachronic version of model 2.b. above, that of motivational failure. Indeed, Gideon Yaffe has argued that addicts oscillate between what it is they most want and most value: most of the time they most want not to take drugs and most of the time they value such abstinence over use; but on the occasions when they do take drugs, their motivational and evaluative balances shift, so that at the moment of consumption they value and want most to use drugs. Gideon Yaffe, *Are Addicts Akratic? Interpreting the Neuroscience of Reward*, *in* ADDICTION AND SELF-CONTROL 190–213 (Neil Levy ed., 2013). And then shortly thereafter, they revert to their long-term balances of wants and values leading to regret at their earlier decision. As with the diachronic conceptualization of weakness of will, these diachronic conceptualizations avoids the puzzles raised by acting against what one most wants, most values, or intends at the time one acts.

it is neither an actual nor a sublimated desire of childhood to have it.[98] In any case, one's actions are fully rational only if what satisfies one is also what one wants.

This is also true with respect to what one values. Although Kant famously held that "virtue and doing the right thing" was no guarantee of happiness—that was the role for god in Kant's ethics—nonetheless (and despite Kant's clumsy attempt to give god some work to do in ethics) achieving what one judges to be valuable should bring with it the satisfaction that comes with the completion of worthwhile projects.[99] If it doesn't, this too is a failure of practical rationality.

Berridge and Holton have argued that addicts have these kinds of mismatch between their wants and what they truly like.[100] They argue that for such addicts getting high gives no pleasure even if such state of intoxication is wanted in advance of acting to achieve it.[101] Worse, such addicts may repeat the action that disappointed them in the past full well knowing that it will disappoint them again. This is a failure of practical rationality for such addicts, one that they are afflicted by a class of desires that do not do what desires are supposed to do, viz., produce pleasure in their satisfaction.

D. Combining These Explanations into One Overall Folk-Psychological Explanation of Why Addicts Use and Acquire Drugs?

If one counts carefully in the above exposition, there are twelve ways in which the unwilling addict continues to use drugs that are criticizably irrational. It is tempting to consolidate these twelve ways into one kind of failure. After all, what the unwilling addict does by continuing to use drugs is to bring about states of affairs: that he does not like (in the sense of find pleasurable); or that he did not choose (intend) to bring about; or that he does not (overall) want; or that he does not (overall) value; or that he does not like, intend, want, or value because he is misled by his irrational beliefs; or that are the products of automatic behaviors by him that are only marginally actions. The unity lies in the fact that the unwilling addict acts contrary to one or more of the ingredients to behavior that marks that behavior as a rational action. Yet it would be unduly syncretistic in this context to abstract away the differences here. For our ultimate purpose here is to explain the behavior of addicts in a way congenial to our later discussion of excuse. And unless the irrationalities discussed above add up to a status excuse of nonrationality (as I have long urged the irrationalities of severe mental illness to do in the defense of insanity)[102]—which they do not—then one wants to keep

98. Sigmund Freud, Letter to Fliess of January 16, 1898, https://www.pep-web.org/document.php?id=zbk.042.0294a.

99. IMMANUEL KANT, CRITIQUE OF PRACTICAL REASON (Mary Gregor, ed. & trans., 1797) (2015).

100. Holton & Berridge, *supra* note 72, at 249.

101. *Id.* at 241, 246.

102. *See* MICHAEL S. MOORE, LAW AND PSYCHIATRY: RETHINKING THE RELATIONSHIP 217 (1984); Michael S. Moore, *Mental Illness and Responsibility*, 39 BULL. OF THE MENNINGER CLINIC 308 (1975); Michael S. Moore, *Seeking a Responsible Responsibility Test: Norwegian Insanity Law After Breivik*, 9 CRIM. L. & PHIL. 645 (2015).

these explanations of addicts' behavior separate so that their potential for excuse can be examined separately.

If we do keep these items discrete and separate, it would be a mistake to think that any one of these views predominate to the exclusion of the others in the folk-psychological explanation of the behavior of addicts. Sometimes addicts act as they do for several (or even all) of these reasons in combination. It would equally be a mistake to think that all addictive behavior is to be explained in all addicts by some one combination of these explanations. Addicts can differ between each other, and even between themselves on different occasions, as to what factors are at work and in what combination.

These two facts make the folk-psychological explanation of addictive behavior quite context-sensitive and therefore complicated. But such multifactored complexity does not require that we regard all such explanations as equally prominent parts of an overall explanation of addiction. It is an empirical question, resolvable by ordinary social science, as to which factors are more commonly important than others. Even in the absence of such scientific studies, however, some of these models of addiction seem more often instantiated by addicts than others. My own armchair bets, for example, are that automatic and synchronically weak-of-will addicts are few and far between, and that these factors therefore are of less importance in an overall, general explanation of how addiction causes behavior.

The context-sensitive and multi-factored nature of any complete explanation of behavior by addicts does not make such explanation unusable for either of my purposes here. One of those purposes, pursued next in this Article, is to use such explanation to probe into the question of whether addicts are excused from their addiction-caused behaviors. Given the multifactored nature of the explanation, this necessarily will be a "one factor at a time" kind of inquiry but still a possible one to do. The other of those purposes, pursued in Part V of this Article, is to see where neuroscience could aid in the deepening or the fleshing out of these folk psychological explanations and thus of excuses that depend on them. One of the ways that neuroscience might be helpful here is in resolving which of these factors predominates over the others, a task for which the separateness of the factors is essential.

IV. ADDICTION AS A MORAL EXCUSE AND LEGAL DEFENSE

We turn from the explanation of addiction to the moral question of whether addiction excuses those who act because of their addiction. Also of interest is the question of whether addiction should be considered a defense or mitigating factor in the criminal law. Because on my view of criminal law the answer to the question of whether there should be a legal defense follows the answer to the question of whether there is a moral excuse like an obedient dog, the legal question will be given little separate consideration.[103]

103. *See* MICHAEL S. MOORE, PLACING BLAME: A GENERAL THEORY OF THE CRIMINAL LAW 83–188 (1996).

As with explanation, there are three items for which initially at least we want to ask our moral question: (1) Are addicts responsible for becoming addicts? (2) Are addicts responsible for those behaviors while addicted that are symptomatic of being addicts, viz., the possession and use of drugs, alcohol, etc.? (3) Are addicts responsible for those behaviors as addicts not symptomatic of being addicts but which are caused by their addiction, such as stealing money with which to purchase drugs?

A. Three (More) Ways in Which Not to Talk About This Issue

Discussion of the moral issues here has been clouded by irrelevant distractions that have often featured in the extensive literature that has developed. Clarity is aided if one first exposes such distractions for the irrelevancies that they are before we then settle down to serious consideration of the issues in hand. We already (in Part II above) disposed of one of the leading distractions here, that engendered by the disease model of addiction.[104]

About such a model's crucial equation of diseased and excused, I concluded that there may be considerable overlap between the extensions of "is diseased" and "is excused." But this is merely a kind of interdisciplinary serendipity. Even when one is both diseased and excused, one is never excused *because* one is diseased. Put simply, being diseased—*i.e.*, being the appropriate subject of medical treatment—is empty of normative implications for being morally excused. I next consider three more distractions that we need to put aside in the context of asking the moral question of whether addiction excuses.

1. Addiction Excuses Because Addicts Do Not Have the Capacity Not to Have the Craving Distinctive of Addiction

In Part II earlier I put aside incompatibilist analyses of capacity and incapacity, analyses finding incapacity and excuse whenever they found physical causation of behavior.[105] But even on some compatibilist analysis of capacity, skepticism about the responsibility of addicts, proceeds from the plausible idea that addicts cannot choose not to have the cravings that move them to take drugs. The idea is that even though addicts *can* (in the compatibilist, counterfactual sense of "can" earlier discussed) refrain from taking drugs, and even though addicts *can* choose not to take drugs—compatibilist free action and free will, respectively—still, addicts *cannot* refrain from having the desire (cravings) to take the drugs that they do. And this, it is concluded, excuses addicts because, although their choices control their actions and although they may choose as they most want to choose, still they do not control what governs what they most want, *i.e.*, what they crave. Without this third aspect of control they are, it is said, excused.[106]

104. *See supra* Part II.

105. *See supra* Part II.

106. Philip Pettit, *The Capacity to Have Done Otherwise: An Agent-Centred View*, *in* RELATING TO RESPONSIBILITY: ESSAYS IN HONOUR OF TONY HONORÉ, 25 (Peter Cane & John Gardner, eds., 2001).

The problem for this view lies in its assumption that we must have the capacity to choose or in some other way control the desires that motivate our choices and actions, if we are to be responsible for those choices and actions. Whereas in truth none of us has much capacity to choose or otherwise control what we desire. Willing ourselves to desire something that in fact is distasteful to us would be a bit like willing emotions such as love: saying, "I am trying to love you," is a far cry from loving someone. True, we have some indirect strategies for getting ourselves to desire things that we initially find distasteful, but these are indirect, long-term, and of limited efficacy. A decently compatibilist moral philosophy does not demand that we be able to choose not to have certain desires, in order to be responsible for choosing to act, and acting, on those desires. Put bluntly, compatibilism holds us to be responsible where we both *can* choose one way or another and *can* act one way or another; it does not demand that we have the freedom (or capacity) to desire one way or another. We are responsible for how we choose and act on the desires that we happen to have, however we have happened to have such desires.

This means that what we need to focus on to assess the responsibility of addicts are the capacities to form and to act on intentions not to take drugs when those intentions are called for in light of what addicts most want, most value, and most like. These capacities will indeed be the focus of this part of the Article.

2. Addiction Excuses Because Withdrawal, etc., Makes It More Costly for Addicts Not to Use or Steal Than It Is for Nonaddicted Persons

There is a fundamental misunderstanding of the nature of excuses that has invaded the discussion of the responsibility of addicts. The misunderstanding is perhaps best seen by adverting to the standard excuse of duress. Duress is the doing of a wrongful action required by another who threatens us with adverse consequences if we do not do it. Some (including my former self) think that duress at least sometimes excuses because our opportunities to do the right thing are so constricted by the threat of adverse consequences that we cannot fairly be blamed for yielding to the threat by doing what the threatener wants.[107]

Although there is no threatening second agent in the use or acquisition of drugs by addicts, the analogy drawn is to the "threat" of nature: if the addict does not do what he needs to do to acquire and use drugs, then he will suffer the adverse consequences of withdrawal. Given the painfulness and inevitability of withdrawal following on nonuse of drugs by those addicted to them, the addict is said to have much less opportunity than the nonaddicted person to refrain from using and acquiring drugs.[108]

107. *See, e.g.,* Moore, *Choice, Character, and Excuse, supra* note 75, at 30.

108. Thus, Walter Sinnott-Armstrong concludes that the threat of withdrawal can constitute a kind of loss of control so that "the addict cannot quit . . . ." Walter Sinnott-Armstrong, *Are Addicts Responsible*?, *in* ADDICTION AND SELF-CONTROL 122, 124 (Neil Levy ed., 2013). Since the threat of withdrawal is more plausibly seen as a diminishment of opportunity than an impairment of capacity, more accurate is Jay Wallace's assessment

The misunderstanding common to this loss-of-opportunity version of both duress and addiction is this: although lessened opportunity to avoid doing wrongful acts may lessen blameworthiness, it does not do so by lessening culpability (it does not, in other words, *excuse*). What lesser opportunity can do is lessen the degree of wrong done; it does this, when it does this, by showing that something good came out of the wrongful act in question so that *net*, less wrongdoing was done. Such lessened opportunities thus operate as partial justifications, not as excuses. When such opportunity costs of ordinarily rightful action get high enough, they may not *partially* justify the omission of such ordinary rightful action, they may fully justify it. Thus, duress as a legal defense at common law operates exclusively as a full (or sometimes partial) justification, never as an excuse.

Even with this misunderstanding removed so that withdrawal's potential to lessen blameworthiness is properly categorized, withdrawal-related loss of opportunity does not do much moral work with respect to the responsibility of addicts. Withdrawal does not exist for nondrug or alcohol related addictions; and it is almost nonexistent for some addictive drugs such as cocaine.[109] Even when withdrawal does exist, with perhaps the exception of alcohol it is not that adverse a state to suffer through to constitute much of a diminution of the addict's opportunity set.[110] Some have likened withdrawal from nonuse of addictive drugs to be no worse than having an average case of the flu,[111] and we do not let flu sufferers off the hook for stealing flu medicine that they cannot afford to buy.

Costs of quitting other than withdrawal are also possible candidates for being wrongness-reducers. One's only friends may be addicted, for example, so that quitting imposes a social cost; or certain drugs may be performance-enhancing, so that discontinuing use would cost the addict that heightened performance.[112] Gideon Yaffe advances yet another version of the loss of opportunity argument with respect to addiction that does not depend upon the adverseness of withdrawal.[113] Yaffe contends that addicts who contemplate refraining from use, face the prospect of acting against what, at the time they would use, they most value and most want.[114] Yaffe speculates that addicts must find this prospect so daunting as to constitute a morally relevant diminishment of their opportunity not to take drugs.[115]

My own view of all of these lost opportunities is that they are insufficient for much diminishment of responsibility. Like withdrawal, they are not such costs as would reduce the net wrongs of use or acquisition significantly. Where

of a lesser blameworthiness because of withdrawal: the threat of withdrawal may make the addict's use *permissible* whereas for nonaddicts such use is *impermissible*. Jay Wallace, *Addiction as Defect of the Will: Some Philosophical Reflections,* 18 L. & PHIL. 621, 644 (1999).

109. Holton & Berridge, *supra* note 72, at 244.

110. *Id*.

111. Hanna Pickard & Steve Pearce, *Addiction in Context: Philosophical Lessons from a Personality Disorder Clinic*, *in* ADDICTION AND SELF-CONTROL 165, 171 (Neil Levy ed., 2013).

112. The examples are from Sinnott-Armstrong, *supra* note 108, at 128.

113. *See generally* Yaffe, *Are Addicts Akratic?*, *supra* note 97, at 193–94.

114. *Id.*

115. *Id.* at 211.

the acts are seriously wrong, as is true of some kinds of drug-acquiring actions like robbery or homicide, such lessening of opportunity does de minimis moral work.

In any event, these costs need to be recast as creating incapacities of the agents contemplating such costs, not as lost opportunities, to do any moral work as excuses. One would have to picture the addict frozen in fear of withdrawal if he does not use, much like a person excused because of duress needs to be incapacitated by his fear in order to be excused. And given the relatively nonaversive nature of the consequences threatened if the addict does not use, that is as implausible for all of such supposed costs as it is for withdrawal.

3. Addicts Are Not Excused for Acts as Addicts Because They Are Responsible for Being Addicts in the First Place

There is a well-known tendency in discussions about responsibility to engage what is known in the trade as the "tracing strategy."[116] The general idea is this: if at the time of causing harm to another (call that time "t_2") the actor suffers under some debilitating and excusing condition—an epileptic seizure, say—yet the actor is at fault at some earlier time ("t_1") for getting himself into such a debilitating condition, then a condition that normally excuses does not excuse. The excuse the actor would have had at the later time is forfeited by tracing his fault back to some earlier time when he brought about the conditions for his later excuse.

In the case of the behavior of addicts who acquire and use drugs (t_2), they themselves earlier acted when they were not addicts (t_1) in ways that made them addicts. According to the tracing strategy, therefore, they are responsible for their acts at t_2, no matter how excusing addiction might otherwise be.

This is a terrible argument about the responsibility of addicts because the tracing strategy itself is generally a terrible argument for responsibility. The stunning problem for the tracing strategy lies in its equation of an actor's blameworthiness at t_1 with the blameworthiness that actor would have had at t_2 if he were not in the debilitating condition he culpably (at t_1) caused himself to be in. There is no reason whatsoever to think that such equation is necessarily (or even often) true.

Take duress as an example. Suppose at t_2 defendant badly beats a victim, but he does so because the defendant was threatened with unlawful force against his children unless he did what the threatener told him to do. Suppose the threat at t_2 is sufficiently credible, proximate, and onerous to excuse defendant from most or all blame. The tracing strategy would eliminate this defendant's excuse of duress if at t_1 the defendant culpably placed himself in a position where he

116. *See generally* John Martin Fischer & Neal A. Tognazzini, *The Truth About Tracing*, 43 NOUS 531 (2009); Manuel Vargas, *The Trouble with Tracing*, 29 MIDWEST STUD. PHIL. 269 (2005). Heidi Hurd and I discuss the tracing strategy generally, and then as applied to negligence, in Michael S. Moore & Heidi M. Hurd, *Punishing the Awkward, the Stupid, the Weak, and the Selfish: The Culpability of Negligence,* 5 CRIM. L. & PHIL. 147 (2011).

might be subjected to such a threat. This means that a defendant who is only negligent at t_1 in unreasonably risking that he might be coerced into doing some minor wrong, is to be blamed as an intentional beater at t_2 with no account taken of the ordinarily excusing threat. An unexcused, intentional beating (for which this defendant is blamed under the tracing strategy) is much more blameworthy than a merely negligent risking of some minor wrong being done; yet the tracing strategy equates the two, blaming slightly culpable, minor wrongdoers as if they were seriously culpable, major wrongdoers.

Why does the tracing strategy persist in the face of its obvious potential for disproportionate blame and punishment?[117] Mostly because there is another analysis that is not unjust and with which the tracing strategy is confused.[118] According to this alternative analysis, when someone culpably does some wrong at t_1, the doing of that wrong can cause a further state of affairs to exist at t_2 for which the defendant is properly blamable. Suppose, for example, that the defendant in the above duress hypothetical wanted to beat up the person he did in fact beat up but lacked the courage to do so on his own. So at t_1 he coerces another to coerce him at t_2 to beat up the intended victim. Under the alternative analysis, at t_2 defendant does not lose his excuse as he would under the tracing strategy—at t_2 he was coerced into beating up the victim and he retains that excuse of coercion; but at t_1 defendant's act of (coercing another) causes the threat which causes the defendant to beat the victim up—*i.e.*, at t_1 defendant has intentionally albeit indirectly caused contact on the victim's body with his fists and should be blamed accordingly. He is guilty of assault, but it is a t_1 assault for which he has no excuse, not a t_2 assault where he has the excuse of coercion.

Unlike the tracing strategy this alternative analysis blames and punishes people proportionate to their desert. It recognizes that sometimes one can equally well cause a bad state of affairs to exist at t_2, not by an act at t_2 itself, but by an act as some earlier time t_1, and that when one does so one is blameworthy in proportion to the culpability with which one acted at the earlier time. There is no fictional equation of blameworthiness here, as there is with the tracing strategy.

Let us apply all of this to addiction. The relevant t_1 is when the addict is not yet addicted but takes the drugs that make him addicted. (This of course occurs over an interval of time, not all at once, but this nuance can be ignored for these purposes.) Are nonaddicted users to blame for using the drugs that make them

117. It persists at least in the common law of crimes, where it has its own doctrinal name, "actio in libera causa' (an act that is free (and responsible) in its cause even if not free and responsible in itself). The Model Penal Code rejects the doctrine, with partial exceptions for voluntary intoxication and duress. *See*, *e.g.*, MODEL PENAL CODE § 2.09 (A.L.I. 2018).

118. The alternative analysis is laid out in Paul H. Robinson, *Causing the Conditions of One's Own Defense: A Study in the Limits of Theory in Criminal Law Doctrine*, 71 VA. L. REV. 1 (1985). Sometimes adherence to a tracing strategy is not due to confusing it with the alternative analysis explored in the text. Rather, the tracing strategist applies a kind of forfeiture morality: if someone is doing something they should not be doing in the first place (like shooting up when not yet an addict), then they are responsible for all the effects of that initial wrong-doing no matter how unintended, unforeseen, or unforeseeable those effects might be. Such wrongdoers' initial wrongdoing is said to "forfeit" our normal concerns to grade their culpability by their actual mental states at the time they act. One sees this crude, forfeiture view on vivid display with the notorious felony-murder rule in Anglo American criminal law.

into addicts? Surely in many cases that answer is "yes," although it is a qualified "yes." It is yes because the conditions of culpability are often met at t_1. Some users may intend to become addicted (think Timothy Leary types); or, more often, they may know that they will or that they might become addicted. Or they should know of such a risk even if they do not in fact know. In all such cases, such nonaddicted users satisfy the conditions of culpability with respect to the consequence of being an addict.

The qualification to this "yes" lies in the aspect of blameworthiness known as wrongdoing, for one might well doubt (as do I)[119] that becoming an addict is a wrong at all (or alternatively, it is a wrong that one had a right to do). The worry is that perhaps ruining one's own life prospects, abusing one's own talents, etc., is one's own business, and that is not wrong to do because it is not a wrong to someone other than the actor. True, the criminal laws currently on the books criminalize the acts of use that produce addiction; but one might well doubt that laws have the capacity to make morally wrong behaviors that were not, prior to the law's enactment, antecedently wrong.

For purposes of reaching the issue examined here, we should concede arguendo that at t_1 nonaddicted users who use sufficiently to addict themselves are both culpable and wrongdoers, *i.e.*, blameworthy. Can one use this moral fact as sufficient to find them blameworthy at t_2 for acquiring and using drugs when they are addicted? The tracing strategy would answer affirmatively, but that simply illustrates the general injustice of the tracing strategy. Nonaddicted drug users' culpability and wrongdoing at t_1 need bear no relationship to the culpability and wrongdoing that they would have had if they robbed, stole, or used drugs at t_2 in their (counterfactually) nonaddicted state.

