Georgia State University
ScholarWorks @ Georgia State University

Philosophy Theses | Department of Philosophy

4-4-2009

Drug Addiction and Personal Responsibility

Andrew Shawn Reagan
Georgia State University

Follow this and additional works at: http://scholarworks.gsu.edu/philosophy_theses

Recommended Citation
Reagan, Andrew Shawn, "Drug Addiction and Personal Responsibility." Thesis, Georgia State University, 2009.
http://scholarworks.gsu.edu/philosophy_theses/63

This Thesis is brought to you for free and open access by the Department of Philosophy at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Philosophy Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

DRUG ADDICTION AND PERSONAL RESPONSIBILITY

by

ANDREW REAGAN

Under the Direction of George Graham

ABSTRACT

This project examines drug addiction and personal responsibility from the perspective of three different types of theories of addiction: full responsibility, diminished responsibility, and no responsibility. The rational theory of addiction is the fully responsibility theory. The philosophical insights by R. Jay Wallace and George Graham are the diminished responsibility theories. Berridge and Robinson's Incentive Salience theory of drug addiction is the no responsibility theory examined. My conclusion is that diminished responsibility frameworks are the most suitable in a therapeutic context because they are most sensitive to relevant normative aspects.

INDEX WORDS: Addiction, Personal responsibility, Therapy, Normative, Incentive salience

DRUG ADDICTION AND PERSONAL RESPONSIBILITY

by

ANDREW REAGAN

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2010

Copyright by
Andrew Reagan
2010

DRUG ADDICTION AND PERSONAL RESPONSIBILITY

by

ANDREW REAGAN

Committee Chair: George Graham

Committee: Andrea Scarantino

AJ Cohen

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2010

DEDICATION

I am dedicating this thesis and my work in the academic world in general to my family. They have stood by me and helped me through hard times in my life, without which I would probably not be here in the shape and form that I am. I also dedicate my effort here to all ten directions around me, spreading out and never ending, finding and creating meaning for me in this world.

TABLE OF CONTENTS

INTRODUCTION 1

CHAPTER

1. THE RATIONAL THEORY OF ADDICTION 8

2. INCENTIVE SALIENCE THEORY OF ADDICTION 21

3. DIMINISHED RESPONSIBILITY THEORIES 36

WORKS CITED 44

INTRODUCTION

My aim here is to provide an account of various theories of addiction and show how these theories are committed to assigning a degree of personal responsibility to the addict for their actions. I will focus on *drug addiction*, since I take drug addiction as the prototypical case of addiction. My assumption here is that what can be said about the degree of personal responsibility in a drug addict will be applicable to other types of addicts (gambling, food, sex, etc.).

The substantive portion of my thesis is an attempt to show how a diminished responsibility view of addiction is the most appropriate for a therapeutic context. I show this by first rejecting full and no responsibility accounts of addiction as appropriate for therapy (for separate reasons for each). I then show how diminished responsibility theories are appropriate.

It is not necessary that a theory of addiction benefits therapy (one can imagine a more strict neurobiological theory of addiction remaining disconnected from a therapeutic context). It is a contingent fact, however, that all of the theories examined in this thesis are either constructed with the intent to improve addiction therapy or have been (after some time in the academic sphere of analysis) come to impact addiction therapy indirectly by influencing the thoughts of people involved in therapy (other addiction theorists, therapists, etc.). The scope of my thesis, therefore, will be restricted to the therapeutic context. I am not concerned at all here with some hypothetical sphere of ‘pure description’ for the term ‘addict’. If such a sphere exists at all (and I am not certain it does), it is not my business here to examine the ideas and statements of people working within that sphere as I see no likelihood with that work for improving the welfare of people within our society that suffer from addiction.

The focus of this thesis is determining to what degree, relative to each particular theory, addicts are *personally* responsible for their actions. The theories are then examined for problems and inconsistencies. There are three broad categories of personal responsibility that an addiction theory can be committed to: full responsibility, diminished responsibility, and no responsibility. The concept of personal responsibility is a complex concept (meaning it is composed of other complex and/or simple concepts). One of the primary constituents of the concept (and the aspect I will focus on most heavily) is the concept of *personal control*. Personal control over one's actions means having the ability for *self-management* regarding the actions one produces. A person that lacks personal control would lack the ability to self-manage their behavior. This type of person would be a prime candidate for therapy that externalizes control (such as hospitalization/institutionalization). The remaining constituents that compose the concept are concepts of *autonomy* and basic concepts of *personhood*.