So using the tracing strategy is out, here as it is generally. Does the alternative analysis outlined above show that addicts are blameworthy for their acts of theft and use at t_2 by virtue of their culpable acts of using at t_1 having caused these later bad actions at t_2? It does not. First of all, the causal connection—between the much earlier acts of use that made an addict an addict, and the later acts of acquisition and use of drugs—is much too attenuated to support responsibility. Even if a later theft at t_2 for example, counterfactually depended on earlier, addiction-producing acts of use at t_1, the t_1 using is not the proximate cause of a t_2 stealing. Second, the culpability needed for serious wrongs like theft or robbery is lacking at t_1 when the soon-to-be addict uses drugs. Such a user at t_1 at most might be aware of a risk that if his present use leads to addiction, he might later resort to theft or robbery to support his habit. Such recklessness is a lesser culpability than that of an intentional thief.

The upshot is that there is no legitimate basis for holding addicts responsible for their acts as addicts because they are (arguably) responsible for becoming addicts in the first place. The tracing strategy is unavailable because generally unjust, and the alternative analysis does not justify such responsibility in the par-

119. Michael Moore, *Liberty and Drugs*, *in* DRUGS AND THE LIMITS OF LIBERALISM: MORAL AND LEGAL ISSUES (Pablo De Greiff ed., 1999).

ticular case of addiction. The result is that we must explore the moral responsibility of addicts for their acts of use and acquisition on its own terms, unaffected by whatever responsibility addicts might have for becoming addicts.

B. The Main Normative Question: Are Addicts Partially or Wholly Excused by Their Addiction for Acquiring and Using Drugs?

Despite my having earlier included the normative question of whether addicts are blameworthy for becoming addicts, our discussion of tracing shows that that is not the interesting normative question about addiction. My own view of liberty is that if people want to go to hell in a handbasket, they should be allowed to do so as long as they do not transgress other's rights.[120] But however one comes out on this normative question, the folk-psychological nature of addiction does not figure into the answer in the relevant way. Such a normative question only involves addiction as a bad state of affairs that one's actions of nonaddicted use might cause; addiction does not enter in as a potential excuse because one is not yet an addict when one does the acts that make one an addict. It is the doing of actions while addicted that allows addiction, potentially at least, to play an excusing role, and it is to that question that I will direct my attention.

Even within this category it is common to distinguish addicted acts of use from addicted acts to acquire possession of drugs.[121] But no such distinction is needed for our purposes here. As Thurgood Marshall pointed out in his plurality opinion in *Powell*, if addicts are compelled by their addiction to use drugs, they are equally compelled by that same addiction to do what they have to do (such as lie, cheat, and steal) to acquire the drugs they "have to" use.[122] True enough, for such degree of compulsion to amount to an excuse depends on the degree of wrong done; as J.L. Austin once said, it takes a lot better excuse to excuse stepping on a baby than to excuse stepping on a snail.[123] And killing to get drugs is worse than stealing to get drugs is worse than using drugs—so actual degree of excuse can vary with respect to the seriousness of these different wrongs. But Marshall's point is still secure: the degree of compulsion and incapacitation occasioned by the defects of practical rationality earlier discussed, need not vary between these different kinds of acts.[124] In what follows, I thus lump addicts' acts of use with addicts' acts of acquisition. In each case the question is the same: do any of the "models" (folk-psychological explanations) earlier explored instantiate some plausible version(s) of moral excuse?

120. Michael S. Moore, *Liberty and the Constitution*, 21 LEGAL THEORY 156 (2015); Moore, *Liberty and Drugs*, *supra* note 119.

121. The dissenting Justices in *Powell v. Texas* tried to urge that the two acts must be distinguished with respect to the excuse of compulsion because addiction could compel use even though addiction could not compel stealing or other wrongful acts of acquisition of drugs by addicts. 392 U.S. 514, 569–70 (1968) (Fortas, J., dissenting). Gideon Yaffe seems to buy into this view in his regarding use and possession as constituting addiction (by being symptoms of addiction) whereas theft etc. can be caused by addiction but are not constituted by addiction. *See* Yaffe, *Are Addicts Akratic?*, *supra* note 97, at 210.

122. *Powell*, 392 U.S. at 534.

123. *See* Austin, *supra* note 93, at 20.

124. *Powell*, 392 U.S. at 534.

1. The Fully Rational Addict

If addicts are people who do a lot of the behavior to which they are addicted (*e.g.*, use drugs) because: that is what they choose to do; such choices conform to what they most want and most value (even though they may have conflicting desires and conflicting evaluative beliefs); such overall wants and overall evaluations are in character for them; and doing such actions brings precisely the satisfactions they anticipated when they chose to do them—if, in other words, full practical rationality holds for addicts in their use and acquisition of drugs, then pretty obviously addicts have no excuse because of their addiction. After all, on this view such behaviors by addicts differ not at all from the behaviors for which we rightfully praise and blame nonaddicted people all of the time.

The moral implications of the rational choice model of addiction are thus univocal and clear. The problem is whether many of those properly regarded as addicts conform to this model in their psychology. That some do I take to be uncontroversial; whether all do is another matter. True enough, there is a large social science literature, documenting that spontaneous remission for drug addicts is common,[125] and that age and experience alone seems to lead many addicts to quit.[126] The rational choice model thus has more respectability than merely being a mantra of the conservative right in American politics. Yet these results of social science do not show that the behaviors, choices, wants, values, characters, and enjoyments of addicts are as is predicted by the model of full practical rationality. That age, for example, matters to continued addiction could be because as one matures the cravings distinctive of addiction lessen or disappear; or it could be that increased maturity brings with it greater will power, *i.e.*, capacity to overcome temptation. It need not be because overcoming addiction when one gets older shows that one could have overcome it when one was younger.

Similarly, the facts about successful spontaneous remission at any age do not necessarily support the universality of the rational choice explanation. True enough, that an addict recovers from his addiction without outside help requires that at some point he must have had the wherewithal to "just say no" to the use of drugs on some occasion and the wherewithal to keep saying no on the occasions that followed. Yet three points: First, that there was a real possibility for success for certain individuals does not mean that equal possibilities of success exists for others; second, even for a given, successfully remitting addict, it may be that success requires that a particularly propitious set of environmental and psychological circumstances co-occur, and those circumstances may well not be present on other occasions of use or acquisition for which the addict is being held responsible; third, mere statistical rates at which people behave in certain ways

125. See the summary in GENE HEYMAN, ADDICTION: A DISORDER OF CHOICE 132–33 (2009).
126. *Id*. at 70–71.

(such as quitting an addiction without help) are not determinative of the degree of difficulty there may be in behaving in those ways.[127]

2. The Addicts Who Act on "Automatic Pilot"

a. Habits and Preconscious Actions

If one took the automaticity model of addiction literally, addicts would not be responsible for their addiction-motivated behaviors. This would not be because addicts were *excused* for such behaviors; rather, it would be because these behaviors were not actions and so there would be nothing to excuse. The cravings of addicts would cause drug use, etc., in the same way that: (1) a desire to kill my old enemy causes my foot to slip from brake to accelerator when I get excited at seeing him unexpectedly in front of my car; (2) my desire to kill my old enemy causes me to dream of his death, or even to kill him in my sleep or under post-hypnotic suggestion; or (3) such desires cause me, when in shock from being shot or being in a hypoglycemic episode, to kill him. For addicts as for all of such cases there would be no actions for which one is responsible, however much those desires resulted in just the behavior that satisfied them.

But no one (I think) thinks that the supposed automaticity of addictive behavior is to be taken so literally. Addicts *act* when they use drugs and when they steal so that they can acquire drugs. Addictive automaticity is thus much more like the preconscious actions that were described before, where habits and skills have developed to the point that *conscious* intention and choice is not needed to successfully execute these routines.[128] Yet such preconscious actions are ones for which the actor is fully responsible because within the control of the actor, as mentioned before. So if addicted behaviors are simply the by-passing of conscious intentions, no diminishment of responsibility is to be found here.

Does the moral conclusion change for that subclass of preconscious actions where our behavior surprises us when we do find ourselves engaged in it—such as finding ourselves nibbling on cake when we had resolved not to have dessert today, or speeding through traffic when had decided there was no reason to hurry today? Wittgenstein once famously quipped that actions are marked "by an absence of surprise"; should we reverse that and think that the presence of surprise marks a bit of behavior as being a nonaction?[129] Surely not. When I munch or speed preconsciously and am surprised and irritated at myself when I see that that is what I have been doing, I rightly regard this as my fault, something for which I can justly be blamed. We control these behaviors as we control the more

127. On this last point, consider the issue of excuse raised in Regina v. Dudley & Stephens (1884) 14 QBD 273 DC. Lord Coleridge speculates that few of us could have resisted the temptations facing Dudley and Stephens but concludes nonetheless that there was no excuse; conversely suppose that quite a few of us would be like the able seaman, Brooks, who did not yield to this temptation – even so, the duress of the circumstances might still excuse because the Brooks-like overcoming of temptation is still very difficult even if more common than one would have thought.

128. *See supra* note 72 and accompanying text.

129. WITTGENSTEIN, *supra* note 16, at 162e.

typical routines and skills which do not surprise us when we pay attention to them.

To the extent that addicted behavior is automatic in the absent-minded way of preconscious behaviors, no excuse is to be found for it. Does it change this moral conclusion if we add strong emotion (such as the cravings of an addict) to the mix? Do actions that are automatic in the sense relevant here—that by-pass conscious intention—become more excusable to the extent that this automaticity is due to strong emotions?

b. Emotion Caused Automaticity

Surely strong emotionality as the genesis of action does not in general tend to diminish responsibility. Being passionate about one's commitments, whether good or bad, does not diminish one's praiseworthiness/blameworthiness for acting on those commitments. Indeed, save for some Kantian fantasies about willing against all inclination, it would seem that some emotionality must be built into all motivated action.[130]

It is more specifically the unhinging of will by emotion that gives rise to some intuition of excuse here. Emotions can incapacitate the will, and, more commonly, make it more difficult to exert self-control. Consider this (true) story. While leading a pitch of near vertical rock that was indeed scary—not because the pitch was particularly difficult but because at the bottom was a swiftly flowing river going immediately under a dark and forbidding snow cave where a certain, cold, and dark death awaited anyone swept therein—the down rope climber was "overcome with fear." She was unable to rid herself of the thought of disappearing into that dark void and its miserable death, and she was unwilling to move further upward. The story has a happy ending—I told her a joke, released the tension and got her to climb to the top of the cliff safely (whereupon she collapsed in a paroxysm of tears as she let her fear have full expression).[131] But suppose the story had no such happy ending and that serious injuries resulted from her freezing mid-cliff—would her fear excuse or at least diminish her responsibility for the harms that would have happened if she had not been successful in overcoming her fear? Plausibly, I think, the answer is yes.

The criminal law is thus not wrong when it eliminates or reduces responsibility for those who do wrongful acts because of their fear of the threat of others (duress when it operates as an excuse), their fear of the "threats" in nature (necessity when it operates as an excuse), and their craving to preserve their own life (self-defense when it operates as an excuse), their anger at the provoking done by some act of the victim (provocation when it operates as an excuse).[132]

130. Moore, *Responsibility and the Unconscious*, *supra* note 69, at 1564, 1668.

131. The joke was from the film, Butch Cassidy and the Sundance Kid, where Butch convinces Sundance to overcome his fear of drowning in the river below the cliff on which they were trapped by admonishing the Kid, "Are you crazy? The fall will probably kill ya." BUTCH CASSIDY AND THE SUNDANCE KID (20th Century Fox 1969). The actual incident was in the Himalayas of South Baltistan, 1988.

132. Moore, *Responsibility and the Unconscious*, *supra* note 69, at 1664.

Might one say the same about the excusability of actions done because of strong cravings for drugs in addiction?

Rather than answer that last question directly, we should distinguish two ways that strong emotionality might do its excusing work. The one way that is relevant to automaticity is the by-passing of intention by emotions directly causing wrongful actions. This is the view that emotions can literally unhinge the will and that when they do the actor does not in any sense—consciously, preconsciously, or unconsciously—choose (intend) to do the act he does.[133]

The first thing to notice about this view is that emotions typically do not do their excusing work in this way. Rather, emotions incline us to do acts we either do not (most) want to do or do not think that we should do.[134] The will is not bypassed in such cases of potential excuse—we typically act intentionally and with full awareness when we yield to the fear aroused by a threat and do what the threatener wants; likewise when we yield to our anger and direct our aggressions to the source of such anger; likewise when we yield to our cravings and shoot up with the drugs that we crave. Emotionality is mostly relevant to reduce responsibility through the conflict it produces behind these choices, which is why I examine this mode of responsibility-diminishment later. When it excuses, emotionality does not typically excuse by eliminating such choices. But what about true by-passing of the will cases, where emotions do "unhinge the will" in the sense that they directly cause wrongful behavior without the mediation of choice or intention? Unless nonexistent, the moral status of such emotionally direct actions is worth assaying.

Suppose one is in a situation where it is wrong to raise one's voice—one is in a library, for example, or on a submarine in war time, or is surrounded by hostiles looking for one's location. Suppose one "raises his voice in anger," *i.e.*, one shouts because one was angry but not in order to show anyone (or otherwise to express) that one was angry. I take it that there need be no choice to raise one's voice, that intention can indeed be by-passed in such a case. Even so, there is surely some blameworthiness to be attached to the actor for shouting when he really needed not to. Not the blameworthiness for intended or intentional wrongdoing—for having bypassed intention, these are not cases of intentional voice raising. But something like the judgment of negligence, not in the sense of not adverting to some risk that a reasonable person would have adverted to, but in the sense of not maintaining the vigilance over oneself required in these situations.

To the extent addicts use drugs because their cravings directly cause their usings in this intention-by-passing way, they too would have a lesser, but not no, responsibility for their actions. Still, the stubborn psychological facts make this moral conclusion of limited interest in this context. Those facts are that most of the acts of addicts in which we are interested, are not by-passing of intention cases. Not of the forgetful kind earlier discussed, and not of the emotion-driven

133. *Id.* at 1668
134. *Id.* at 1668–69.

kind discussed here. Most addicts use drugs and do what they have to do to acquire drugs, by consciously choosing to do such acts of use and of acquisition.

c. Dual Intentions Automaticity

The psychological rarity of the third model of automaticity that we discussed—the dual intention model—should also preclude too extended a discussion of the moral status of such dual intention, automatic addicts. Briefly: we should think of such rare specimens as we do encounter as suffering from what might be called, "conative dissonance." Just as cognitive dissonance is the simultaneous maintenance of two conflicting beliefs, so conative dissonance is the simultaneous maintenance of two conflicting intentions.[135] We might thus analyze how responsible the latter actors are by seeing how responsible are their cognitive analogues.

When an actor believes that there must be controlled substances hidden somewhere in the car that he is driving across the border, and yet he simultaneously believes that there cannot be such drugs in the car (because, for example, he has looked at every possible hiding place he can think of and found nothing), is he a blameworthy smuggler when it turns out that there were such drugs in the car? Surely someone who does a wrongful act while in a state of such cognitive dissonance is at least reckless with respect to the aspects of his action that make it wrongful. He does, after all, appreciate that there is a substantial risk that there might be drugs in the car. Notice that to reach this moral conclusion—of some but not the greatest culpability—*we* (as observers/evaluators) have to do what the cognitively dissonant subject did not do, namely, integrate his two conflicting beliefs ("there must be marijuana in the car," and "there cannot be marijuana in the car") into an overall belief ("there is a risk there might be marijuana in the car").

Can we do the same thing to the intentions of the conatively dissonant addict? It is not obvious how. Absent degrees of intention (as there are degrees of belief), how does one integrate (or net out against one another) an intention to take drugs now and an intention not to take drugs now? Perhaps the best one can do here is to describe the conatively dissonant actor's mental state as being "sort of" (or "half of") of an intention, reminiscent of Holton and Berridge.[136] And then, having scalarized intentions, match up a "half-way intention" to a partial responsibility? I find this pretty frothy and murky stuff. Fortunately, given the rarity of this kind of addiction, also stuff that it is academic in a pejorative sense to pursue further.

135. *See Conation*, APA DICTIONARY OF PSYCHOLOGY (2018), https://dictionary.apa.org/conation.

136. *See* Holton & Berridge, *supra* note 72, at 261.

3. The Addicts Who (Unlike Addicts on Automatic Pilot) Do Choose to Take Drugs But Whose Choices Do Not Match What They Most Want or Most Value

Here we examine how we should assess the responsibility of those who, through wishful thinking, conflicts of desire, or conflicts between desire and evaluative beliefs, do not choose to do what they most want to do or most value doing when they choose to use drugs. There are all defects of rationality, as we earlier explored. Are they also excusing?

a. The Moral Relevance of Cognitive Failures by the Wish-Caused Erosion of Rational Beliefs

Let us begin with the addict whose cognitions have been eroded by his craving for some drug. Some addicts' craving seemingly cause them to believe: (i) that there is no conflict between what one most wants and most values, on the one hand, and taking drugs now, on the other hand, even though there is such a conflict; (ii) that this time one will find the drug experience pleasurable, even though one's past experiences should tell one that in fact it will not be pleasurable; or (iii) that one will fail if he tries to quit, even though there is good evidence available to him supporting the opposite belief.

Although one can say, very generally and paraphrasing Aristotle, that ignorance joins compulsion as the other of two kinds of excuse, in point of fact only ignorance (or mistake) about certain things lessens one's responsibility.[137] To lessen responsibility, such ignorance must be about some wrong-making characteristics of one's actions. To be ignorant that there is poison in the drink one is serving, or to mistake a man for a stump at which one is shooting, is relevant to the diminishment/elimination of responsibility, for these are mistakes about the causal properties of an act that make it wrong. To be mistaken about the color of the hair of one's intended rape victim, or to not know her parentage, is not relevant at all to our responsibility assessments.

This negative verdict—about the irrelevance of immaterial mistakes to responsibility—does not change when the ignorance or mistake was a necessary condition of the actor doing the wrong that he did. Suppose a rapist would not have raped the victim he did had he known that she was pregnant, or had he known that she was not a natural blond. Ignorance or mistake about such matters was thus necessary for him to have done what he did, because the presence or absence of these factors happened to be motivationally significant for this rapist. Even so, such mistakes/ignorance about facts that are immaterial to the wrongness of what was done—even if material to this particular individual's motivation—are in no way diminishing of responsibility.

We can thus put aside worries about the excusability of addicts whose mistakes were necessary for them to use and acquire drugs. Mistakes about whether

137. ARISTOTLE, NICOMACHEAN ETHICS (Jeffrey Henderson ed., H. Rackham trans., Loeb Classical Library, 1934), BK. III Ch. 1.

drug use will give them pleasure, whether they can or cannot quit, or whether their cravings are or are not consistent with what they most want or most value, are all by-the-by for assessment of their responsibility.

Some would resist this conclusion by thinking that the cravings of addicts interfere with the processes of rational belief formation; and for mistaken beliefs that are necessary for addicts to act as they do in using and acquiring drugs, such interference with belief is therefore a causing of a lessening of the addict's capacity for rational action.[138] The paradigm here might be the interference in belief formation done to Patty Hearst by the coercive indoctrination ("brainwashing") of her kidnappers, the Simbionese Liberation Army.[139] The SLA indoctrinated Patty to believe that her wealthy parents had disowned her and abandoned her to her fate.[140] Suppose this mistaken belief by Patty, although immaterial to the wrong she did (bank robbery), nonetheless was motivationally significant for her, *i.e.*, she would not have robbed the Hibernia Branch of the Bank of America in San Francisco had she known the truth about her parents. The analogy for addicts would be to liken the disturbance in the formation of motivationally significant beliefs done by addictive cravings to the disturbance done by brainwashing. In each case the argument is simple: the actor would not have done the wrong she did if she had not been mistaken about some facts, and that mistake was not her fault because such mistake was caused by interfering factors outside her control.

The error in this argument lies in its assumption that cravings (or brainwashings for that matter) interfere in a morally relevant way with belief formation. The deeper error here is to hold belief formation up to some hyper rational standard whereby no factor that is itself not probative evidence for the truth of a belief, can have any causal role in the production of that belief if that belief is to be rationally held. This is an error because it ignores the commonness of beliefs being influenced by all kinds of factors having little to do with the probative evidence supporting of their truth. Who we are taught by, what happens to grab our attention, what mood (of receptivity) we happen to be in when we confront evidence, etc., etc., all have influence on what we believe. Even though the standard by which we adjudge a belief to be rational lies in a proportionality of degree of belief to the evidence available to support it, many beliefs that are rational by that standard are not formed through a process focusing exclusively on such evidence. Finding excuse for addicts in this cognitive locale would be to romanticize the processes by which rational beliefs are formed.

One last thought about irrationality in the process by which addicts acquire factual beliefs motivating of their acts of drug use and acquisition: wishful thinking is an irrational mode of belief acquisition. And to the extent addicts arrive at their motivationally significant mistakes by such wishful thinking (in the service

138. *See* Holton & Berridge, *supra* note 72, at 265.

139. For those unfamiliar with the then very famous 1974 kidnapping of the Hearst heiress, Patty Hearst, see *Radicals: Patty's Twisted Journey*, TIME (Sept. 29, 1975), http://content.time.com/time/subscriber/article/0,33009,913456-3,00.html.

140. *See id.*

of their cravings), their belief acquisition process is in that way irrational. Yet wishful thinking, like self-deception generally, is hardly an excusing kind of irrationality. (In this it is like weakness of will, which, although deeply irrational, is not as such excusing, as we will shortly discuss.) For mistakes that are the product of self-deception have an active flavor to them lacking in ordinary mistakes. Self-deception requires an active deceiver as much as a passive deceived. Whether this activeness *increases* responsibility for acts and mistaken beliefs resulting from it is a much-debated question.[141] But surely such activeness does not *decrease* one's responsibility for such mistakes.

b. Motivational Failures to Form an Intention That Matches What One Most Wants

The moral issue we face here about desires bears some resemblance to the issue just resolved about beliefs, namely, is there some rational mode of (now desire rather than belief) formation that addictive cravings unduly interfere with, which interference gives rise to excuse? On both the synchronic and the diachronic versions of this model of addiction, cravings are said to monopolize attention, freezing stronger desires out from competition with them, and resulting in choices (to use and acquire drugs) that are not in accordance with what one most wants.[142] Does this diminish the addict's responsibility?

As with belief-formation, this forces us to confront whether responsibility depends on some normal mode of both (component) desire acquisition and (overall) want formation existing such that deviation from this mode is excusing. As was said before, there is a rationalistic picture of this mode that is hopelessly romantic about human capacities and thus that must be put aside. This is the view that we form evaluative beliefs about what is desirable and then choose desires to conform to such evaluative beliefs. Yet like factual beliefs, our desires—both componential and overall—are not (much) up to our choice and our will in this way. Our desires arise within us more than proceed from our choosing them. So it cannot be that addicts are to be excused because their desires do not proceed from their evaluative beliefs (whether the fact that addicts' desires do not conform to, and are not controlled by, addicts' evaluative beliefs, is the topic of the next succeeding Subsection). For all of us, nonaddicts, and addicts alike, would be excused if this were the case.

Holton and Berridge have examined addicts' process of overall want formation in more detail than anyone else; their account repays careful attention.[143] They hypothesize: (1) that addicts component desires for drugs, because of their addiction, spike into those emotional states we know as cravings; (2) that such cravings do not compete with ordinary desires (like keeping one's job, saving one's marriage, etc.) but rather bypass such desires' integration into overall

141. *See* Hurd & Moore, *supra* note 116, at 155–56.
142. Or: wants most of the time, on diachronic versions.
143. *See* Holton & Berridge, *supra* note 72, at 239–68.

wants and intentions in causing drug-seeking behavior directly; (3) that such bypassing is done by the monopolizing of attention on the attractiveness of drug use, robbing all the competing considerations that urge against drug use from getting the attention needed to properly enter into the formation of what one overall wants; and (4) yet that monopolization is not complete but is subject to the addict's efforts of will to focus on (and thus to form and be motivated by) an all-out want that includes these competing desires.[144] From this account Holton and Berridge conclude that "something goes badly wrong with the process by which substance addicts . . . form their desires . . . substance addiction results from the malfunctioning of a normally rational system for creating intrinsic desires."[145]

One aspect of this malfunctioning of rational desires as analyzed by Holton and Berridge—the absence of liking in the process of the desire formation of addicts—I shall defer discussion of until Subsection 5 below. For now consider just aspects (1–4) above.

Surely the drug-induced spiking of desire into the kind of intensively experienced desire we call a craving in (1) above, does no excusing work. Strength of desire, as Aristotle remarked long ago, is surely not on its face excusing of action done to satisfy such a desire.[146] And this remains true even when such desire arose from irrational processes and not because the actor desired to have such a desire and chose to create it (the romantic picture we put aside earlier). Consider in this regard the well-known case of Mr. Ott, a Virginia school teacher who was accused of inappropriate sexual contact with his fourteen-year-old step-daughter.[147] The teacher's sexual desires were seen by the neurologists to be caused by a brain tumor that in no sense was Mr. Ott's fault. Even given this faultless and sudden origin of the desires that motivated his wrongful acts, if Mr. Ott chose to act on such desires when he could have chosen not to do so, surely he is blameworthy nonetheless. Blameworthiness of choice and action does not depend on blameworthiness in desire origination.

As to the "asociability" of desires in (2) above, whether this is excusing depends on whether one had some fair opportunity to integrate such component desires into an overall want.[148] And such fair opportunity, on the Holton/Berridge account, depends on what control one has of the monopolization of attention mentioned in steps (3) and (4) of the Holton/Berridge account.

144. *Id*. at 260–65.

145. *Id*. at 265.

146. ARISTOTLE, *supra* note 137.

147. The case is described in Jeffrey M. Burns & Russell H. Swerdlow, *Right Orbitofrontal Tumor with Pedophilia Symptom and Constructional Apraxia Sign*, 60 ARCHIVES OF NEUROLOGY 437 (2003).

148. The same point was often made about the supposed "implantation" of evaluative beliefs in cases of brainwashing like that of Patty Hearst. Even if such beliefs were suddenly arising through no act, choice, or fault of Patty, still, after the passage of enough time (Dan Dennett gave her about two weeks, if memory serves) in which Patty could integrate such beliefs into her evaluative system one way or the other, she was responsible for acting on such beliefs. Whether Patty had such a fair opportunity to accept or reject such implanted beliefs depended not just on the amount of time but also on whether she was in some fugue-like, disassociated state making it difficult or impossible to compare her implanted beliefs with her contrary beliefs.

Everyday life confirms the limited control we have of obsessional thoughts. The melody that we cannot get out of our head ("It's a small world after all . . ."), the insult we cannot forget, the image of a loved one that will not disappear, are familiar experiences of obsessional thoughts by persons who are neither addicts nor obsessive-compulsive neurotics. Such experience confirms two truths about this phenomenology of addicts here: (1) It can be difficult to be rid of the thought of using drugs once that thought has come into existence by some drug-related cue in the environment; and (2) yet it is not impossible to do, mostly via indirect strategies of distraction and supplantation rather than a direct strategy of willing away. The first of these truths is enough to reduce the responsibility a little in line with the small break we accord to obsessional neurotics; the second is enough not to eliminate such responsibility entirely.

c. Normative Failure to Form an Intention that Matches What One Most Values

As we have seen, it is possible that an addict, at the time she acts,[149] to most want to take drugs despite the costs to jobs, relationships, etc., such use will cost her. Yet the striking experience of many addicts—the ones Frankfurt called "unwilling addicts"—is that such most wants (and the choices and actions they lead to) go against what, at the time they act, they judge to be of most value.[150] These are the addicts whose wants, choices, and acts fly in the face of their better judgment.

Acting against one's own best judgment is a phenomenon to be found in a far broader range of cases than just those of addicted behavior. It is the favored form of locution of many spurious claims of excuse, from the pathetic males who cannot take a rejection by their supposed love object and therefore kill her even though they knew it was wrong to do so, through the righteous religious zealots who proclaim their sinfulness yet repeatedly choose to do what their evaluative beliefs tell them they should not do, to the childish, "the Devil made me do it." The interesting question is whether there are any genuine forms of such an excuse, and, if there are, whether any cases of addictive behavior presents a plausible instance of such excuse.

The weakness of the case for there being any excuse to be found in these environs, can be glimpsed by noting that some recent moral theorizing regards such cases as the very paradigm of responsible and blameworthy action. Gideon Rosen, Michael Zimmerman, and Douglas Husak have all urged recently that those who choose contrary to their own best moral judgment are the *most* blameworthy of wrongdoers, not the least.[151] Their point: Acting in the face of knowing

149. The qualifier as a reminder of Gideon Yaffe's observation that such an overall want may well be temporary in the sense that shortly before and shortly after the act of use the addict may well most want not to take (or not to have taken) the drug. *See* Yaffe, *supra* note 97, at 206.

150. Harry G. Frankfurt, *Free Will and the Concept of a Person*, 68 J. PHIL. 5, 12 (1971).

151. DOUGLAS HUSAK, IGNORANCE OF LAW: A PHILOSOPHICAL INQUIRY 170 (2016); MICHAEL ZIMMERMAN, LIVING WITH UNCERTAINTY: THE MORAL SIGNIFICANCE OF IGNORANCE (2008); Gideon Rosen, *Culpability and Ignorance*, 103 PROCEEDINGS OF THE ARISTOTELIAN SOC'Y, 61, 83 (2002).

the nonmoral facts by virtue of which some act is wrong is bad enough because it reveals a will that is nonresponsive to the moral reasons to which such facts give rise; but to act in the face not only of that knowledge, but also in the face of knowledge that what one is doing is morally wrong—really wrong, not just believed by others or by convention to be wrong—reveals a will that is truly vicious.[152] Acting contrary to one's better judgment on this view, becomes a harbinger of culpability, not a defeater of it.

Those who find excuse for addicts in these quarters seek to distinguish the addicts' kind of choosing against judgment from the cases envisioned by Husak, Rosen, and Zimmerman. Addicts, they say, are not like the proud Nietzchean uber-mensch who overcomes his own values in his choosings and who identifies with those choosings more than with his own evaluations; rather, addicts identify with their evaluative beliefs and so regard those choosings that are against such beliefs as if they were made by someone or something else. In the terms of later Frankfurt, addicts wholeheartedly endorse their evaluations of what is desirable, right and good, and are disappointed in those aspects of themselves "that are not really me" when those aspects chose as they did.[153]

Such phenomenology is certainly a possible one. It is also plausible that it is indeed the phenomenology of at least some addicts. But it is also the phenomenology of many others who deeply regret their wrongful decisions, who say things like, "I don't know what came over me," who view their decisions as not fully their own but as having been made by some other agency within them. As Freudians say, such people see their choices, and the cravings behind them, as "ego-alien," as belonging to a "not-me," an alien thing, an it, an "id."[154]

There are occasions where I think such experiences are excusing. These are the occasions where there is a breach in the continuity of the consciousness of the actor, so that the "not-me" decision is made in an altered state of consciousness (one is asleep, unconscious because in shock, has the amnesia between the personalities of a multiple personalitied person, is under hypnosis or posthypnotic suggestion, is in some fugue state induced by torture or brainwashing, etc.). But where there is no such break in consciousness, those items that are "not-me" are identified solely by the sense of identification of the actor. This is troublesome for reasons I expressed in earlier work (not particularly in relation to addiction):

> [T]he worry [is] that we as moral agents have limited normative power to map out the domain of excuse for ourselves by our self-identifications . . . that our own self-identifications . . . can make us excused . . . is troubling. Seemingly the size and boundaries of our moral agency is not up to us in the way or to the degree that this doctrine of excuse suggests.[155]

152. *See, e.g.*, HUSAK, *supra* note 151, at 170.

153. *Frankfurt*, *supra* note 88, at 33.

154. Morris Eagle, *Anatomy of the Self in Psychoanalytic Theory*, *in* 2 NATURE ANIMATED: WESTERN ONTARIO SERIES IN THE PHILOSOPHY OF SCIENCE 133, 146 (Michael Ruse ed., 1983).

155. Moore, *The Neuroscience of Volitional Excuse*, *supra* note 67, at 200.

It is too plainly self-serving to declare that one's wrongful choices are due to some bad agency within us that we do not identify as being part of our self. Even Freud himself once scornfully remarked that he would "leave it to the jurist to construct for social purposes a responsibility that is artificially limited to the metapsychological ego" and that this would be to "disregard the evil in the id" and "not make my ego responsible for it."[156] True responsibility is for the choices of the entire self, and that self includes both the choices and desires behind them that go against what one most values.

Even if we were to credit this "divided self" approach to excuse, we would still face the issue of self-control by this narrow self (the self-identified with evaluative belief). This judgment (by the narrow self-consisting of our evaluative beliefs) surely has some power to restrain the cravings and to issue in choices to refrain from using drugs; we would thus need to assay whether addiction diminishes this power, and if so, whether that diminishment can be the basis of moral excuse. Because there are precisely the issues we have to confront in judging responsibility for addicts who are weak of will, I shall discuss them in that context.

4. The Akratic Addicts Who Act Against Their Own Intentions Not to Take Drugs.

The akratic addict (in the narrow sense of "akratic" I earlier stipulated) exhibits no defects of rationality up to and including his choices. That is, he believes that it is inconsistent with much that he desires to take drugs, he most wants not to take drugs, he firmly evaluates that such abstinence is the right course of action, and he does not act automatically but rather, chooses not to take drugs, which choice matches in content both which he most wants and most values. And yet the akratic addict intentionally takes drugs nonetheless. The synchronically akratic addict takes such drugs at the very time at which he has the beliefs, wants, values, and intentions directing him not to; the diachronically akratic addict takes such drugs during a temporary reversal and replacement of his beliefs, wants, values, and intentions with those of opposite content. Our current question is whether such gross and plain irrationality excuses.

As a first cut at answering this moral question, surely the intuitive answer is no. Such weakness is not only not excusing; it is itself a form of moral shortcoming. When St. Paul complains in Romans vii that "the good which I want to do, I fail to do" and that "what I do is the wrong which is against my will . . . ," he was not exonerating himself, he was blaming himself.[157] Such weakness to

156. Sigmund Freud, *Moral Responsibility for the Content of Dreams*, *in* STANDARD EDITION OF THE WORKS OF SIGMUND FREUD 133–34 (1961).

157. Romans 7:19–20.

do what one knows is right has perhaps a contemptible cast to it that fully affirmed and willed evil does not,[158] but both on their face are morally condemnable, not excusing.[159] True, we have limited capacities to strengthen our will in general by will-power-building exercises, and even less strengthening capacity in particular cases by willing our self to be stronger of will. But that is true of our ability to shape our desires too–yet no one thinks that my insufficient concern for others, my hatred of some virtuous person, or my fondness for watching others suffer, excuses me just because these attitudes, desires, or emotions are difficult to eliminate or even substantially change very much. Some aspects of who we are ground our blameworthiness for our actions even when those aspects are not subject to our willing them to be otherwise.

So as a first cut a rejection of there being any general excuse of lack of will-power seems appropriate. I take it that Anglo-American criminal law recognizes this moral truth in its doctrines of duress and provocation.[160] The Model Penal Code allows the excuse of duress only when the threats are such that a "person of reasonable firmness" would have been unable to resist them.[161] Such a restriction seemingly eliminates weakness of will as a legal excuse. Similarly, the common law's partial, provocation defense to murder requires that the provoking act of the victim be such as would make a "reasonable person" lose his powers of self-control over his anger.[162] One of the attributes that makes a person "reasonable" in this context is that he has the power to control emotions (like anger) possessed by a person of reasonable firmness. The hot-tempered, impulsive, pugnacious, emotionally explosive, unthinking brutes get no excuse under such a standard, no matter how deeply and how demonstrably they lack the power to control their emotions because their will to do so is weak.

Yet apart from criminal law's confirming morality's denial that there is any *general* excuse of weakness of will, the criminal law more interestingly evidences a more subtle moral truth: for sometimes weakness of will—inability to effectuate one's intentions formed at an earlier time—is an excuse. For sometimes the lack of will power is *not* a moral defect in the person who lacks it. An easy example is intoxication. It is common for intoxication to loosen the inhibitions of the intoxicated person. In such a state he has less control over his emotions of fear or of anger, and of the desires that they spawn, with the result that he can maintain his resolve (earlier intention) less successfully over time. When the intoxicated state is not his fault—as it is not in cases of involuntary intoxication—then he has a more plausible, perhaps partial excuse of weakness of will.

Youth is another easy example. The time-discounting is steep, the impulse control poor, for young people as opposed to adults. And this is not their fault—

158. R.A. Duff, *Virtue, Vice, and Criminal Liability: Do We Want an Aristotelian Criminal Law?*, 6 BUFF. CRIM. L. REV., 147, 164–65 (2002).

159. *See* HILL, *supra* note 95, at 135–37.

160. *See* Alan Reed, *Duress and Provocation as Excuses to Murder: Salutary Lessons from Recent Anglo-American Jurisprudence*, 6 J. TRANSNAT'L L. & POL'Y 51, 51–52 (1996).

161. MODEL PENAL CODE § 2.09.

162. R.A. Duff, *The Virtues and Vices of Virtue Jurisprudence*, *in* VALUES AND VIRTUES: ARISTOTELIANISM IN CONTEMPORARY ETHICS 90, 95–96 (Timothy Chappell ed., 2006).

being young is not a moral defect (no matter how much one might have aesthetic complaints about teen-agers). The chronologically immature have not yet had a fair opportunity to develop into adults of whom we may fairly expect a higher standard of self-control. So young wrongdoers too have some excuse of weakness of will.

It is not that the involuntarily intoxicated, the young, and others with blamelessly lowered abilities to maintain their resolve in the face of fear, anger, cravings, or other emotional states, get a complete pass. For they only have a *lesser* capacity to control themselves, not *no* capacity. Anglo-American criminal law recognizes this last fact by asking whether a particular young or intoxicated person did as well as could fairly be expected of one with the lessened capacity typical of those similarly young or drunk.[163] But where there is no unexercised capacity as judged by this lesser standard, then there is excuse.

So there is some room for a viable excuse of weakness of will. How much depends on how many conditions there are where two things are true: (1) The power of self-control is lessened from what we normally demand of persons generally; and (2) it is not a moral defect in such persons to have such lessened powers of self-control.

It is a common feature of how addiction is experienced by addicts that they have a diminished capacity to maintain their resolve (not to take drugs) in the face of opportunities to use drugs.[164] This "stickiness of intention" as part of self-control is poor for them as a class. The more difficult question is that of fault for having this lessened ability to control themselves. For almost all addicts are voluntary consumers of the drugs that eventually addict them, and it is those early choices that make their loss of control later as addicts, in some sense their fault.

I earlier concluded that addicts were not to be blamed for their later acts of use and acquisition of drugs, just because they were at fault for becoming addicts.[165] I put aside the tracing strategy as generally unacceptable, and I interpreted the alternative analysis not to yield justified judgments of blame for the later, more wrongful acts as addicts. Here we encounter a third use of the fault of addicts for becoming addicts: when we address the question of excuse at the later time ("t_2"), we are right to do so because the tracing strategy does not mandate a forfeiture of excuses because of earlier fault. Yet *this* excuse—lessened control capacity—has a no-fault condition built into the very nature of the excuse. This condition is sometimes called "the moral baseline" requirement for compulsion excuses generally, including duress.[166] The requirement is that one's character (by virtue of which one has lessened control capacity) not itself be aretaically blameworthy, for if it is no amount of control incapacity will excuse.

163. The special standard accorded various classes of wrongdoers has received much attention from criminal law scholars. *See, e.g.*, TADROS, *supra* note 91, at 349–58.

164. *See* Holton & Berridge, *supra* note 72, at 241.

165. *See supra* Section IV.A.2.

166. *See, e.g.*, Claire Finkelstein, *Duress: A Philosophical Account of the Defense in Law*, 37 ARIZ. L. REV. *passim* (1995).

Consider the extremely short-tempered and pugnacious man, one who with little provocation goes off the handle and answers modest insults with violence. We hold him not to be excused even if he did as well as could be expected of such short-tempered, pugnacious men; his lousy temper and his violent disposition are settled character traits of his; and they are vicious.[167] He undoubtedly acquired such traits by a lifetime of choices and acts of hot-tempered, hair-triggered responses to others. But notice the unavailability of excuse is not due to tracing his pugnaciousness to those earlier acts and choices and forfeiting an excuse (of volitional incapacity due to anger) because of the earlier bad choices. Nor do those earlier acts stand in direct enough causal relationship of the later harms done by the pugnacious man to his victims, to hold him liable to blame because of the alternative analysis earlier explored. Rather, the pugnacious man has no excuse because his settled character is that of viciousness—pugnaciousness—and the nonviciousness of the trait that incapacitates the will is a condition of the excuse of volitional impairment.

This third way of taking earlier fault into account in assessing later responsibility, requires that we face squarely the question largely ducked before: how blameworthy are addicts for being addicts? Is being an addict a vicious trait whose incapacities of will are thus ineligible to excuse in the way that the incapacities of pugnaciousness are ineligible to excuse? With regard to the earlier acts of nonaddicted use that cause one to become an addict, my own view, defended elsewhere, is that use of drugs and alcohol is either not a wrong at all, or is a wrong no one has a right that one not do.[168] But since the issue is not that of holding addicts responsible because of their earlier acts of use—as it is for both the tracing strategy and the alternative analysis—the nonblamability of earlier acts of nonaddicted use does not fully answer our present moral question.

Suppose that someone has become a very lazy person through a lifetime of choices that always shunned ambitious projects. Surely even the laziest of persons can rouse themselves out of their laziness if the occasion is dire enough to demand it, but even so, equally surely it is harder for lazy people to overcome their inertia and actually help others in need on occasions where it requires some effort to do so. My own judgments here are these: (1) Lazy people are entitled to the lazy choices they make; even if they have the natural talents of a Mozart, they are entitled—they do no wrong—in wasting their time and their talent in idle pursuits. (2) Despite the foregoing, laziness is still a vicious disposition in the sense that it results in a life poorly lived. Mozart's life as a composer was better than a Mozart life as a dissolute. (3) The viciousness of laziness precludes the incapacities of laziness to be used as an excuse when the lazy person fails to do what he should do such as buy food for his children.

I would make the same judgments about addiction. Even though nonaddicted use of drugs is not a wrong (or is a wrong one has a right to do), the life

167. *See also supra* note 94 and accompanying text.

168. Moore, *Liberty and Drugs*, *supra* note 119.

of an addict is a poor life. The incapacities of will that this nonvirtuous trait engenders are thus unavailable to the addict to excuse his wrongful behaviors while addicted.

5. Addicts Who Most Want or Most Value What They Do Not Like

Isolated instances where achieving what we most wanted to achieve fails to bring psychological satisfaction, do not on their face tempt one toward excuse. Andre Gide's Lafcadio is disquieted by his pushing an old man out of a moving train to that man's death, whereas he had anticipated some feelings of satisfaction at proving his surmounting of conventional morality; Lafcadio's disappointment in no way excuses his gratuitous killing of an innocent.