Having full personal responsibility over one's actions means being able to control what one does in a given situation and operate as an autonomous person through one's actions. Having partially diminished responsibility over one's actions means being able to partially control what one does in a given situation, but also occasionally succumbing to personal biases, cravings, or some other sort of internal influence that may run counter to what one *believes*[1] one ought to do, thereby potentially lowering the level of autonomy of the agent (by potential harm caused by operating in a state of diminished personal responsibility). No personal responsibility over one's actions means not being able to control oneself in a given situation; the prototypical version of this would be a compulsive drug user; a person who has no personal control over the drugs they consume.

[1] This *believing* can either be in the immediate moment *or* the beliefs the organism would entertain in a more sober/rational frame of mind.

The three degrees of *personal* responsibility closely track comments made by Gideon Yaffe in a recent (2002) article where he discusses *moral* responsibility and addiction. I am bracketing off some of his remarks here for this project, since I am *not* working with moral responsibility (which would require a separate theory of moral responsibility beyond the conception of personal responsibility I am offering). Personal responsibility is ontologically a more basic responsibility than moral responsibility. Personal responsibility is required in order to posit moral responsibility, but one can fruitfully discuss personal responsibility without discussing moral responsibility (but obviously not vice versa). I remain agnostic in this project whether the degree of personal responsibility an addict has is identical to the degree of moral responsibility. The assumption by many is that they are identical degree properties (the degree of moral responsibility obtains from the degree of personal responsibility). However, to assent to this claim would require a substantive theory of moral responsibility.

What is drug addiction? For the purposes of this thesis, I will follow Robert West and begin with defining addiction as 'impaired control over reward-seeking (drug-taking) behavior from which harm ensues' (2002, 3). This definition allows for all degrees of personal responsibility (even full responsibility is not ruled out *a priori* based on this definition).[2]

Compare to the DSM-IV definition of drug addiction: (1) compulsion to seek and take the drug, (2) a loss of control in limiting intake, and (3) emergence of a negative emotional state when access to the drug is prevented (APA, 1994). Defining addiction as compulsion allows a theorist to rule out *a priori* the possibility for full personal responsibility theories of addiction.

[2] How could a theory of full responsibility be included under this definition (that begins with "impaired control")? Rational theorist have answered this before by claiming that the impaired control does not have to occur in the faculties of reasoning, but rather can be subsumed under the knowledge base of the addict. The addict just might not know certain things about the effects of the drug use, which can impair their ability to fully consider/reflect on the safety of their use. Their reasoning faculties are not impaired in this case (compare to the diminished responsibility theorist where the impairment does come in the decision-making faculties of the person, West, 2002).

On the face of it, however, full personal responsibility theories of addiction have some merit and should not be ruled out *a priori*. So, the definition of addiction that rules these theories out must not be a good definition to start with. Therefore, West's definition is more suitable for our purposes.

Quite a few addiction theorists also agree about the importance of an explanation of the neurobiological mechanisms that underpin the transitions of a non-addicted person into an addicted person. It is my view that there are certain facts[3] about addiction that cannot be determined without empirical evidence. We need neurobiological evidence to weigh in and support the claims that theorists are making. In a recent textbook on addiction, Koob and Le Moal claim:

> "An important goal of current neurobiological research on addiction is to understand the neuropharmacological and neuroadaptive mechanisms within specific neurocircuits that mediate the transition between occasional, controlled drug use and the loss of behavioral control over drug-seeking and drug-taking that defines chronic addiction[4]" (Koob and Le Moal, 2006).

It is for this reason that Berridge and Robinson's neuroadaptationist Incentive Salience theory of addiction is so widely regarded as useful and correct (to be reviewed critically in chapter 3). All things being equal, a theory with more empirical evidence weighing in its favor is the better theory.