Holton and Berridge observe that addicts do not suffer isolated instances where the (logical) satisfaction of their desires to use drugs results in no (psychological) satisfaction; rather, addicts recurringly desire to use drugs despite their prior experience of doing so giving them so little pleasure.[169] Holton and Berridge paint this as part of the "bruteness" of addicts' cravings for drugs, such cravings being isolated from both anticipations of pleasure and judgment of value.[170] This is a defect of practical rationality; is it also an excuse?

To think that there is excuse to be found here would require that one thinks that the formation of the component desires that make up overall wants must include judgments of anticipated pleasure on the satisfaction of such desires. Yet as we have seen, desire formation is largely an irrational process that is beyond the powers of the will even for normal, nonaddicted people. We cannot choose our desires nearly to the extent that we can choose whether to act on them. Responsibility thus doesn't depend on some normal route of desire-acquisition, a route that addicts can then be said not to follow. This was (I argued earlier) true of implanted desires and implanted evaluative beliefs; it is also thus true of desires arising in ways not responsive to anticipations of pleasure or displeasure.[171]

I did allow that where desires are not affected by what one most values, one might be tempted to identify one's "real self" with the evaluative belief rather than with the desires and the choice that they lead to. I was suspicious of that temptation towards excuse earlier, but in this context there is not even a temptation towards excuse. Our sense of self-identity is not built on our likings as it more arguably is on our valuings, so no ego-alien characterizations of addictive cravings arise simply from the fact that satisfaction of such cravings brings no pleasure.

169. Holton & Berridge, *supra* note 72, at 241.
170. *Id.* at 264–65.
171. *See supra* Section III.C.1.

V. THE PROMISE OF NEUROSCIENCE TO DEEPEN OUR EXPLANATORY AND EVALUATIVE UNDERSTANDINGS OF ADDICTIVE BEHAVIOR

A. The Two Potentials for Neuroscience: Changing (Broadening, Deepening, Correcting) Our Folk Psychological Explanation of Addiction and Changing or Justifying our Present Doctrines of Moral and Legal Excuse

Neuroscientific research has the potential to advance our understanding of addiction in two dimensions. One, it can change our explanation of addiction. This it can do in a variety of ways: (1) It can *deepen* our folk psychological explanations by showing how the variables in such explanations are underlain by the mechanisms of brain function; (2) it can *precisify* the folk explanations by making the folk psychological states more precise in their boundaries or more precise in the modes of their combination; (3) it can *correct* mistakes in the folk psychological explanation; and/or (4) it can *broaden* those explanations by supplementing them with explanations couched in the terms and variables of cognitive psychology, genetics, and neuroscience. Secondly, such research has the potential to change how we evaluate the behavior of addicts. It might show that we should excuse where currently we do not or that we should not excuse when currently we do. Alternatively, our present evaluations of excuse could remain unchanged but they could be supported and justified by neuroscientific explanations, showing us that addicts are incapacitated to the point of excuse just where and to the extent that we currently think that they are.

I shall assess each of these potentials for neuroscientific research in this, the fifth part of this Article; but the beginning of wisdom here is to keep the potential *explanatory* work done by neuroscience, separate from the potential *excusing* work done by neuroscience. These two uses of neuroscientific research can be related—for a true excuse does depend on a certain form of explanation (of the behavior being excused) being true—yet they are not related in the simple-minded ways that are so rampant in the neuroscience literature and that we have discussed before.

One of these simple-minded ways (of moving too quickly from explanation to excuse) is that of the incompatibilist. Such a person believes that mechanistic, causal explanations necessarily excuse because they show the actor had no freedom to do other than he did, such as take drugs.[172] On such a view, to show how unusually large releases of dopamine in certain areas of the brain cause early drug use, or to show how decreased releases of dopamine cause continued drug use by addicts, is to show such drug usages to be excused.[173] On such a view there is no break between explaining and excusing: to (causally) explain is *ipso facto* to excuse. We all need to be more patient here, asking what neuroscience adds to the explanation of addiction, and then asking whether the explanations given by neuroscience shows there to be the kind of incapacity that excuses.

172. Ingo Willuhn et al., *Excessive Cocaine Use Results from Decreased Phasic Dopamine Signaling of the Striatum*, 17(5) NATURE NEUROSCIENCE 704, 710 (2014).

173. *See, e.g.*, *id*.

More insidiously present in the literature in neuroscience explaining addiction, is a second of the confusions that we explored before in Part II. This confusion is evident in the way that such neuroscientific literature orients neuroscientific research findings around the verification of "the disease model of addiction." This may be fine as a bit of medical science, but as we saw in Part II such orientation is without import for questions of moral and legal excuse. Thus, the leading researchers in this area may well set for themselves the medical-scientific task of reviewing "recent advances in the neurobiology of addiction to clarify the link between addiction and brain function and to broaden the understanding of addiction as a brain disease."[174] And it may be good medical science to conclude "that neuroscience continues to support the brain disease model of addiction"[175] because this may well advance the medical goals of finding "new opportunities for the prevention and treatment of substance addictions."[176] What such medical characterization of the research (as "addictions are brain diseases") does not do, is to call into question "deeply ingrained values about self-determination and personal responsibility."[177] Moralists and lawyers thus have no reason to join medical researchers in organizing neuroscientific findings by whether or not they support the view that addictions are diseases of the brain.

Part and parcel of the disease-orientation in characterizing neuroscientific research into addiction, is the characterization of the findings of that research into the disease criteria of dysfunction and disability that we discussed in Part II. One might (with apologies to Nancy Andreasen) call this the "broken-brain" way of characterizing the findings of neuroscience.[178] Thus, some statistically dominant mode of functioning is characterized as "normal" and "proper," and addiction is presented as dysfunctional and improper. Thus: "we have learned that addiction is characterized by an expanding cycle of *dysfunction* in the brain."[179] More specifically, "we introduce the key brain circuits that are affected by the chronic abuse of drugs and then present a coherent model [in terms of four such circuits] according to which addiction emerges as the net result of *imbalanced* information processing in and among these circuits,"[180] each of which changes in ways characterized as "faltering."[181] One of these circuits is the reward circuit, where a decrease in sensitivity to reward caused by abuse of drugs is characterized as a "dysfunction"[182] a "disruption"[183] and a "repeated perturbation."[184]

174. Volkow et al., *supra* note 35, at 363.

175. *Id.*; *see also* Nora D. Volkow & Marisela Morales, *The Brain on Drugs: From Reward to Addiction*, 162 CELL 712, 720 (2015) ("In conclusion, uncovering the neurobiology underlying drug abuse has led to the recognition of addiction as a chronic disease of the brain.").

176. Volkow et al., *supra* note 35, at 363.

177. *Id*. at 364.

178. *See generally* NANCY ANDREASEN, THE BROKEN BRAIN (1984) (dealing with mental disease generally, not addiction specifically).

179. Nora D. Volkow et al., *Addiction: Decreased Reward Sensitivity and Increased Expectation Sensitivity Conspire to Overwhelm the Brain's Control Circuit*, 32 BIOESSAYS 748, 748 (2010) (emphasis added).

180. *Id.* at 748–49 (emphasis added).

181. *Id.* at 748.

182. *Id*.

183. *Id*.

184. *Id*. at 750.

Likewise, the disconnect between normal patterns of dopamine release in the synapses of the ventral striatum is characterized as an "unrestrained hyperactivation of the motivation/[reward] circuit"[185] and as an "improper regulation of brain activity by [those] prefrontal brain regions [part of the control circuit]"[186]

Again, there is nothing improper for medical purposes in characterizing the findings of neuroscience in terms of the dysfunction and disability that makes the disease classification plausible. The mistake is to infer from these criteria of disease being met by addiction, that such (medically classified) dysfunction "results in the compulsive drug intake that characterizes addiction" or that drug use by addicts is "an automatic compulsive behavior."[187] The incapacities that lead to the excuse of compulsion are not necessarily the same as, and do not necessarily follow on, the dysfunctions and disabilities that rightly characterize disease.

With these cautions in mind, I turn to the explanation of addiction offered up by the neuroscience of the last sixty-four years.

B. The Neuroscientific Explanation of Addiction

Despite the above separation of explanation from excuse, our ultimate interest here is the excusing effect (or lack of it) of drug addiction. So it is useful to organize explanations in a way that is congenial to later discussions of excuse.[188] We should thus separate the explanations proffered up by neuroscience into two camps (for this will track an earlier moral distinction). The first camp will be to explain why *nonaddicted* persons use drugs and thus (in some cases) become addicts; the second camp will be to explain why those who are addicts use and (do what they have to do to) acquire drugs.

185. *Id*. at 748.

186. *Id*. at 751.

187. *Id*. at 748.

188. In truth some of the neuroscience literature on addiction is already organized with an eye to excusing addicted drug taking. Consider this summary of her explanation of drug use by addicts by Nora Volkow and her associates:

> Some of the most pernicious features of drug addiction are the overwhelming craving to take drugs . . . and the severely compromised ability of addicted individuals to inhibit drug seeking once the craving erupts . . . . During addiction, the enhanced value of drug cues in the memory circuit drives reward expectation and enhances the motivation to consume the drug, overcoming the inhibitory control exerted by the already dysfunctional PFC [prefrontal cortex] . . . . At the same time, addiction is likely to also recalibrate the circuits that instantiate mood and conscious awareness . . . in ways that . . . would further tilt the balance away from inhibitory control and towards craving and impulsive drug taking.

Id. at 753–54. Such an explanation of why addicts consume drugs is organized as an imbalance between two opposing brain systems, one (the craving, reward-expecting, motivating one) pushing for drug consumption, and the other (the inhibitory, control system) pushing for nonuse. The strength of the one, and the weakness of the other, explain addictive use in a way most congenial to excuse.

1. The Explanation for Nonaddicted Drug Use that Risks and Sometimes Causes Addiction

Within the category of nonaddicted use, let us start with the question of why people use addictive, opioid drugs for the first time. At the folk-psychological level, surely there are a variety of familiar but mundane explanations for this risky behavior: the user is curious, wants to fit in, feels peer pressure, wants to enhance his variety of experiences, wants to take some supposed "voyage of discovery," wants to socialize with his friends, wants to relieve anxiety or stress, wants to relax, wants to do better in some upcoming performance, etc. Whatever this reason-giving account might be, it will be an account framed in the folk-psychological concepts of belief, desire, and intention and an account such that such initial drug use looks like every other voluntary, intentional choice for which we ordinarily hold people fully responsible.

There is very little in the neuroscience literature of addiction that challenges these folk-psychological explanations of initial use, whichever they might be on different occasions.[189] Nonaddicted first users choose to take drugs in the same way and with the same responsibility as they chose what food to eat, what kind of sex to have, what vacations to take, etc. There are of course the general challenges that are the subject of my book, challenges according to which *all* folk-psychological explanations are suspect (reductionism, determinism, epiphenomenalism, and fallibilism).[190] But these are the supposed implications of general aspects of neuroscientific explanation; they are not based on neuroscientific explanations *of drug use and addiction,* our present interest. (Besides, I have in the book hopefully blunted these general challenges, so that when we do have a comprehensive neuroscience of BDI psychology it will support conclusions of responsibility, not challenge them.)

There are also of course the familiar kinds of explanations offered up by behaviorist psychologists and geneticists that also seek to undermine the folk-psychological accounts on which responsibility for initial, nonaddicted drug use are based. Such behavioral or genetic explanations refer to stressful environments such as poverty, lack of opportunity, emotional abuse, impulsiveness, risk preferences, steep time discounting of the young, and the like, all of which are taken to explain why first-time drug users turn to drugs. Yet these, too, are the familiar competing explanations offered up by science generally to folk psychological explanation. If they defeat rather than support and supplement folk psychological explanations, that will not be because of any weaknesses unique to the folk psychological explanation of why nonaddicts use the drugs that make them addicted.

I thus take the folk-psychological explanation of why first-time users take drugs to be secure and not even challenged by the neuroscience of drug use and

189. *See* Volkow & Morales, *supra* note 175, at 712 ("[I]nitial drug experimentation is largely a voluntary behavior . . . .").

190. MICHAEL S. MOORE, MECHANICAL CHOICES: THE RESPONSIBILITY OF THE HUMAN MACHINE (Oxford Univ. Press) (forthcoming 2020).

addiction. Let us then move to explanations of the acquisitive and using behaviors of nonaddicts who are further down the road to addiction but are not yet there. Such behaviors too will have a number of folk-psychological accounts of them that will be plausible. But for these explanations, unlike for those explaining initial drug use, the neuroscience of drug use and addiction will propose truly competing and thus debunking explanations framed in terms of the causal effects of prior drug use in the brain.

The older neuroscience literature told a story here that came to be characterized as the "demon drug" story.[191] Drug use by anyone, addicts, nonaddicts, and first time users alike, was said to "hijack" the mesolimbic system.[192] The mesolimbic system was thought to be the reward system in the human brain. It includes the ventral striatum and particularly the nucleus accumbens within that striatum. Although no serious researcher on addiction today buys the complete story I am about to tell,[193] there are several reasons to tell the story anyway. One, it still grips the popular imagination of some sizable portion of the public when it thinks of the responsibility of drug users.[194] Two, the story's problems lead naturally into succeeding accounts of the effects of drug use that better fit the evidence. Three, there is much in the story that is true and on which subsequent accounts are based. And four, the story illustrates the kind of dramatic impact neuroscience might have on our ordinary, folk explanations of recreational drug use and on our willingness to excuse such use.

The story begins in 1954. In that year James Olds and Peter Milner published the results of their research discovering that rats had a distinct part of the

191. This is the intentionally unflattering name given the story by one of its critics, Bruce Alexander. Bruce K. Alexander & Linda S. Wong, *The Myth of Drug-Induced Addiction*, BRUCE K. ALEXANDER'S GLOBALIZATION OF ADDICTION WEBSITE: ARTICLES & SPEECHES, https://www.brucekalexander.com/articles-speeches/demon-drug-myths/164-myth-drug-induced (last visited Jan. 21, 2020); Bruce K. Alexander, *Rat Park Versus the New York Times,* BRUCE K. ALEXANDER'S GLOBALIZATION OF ADDICTION WEBSITE: ARTICLES & SPEECHES, https://www.brucekalexander.com/articles-speeches/281-rat-park-versus-the-new-york-times (last visited Jan. 21, 2020). Alexander became well known in the late 1970s for his controversial "rat park" experiments which purported to show that stressful environment (living in cages versus living in a kind of rat paradise) had more to do with excessive drug use by rats than did the rewards of prior use. Bruce K. Alexander et al., *The Effect of Housing and Gender on Morphine Self-Administration in Rats*, 58 PSYCHOPHARMACOLOGY 175, 178 (1978). Popular interest in Alexander's rat park experiments was briefly revived by the (even more controversial) journalist Johann Hari's much watched TED talk of 2015, "Everything You Thought You Knew About Addiction Is Wrong," based on his book, JOHANN HARI, CHASING THE SCREAM (2015); Johann Hari, *Everything You Think You Know About Addiction Is Wrong*, TED CONFERENCE: TEDGLOBALLONDON (June 2015), https://www.ted.com/talks/johann_hari_everything_you_think_you_know_about_addiction_is_wrong?language=en.

192. Leshner's now much adopted term. Leshner, *supra* note 20, at 46.

193. Not in print anyway. The website for the National Institute for Drug Abuse, "Understanding Drug Use and Addiction," as of March 20, 2019, still tells the crucial part of the demon drug story, the dopamine pleasure hypothesis, even though that is known to be false. *Understanding Drug Use and Addiction*, NATIONAL INSTITUTE ON DRUG ABUSE (June 2018), https://www.drugabuse.gov/publications/drugfacts/understanding-drug-use-addiction.

194. As Alexander also notes: "In popular culture the Demon Drug Myth has survived almost intact. In its most recent incarnations, it says that all or most people who take one of the demon drugs . . . lose their will power and are converted into hopeless addicts . . . . [Such people] are still said to have been robbed of their will power, as if the drug had 'flipped a switch' in their brain." Alexander, *supra* note 191.

brain associated with pleasure and reward.[195] They discovered this almost by accident, doing a general mapping of brain regions and not looking specifically for a reward center (the conventional wisdom being at the time that rewards would be system-wide brain functions, not locatable in any particular area of the brain).[196] They found that electrodes initially placed into the septal area of the brain (and later into the nucleus accumbens located adjacent to the septal area) would induce rats to repeatedly press a bar that activated those electrodes.[197] Indeed, the reward the rats seemingly experienced from such stimulation dominated other naturally occurring rewards rats otherwise sought.[198] As they put it in their article, "the control exercised over the animal's behavior by means of this reward is extreme."[199] Some rats would press the bar nearly 2000 times per hour over a 24 hour period even if they were hungry and food was available.[200] As Olds concluded, "a hungry animal often ignored available food in favor of the pleasure of stimulating itself electrically."[201]

Parallel discoveries were made in the early 1960s at the University of Michigan about the proclivity of rats to press bars delivering various drugs (such as amphetamines) directly to the nucleus accumbens: the rats would self-stimulate their pleasure center (the nucleus accumbens) with drugs as single-mindedly as with electrical stimulation.[202] And then, in 1975, the mechanism common to both electrical stimulation of the nucleus accumbens and direct delivery of drugs to that brain structure, was discovered: the increase of the dopamine neurotransmitter in the synaptic clefts of the nucleus accumbens.[203] Wise and Yokel concluded that it was the levels of dopamine in this brain location that was affected by both electrical stimulation and drug delivery there and that produced like behaviors in rats: "We offer the hypothesis that normal functioning of a dopaminergic mechanism is essential for the perception of the rewarding consequence of amphetamine and intracranial electrical stimulation."[204] They also ventured the speculation that it was this mechanism that was essential for all forms of experienced pleasure, naturally occurring as well as drug or electrically induced: "The same mechanism may also be involved in the perception of the reward properties of naturally occurring reinforcers."[205] As the popular press was later to summarize their speculation, now extended to the human brain: "At a purely chemical level, every experience humans find enjoyable—whether listening to music, embracing

195. James Olds & Peter Milner, *Positive Reinforcement Produced by Electrical Stimulation of Septal Area and Other Regions of Rat Brain*, 47 J. OF COMP. PHYSIOLOGICAL PSYCHOL. 419, 426 (1954).

196. *Id.* at 419–21.

197. *Id.* at 421, 425.

198. *Id.* at 421–22.

199. *Id*. at 426.

200. James Olds, *Pleasure Centers in the Brain*, 195 SCI. AM. 105, 116 (1956).

201. *Id.* at 114, 116.

202. See the summary in James H. Woods, *Behavioral Pharmacology of Drug Self-Administration*, *in* PSYCHOPHARMACOLGY: A DECADE OF PROGRESS 595, 595 (Morris A. Lipton et al. eds., 1978).

203. Roy Wise & Robert Yokel, *Increased Lever-Pressing for Amphetamine After Pimozide in Rats: Implications for a Dopamine Theory of Reward*, 187 SCIENCE 547, 547–49 (1975).

204. *Id.* at 548.

205. *Id*.

a lover, or savoring chocolate—amounts to little more than an explosion of dopamine in the nucleus accumbens."[206]

Later research showed that drugs effected their "explosion of dopamine in the nucleus accumbens" in a variety of ways, some favored more than others by particular drugs.[207] Some drugs stimulate release of the vesicles containing dopamine molecules from the terminals on the presynaptic neuron; others allow such release by inhibiting the release of GABA molecules from near-by terminals on the presynaptic neuron (which GABA would otherwise inhibit the release of the dopamine vesicles); some drugs work by blocking the transporters that transport dopamine back into the presynaptic neuron, preventing the re-uptake of such dopamine and thus increasing the amount of it left in the cleft; some drugs block the enzyme molecule that otherwise would remove dopamine from the cleft; and some drugs occupy the receptor sites on the postsynaptic neuron and (if they are agonists rather than antagonists) release the ion gates on the postsynaptic neuron without use of dopamine, meaning that the dopamine is not absorbed into that neuron but remains in the synaptic cleft.[208] The effect that all of these mechanisms have in common is to have more dopamine in these synaptic clefts than would have been there without the drug, thus making easier the mode of transmission across the cleft unique to dopamine neurotransmitter molecules.[209]

The assumption was that the message transmitted by excess dopamine was a pleasure message, so that "at a purely chemical level" pleasure just is lots of dopamine in these clefts. As an explanation of addiction, this came to be known as the positive-reinforcer theory of addiction by both its proponents and its critics.[210] As the literature of the day concluded, "drugs are addicting (established compulsive habits) because they produce euphoria."[211] But the view had broader implications than just to explain addiction; the view also purported to explain drug use by anyone, nonaddicts and addicts alike, once they had tasted the forbidden fruit of the demon drugs. All human drug users were likened to the laboratory rats pressing their bars to the exclusion of anything else: "It is the 'taste' of the drug or the experience of stimuli that—through Pavlovian conditioning—cause drug-like central effects that motivate drug intake."[212]

206. J. Madeleine Nash, *Addicted: Why Do People Get Hooked?*, TIME (May 5, 1997), http://content.time.com/time/printout/0,8816,986282,00.html.

207. *Id.*

208. *See generally* R.A. Wise & M.A. Bozarth, *A Psychomotor Stimulant Theory of Addiction*, 94 PSYCHOL. REV. 469 (1987).

209. *See, e.g.*, *id.* at 481 (noting that amphetamine and cocaine "increase[] concentrations of dopamine in synapses of nucleus accumbens, and the stereotypy seems to derive from increased concentrations of dopamine in synapses of the caudate nucleus").

210. *See generally* Roy Wise & George Koob, *The Development and Maintenance of Drug Addiction*, 39 NEUROPSYCHOPHARMACOLOGY 254, 254–62 (2014).

211. Wise & Bozarth, *supra* note 208, at 474.

212. J. Stewart & R.A. Wise, *Reinstatement of Heroin Self-administration Habits; Morphine Prompts and Naltrexone Discourages Renewed Responding After Extinction*, 108 PSYCHOPHARMACOLOGY 79, 80 (1992).

On the demon drug view of things, it is pretty easy to see why one might think that drug users should be excused for any continued use of drugs after initial use. For this is a story of how drug users become "possessed," not by a demon to be sure, but by something likened to a demon; their will, their ability to control themselves, their choice of which desires they should value and act upon, have been "hijacked" by the effects of their prior drug use.[213] Like the ancients of old, they have tasted of the lotus fruit, and once tasted they can never ignore its siren call.[214]

There are six senses of "hijacking" in this story that we need to distinguish. One admittedly modest sense of "hijacking" here lies in the artificial nature of the pleasures induced by drug (or electrical) stimulation.[215] Ordinary pleasures are the by-products of activities we engage in for reasons other than to produce the pleasure they do produce. As is often said in criticism of hedonistic ethics in philosophy, pleasure is rarely an end in itself that we seek; indeed, to seek it directly is usually to fail to achieve it. We value other activities and states of affairs like reading a good book, enjoying a good bottle of wine, taking in the beauty of nature; and although achieving them is pleasurable that is not what motivates us to do them. Whereas with the artificial stimulations of our pleasure center with electrical probes and drugs, we do seek pleasure for its own sake. We do not earn the pleasure by achieving something we think is worthwhile like writing a novel or reading one; for being stimulated by an electrical pulse or giving ourselves a shot are not activities we value for their own sake. Rather, we shortcut the route to pleasure by skipping the pursuit of any activity that gives pleasure and go straight for the pleasure. Drug and electrical stimulation of the pleasure center is the kind of "experience machine" that philosophers have long used in their thought experiments testing what it is we really value, a machine that duplicates the pleasant sensations attendant upon any achievement but without those sensations being caused by any such (nonexistent) achievement. Such machines "hijack" the reality that is supposed to lie behind our achievements and our sentiments substituting pleasure for the goodness and beauty that does and should normally motivate us.

The second aspect of "hijacking" done by drugs in this story is the intensity of the pleasure drugs can give as compared to the ordinary pleasures of life.[216] If dopamine is the chemical basis and marker of pleasure, then more of it in the nucleus accumbens should mean more pleasure is being experienced. Given that

213. Although some patients experience their drug cravings as "demon." Consider this testimony of one crack addict during treatment: "There's a strong-ass demon that messes you up." Fran Smith, *How Science is Unlocking the Secrets of Addiction*, NATIONAL GEOGRAPHIC (Sept. 2017), https://www.nationalgeographic.com/magazine/2017/09/the-addicted-brain/.

214. *See* HOMER, THE ODYSSEY ("[W]e reached the land of the Lotus-eater, who live on a food that comes from a kind of flower . . . the Lotus-eaters . . . but gave them to eat of the lotus, which was so delicious that those who ate of it left off caring about home . . . but were for staying and munching lotus with the Lotus-eater without thinking further of their return.").

215. Smith, *How Science Is Unlocking the Secrets of Addiction*, *supra* note 213.

216. *Id.*

studies show that drugs release much more dopamine than do other, natural rewards, drugs must be that much more pleasurable to experience. On such a view, drugs truly are the lotus fruit of old.

The second kind of "hijacking" was thought to lead to a third and fourth kind. Experiencing the euphoric "highs" or "blasts" unique to drugs was thought to give rise both to the phenomenology of craving for them over all other wants (the third sense), and to the behavioral dominance of drug-seeking behaviors over all other pursuits (the fourth sense).[217] The assumption for the first of these additional kinds of hijacking was that it was the reward of pleasure (often described as the "liking" of drugs) that gave rise to the distinct *wanting* of drugs. This seemed a natural enough assumption because ordinarily much of what we want is formed out of judgments of what we like ("like" in the sense of, "experience as pleasurable"). The assumption of the second of these additional kinds of hijackings was of the extraordinary motivating power of such wants formed from these likings: in the language of Olds and Milner earlier quoted, the control exercised over our behavior by such wants is "extreme."[218] Nonaddicted drug users are like their rats pressing the bar of pleasure stimulation to the exclusion of all else, once such users have discovered the "bar" by initial use.

The fourth and motivational sense of "hijacking" lead to a fifth sense having to do with the mode of our decision-making when nonaddicted users decide to use drugs. Many Nineteenth and Twentieth Century psychology has foundered on some monistic drive theory. A famous example was Freud: all we ever really want when we solve problems in logic, go to the beach, or help our friends, etc., is some form of either sex or aggression.[219] In point of fact our decisions are too various in their motivations and outcomes for those drive theories to have ever been anything more than myths. It is not a familiar feature of our psychology that any one thing dominates our choices. So when the demon drug story posits that our pleasure-driven wants for drugs always win over all the other things we want once drugs have been tried, to be plausible it posits the absence of any real choosing by us to take drugs. Rather, the connection between our wanting drugs and our acting to acquire and use them in seen as a kind of Pavlovian-conditioned response (as Stewart and Wise were quoted earlier).[220] We are seen as being on automatic pilot when our acts are the products of such wants.

A sixth aspect of the demon drug story, one giving a sixth sense in which drugs may be said to "hijack" us, only came to light late in the telling of the

217. *Id.*

218. Olds & Milner, *supra* note 195, at 426. The Olds & Milner experiments with regard to the extent of self-stimulation of the pleasure center were extended to humans in R.G. Heath, *Electrical Self-Stimulation of the Brain in Man*, 120 AM. J. OF PSYCHIATRY 571 (1963). In a subsequent study Heath reported that one of his patients would press the button electrically stimulating the septal region of his brain up to 1,500 times over a three-hour period. R.G. Heath, *Pleasure and Brain Activity in Man: Deep and Surface Electroencephalograms During Orgasm*, 154 J. NERVOUS & MENTAL DISEASE 3, 6 (1972).

219. Freud's "dynamic metapsychology" (*i.e.*, his theory of instinctual drives) is sketched by me in MOORE, LAW AND PSYCHIATRY, *supra* note 110, at 134–40.

220. Jane Stewart & Roy A. Wise, *Reinstatement of Heroin Self-administration Habits; Morphine Prompts and Naltrexone Discourages Renewed Responding After Extinction*, 108 PSYCHOPHARMACOLOGY 79, 80 (1992).

story.[221] This is the fact, surprising to those who told this story, that onset of a new *expectation* of a reward correlates with increase of dopamine in the nucleus accumbens. Indeed, for normal (*i.e.*, nondrug-induced) pleasures, it is *only* the onset of a new expectation of reward that releases the dopamine; actual pleasurable experience when the reward is received later releases no more dopamine.[222] Yet drugs are here different than ordinary rewards. During the early use of drugs before addiction, dopamine is released both when the drugs are first expected *and* when they are received. This has led to the view that drugs lack the potential for satiety of other forms of reward. Ordinary pleasures are sated by receiving the reward that is expected, such satiety being marked by no new dopamine release upon receipt of the reward; but with drugs the second release of dopamine on receipt can be interpreted as yet another onset of a new expectation that even more pleasure is on the way, beyond what was expected prior to receipt of the reward.[223] "Pleasure increasing without end" is the promise of drug use, according to this nonsatiability interpretation of secondary dopamine release during consumption of drugs.

Seemingly the demon drug story, if true, could incline one towards excusing drug users for what they do to acquire or use drugs after their initial use. So long as their responsibility for that initial use is not carried over to make them responsible for those later acts of acquisition and use (by tracing or otherwise),[224] drug users as depicted in the story seemingly have had their minds "hijacked" in the six senses mentioned so that their minds' behavioral outputs are no longer their responsibility.[225] But is the story true?

The heart of the demon drug story just told lies with mechanisms in the brain that the story holds to make drugs the new lotus fruit that is irresistable once tasted. These mechanisms have to do with the "explosion of dopamine" in the nucleus accumbens drug use causes. The thesis of that crucial part of the story—known as the "dopamine pleasure hypothesis"—is that the experience of pleasure either is identical to, is constituted by, or is caused by, the increase of dopamine in the nucleus accumbens.[226] Yet this hypothesis has seemingly been falsified by the research of the past two decades.[227] As one recent survey of that research put it: despite the fact that the "mesolimbic dopamine system has been

221. *See* R. de la Fuente-Fernández et al., *Dopamine Release in Human Ventral Striatum and Expectation of Reward*, 136 BEHAV. & BRAIN RES. 359 (2002).

222. *See* W. Schultz, P. Dayan & P.R. Montague, *A Neural Substrate of Prediction and Reward*, 275 SCIENCE 1593 (1997).

223. *See* Volkow & Morales, *The Brain on Drugs*, *supra* note 175, at 714 ("[i]n sharp contrast" to natural reinforcers such as food and sex, "the response to drugs of abuse . . . continue increasing DA [dopamine] release during their consumption . . . . This may explain why drugs are more likely to result in compulsive patterns of administration than natural reinforcers.").

224. Since one implication of the story is that even one drug use can be the taste of the lotus that leads the taster down the road of a continued use that is inevitable, one might think that the tracing of responsibility (from later acts to the first act of use) is justifiable because the causal chains are short.

225. Whether there actually is excuse lurking in the demon drug story is a question I do not pursue; this, because the story itself is untrue in certain respects crucial to the seemingly excusing force of the story.

226. *See generally* Roy Wise, *The Dopamine Synapse and the Notion of 'Pleasure Centers' in the Brain*, 3 TRENDS IN NEUROSCIENCES 91, 91–95 (1980).

227. Kent Berridge & Morten Kringelbach, *Pleasure Systems in the Brain*, 86 NEURON 646, 656 (2015).

the most famous neurochemical candidate in the past half century for a pleasure generator in the brain," and despite the wide acceptance of this for the forty years following the work of Olds and Milner in the mid-1950s, "today relatively few neuroscientists who study dopamine in reward appear to assert in print that dopamine causes pleasure."[228]

The seeds for questioning dopamine's role in constituting or causing pleasure were sown in the last bit of the demon drug story itself. The fact that ordinary pleasures do not release dopamine when experienced but only when first anticipated, suggests a quite different message (that dopamine-aided synaptic transmission carries) than that of pleasure. It suggests a nonsensory, more cognitive message, namely, a predictive *belief* that either the reward that gives pleasure, or pleasure itself, is on the way. Moreover, such findings seem to go against there being any strong correlation between pleasurable experience and dopamine release: where there is such experience at t_2 for ordinary pleasures, dopamine is not released, and where there is not yet such experience at t_1 (although it is anticipated), dopamine is released. Seemingly from these facts alone dopamine release is shown to be neither sufficient nor necessary for the experience of pleasure.[229]

Once the tie of dopamine release and pleasure is shown not to exist, then the flood of dopamine caused by drug use cannot be taken as a measure of the degree of experienced pleasure given by drugs.[230] In particular, the dramatic comparisons of the amount of dopamine produced by nonaddicted use of drugs versus ordinary pleasure cannot be taken as establishing the lotus fruit character of drugs.[231] Moreover, if the synaptic transmissions made possible by the surge of dopamine following upon drug use do not message pleasure, that leaves open the question of what they do message. They apparently message the onset of a predictive judgment that reward is coming. Do they also message that that reward is wanted? Valued? Wanted and valued?[232] Such interpretations of the message

228. *Id.* For an earlier but more extensive review reaching a like conclusion, see Berridge, *The Debate over Dopamine's Role in Reward: The Case for Incentive Salience*, 191 PSYCHOPHARMACOLOGY 391, 391 (2007).

229. A defender of the dopamine pleasure hypothesis could try to rescue the hypothesis by urging that the anticipation of pleasure is itself a pleasurable experience, and perhaps even that excess dopamine remains from its generation during the anticipation of pleasure and so is present as pleasure is experienced even though no more is released. Yet no such *ad hoc* end runs are possible in the many situations in which dopamine and pleasure do not co-vary. *See* Kent C. Berridge & Morten L. Kringelbach, *Pleasure Systems in the Brain*, 86 NEURON 646, 659 (2015). For summaries of the relevant findings, see Berridge, *The Debate over Dopamine's Role*, *supra* note 228, at 396.

230. As the foremost proponent of the dopamine pleasure hypothesis, Roy Wise, came to accept: "I no longer believe that the amount of pleasure felt is proportional to the amount of dopamine found floating around in the brain." Berridge, *The Debate Over Dopamine's Role*, *supra* note 228, at 397 (quoting Roy Wise).

231. It is of course possible and even likely that some other neurotransmitter(s) will be found to cause or constitute pleasurable experiences when released in the synapses of the "hedonic hot spots" of the brain, including such hot spots in parts of the nucleus accumbens. It is also possible (although I have found no reports of evidence of this) that such other neurotransmitter(s) will be increased in the sudden and dramatic manner of dopamine increase upon using drugs and even that such releases could be epiphenomenal with such dramatic dopamine increases.

232. *See* G. Di Chiara & V. Bassareo, *Reward System and Addiction: What Dopamine Does and Doesn't Do*, 7 CURRENT OPINIONS IN PHARMACOLOGY 69, 71 (2007).

of dopamine seem less "hijacking" than the interpretation making drugs the lotus fruit of pleasure.

Even if the message of dopamine-aided transmission across the synapses in the ventral striatum were the experiencing of pleasure, the dopamine pleasure hypothesis ran into trouble in explaining drug use by addicted users. Naturally enough, the researchers promoting the dopamine pleasure hypothesis such as Roy Wise expected that the flood of pleasure (and of dopamine) could be extended from explaining why nonaddicted users used drugs, to explaining why addicted users continue to use the drugs whose earlier use addicted them.[233] The same mechanism might reasonably be expected to be at work in both cases, *i.e.*, both sets of users should presumably be motivated to seek the promise of ever greater pleasure offered by the lotus fruit of drugs. Yet the facts about the mechanisms underlying addicted use are plainly otherwise. Both phasic dopamine increases and experienced pleasure decrease with the drug use of addicts (as compared to the levels of both preaddiction).[234] As many addicts say, the "high" just is not that great anymore and, indeed, it is not so much a high that is sought but rather an escape from a low and an allowance of one to just feel normal again.

Apart from the questioning of whether the "explosion" of dopamine in the nucleus accumbens represents pleasure, the demon drug story also runs into trouble in accommodating certain behavioral and phenomenological facts. Take the behavioral facts first. If the demon drug story were literally true, addiction should be both instantaneous and universal upon first use of drugs. After all, the story is that there is nothing like the explosion of pleasure given by drugs and, like the mythical lotus fruit, once tasted all other pleasures pale by comparison. The problem is that neither of these facts seems to be true. True enough, the propaganda used during the "war on drugs" of the 1980s in America promoted both of these myths.[235] But the behavioral facts never supported them. Hardly anyone gets addicted by one use of any drug; and most users of recreational drugs—even regular users—do not become addicted to those drugs.[236] Indeed, some studies indicate that for both animals and humans only 10% of those exposed to the lure of

233. *See* Wise & Bozarth, *supra* note 208, at 469.

234. *See* Ingo Willuhn, Lauren Burgeno, Peter Groblewski & Paul Phillips, *Excessive Cocaine Use Results from Decreased Phasic Dopamine Signaling in the Striatum*," 17 NATURE NEUROSCIENCE 704, 708 (2014). A "phasic" dopamine increase is the kind of sudden and dramatic increase of dopamine associated with pleasure by the dopamine pleasure hypothesis. The slower, less dramatic, "tonic" dopamine increases were never regarded as being associated with the pleasure experience by the dopamine pleasure hypothesis

235. *See, e.g.*, German Lopez, *How the Internet Freed America from Ridiculous Anti-Drug Propaganda*, VOX (Dec. 22, 2015, 8:00 AM), https://www.vox.com/2015/12/22/10621810/internet-marijuana-legalization-drugs ("The 1980s, boosted by Nancy Reagan's 'Just Say No' campaign, were filled with hyperbolic anti-drug ads—including an infamous ad that suggested a person's brain would be scrambled after any drug use, and one that compared drug abuse to literal slavery.").

236. *See generally Is Marijuana a Gateway Drug?*, NAT'L INST. DRUG ABUSE, https://www.drugabuse.gov/publications/research-reports/marijuana/marijuana-gateway-drug (last updated Dec. 2019). This latter fact could perhaps be explained in terms of there being an "addiction gene" or some other factor, present only in some humans, that predisposes such humans to become addicts; in which case the modified demon drug story would be that for *addiction-prone* drug users, the story is true even if it is not true for most people. The jury is still out on the existence of such vulnerabilities to addiction, and particularly on the degree of predisposition of such vulnerabilities for such classes of users.

drugs by use become addicted.[237] Nonaddicted users thus behave like actors who choose to go to concerts, watch television, have sex, or work late: none of these activities predominate in their lives, none of them produce an explosion of pleasure with which no other activity can compete.[238]

The phenomenology predicted by the demon drug story also does not seem to fit the facts. First of all, there is the fact reported before: addicted users report less and less euphoria from their drug use. But even for nonaddicted users, for the demon drug story to be true, "the subjective pleasurable effects of drugs must be enormous. Indeed, the subjective pleasurable effects of drugs would have to be so potent that just the memory of drug experiences would be sufficient to evoke compulsive drug-seeking and drug-taking behavior."[239] For those of us who are the proverbial "children of the sixties," such predicted phenomenology does not ring true. Moreover, if we turn to the decision-making process by which nonaddicted drug users come to their decision about drug use, common experience also does not square with the demon drug story. Nonaddicted users hardly seem to be on automatic pilot when they decide to take drugs; they are not even to be likened to those with bad habits such as absent-minded munchers of snacks while their mind is on something else. Rather, their choice about use is experienced like other choices they make when nothing constrains them from doing what they would most like to do.

This is worlds apart from the phenomenology pictured by the demon drug story. An exaggerated version of that story can be seen in the 1936 propaganda film, "Reefer Madness."[240] The film depicts drug users (marijuana in this case) as automatons who in their drug-induced behaviors do things such as rape that they otherwise would not have done but for the demon drug.[241] The reason that the film became such a cult classic in the late 1960s/early 1970s, at least in Berkeley where I was teaching, was that it was so obviously wrong in its depiction of the phenomenology of drug use. The drug users watching the film (typically using so heavily while watching that it was difficult to see the film through the haze of marijuana smoke in the theater) found humor in what they knew to be an inaccurate depiction of what they were like when they used drugs.

For these neuroscientific, behavioral, and phenomenological reasons, the contributions of neuroscience to the understanding and evaluation of nonaddicted drug use cannot lie in the "hijacking of the brain" story that began in the

237. Terry E. Robinson & Kent C. Berridge, *The Incentive Sensitization Theory of Addiction: Some Current Issues*, 363 PHIL. TRANSACTIONS ROYAL SOC'Y B 3138 (2008).

238. Robinson & Berridge, *supra* note 61, at 272. Nothing in the human self-stimulation studies modelled on the rat self-stimulation studies calls this conclusion into question. Robert Heath, cited *supra*, gave his patients a highly limited, three valued choice set: they could push one of three buttons, only one of which was connected to an electrode placed in a pleasure center. No surprise, given the limited choice set, that his patients kept hitting the button giving pleasure as opposed to those buttons that did not. *See* Heath, *Electrical Self-Stimulation of the Brain in Man*, *supra* note 218, at 573–75. Fortunately for humans in real life, administration of drugs has to compete with appreciation of music, good food, sexual intimacy, and good conversation, not just button-pushing behaviors of no consequence to anyone.

239. Robinson & Berridge, *supra* note 61, at 252.

240. REEFER MADNESS (Dwain Esper & George Hirliman 1936).

241. *Id.*

mid-fifties. Had the story been true, neuroscientists could rightly take pride in having significantly advanced our understanding of why addicts and nonaddicts alike use, and could perhaps be excused for using, drugs. As I understand contemporary neuroscience's ambitions, it has narrowed what it now seeks to explain by focusing now on *addicted* drug use. That is because what is doing the explaining are now not the properties of all drug use, first time as well as addicted—properties like the "deliciousness" Homer attributed to the lotus-fruit or the explosion of dopamine in the nucleus accumbens; rather, what is doing the explaining are the properties of systems or circuits in the brain that have been altered by *long term* drug use. It is thus to the explanation of drug use by those already addicted to the drugs they use to which I now turn.

2. The Explanation of the Continued Use of Drugs by Addicts

There is a cacophony of views on addiction having currency within the neuroscience of the last two decades. A recent survey of this literature groups this cacophony into seven kinds of theories, viewing each theory as competing with the others to explain why addicts continue to use the drugs to which they are addicted.[242] My own take on this literature is that the seven kinds of theories are largely (but not entirely) complementary with one another. I thus seek here to extract what an overall theory might look like, incorporating bits from different sources as I go.

Despite my ecumenical impulses, I eschew paying much attention to the first of these seven theories, what are often called the "learning or habit" theories of addicted drug use. These are theories according to which one explains:

> [T]he consistent self-defeating patterns of behavior seen in addiction by appealing to the changes in associative learning mechanisms purported to underlie the drug abuse and dependence. That is, the transition from-goal-directed behavior to habitual responding means that unhealthy or self-defeating patterns of behavior become entrenched and semi-automated due to the repeated and extended patterns of reinforcement.[243]

The key to such explanations of addicted drug use lies in the effect of prior long-term drug use on the memory and learning systems of the brain. These are said to be "pathologically subverted" so that "drug seeking is a simple response habit elicited by environmental and drug-associated stimuli . . . known as habit learning."[244]

This is not so much wrong as it is incomplete. For left out are the mental states of belief, desire, intention, satisfaction, evaluation, and emotion in terms of which a moral appraisal of the behavior of addicts can be reached. Showing that drug related cues have a larger-than-normal effect on behavior does not tell us how perception of those cues is processed in the mental life of addicts. True

242. For a helpful overview, see Warren Bickel et. al, *21st Century Neurobehavioral Theories of Decision-Making in Addiction: Review and Evaluation*, 164 PHARMACOLOGY, BIOCHEMISTRY, & BEHAV. 4, 4–21 (2018).