The next section is an overview of where this thesis is headed, chapter by chapter and outlines the major arguments/revisions I will suggest in regards to the specific theories I am analyzing.

---

[3] i.e. the transition from previously widely accepted positive hedonic models of addiction to the now contemporarily widely accepted incentive salience model of addiction. This transition required empirical studies to prove that addicts were not operating based on a drive toward pleasure, but rather were operating based on the 'wanting' (Berridge and Robinson, 2000; 1993).

[4] Their use of 'chronic addiction' here instead of simply 'addiction' is due to their assumption that nobody can become addicted without a somewhat extensive history of use. It takes time to become fully addicted (i.e. the conformation of certain key brain areas).

1.1 Overview of Thesis Chapters

In chapter 2 I discuss the Rational Theory of Addiction (RTA) by Becker and Murphy (1988). Their theory (as cited above) is a full responsibility theory of addiction. They believe that addiction is the result of rational processes that are fully under the control of the individual (hence the full responsibility view they are committed to). The same kind of explanation that would be given for why a person decided to eat at this restaurant rather than this other one, or a why a person decided to postpone work for recreation, could be given for an addict that decides to consume crack cocaine every night. I outline the mathematical formulae involved in the theory and then show deficits in their ability to explain aspects of addiction are a result of what I call the 'diagnostic conditions' for individual addicts. The main problem with RTA is its inability to provide meaningful predictions about the nature of relapse, narrowly, and more broadly speaking provide meaningful predictions about the internal states an addict goes through that will affect the type of therapy that would be appropriate to give them.[5,6]

In chapter 3 I turn my focus to Berridge and Robinson's theory of addiction: the Incentive Salience theory (1993; 2000). They construe addiction as *compulsion* that arises after complex neurobiological adaptations in the mesotelencephalic dopamine system. Basically, their view is that this system sensitizes to a set of drug 'cues' or stimuli that provide motivation (incentive) to the organism to consume the drug. If the sensitization reaches a high level then the incentive can be so strong that the organism is *compelled* to follow through with consumption (despite perhaps having a strong belief that they should *not* consume).

---

[5] I am not the first person to notice this fact about RTA. See West, 2002 for more discussion of the deficits in RTA.
[6] Relapse is a reinstatement of drug-seeking and drug-consuming behavior after the addict has quit that behavior. Note, if one's supply of the drug is cut off (for instance, one's dealer is incarcerated) and the addict is forced into withdrawal and a temporary break and then shortly afterwards (days, weeks, it does not matter) the addict uses again: this is not relapse. Relapse connotes an attempt (whether entirely autonomously or with the support of family, friends, or professionals) to quit the addiction. So, relapse is a failure to successfully quit the addiction.

The major problem I find with this theory is that Berridge and Robinson do not necessarily have to commit themselves to such a strong view of addiction (that it is coextensive with drug compulsion) and in doing so have effectively described the set of addicts as persons without autonomy[7] (and in describing them this way have provided constraints on the imagination of the addict during the recovery process). If it is not necessarily true that addiction is compulsion (and the evidence they present to support that view also equally supports a diminished responsibility view of addiction), then I suggest a revision to their theory: construe addiction as being 'strongly urged'[8] to consume and one can effectively circumvent the normative implications cited above while still preserving the spirit of the project and the validity/soundness of the theory given the set of evidence they provide for the truth of the theory.

The situation here is that we have two conceptual claims: 1) addiction is compulsion as a result of a-rational motivational forces, and 2) addiction is being 'strongly urged' by a-rational motivational forces. If the evidence the theorists provide equally supports both claims, then there is no way to decide given the restriction to empirical data. Unless we are to throw up our hands and claim indeterminacy here, I suggest, since we are working with theories that are intended for a therapeutic context, that we turn to the normative implications of the statements and determine whether these implications weigh in favor of one claim or the other. Once we bear the normative implications in mind, it is my view that 2 (the 'strongly urged' revision I propose) is the better conceptual claim to assent to given that the theory will be incorporated within a therapeutic context where the descriptions of the terms used will bear on the imagination of addicts.

---

[7] What I mean here with the term autonomy is the ability to personally control one's own actions. A person who lacks autonomy cannot personally control their actions and is either out of control or under the control of somebody else or some institution (i.e. mental asylum).

[8] What I mean by 'strongly urged' is very close to the diminished responsibility theorist's talk of being influenced by desires/cravings/impulses. These incentive attribution structures activate and 'attune' the organism to what is important in their environment. The process of being 'strongly urged' is ignoring (or truncating) parts of their rationality.

In chapter 4 I discuss a few different philosophical approaches to diminished responsibility theories of addiction. I begin with Harry Frankfurt (1971). I also cover R. Jay Wallace's theory of addiction (as impairment to volition from conflict with intensely phenomenological a-desires, 2006) and George Graham's view of addiction (Forthcoming, The Disordered Mind).

I then show how once Berridge and Robinson's theory has been revised to back the conceptual claim from compulsion to 'strongly urged' that their neuropsychological adaptationist theory of addiction is consistent with these more purely philosophical theories. It is my view that Berridge and Robinson's theory (with my suggested revision) is the best possible neuropsychological theory of addiction available to us now and that the philosophical theories shed light on some of the more nuanced conceptual issues at work within their theoretical framework.

CHAPTER 1

THE RATIONAL THEORY OF ADDICTION

1.0 Introduction

In this chapter I examine the Rational Theory of Addiction proposed by Becker and Murphy in 1988. Their theory is an extension of rational choice theory to explain and predict aspects of addiction. Becker and Murphy conceive of addiction as a process of deciding to use (or abstain) based on a rationally well-informed decision that consumption (or non-consumption of) the drug will maximize benefit over costs.

In section 1, I outline the Rational Theory of Addiction (RTA hereafter). In section 2 I cite the primary explanatory problem with RTA: they are not able to differentiate *in kind* between an addict that wills their use to continue and an addict that wills their use to stop (yet continues despite their will, Frankfurt 1971). In 2.1 I propose formal 'diagnostic conditions'[9] for RTA and the term 'addict' that I use to illustrate the explanatory deficit in RTA. The explanatory problem is serious for RTA because the inability for differentiation in kind means that it would be poorly suited in any therapeutic context (insofar as these two different types of addicts ought to be treated very differently in therapy).

2.0 Overview of the Rational Theory of Addiction

The fundamental idea of RTA is that addicts make choices in pursuing and using their addictive good with forward-looking, optimal[10] behavior (Becker and Murphy, 1988). RTA theorists maintain that addictive behavior can *always* be explained by appealing to the rationality

[9] By 'diagnostic conditions' here I mean what are the conditions according to the RTA theory that make a person an addict. I will propose in section 2 a set of variables that represents each individual addict. These variables are the aspects of analysis from the RTA theory.

[10] Optimal according to the 'epistemic boundary' of the individual (i.e. the boundary according to what the individual knows). Optimality here might be construed along two dimensions. Dimension 1: relative to the individual's belief/desire set. Dimension 2: relative to a theoretical individual possessing a normatively non-deviant belief/desire set. It is the former upon which optimality in the decision-making process for using addictive goods is based (i.e. on the individual). As a sidenote, the latter dimension, I take it, is the dimension upon which a RTA theorist would base a definition of the *pathology* to addiction (i.e. the pathology rests in deviating in a certain way which results in harm from the normative basis of the non-deviant belief/desire set).

of the agent. RTA holds that an addict's *choice* to consume a particular 'addictive good' is based on a rationally well-informed decision that using the drug will maximize benefit relative to the future costs likely to incur.

How did RTA theorists (Becker and Murphy in particular) arrive at this theoretical commitment? Before Becker and Murphy's (BM hereafter) model, the predominant models of addiction were chemical dependence models based on animal studies and focusing on past consumption as the primary causal factor. Past consumption reinforced future consumption through the plasticity of brute dopaminergic mechanisms (which were thought to be the reinforced 'reward' pathways).[11,12] While reinforcement in drug use behavior in the sense that the animal model of chemical dependence elucidates is a necessary aspect of addiction, the scope of the conclusions found within the animal model of addiction is limited when applied to complex rational agents such as ourselves.