243. *Id*. at 10.

244. *Id*. at 7.

enough, if the claims of these theories were that addicts are not just "semi-automated" but are on the "automatic pilot" of some fugue state, then the theories would be morally relevant even without filling in the gap between stimulus and response with intervening mental states. But as I have adverted to earlier, the "automation" picture of addicted drug use just does not square with the experience of addicts. They are not intention by-passing, tropistically wired, jerky, zombie-like, and stumbling robots when they acquire and use drugs.

It is also true that the current learning/habit theories of addiction are not incomplete in the sense that they are without some findings (or at least speculation) about the neural mechanisms by virtue of which cue-driven stimuli do their behavioral work. These are not throwbacks to the truly "black box" approaches of the older behaviorisms. We are told, for example, that long term drug abuse causes changes in both the cortical and the striatal structures of the brain; that this is important because "the change from voluntary drug use (goal-directed processes) to more habitual and compulsive drug use (habit learning processes) represents a transition at the neural level from prefrontal cortical to striatal control over drug seeking and drug taking . . .";[245] that within the striatum, addicted drug use releases excessive dopamine more in the dorsal domain as opposed to the ventral domain of the striatum (where nonaddicted drug use is accompanied by excessive dopamine release).[246] Yet these findings do not wear their mentalistic interpretations on their face. They do not tell us whether the transmissions across the synaptic clefts in the dorsal striatum (when made by release of excessive dopamine), represents a wanting ("craving") for the drug, a liking of it, an evaluative belief about taking it, a wish-caused erosion of factual belief, or whatever. Such uninterpreted, neural-level-only filling in of the black box between stimulus and response, thus does little to help us in our evaluation of addicted drug use.

Habit and learning theorists of addiction might feel that they are entitled to ignore the mentalistic filling-in between the stimulus of perception of a drug cue and the response of drug-seeking behavior. They might think this because the behavior (drug-seeking) is so regular in its following upon presentation of drug cues to addicts. Yet notice that this would only be an explanatory strategy, one preferring parsimony to completeness, not an ontological conclusion about the absence of intervening mental states. That some behavior regularly—even always—follows upon some cue does not justify any conclusion about the non-goal-directed, habitual or "semi-automatic" nature of the behavior. Hume gave this example centuries ago: if a pot of gold is left in plain sight at Charring Cross Station in London, Hume said, it will regularly be picked up by the next persons passing by who see it.[247] Or try this example: two good friends of mine, husband and wife, regularly went to Venice in March for their annual vacation; they did this for a number of years after debating on each occasion about where and when they should take their vacation, but eventually they just adopted a rule to save

245. *Id*.
246. *Id*.
247. DAVID HUME, AN ENQUIRY CONCERNING HUMAN UNDERSTANDING 128 (Antony Flew ed., 1988).

the wasted debate: always go to Venice in March. Such decisions—to steal a pot of gold or to take a vacation in Venice in March—can be as regular as you please but that will not make them into habits or anything other than in fact what they were, fully voluntary, goal-directed decisions. (This is true of my friends even after they adopted their rule, for that kind of rule following need not be habitual either.)

I thus leave the purely learning/habit accounts of addiction aside and turn to accounts that do seek to explain the drug use of addicts in ways more congenial to moral evaluation of the possibly excusing character of addiction. I would integrate the various explanations of addicted drug use by first dividing the theories into two camps, each camp corresponding to one side of a balance which balance will then be seen as determinative of drug seeking behavior. One side of the balance is in terms of the brain processes that underlie what moves addicts to use drugs: their cravings/wanting of drugs, their likings of drugs, their wished-caused factual beliefs making drug-taking seem rational to them, their positive evaluative beliefs about the desirability of continued drug use. The other side of the balance is in terms of the brain processes that underlie control of such prima facie motivating states, such controlling states including negative evaluative beliefs, choice, and will power. This integrative approach then explains the drug seeking behavior of addicts, in terms of an overbalance of the former factors over the latter factors, which imbalance is in turn explained by drug-caused changes (a strengthening and a weakening, respectively) in the processes that realize these factors in the structure of the human brain.[248]

Neuroscientists themselves often see their theories in terms of the balance just described. Nora Volkow and her associates have presented addicted drug usage as an "overwhelming" of the control circuits by the reward circuits.[249] And Warren Bickel presents his own "competing neurobehavioral decision systems theory" in terms of such a balance:

> The key concepts of the dual-decisions systems approach . . . begin with the notion that behavior emerges from relative control between two systems: (1) the reward-driven impulsive system . . . and (2) the evolutionarily newer executive system which governs self-regulatory pro-

248. The mode of explaining molar behavior as being the outcome of a balance between two competing systems is no doubt an overused trope in psychology. Recalling Nietzsche's "Appollinean versus Dionysian" subsystems, Freud's three metapsychological oppositions (topographical, structural, and dynamic), and Karl Meinniger's "vital balance," all illustrate such overuse. *See generally* Michael Moore, *Mind, Brain, and Unconscious*, *in* MIND, PSYCHOANALYSIS, AND SCIENCE 141 (Peter Clarke & Crispin Wright eds., 1988); Michael Moore, *The Unity of the Self*, *in* 2 NATURE ANIMATED 163 (Michael Ruse ed., 1983). Yet unlike these overblown and under-evidenced oppositions, the phenomenology of addicted drug use suggests the "sub-stems in conflict" model of explanation. Addicts often experience a conflict between their desires inter se, between what they like and what they want, between what they value and what they want, etc. Although hardly an infallible heuristic, the phenomenology here warrants use of a balance-between-two-competing-subsystems model of explanation.

249. Volkow et al., *supra* note 179, at 748. Although Volkow's theory is generally considered a kind of "imbalance" theory, her full statement of the theory implicates *six* brain circuits being out of balance with each other to explain addicted drug use. *See* Bickel et al., *supra* note 242, at 8–9.

> cesses . . . . These two systems compete for relative control during decision-making . . . when the regulatory balance between the two systems is disrupted, pathology (e.g., addiction) may result . . . ."[250]

I start on the pro-drug use side of the balance. One of the best known theories on the pro drug-use side of this balance, is the "incentive-salience" theory of Kent Berridge and others.[251] Incentive salience takes its name from the comparatively greater motivating force attributed to drug rewards by addicts over other, more usual rewards that normally motivate their lives. To see what is being claimed here, it is fruitful to see what is not being claimed. One of the things not being claimed is that drugs produce such great pleasure for addicts and nonaddicts alike that such pleasure motivates continued use. That was the thesis of the positive pleasure theories of the 1980s we earlier rejected. One of the grounds for that rejection was that dopamine does not seem to function as the common currency of pleasure, as was once thought. Still, there is a burgeoning neuroscience of pleasure, even though it is not based on either dopamine or (exclusively) on activity within the nucleus accumbens.[252] Both the neurotransmitters and the brain regions involved with the subjective experience of pleasure appear to lie elsewhere. As to the latter, there appear to be five "hedonic hot spots" involved in pleasure, only one of which lies within the nucleus accumbens (and which occupies only a tiny fraction of the tissue of that region). As Berridge and Kringelbach summarize the literature here, the data seem to "suggest hedonic functions to be reiteratively represented at multiple levels of the brain," including hedonic hot spots in the ventral pallidum, the orbitofrontal cortex, the insula cortex, and even the brain stem, in addition to the nucleus accumbens.[253] Flooding these hedonic hot spots with dopamine does not cause or correlate with experienced pleasure; but flooding them with opoid and endocannabinoid neurotransmitters does do so.[254] "[The] neurochemical mode is clearly as important as the anatomical site."[255]

So the long sought after pleasure center(s) and common currency of pleasure do exist, even if these are not (exclusively or the entire) nucleus accumbens and dopamine, respectively. Yet the incentive salience theory is not a new lotus fruit theory (with but a new mechanism for how the lotus fruit does its motivating work). For drugs do not cause an "explosion" of opoid and endocannabinoid neurotransmitters in the hedonic hot spots of addicts in the way that they do cause a dramatic increase of dopamine in the ventral striatum of nonaddicted users. For that matter, the experiments using electrical stimulation of the ventral striatum are now not thought to show either experienced pleasure in their subjects or an

250. Bickel et al., *supra* note 242, at 14.

251. The classic cite here is to the 1993 article by Terry Robinson & Kent Berridge, *supra* note 61, at 249. *See also* Berridge, *supra* note 228, at 391; Berridge & Kringelbach, *supra* note 227; Terry E. Robinson & Kent C. Berridge, *The Incentive Salience Theory of Addiction: Some Current Issues*, 363 PHIL. TRANSACTIONS ROYAL SOC'Y B 3137, 3137–46 (2008).

252. See the survey article by Berridge & Kringelbach, *supra* note 227.

253. *Id*. at 652.

254. *Id.* at 652–58.

255. *Id*.

explosion of opoid and endocannabinoid in the hedonic hot spots of their brains.[256]

The older experiments delivering electrical stimulation or direct delivery of drugs to the nucleus accumbens glossed over a distinction that is at the heart of the incentive salience theory. For those experiments (done in the heyday of Skinnerian behaviorism) took the bar-pressing behavior of rats and people to evidence pleasure being experienced by such subjects. Missed was the possibility that the drug or electrically caused flood of dopamine caused the intense wanting ("craving") for drugs and for the action of taking drugs without influencing the amount of pleasure the taker experienced or anticipated experiencing by taking them. In the language of the incentive salience theory, what dopamine in the ventral striatum causes is wanting something without liking it. Normally we like what we want, and we want what we like. Yet these are contingent, psychological truths; they are not analytic truths. We sometimes want what, when we get it, does not please us; and we sometime find pleasure in states that we antecedently did not want. Even so, normally our wants are educated by our likes. Put simply, normally our likes motivate us by causing us to want what will give us pleasure when received.

The heart of the incentive salience theory lies in its thesis that drugs break this normal connection for drug addicts. The flood of dopamine causes the uptick of wanting that is experienced by addicts as a craving and that motivates behavior satisfying such wanting-craving, all without there being (again, for *addicted* drug users) any pleasure in the offing.

It may seem that the incentive salience theory runs afoul of the same fact that doomed the dopamine pleasure theory, namely, that the dopamine increase due to drug use *decreases* for addicts. We asked of the dopamine pleasure theory, why then does not the liking (pleasure) also decrease? Now we should ask the analogous question of the incentive salience theory: why then does not the wanting (incentive salience experienced as craving) decrease? Defenders of the incentive salience theory here engage in some fancy footwork. They claim that "the current literature contains conflicting results about brain dopamine changes in addicts" in that "detoxified cocaine addicts show a decrease in evoked dopamine release rather than the sensitized increase described above [by the incentive sensitization theory]."[257] They then claim that "the role of context is crucial in gating the expression of sensitization in general, and thus of sensitized increases in dopamine release," and that therefore the apparently contrary results "must be interpreted with caution."[258] Perhaps this is right, but more honest is the confession by the same proponents of the incentive sensitization theory that "we are unsure at this point about exactly which of the many changes in the brain produced by drugs underlie the psychological changes of incentive sensitization."[259] So if we are going to speculate here about the changes worked by the flood of

256. *Id*. at 20–23.
257. Robinson & Berridge, *supra* note 237, at 3140.
258. *Id*.
259. *Id*. at 3139.

dopamine released in the ventral striatum during nonaddicted drug use, why not speculate (and then of course test) this way: The flood of dopamine brought about by nonaddicted use of drugs changes the brain circuits, not only by decoupling the liking (pleasure) system from the wanting (motivational) system, but also by sensitizing the wanting system to the stimulation of less and less dopamine. Sensitization then is the process of altering the brain so that there is more "bang for the buck," *i.e.*, more intensely felt and more intensely motivating wanting for less dopamine?[260] One cannot then identify wanting with dopamine release in the ventral striatum, no more than the older theories could identify the experience of pleasure (liking) with such dopamine release. But such dopamine release could still *cause* greater wanting later by altering the motivational system in the brain.

It is unclear to me that this last bit of speculation by me changes the incentive salience theory very much if at all. Sensitization is generally defined by the theory "an increase in a drug effect caused by repeated drug administration."[261] More specifically:

> The central thesis of the incentive sensitization theory of addiction is that repeated exposure to potentially addictive drugs can . . . persistently change brain cells and circuits that normally regulate the attribution of incentive salience to stimuli, a psychological process involved in motivated behavior. The nature of these 'neuroadaptations' is to render these brain circuits hypersensitive ('sensitized') in a way that results in pathological levels of incentive salience being attributed to drugs and drug-associated cues.[262]

Such a psychological theory does not depend in any essential way on there being any particular role for dopamine release as the realizer in the brain of this sensitization process. Understandably, of course, any theory of addiction should want to use this dramatic effect of nonaddicted drug use in the brain as part of its neuronal level explanation of addicted drug use.[263] But the causal role for dopamine release above sketched respects this desideratum as well as does a theory postulating an identity (or some other relation supporting proportionality) between the amount of dopamine and the degree of wanting.

It is widely appreciated that the incentive salience account overlaps significantly with the habit and learning accounts earlier put aside in that both kinds of accounts emphasize how drug use causes changes in the brain such that drug cues in the environment produce an unusually vigorous response in addicted subjects. This is a kind of learning posited to exist by both sorts of theories.[264] Yet

260. Robinson & Berridge, *supra* note 61, at 249. The mechanism for this "greater bang for the buck" presumably would consist of fewer dopamine molecules being required to open/keep open the ion channels in the receptors on the postsynaptic neurons of the relevant brain areas. How exactly this works is a mystery, at least to me.

261. *Id*. at 3139.

262. *Id.* at 3140.

263. Thus, incentive salience theorists seek to downplay or displace the prediction-error role for dopamine in learning and habit theories, in favor of their hypothesized role for dopamine in entrenching the wanting for drugs. For evidence in favor of the incentive salience theorists here, see generally Shelly B. Flagel et al., *A Selective Role for Dopamine in Stimulus-Reward Learning*, 469 NATURE 53 (2011).

264. *See, e.g.*, Bickel et al., *supra* note 242, at 8.

the response posited to exist by incentive salience theories is not directly the drug-seeking behavior of addicts as it is in the habit/learning theories; rather, the direct response to drug cues as stimuli is the *craving* to use drugs, not the use of the drugs themselves. True enough, that craving often then causes drug-seeking behavior; but not directly as in habit theories, only indirectly through the causal intermediaries of wanting and choice.[265] The phenomenology of addicted drug use better fits this not-automatic, intentional model of drug use, as was argued before.[266]

An account that is both complementary and competitive with the incentive salience theory of addicted motivation, is what George Koob and his associates somewhat darkly call the "dark side of addiction."[267] The dark side of addiction is an alternative theory about what motivates addicts to continue to use drugs. It has long been speculated that there were two possible kinds of motivation for this behavior: positive motivations such as both the old pleasure-reward theory and the incentive salience theory, and negative motivations consisting not of seeking of something good such as pleasure but rather, of avoiding something bad such as the dysphoria of withdrawal. Indeed, when Roy Wise first proposed his pleasure reward theory in the 1970s he took himself to be arguing against negative theories of motivation.[268] Koob's dark side theory is an attempt to revive that older, negative approach to the motivations of addicts.[269]

Throughout his career Koob has proposed that we should break up drug use by addicts into three temporal stages: the first is binge/intoxication; the second is withdrawal/negative affect; and the third is preoccupation/anticipation/craving.[270] As Koob recognizes, his dark side theory focuses on "one part of the addiction cycle—the withdrawal/negative affect stage—which has been largely neglected."[271] The basic picture is this: drug use by nonaddicts starts out as a pleasure seeking activity; yet as regular use becomes addictive, the motivation to use changes from the seeking of positive rewards to the avoidance of negative effects. Those negative effects are not just the effects we would class as withdrawal, but include the depression, irritability, purposelessness, anxiety, and

265. In distinguishing the two kinds of theories, Robinson & Berridge concede that "learning specifies the object of desire" but then urge that "learning per se is not enough for pathological motivation to take drugs." Robinson & Berridge, *supra* note 237, at 3138. As we saw in Part II above, except for automatic behaviors that are not really even actions, motivation requires some kind of "pro-attitude" such as wanting or valuing, the cognitive state of believing not being enough.

266. *Compare id.*, *with* Bickel et al., *supra* note 242, at 8.

267. George F. Koob & Michel Le Moal, *Drug, Addiction, Dysregulation of Reward, and Allostasis*, 24 NEUROPSYCHOPHARMACOLOGY 97, 97–129 (2001). Koob helpfully summarizes his theory in George F. Koob & Michel Le Moal, *Plasticity of Reward Neurocircuitry and the 'Dark Side' of Drug addiction*, 8 NATURE NEUROSCIENCE 1442, 1442–44 (2005) and updates his summary in the early pages of George F. Koob & Barbara J. Mason, *Existing and Future Drugs for the Treatment of the Dark Side of Addiction*, 56 ANN. REV. PHARMACOLOGY & TOXICOLOGY 299, 299–322 (2015).

268. *See* Wise & Yokel, *supra* note 203, at 547.

269. *See generally* Koob & Mason, *supra* note 267.

270. *See, e.g.*, *id.* at 299. Koob's colleague at the National Institute of Health, Nora Volkow, also adopts Koob's tripartite division. *See* Nora D. Volkow et al., *Neurobiologic Advances from the Brain Disease Model of Addiction*, *supra* note 35, at 364–67.

271. Koob & Mason, *supra* note 267, at 300.

stress addicts often report when they hit a dry spell in their use of drugs. As such addicts say, they are not trying to achieve euphoria with drugs, just trying to feel normal again.[272]

So far this might suggest a kind of all-inclusive ecumenicalism on Koob's part: his first stage of use by nonaddicts is motivated by the goal of pleasure whereas by the third stage this goal of pleasure has transformed itself into the goal-less wanting for drugs and for drug-seeking behavior well described by the incentive salience theory; while intervening between the two at the second stage is the negative motivation of the dark side theory. This is such a rosy picture because it gives each theory a little data to call its own to explain.

I am pretty sure that Koob would be appalled by this ecumenicalism on the cheap. For what the dark side theory posits at its core runs against this temporal accommodation. Koob's theory is often called (by himself as well as others) the "opponent process" theory.[273] The idea behind the label is that from the beginning when drugs are first used the reward system is opposed by what Koob calls an "anti-reward system."[274] "The hypothesis is that there are brain systems in place to limit reward . . . , an 'opponent process' concept that is a general feature of biological systems."[275]

Such a system operates out of its own anatomical sites in the brain different from the hedonic hot spots and the rest of the mesolimbic system in which pleasure, reward, and wanting are located. Such a system also uses neurotransmitters different from the opoid and dopamine neurotransmitters used by pleasure and wanting respectively. Remarkably, the thesis is that excitation of the positive reward system (pleasure and wanting) causes excitation in the anti-reward system.[276] (I find this remarkable because this means the brain operates not just on a pleasure principle but also on a principle well-described by the lyrics of a country western song, "You can't have too much fun.")[277] In the early stages of drug use the reward systems predominate to motivate continued use; but as addiction sets in and deepens, the brain's circuitry is changed so that two things happen: the reward system degrades so that positive reinforcers do less and less work to motivate continued drug use, and the anti-reward system has been strengthened by the continued work given it by the prolonged use of drugs. This lessening of positive reward and strengthening of negative avoidance results in an explanation of continued drug use by addicts that is in the end entirely negative: they use drugs not to achieve highs or even for their own sake but rather to avoid the lows of nonuse. This is not to say that this is a rational strategy by addicted users—for more use generates further activation of the stress circuitry that will in the future

272. *The Science of Drug Use: Discussion Points*, NAT'L INST. DRUG ABUSE, https://www.drugabuse.gov/related-topics/criminal-justice/science-drug-use-discussion-points (last updated Feb. 2017).

273. Koob & Le Moal, *Plasticity of Reward Neurocircuitry and the 'Dark Side' of Drug Addiction*, *supra* note 267, at 1442.

274. *See generally id.*

275. *Id.* at 1442.

276. Koob & Mason, *supra* note 267, at 300 ("There is accumulating evidence that neurobiologically excessive activation of the reward system is a causal mechanism for activation of the brain stress systems.").

277. DARYL SINGLETARY, TOO MUCH FUN (Giant Records 1995).

cause further dysphoria—but the hypothesis is that addicts are nonetheless so motivated.

Notice that the motivations tendered by the incentive salience and the dark side theories are different in two dimensions. One we have remarked on before: the former is positive whereas the latter is negative. But perhaps more important is this second difference: the want that emerges to motivate continued drug use by addicts is an intrinsic desire according to the incentive salience theory whereas this want is an instrumental desire for the dark side theory. That is, according to the incentive salience theory what was originally (prior to addiction) an instrumental, goal-promoting desire (to take drugs in the service of the pleasure they gave) becomes with addiction's rewiring of the brain an intrinsic desire whereby drugs (and drug seeking behavior) are just wanted for their own sakes. By contrast, according to the dark side theory the want that motivates drug use starts and remains an instrumental desire throughout the process of becoming addicted; just the goal such want serves changes, from attainment of positive reward to avoidance of negative affect.

Despite these differences, can these two theories be friends? At the psychological level I do not see why not. Operating for mixed motives is a familiar feature of human psychology, and this remains true even in cases where one of the "motives" mixing in is not a further motive at all but is rather an intrinsic desire to do the behavior in question. Further, each theory is supported by an aspect of the phenomenology of drug addiction. Still, the combinability of the theories, as well as each theory's viability individually considered, ultimately rests on how the details of the brain circuitry they posit plays out. From my reading of the literature, it is premature to render a verdict here. So I shall pursue my ecumenical stance: assume that each theory partly explains why and how addicts are motivated to continue using drugs.[278] And then ask about the other side of the balance with which we began: does abuse of drugs also cause changes in the brain circuits that account for executive control functions? Does deterioration there also help to explain why addicts continue to use drugs? Explain such behaviors in a way that tends to excuse them by supplementing the strength of motivation stories above with an incapacitation of control story?

Almost all theories about the motivations of addicts to continue using drugs pay lip service to the idea that the overall explanation of addicted drug use will include not only their favored explanation in terms of why the motive to take drugs is so strong, but also an explanation in terms of how and why the executive control functions of addicts are so weak. Kent Berridge, for example, often throws in such a nod to this control side of the balance.[279] But from my review of the literature, there is very little meat put on the bones of an explanation all

278. A like ecumenicalism is displayed in David Belin et al., *Addiction: Failure of Control Over Maladaptive Incentive Habits*, 23 CURRENT OPINION NEUROBIOLOGY 564, 564–72 (2013), where the authors assume that the incentive salience and the opposed process accounts describe "two parallel and likely interactive motivational processes."

279. *See generally* Robinson & Berridge, *supra* note 251, at 3138.

see as possible. Often there is no more than a conclusory statement that of course addiction also weakens the executive control circuits, period.

fMRI studies have done much to increase our general knowledge as to the anatomical locations for executive control functions such as resisting temptations on diets, delaying gratifications, focusing attention, or making oneself get out of a warm bed on a cold morning.[280] In one well known set of studies involving choices of foods by dieters, for example, researchers have isolated activation of the dorsolateral prefrontal cortex (dlPFC) as crucial if dieters are going to successfully resist the blandishments of unhealthy but tasty foods.[281] More generally, it is known that regions of the anterior cingulate (ACC) and the orbitofrontal

280. The studies I have in mind are those showing: that long term goals (*i.e.*, controlling desires) do their work in counteracting tempting desires by inhibiting the behavior satisfying the latter, only where there is a negative functional interaction of anteroventral prefrontal cortex with nucleus accumbens and ventral tegmental area, Esther K. Diekhof & Oliver Gruber, *When Desire Collides with Reason: Functional Interactions Between Anteroventral Prefrontal Cortex and Nucleus Accumbens Underlie the Human Ability to Resist Impulsive Desires*, 30 J. Neuroscience 1488 (2010); that successful modulation of cravings (for cigarettes, at least) by a controlling desire (to be healthy and to live) is associated both with heightened activity in those regions associated with controlling emotion in general (dorsomedial, dorsolateral, and ventrolateral prefrontal cortices), and with lesser activity in those regions generally associated with the presence of cravings (ventral straitum, subgenual cingulate, amygdala, and ventral tegmental area), Hedy Kober et al., *Prefrontal-Striatal Pathway Underlies Cognitive Regulation of Craving*, 107 PROC. NAT'L ACAD. SCI. U.S. 14811 (2010); that activations of the right ventrolateral prefrontal cortex is needed for the six most common forms of self-control (delaying gratification, regulating emotion, suppressing risky behavior, motor response inhibition, memory inhibition, and thought suppression), Jessica R. Cohen & Matthew Lieberman, *The Common Neural Basis of Exerting Self-Control in Multiple Domains*, *in* SELF-CONTROL IN SOCIETY, MIND, AND BRAIN 141, 141–60 (Ran R. Hassain et al., eds., 2010); that the normal execution of desires by intentions to do the actions immediately (proximal rather than distal intentions) takes place in the presupplementary motor area, the supplementary motor area, and the cingulate motor area, Patrick Haggard, *Human Volition: Towards a Neuroscience of Will*, 23 NATURE REVS. NEUROSCIENCE 934, 934–46 (2008); that the *inhibition* of such intentional actions requires the activation of brain areas distinct from those activated in the initiation of intentional actions, namely, the dorsal fronto-median cortex, the left and right anterior ventral insula, and the right superior sulcus, Marcel Brass & Patrick Haggard, *To Do or Not to Do: The Neural Signature of Self-Control*, 27 J. NEUROSCIENCE 9141, 9141–45 (2007); that three main circuits in the prefrontal cortex are necessary for the selection and initiation of actions, to compare actions done with actions intended, to control emotion and behavioral impulse, and to the action-guiding function of intentions, namely, the dorsolateral prefrontal cortex, the ventromedial-orbitofrontal cortex, and the anterior cingulate cortex, Jane F. Banfield et al., *The Cognitive Neuroscience of Self-Regulation*, *in* HANDBOOK OF SELF-REGULATION: RESEARCH, THEORY, AND APPLICATIONS (Roy F. Baumeister & Kathleen D. Vohs eds., 2004); that the anterior cingulate cortex in particular seems to be the locale where comparisons of acts done to acts intended takes place, Angus W. MacDonald et al., *Dissociating the Role of the Dorsolateral Prefrontal and Anterior Cingulate Cortex*, 288 SCIENCE 1835, 1835–98 (2000); that the anterior cingulate is activated when a particular thought is suppressed (although less uniquely so when thoughts in general are suppressed), Carrie L. Wyland et al., *Neural Correlates of Thought Suppression*, 41 NEUROPSYCHOLOGIA 1863, 1863–67 (2003); that response inhibition in go/no go tasks most prominently activates the right lateral orbitofrontal cortex (although four other areas also show some activation), and that greater activation of the right lateral orbitofrontal cortex was observed in characteristically impulsive individuals who achieved the same level of accuracy on go/no go tasks as nonimpulsive individuals, N. R. Horn et al., *Response Inhibition and Impulsivity: An fMRI Study*, 41 NEUROPSYCHOLOGIA 1959, 1959–66 (2003). The insight generally motivating this work is the insight animating much fMRI work in neuroscience. As stated by Tony Damasio and his colleagues, it is that "different sectors of the human prefrontal cortex are involved in distinctive cognitive and behavioral operations." Antoine Bechara et al., *Emotion, Decision Making and the Orbitofrontal Cortex*, 10 CEREBRAL CORTEX 295, 295–307 (2000).

281. *See* Michael Camus et al., *Repetitive Transcranial Magnetic Simulation Over the Right Dorsolateral Prefrontal Cortex Decreases Valuations During Food Choices*, 30 EUR. J. NEUROSCIENCE 1980, 1980–88 (2009); Todd Hare et al., *Focusing Attention on the Health Aspects of Food Changes Value Signals in vmPFC and Im-*

cortex (OFC) are also involved in such executive control tasks.[282] This knowledge has encouraged addiction researchers to conclude that there must be some *impairment* in the "executive control circuit" because these regions of the brain are not fully activated when addicts choose to use drugs.[283]

"Impairment," like near synonyms such as "disable," "broken," "dysfunctional," "unusable," is a functional term. Mere structural deviation of a part or a process from some statistically normal state is not necessarily an impairment of that part or that process. After all, the function normally performed by that part or that process might well be alternatively realized by that part or process in its abnormal, altered physical state. What must be stopped for a part or process to be impaired is the ability of the part or process to perform its function. We thus need to understand both how functions are assigned to parts or processes, and what it means to say that such part or process is disabled (lacks the ability) from performing its function. The function of a part or process in some larger system is an effect caused by that part or process, which effect itself causally contributes to the maintenance or achievement of some overall end state of the system that is valued or is otherwise the subject of interest.[284] The end state may or may not be one maintained in a homeostatic state by nature such as body temperature; it being a homeostatic state is one reason that can draw our interest to that end state so that we organize causal information around what contributes to the maintenance of that state, but such naturally occurring homeostasis is not the only way to draw our interest to such an end state.[285] The end state that organizes causal information in the human body (including the brain) is some ideal of health.[286] In the case of the brain specifically, it is the mental health that makes us successful negotiators of our environment, a health marked by theoretical and practical rationality. Thus, when one seeks the function of dopamine being released in the nucleus accumbens, for example, one is seeking some effect of that release that itself contributes to human health—sensations of pleasure, firming up of the wants that motivate us, learning what to do in the future, etc.

Impairment of a body part or process means that that part or process is not performing its function, viz., not contributing to human well-being. Yet in this

proves Dietary Choices, 31 J. NEUROSCIENCE 11077, 11077–87 (2011); Todd Hare et al., *Self-Control in Decision-Making Involves Modulation of the vmPFC Valuation System*, 324 SCIENCE 646, 646–48 (2009). Since their original study in *Science* in 2009, this team (or its overlapping teams) has been busy in verifying and clarifying the roles of the vmPFC and the dlPFC; Cendri A. Hutcherson et al., *Cognitive Regulation During Decision Making Shifts Behavioral Control Between Ventromedial and Dorsolateral Prefrontal Value Systems*, 32 J. NEUROSCIENCE 13543, 13545–54 (2012); Hilke Plassmann et al., *Appetitive and Aversive Goal Values Are Encoded in the Medial Orbitofrontal Cortex at the Time of Decision Making*, 30 J. NEUROSCIENCE 10799, 10799–808 (2010); Peter Sokol-Hessner et al., *Decision Value Computation in DLPFC and VMPFC Adjusts to the Available Decision Time*, 35 EUR. J. NEUROSCIENCE 1065, 1065–97 (2012).

282. R. Hester et al., *The Role of Executive Control in Human Drug Addiction*, 3 BEHAVIORAL NEUROSCIENCE OF DRUG ADDICTION 301, 306 (2009).

283. *See id.* at 302.

284. I discuss the logic of function assignments and functional explanations in MICHAEL S. MOORE, LAW AND PSYCHIATRY: RETHINKING THE RELATIONSHIP 9–43 (1984).

285. A debate between myself and Christopher Boorse in our correspondence in the 1980s about function assignment. *Compare id. with* Christopher Boorse, *Wright on Functions*, 85 PHIL. REV. 70, 70–86 (1976).

286. MOORE, LAW AND PSYCHIATRY, *supra* note 102, at 190.

context it is crucial to distinguish the simple behavioral fact that the part or process is not working on some given occasion(s), from an explanation of that fact in terms of disability, *i.e.*, that the thing *cannot* work, that it is broken, disabled. An area of the brain such as the dlPFC on a given occasion may be quiescent, *i.e.*, not activated, but that does not mean that it could not have been activated, that it could not have worked on that occasion. Disability requires more that the absence of success on a given occasion. It requires the modal judgment that there could not have been such success on that occasion because, in general, the part or process could not do what it normally is its function to do.

Analyzing what we mean when we talk of ability and disability, when we talk of what something can and cannot do, has been called the hardest problem in all of philosophy.[287] My own take is what is usually called the counterfactual analysis of ability.[288] The crucial idea is that every ability has its success conditions, that is, has conditions in which the possessor of the ability would succeed in doing whatever he, she, or it can properly be said to have the ability to do. Things that will never do A under any conditions do not have the ability to do A. This is a counterfactual idea of ability because it translates, "X could have done A" (when X in fact did not do A) as being true if and only if the following counterfactual is true: "If, contrary to fact, conditions C had been present, then X would have done A." This generic notion of ability applies to the abilities of bridges, body parts, mental processes, as well as the abilities of whole persons. When a structural engineer explains in detail why the I-35 bridge over the Mississippi at Minneapolis collapsed, and then adds, "But it could have held," he is not saying that the conditions he used to explain the bridge's collapse were not sufficient on that occasion for that collapse; rather, he is saying that if, contrary to fact, certain of those conditions had been slightly other than they were—the construction equipment on the bridge had been placed elsewhere, the temperature were higher, the morning traffic less congested, for example—the bridge would have held.

On the counterfactual account of them, context is all important in assigning abilities. For it is context that makes appropriate certain conditions and not others that if changed would have resulted in success rather than failure. It is also true that, independent of context, such conditions as we imagine to have been other than they were, must have been "close" to the conditions actually prevailing. What is possible is like a close-fitting halo around what is actual, in the sense that for something that did not occur to have been possible to have occurred it must have been just a little bit different from what did occur. To say that those whose eyes have been gouged out can see because if they had eyes they would see, is to make little sense. This is what logicians mean when they say that counterfactuals like those giving meaning to ability are to be tested in "close possible worlds." Even so, exactly which close conditions we are to imagine being

287. Gary Watson, *Free Agency*, 72 J. PHIL. 205 (1975).

288. *See* Michael S. Moore, *Compatibilism(s) for Neuroscientists*, *in* LAW AND THE PHILOSOPHY OF ACTION 1, 41 (Enrique Villanueva ed., 2014).

changed as we ask whether a given result could have been other than it was, depends on the context of utterance.

In the context of explaining why addicts use drugs, the nonfunctioning of various parts of the PFC can rightly be characterized as an inability of the PFC to do its job, *i.e.*, as an impairment. For, according to certain neuroscientific findings that we will shortly describe, these areas of the PFC cannot do their job because they are caused to be innervated by the certain goings-on the ventral striatum of addicted drug users. The hypothetical removal of these goings-on so that the PFC would do its job, does not constitute a close possible world in the context of explaining addicted drug use; for such drug use causes these goings-on to occur and imagining that either that they do not occur or that their occurrence does not cause what it does cause by way of PFC innervation, would be too "miraculous" (contrary to the evidence) to be considered "close" to reality.[289] In the relevant explanatory context, the PFC of addicted drug users thus *cannot* perform its function, and the drugs users themselves can rightly be classed by medicine as being both disabled and because of this, diseased.[290]

In completing this "failure of executive control functions" part of the explanation why addicts continue to use the drugs to which they are addicted, it remains to detail current neuroscientific findings and speculation as to the "goings-on" in the ventral striatum that cause innervation of the ACC, the dlPFC, and the OFC. This detailing involves a continuation of the dopamine story with which we began some time ago. One of the mechanisms by which dopamine remains in the synaptic clefts in the ventral striatum is the blockading of the dopamine receptors on the postsynaptic neuron. Researchers now distinguish at least four kinds of such dopamine receptors, labeled "D-1," "D-2," "D-3," and "D-4" receptors. The D-1 receptors have to do with the various functions of dopamine we have discussed before—the supposed experiencing of pleasure, the prediction-error in anticipation of pleasure, the wiring in of motivating cravings. Yet it is through the D-2 receptors that signals are sent (via the "indirect pathway") to areas of the PFC. The apparent finding/speculation is that addicted drug use causes decreased activity in these areas of the PFC:

> Repeated exposure to different types of drugs has been associated with downregulation of D2R in striatum . . . . Low levels of D2R in the striatum will result in reduced DA [dopamine] inhibition of the indirect pathway . . . . Reduced D2R-mediated DA inhibition of the indirect pathway will lead to reduced thalamo-cortical stimulation and consequently reduced activity in PFC regions. Indeed, the reductions in striatal D2R (dorsal and

289. David Lewis posits that in judging closeness of possible worlds to the actual world, we must eschew major miracles; minor miracles on the other hand are inevitable as we imagine a world different is some respects from the actual world. *See* DAVID LEWIS, COUNTERFACTUALS 75–77 (1973); David Lewis, *Counterfactual Dependence and Time's Arrow*, 13 NOUS 455, 472 (1979).

290. I shall reserve for later discussion why this scientific and medical conclusion does not warrant an inference of moral excuse here; for those who cannot stand the suspense, and in a nutshell, in the differing context of moral evaluation, the relevant counterfactuals will be differently framed and answered so that addicted drugs users may well have the ability to refrain from using drugs even though their PFC's are innervated in just the ways and for just the reasons that neuroscience says.

> ventral) in drug abusers have been associated with decreased activity in the PFC, including anterior cingulate (ACC) and orbitofrontal (OFC) cortical regions. The ACC and OFC are necessary for self control . . . .[291]

There are undoubtedly alternative models for how prolonged drug abuse impedes the executive control functions of addicts.[292] However this side of the balance is worked out, the overall explanation of continued drug use by addicts lies in the combined strength of the motivating factors outweighing (or "overwhelming") the reduced strength of the executive control functions of addicts. *E.g.*, Nora Volkow's most recent description of this explanatory imbalance:

> In a brain not affected by addiction, the circuits controlling desire for a drug are held in check by prefrontal cortical regions that underlie executive functions . . . [such that] the individual is able to make a reasonable choice and carry it through. However, when the prefrontal cortical circuits underlying executive function are hypofunctional—as a result of repeated drug exposure or from an underlying vulnerability—and the limbic circuits underlying conditioned response and stress reactivity are hyperactive—as a result of drug withdrawal and long term neuroadaptions that downregulate sensitivity to nondrug rewards—the addicted individual is at a tremendous disadvantage in opposing the motivation to take the drug. This explains the difficulty addicted individuals face when trying to stop taking drugs even when they experience negative consequences and have become tolerant to the drug's pleasurable effects.[293]

This comes close to saying that while nonaddicted individuals usually have the ability to refrain from taking drugs, addicted individuals have no such ability—or at least, their ability in this regard is compromised by the disadvantages and difficulties above described. And this may sound like the disability created by this imbalance is the kind of incapacitation that should ground a moral and legal excuse.

---

291. Volkow & Morales, *The Brain on Drugs*, *supra* note 175, at 716.

292. I put aside those "models" for the loss of executive control function that are simply a return to a habits/learning account whereby choice is by-passed. *See, e.g.*, Belin et al., *Addiction: Failure of Control*, *supra* note 278. I would include, however, the very simple model that Warren Bickel calls the "depletion-strength model of self-control failure." Bickel, *21st Century Neurobiological Theories of Decision-Making in Addiction*, *supra* note 242, at 11. This model builds on the well-known but not universally accepted work of Roy Baumeister on how willpower is a depletable resource so that use of it on one occasion weakens its availability on some other, immediately following occasion. (Baumeister's work is cited and discussed in Moore, *The Neuroscience of Volitional Excuse*, *supra* note 67, at 212–13 n.88). This depletion is to be observed at both the psychological/behavioral level and at neuronal level, being tied to sucrose levels in the brain. The application of Baumeister's general work to drug use by addicts works off of the incessant demand to use drugs faced by addicts in the craving stage of their addiction – the addict's will-power "muscle" eventually gets tired and can no longer combat the non-depleted, full strength craving of the addict. This is not so much a "broken brain" hypothesis about impaired executive control functions as it is a story about weakened but normal executive control functions.

293. Nora Volkow & Maureen Boyle, *Neuroscience of Addiction: Relevance to Prevention and Treatment*, 175 AM. J. PSYCHIATRY 729, 731 (2018).

C. Basing an Expanded Moral Excuse and Legal Defense on the Neuroscientific Explanation of Addiction

We now need to return to the distinction previewed earlier in Part I of this Article, that between the disability that makes for disease from the incapacitation that makes for excuse.[294] One might well conclude from the evidence reviewed earlier that for addicts: Drugs have lost their ability to cause pleasure; that the liking system has lost its ability to influence the wanting system; that the PFC has lost its ability to restrain the cravings to take drugs; etc. I earlier said that these and other statements of disability would be warranted if the evidence on which they are based is true.[295] But these disabilities do not add up to an incapacitation of addicts that excuses their behavior as addicts. This is because the abilities on which responsibility is based are different than the abilities just mentioned.

I have elsewhere argued that possession of two abilities is crucial to an agent being morally culpable for some wrongful act ("A") that he has done: he must have had the ability to have not done A; and he must have had the ability to have chosen not to do A.[296] These are the *sine qua non* of free action and free will, respectively.[297] On the counterfactual analysis of ability, these two abilities are present when but only when: if the actor had chosen not to do A, he would have not done A; and: if the actor had wanted and valued not doing A very much, then he would not have chosen to have done A. Actors are responsible for their wrongful actions when they possess these two abilities, because their choices control their actions and their desires and because their choices get them what they most want and most value having.

The important point here is that these abilities can be present even though the disabilities just mentioned in neuroscience are also present. This is because these abilities of whole persons are analyzed by different counterfactuals, themselves tested in different possible worlds, from the abilities of subsystems within persons. For the ability needed for free action by persons, the relevant counterfactual requires that we assume that the addict chose not to take drugs and then ask, would he still have taken them? For the ability needed for free will by persons, the relevant counterfactual requires that we assume the addict most wanted and most valued not taking drugs and then ask, would he still have chosen to take them? If the answer to both question is "no," then he had the ability not to take drugs on that occasion; if the answer to either question is "yes," then he lacked that ability either because his will was not responsive to his wants and values or because his actions were not responsive to his will. The answer to these questions may in fact not be obvious in individual cases, nor need it be the same for all cases; but whatever the answers may be to these questions, such answers will not be determined by the answers to the counterfactual tests for the (dis)abilities of

294. *See supra* Part II.

295. *See supra* Section V.B.

296. Moore, *Compatibilism(s)*, *supra* note 288, at 40–41.

297. *See* KADRI VIHVELIN, CAUSES, LAWS, AND FREE WILL: WHY DETERMINISM DOESN'T MATTER 1, 3 (2013).

the dlPFC, the (dis)abilities of the liking system, etc. That these sub systems *could* not perform their function does not mean that the person whose subsystems they are *could* not have chosen and acted differently than he did—for these are differently meaning "coulds."

To see this more clearly, ask a frequently raised question: "but if the activation of the dlPFC is necessary for a person to choose against his impulses, or if activation of the ACC is necessary for his likings to control his wantings, how can he have the ability to act or choose to act so as not to take drugs?" If these activations are necessary for such choice and such action, how can these latter items exist without such activations? Here we need to attend carefully to the counterfactuals testing the abilities to act and to choose other than one did. When we ask whether the actor could have chosen or acted differently than he did in taking drugs, we need to see whether he would have chosen to take the drugs even if he most wanted and valued not taking them, and whether he would have taken the drugs even if he had chosen not to. In imagining the possible worlds in which we see if either of these things would happen, we need to assume wants, values, and choices that are all against the taking of drugs. Yet we also need to assume that in those possible worlds the identity relations that hold in the actual world also hold there—that is part of what it means to keep possible worlds as *close* to the actual world as we can, given that that possible is changed by the conditions specified in the contrary to fact antecedent of our testing counterfactuals. If choosing, for example, is identical to certain goings-on in the dlPFC, then when we imagine the addict choosing not to take drugs we must also imagine activation of those brain regions constituting such mental state of choosing. And with the dlPFC activated, his choice not to take the drugs may well result in his not taking the drugs—so he could have done other than he did in taking the drugs.

Now the real objection lurking in the question with which we began the preceding paragraph comes to the fore: "but how could we be entitled to assume away the very thing(s) that caused the addict to choose and act so as to take drugs?" This objection is just metaphysical incompatibilism rearing its ugly head once again. For compatibilists—which is every determinist who thinks that there is such a thing as responsibility because we are not all excused by causes of our choices that are themselves unchosen—causation of the choice to take drugs does not erode the ability possessed by the addict to have made the choice not to take drugs. Again, that ability is tested by the two counterfactuals above, both of which make the abilities they test fully compatible with causation of choice and action.

We should be clear of the modesty of the conclusion reached so far in this Section. I have not asserted that all addicts have the abilities required to be responsible for taking drugs while they are addicted; rather, I have only sought to show that the disabilities rightly attributed to subsystems within the brains of addicts does not require us to infer that the addicts themselves lack the abilities required for responsibility. We may well infer that a given addict does lack such

abilities on a given occasion, but that will be on grounds different than the one we have been considering.

So, with this obfuscating brush cleared, we reach the substantive question that motivated this part of the Article: will the neuroscientific explanation of addiction described earlier help us with the moral question of when if ever addicts are excused by their addiction? Take the "imbalance" explanation proffered above one piece at a time, starting with the Berridge thesis.[298] According to Berridge, the brains of addicts have been rewired such that the drugs that initially gave them pleasure (in their hedonic hot spots) no longer do so, yet the wanting for drugs that that pleasure once brought on carries on despite the absence of much or any reinforcement of that wanting by a liking for drugs; further, that a potential for such liking-independent wanting persists so that confronting drug related cues can trigger such cravings long after all drug use has ceased.[299] The findings in neuroscience on which Berridge relies in the giving of this account were not the source of this explanation; phenomenology and behavioral psychology made this plausible before we knew that the mechanisms of pleasure were different in anatomical region and in neurotransmitter usage than the mechanisms of wanting and motivation. Still, the neuroscience helps to validate and refine these initial deliverances of phenomenology and behavioral observation, and it thereby gives us grounds to prefer this explanation to the many competing accounts we earlier explored based just of phenomenology and behavioral observation. It thus makes us more confident in the truth of the explanation vis-à-vis its competitors and this makes us consider more seriously the possibility of excuse lurking in such explanation (although, again, discovering the "mechanical filling" by itself does nothing to increase the excusing potential of such explanation.)

What are the moral implications of a desire (that motivates a wrongful action) being cut off from the normally controlling feedback of the reward circuit? I earlier opined that morality is indifferent to the sources of our desires, that responsibility rests on our abilities to choose to act or not to act on such desires, not on our ability to create or change those desires themselves. In a recent paper Richard Holton appears to dissent from this conclusion, urging that addicts have less responsibility for their choices than nonaddicts because Berridge's incentive salience explanation shows that addicts have "little control, once they encounter a cue, over their cravings."[300] The incentive salience account does indeed seem to show this. The account shows that one of the resources that we can use to tame our desires, bequeathed to us by natural selection, is absent from the tool kits of addicts. That resource was the dislike or at least absence of liking that normally feeds back to dampen the desire for the thing not liked. Addicts who wish to

298. Berridge, *The Debate Over Dopamine's Role in Reward: the Case for Incentive Salience*, *supra* note 228, at 419.

299. Berridge & Kringelbach, *Pleasure Systems in the Brain, supra* note 227, at 656.

300. Richard Holton, *Addiction, Desire, Pleasure, Pollution*, (forthcoming 2019), manuscript at 13 (quoted with permission of the author).

choose not to take drugs are deprived of this resource for dampening the opposing input to that wish in making that choice, according to the incentive salience theory. Yet my original point survives this thought. To be responsible for choosing to act on a desire does not require that we be responsible for the desire or its strength; the abilities relevant to responsibility are the abilities to choose otherwise and to act otherwise, not the ability to desire otherwise.

These same considerations militate against thinking that the alternative account of addicts' motivation to use drugs—the desires to avoid the "dark side" effects of not using drugs that George Koob believes motivated addicts—also have any potential to enlarge the category of those excused because of their addiction.[301] Indeed, these desires have even less potential for excuse because, unlike the desires of the incentive salience theory, the dark side desires: are not intrinsic desires but rather, instrumental desires in the service of avoidance of various forms of dysphoria; are not unmoored from the liking system; are not experienced as cravings. The only feature of these desires inclining towards excuse would be their strength, and strength of desire, as we have seen, is no grounds for excusing choices and actions satisfying of that desire.

Richard Holton seeks to supplement the positive motivational account of drug use by adverting to another feature of addicted drug use (in addition to the separation of wanting from liking Holton uses to ground his conclusion about the undampenability of drug cravings).[302] What Holton adverts to here is certainly a well-established feature suggested by the phenomenology of addicts. This is the fact that the choice to use drugs not only goes against what the addict likes but also against what the addict values. As Holton puts it, there is for addicts a "conflict between beliefs—beliefs broadly about what is valuable—and desires . . . . The crucial feature is that the unwilling addict judges that taking the drug is not the best course . . . ."[303] On this picture, addicts' decisions to take drugs flies in the face not only of what they like but also of what they value ("value" here being a shorthand reference to their evaluative beliefs about what is desirable and worth pursuing).

In my view the neuroscience that backs up this claim for phenomenology is both poorly developed (and incidentally, not distinctive of the incentive salience theory proper). For we do not know the anatomical locations or neurotransmitters distinctive of evaluative beliefs the way we know both of those things for both the reward and the wanting systems. All we have are fMRI studies showing the areas of the PFC activated when such evaluative beliefs are in play to control desires.[304] Still, it is plausible enough to suppose that there are such mechanisms and that one day neuroscience will discover what they are. It is also plausible to

301. *See* Koob & Le Moal, *supra* note 267.

302. *See generally* Holton, *supra* note 300.

303. *Id.* at 17. As was noted earlier, Holton's conclusion here is disputed by Gideon Yaffe. Yaffe, *Are Addicts Akratic?*, *supra* note 97, at 197–204. Yet Yaffe's criticism here is no more than an amendment to the thesis that addicts choose against their own values, amending the thesis to be diachronic rather than synchronic–addicts choose against the values they hold both before and after they choose to take drugs even though at the exact moment that they take drugs they most value doing just that.

304. Holton, *supra* note 300, at 11.

speculate that such discoveries will verify the deliverance of phenomenology here, namely, that addicts often choose against their own values when they choose to take drugs.

Suppose that all of this is well-established fact. Does it make for excuse? I see three routes by which one might think so. One is to go the route of the habit and learning theorists of addiction. One might reason that wants that are (1) intrinsic in the sense that they are not in the service of any further goals; (2) not in accord with what one likes (takes pleasure in); (3) not in accord with what one judges to be worthwhile and desirable; and (4) nonetheless win out in determining behavior, must be by-passing choice and causing drug-seeking behavior directly. The idea would be that a "choice" dictated by a craving with these features must not really be a choice at all, that it could not be since it is so irrational in its behavioral conclusion. Yet the unyielding barrier here is the same one against which the habit and learning theories generally come to grief: drug use by addicts just does not get experienced like nibbling on cake or other habits. Drug use is chosen; there is choice, and the machinery of choosing is not by-passed.

The second route is the one discussed in Part IV earlier: that addicts' cravings have the four features above mentioned might also tempt one to think that they are properly regarded as "ego-alien," *i.e.*, not part of the self.[305] If a want does not bring us pleasure when it is satisfied and we know that it will not, if that want is for something we do not think to be valuable and worth wanting, and yet that want is the one that determines our behavior, this might be taken to blunt the earlier objection I inherited from Freud, viz., that the size of our agency for responsibility purposes is not up to us to fix by our self-identifications. We might then justify feeling like there was indeed one "bad-ass demon" inside of us doing the choosing of behavior satisfying such a want.[306] Yet we know that there really is no such demon in possession of our choosing faculties. There is just each of us as whole persons doing battle with a craving that is admittedly hard to resist. But unless the strength of such a craving, together with its content going against what we like and value, are enough to excuse, our choice cannot be excused either. The decision is ours, and the craving is ours, however much we beat ourselves up about both having it and yielding to it.

The third route to excuse is more promising. What if the anatomical and chemical mechanisms that neuroscience may discover to underlie evaluative beliefs and the means by which those beliefs exert their influence over decisions, are of such a nature that long term drug abuse destroys the brain pathway by which such beliefs influence decisions about drug usage by addicts? That would seem to cut into the addict's ability to choose other than he did when he chose to use drugs. Applying the counterfactual test for free will: The addict could very strongly value all the things dependent on his not using drugs and thus strongly value abstinence over use, and yet those evaluative beliefs would be without causal effect on his choice because he would still choose to use drugs. That he

305. *See supra* Subsection IV.B.3.c.

306. A paraphrase of the cocaine addict quoted before in Smith, *supra* note 213.

would do so in such possible worlds means that he *could* not refrain from doing so in the actual world.

If such pathways of control are damaged such that, at least for choices in the face of drug related cues, evaluative beliefs cannot get any purchase in determining decisions, then such equipment failure would constitute incapacity, the kind of incapacity that should excuse. Indeed, one might imagine that some future neuroscience could likewise discover the pathways by virtue of which desires opposed to drug use—desires for nondrug rewards that seek things that give pleasure if attained and are therefore liked as well as desired—influence choices about drug use. If that science also discovered that such pathways were damaged such that such desires too, along with evaluative beliefs, were incapable of influencing decision, one would have additional grounds for concluding that the addict lacked the ability to choose not to use drugs. For in such a circumstance, no matter how strongly the addict wanted all the things dependent upon his not using drugs, he would still choose to use drugs because those contrary wants were without causal effect.

I conclude that the potential for neuroscience to expand the category of those excused by addiction rests mostly on the discoveries of a neuroscience not yet done. Indeed, such discoveries may depend on facts that do not exist, in which case such discoveries will never be made. The facts about the brain having the most potential for excuse are thus, unfortunately, the least known by the contemporary neuroscience of addiction. Still, that should be grounds motivating future research, not grounds for despair. Neuroscience may yet show us that more addicts are excused by the effects of their addiction than we had thought.

Apart from expanding the categories of offenders excused by addiction, neuroscience also has the potential to verify when those conditions we currently think excuse offenders, are actually present in individual cases. If we had definitive brain markers for the presence or absence of the mental states on which responsibility depends—intentions, factual beliefs, desires, likings, evaluative beliefs—and if we had such markers for the intensity and content of such mental states, together with the interrelationships between them, then we would have a larger and more precise evidential base from which to infer conclusions in individual cases. Such evidence could supplement or even supplant the phenomenological and behavioral evidence on which we now rely. Such a neuroscience may well be a long way off, but we should not lose sight of this potential for neuroscience to be the handmaiden to the law.[307] It need not always be revolutionary or even reformist of the law to be useful to the law.

307. The "handmaiden" characterization I take from the late John Cacioppo, my co-chair of the MacArthur Foundation's Law and Neuroscience Project's Intentions and Decisions Working Group, who at one point expressed his exasperation at my continued harping on the need for legal relevance of our work, "I didn't sign on to this to be your damned handmaiden."

VI. CONCLUSION

Some topics are hard because their subjects are complex and difficult. Sometimes, however, a topic can be hard not because the nature of its subject matter makes it so, but because those who have previously written about it have made it hard.[308] The relation of addiction to responsibility is a hard topic for the first reason, but it is also a difficult topic for the second reason as well. There are now such a number of blind alleys, tangents, and confusions introduced by the now vast literature on this topic that it is often difficult to see the real problems. Part of the ambition for this Article is just to clear these intellectual tangles so that the issues and paths for future research can be seen clearly. I have not organized this Article around these confusions—the discussion is rather organized in the manner described in the introduction as an undertaking of the three-fold tasks of describing, explaining, and evaluating addiction, with a neuroscience redux of these three tasks at the end. Yet during the course of the argument that completes these three tasks, I have sought to defuse a number of mistakes in the literature that would otherwise get in the way of framing the right questions to pursue. These have included:

1. To confuse the adjudicative question of moral excuse and legal defense, with the legislative question of moral wrongdoing and criminalization. Some of those who argue that addiction should be an excuse and a defense really think that the wrongs and crimes of which addicts are mostly charged—possession and use of drugs—are not really moral wrongs and therefore should not be crimes at all. My own libertarianism strongly inclines me to agree with this latter view.[309] But that has nothing to do with the issue of excuse and defense. For that latter issue, the question is whether addicts are compelled to serve their addiction in ways and to an extent that excuses such acts of acquisition and use of drugs; the question assumes that such acts are crimes and thus are moral wrongs—either because prohibited by law or antecedently so—because if that were not so, no issue of excuse could arise. Even if one thinks that the moral part of that assumption is wrong, as do I, one should welcome the separate discussion of excuse because some acts by addicts that are not possession and use but are nonetheless in the service of their addiction, are wrong by anyone's theory of morality; homicide and theft in the acquisition of drugs comes to mind.[310] In addition, however wrongheaded they may be, the present laws under which we live do criminalize drug possession and use, and the issue of legal defense based on moral excuse is thus quite real and pressing.

308. I fear that I have done a bit of that myself in times past. *E.g.*, my law partner Jerry Falk saying to me when I left practice and handed him my litigation files, "Did you just happen to get all the complicated cases in the office or did you make them that way?"

309. Moore, *Liberty and Drugs*, *supra* note 119.

310. An example is provided by the recent Canadian arbitration decision about a nurse who stole her patients' drugs to feed her own addiction. Peter Smith, *Nurse Who Stole Opioids Wins her Job Back Because Addiction Is a Disease, Arbitrator Rules*, NAT'L POST. (Jan. 18, 2019, 2:01 PM), https://nationalpost.com/news/nurse-who-stole-opioids-wins-her-job-back-because-addiction-is-a-disease-arbitrator-rules. I take her theft to be a serious moral wrong that should legally be prohibited and punished. The issue of excuse is thus squarely present in such cases even in legal regimes much more libertarian than our own in their drug policies.

2. To conclude that if retributivism is unjustified as a theory of punishment then the only question to ask about the punishment of addicts is whether the harsh treatment necessarily involved in punishment works to prevent recurrence of use, not whether such harsh treatment is undeserved. Many of those who write about the responsibility of addicts, particularly within medicine and neuroscience, are unsympathetic to retributivism.[311] And if retributivism is rejected generally as they think it should be, they infer that the question of whether addicts are sufficiently responsible to deserve punishment drops away—the only question is one of therapeutic efficacy, namely, is punishment (imprisonment and moral condemnation) a way of curing addiction? And of course, they are skeptical about that. But this ignores one of the major contributions to criminal law theory made by Herbert Hart, namely, the recognition that nonretributivists should care about moral blameworthiness and desert even though they reject retributivism.[312] Even if desert is not a reason to punish as the retributivist believes, it still can be (and is in fact) a reason not to punish when it is absent. The question of moral excuse for addicts is thus a very live question for retributivists and nonretributivists alike.

3. To infer that addicts must be excused whenever neuroscience, medicine, or some other science finds that their criminal behavior is caused by genetic, environmental, and brain factors that they do not control. (A related mistake is to infer excuse whenever a mechanical process in the brain is discovered to underlie or constitute the mental processes of decision by addicts.)[313] The mistake is a mistake because fully responsible choices may be made by addicts like everyone else even though those choices are identical to, or caused by, mechanistic or other causes in the brain. Excuse lies in compulsion, not in causation or in mechanistic reduction.

4. To conclude that addicts must be excused if their addiction can rightly be characterized as an illness or a brain disease. As we saw in Part II, being ill or diseased are medical conclusions having no import for moral or legal issues of excuse.[314] It is very unfortunate that so much of the literature portrays the moral issue of excuse as being determined by the simple dichotomy of "choice versus disease" models of addiction. It is also unfortunate that the issue when cast in these terms has become politicized so that it is now a battle cry rallying the troops to the barricades for either conservatives ("choice") or liberals ("disease").[315]

5. To conclude that addicts are not to be excused for their acts as addicts, no matter how compelled those acts might be, because almost all addicts at some point in time voluntarily ingested the drugs that eventually made them addicts and thereby culpably caused the conditions of their putative excuse. This is the

311. I am not among those. *See, e.g.*, Moore, *The Moral Worth of Retribution*, *in* PLACING BLAME, *supra* note 103, at 110–12.

312. *See* H.L.A. HART, PUNISHMENT AND RESPONSIBILITY (1968).

313. These are two of the four mistakes I unravel in my forthcoming book, MICHAEL S. MOORE, MECHANICAL CHOICES: THE RESPONSIBILITY OF THE HUMAN MACHINE (2020).

314. *See supra* note 26 and accompanying text.

315. *See, e.g.*, Editorial, *If Addiction Is a Disease, Why Is Relapsing a Crime?*, N.Y. TIMES (May 29, 2018), https://www.nytimes.com/2018/05/29/opinion/addiction-relapse-prosecutions.html.

mirror image of the previous mistake in that both purport to provide "Alexandrian" solutions to the knotty problems of addiction,[316] the one by slicing through all complexities by easily concluding that no addict is responsible and the other equally easily concluding that they all are. Each of these mistakes, while purporting to be a shortcut to truth, have instead been blind alleys taking theorists further away from the real issues.

There are other, less general mistakes that have also impeded progress here, such as the mistake of thinking that the loss of opportunity occasioned by withdrawal can provide grounds for excuse, or the mistake of thinking that the kind of control needed for responsibility is not just control of one's actions and of one's choices but that one must also have control of the existence of the desires that motivate choice. In any case, once all of these distracting confusions are put aside so that the complexity of the problem can be seen aright, my conclusions on the merits can be summarized in the paragraphs that follow.

My conclusions from Part II have been: that the common concerns of law, ethics, medicine, and neuroscience converge enough to think that all such disciplines and professions have reason to work with a shared conceptualization of addiction; that that shared conceptualization of addiction with which doctors, lawyers, moralists, and neuroscientists should work is that of continued and excessive craving for, and use of drugs, in the face of knowledge by users that that use causes serious problems to their lives; and that one should not completely separate such conceptualization from the explanations science discovers about addiction, nor from the features of addiction that its moral evaluation as a potential excuse might add, making any such conceptualization tentative and hostage to the future insights of both science and ethics.

Further, my conclusions from Part III were: that the background for understanding the decisions of addicts to use drugs lies in the general schema of the folk psychology by virtue of which we explain the actions of persons in terms of their characters, desires, values, beliefs, intentions, likes, and volitions, all of which are related by our notion of what constitutes rational agency; that the reasoning of that large subclass of addicts we called "unwilling" addicts can display a variety of different breakdowns in this model of practical rationality, breakdowns we rightly characterize as defects of rationality; that there is no one defect of rationality universally present in all choices by addicts to use drugs but that different addicts on different occasions seem to display different irrationalities; that as a matter of folk psychology, limited as that psychology is to behavioral and phenomenological evidence, no one model of addictive reasoning can be seen to predominate.

Further, my conclusions from Part IV have been: that, following on the prolixity of explanations for how and in what ways addicts may be irrational in their choices, there can be no simple, yes-or-no, across the board answer to the question of whether addicts should be excused for their addiction-related behaviors; that nonetheless most of the defects in the rationality of addicted decision-

316. The allusion in the text is to Alexander's famous "solution" to the problem of untying the Gordian Knot, the solution being to cut the Knot through with his sword.

making do nor rise to the level of moral excuse (and thus, not of legal defense), yet even with our present knowledge some defects do generate such an excuse and defense; that when the latter situation obtains, the excuse is usually only a partial one, mitigating but not eliminating responsibility and punishment.

Further, my conclusions from Part V have been: that the neuroscience of the late Twentieth Century produced a lotus-fruit kind of explanation of the behavior of addicts which if true had the potential to excuse virtually all repeat users of drugs, including but not limited to addicted users; but that such "demon drug" explanation turned out not to be true in certain crucial respects; that contemporary, Twenty-First Century neuroscience has deepened our explanation of addicted decision-making and focused that explanation on a few of the many factors part of the folk psychological explanation of addiction; that as yet such neuroscience has not produced the holy grail of explanations for purposes of excusing addicts, namely, an explanation showing that despite appearances there is no choice by addicts to use drugs, or showing how the choices that addicts do make in this regard are inefficacious in causing the actions chosen, or showing how such choices are themselves not the product of what the addict most wants or most values; that the existence of such an explanation must await future discoveries, discoveries that are possible but far from inevitable.