The impetus behind RTA was the idea that humans can be modeled as having conceptually quite complex *reasons* behind their behavior in using addictive goods that accounts for *both* future planning and a stable preference base (as well as reinforcement effects from past consumption). RTA maintains that addicts are not blindly operating based on brute dopaminergic mechanisms, but rather produce behavior as a result of complex reasoning processes. In this sense, RTA is a departure from the no responsibility views of the positive reinforcement chemical dependence theories (or of Berridge and Robinson's Incentive Salience theory). BM's theory is a full responsibility theory of addiction because the addict's *choices* are made from a completely rational basis (there is no recognition of manipulation from irrational impulses or cravings).

---

[11] For an updated theory of these dopaminergic pathways see Berridge and Robinson's (2000) Incentive Salience theory (discussed in chapter 3). The activation of dopaminergic pathways no longer correlates with reward, but rather correlates to the 'wanting' of the reward.

[12] For a good survey of neurological theories of drug dependence and drug habituation, see Kauer and Malenka (2007).

Rational is used within a very particular context here. The RTA theorists do *not* mean rational, i.e. reasonable.[13] Rational is used to particularly connote a type of 'means-end' reasoning that our actions can be analyzed *as if* they are obeying the simple mathematical formulae used to model them (introduced later in this section). Despite the 'common-sense view' that the behavior of addicts seems at times irrational and inexplicable, the basis for explanation in RTA of the behavior an addict produces can be found in the same substrate that would explain a choice a person makes on where to go out to eat tonight or what to do with a couple of spare hours on a Saturday afternoon. There is no difference in the decision-making process of a heroin addict and a 'normal'[14] person. The only difference is that the heroin addict is making decisions about an addictive good and the 'normal' person is making decisions about eating out at a restaurant or going to watch a movie (i.e. about non-addictive goods).

The arguments that Becker and Murphy propose to support their view are completely empirical in kind. For instance, in the 1950s when the surgeon general announced that cigarettes cause cancer, they note that the consumption of cigarettes dropped off in the United States market. They insisted this supported their view that people consume addictive goods according to the rational principles they proposed.[15]

Remember West's cursory definition of addiction that I am using throughout this thesis: 'impaired control over reward-seeking (drug-taking) behavior from which harm ensues.' The

---

[13] The distinction between rational and reasonable I owe to John Rawls. A rational agent has reasons for her actions. A reasonable agent is rational and has *good* reasons for her actions. What is a good reason? John Rawls thought that it had to do with an 'overlapping consensus' between different people (i.e. the reasons we could intersubjectively agree are good). So, a neo-Nazi might be rational (i.e. have reasons for hating some race or whatever it is that they do hate), but they are quite obviously not reasonable (they do not share enough of an overlap with our beliefs).

[14] Scare quotes around normal, since I have absolutely no idea what a good notion of normalcy actually is. Surely it is not derived from the statistical norms of the society (since those can actually be quite abnormal). We do, however, sort of understand what we mean when somebody says 'normal' with scare quotes. We mean something like the guy (or girl), married, maybe with children, no drug addictions, no wild perversions, etc.

[15] They remain agnostic about the nature of the internal mental processes of addicts. They simply provide the mathematical models (and the empirical studies that support their mathematical model) and suggest it can be used as a predictive tool to understand the nature of addiction.

impairment of control according to RTA is in the knowledge base of the addict and not in the deliberative reasoning processes. They may not know certain critical facts about the drug they are consuming (for instance, that it causes cancer) or they may not weight these facts (if they do know them) with high enough priority (they may not be aware of the outcomes of having something like cancer). If I do not know critical facts for helping me make important decisions (for instance about cigarette consumption), then I am impaired in my judgments and decision-making process.

Consider the possibility that cigarettes do not actually cause cancer. If this possibility were true, then it would be understandable (rational) that *more* people would smoke cigarettes (insofar as there would be less of a reason *not* to smoke). Now return to the real world where cigarettes cause mouth, throat, and lung cancer, but focus on an individual who does not know that fact. Relative to this individual's belief set it seems more reasonable to smoke than it does relative to your belief set (that contains the belief cigarettes cause cancer).[16]

The important point here is to note that 'impaired control' is consistent with rational choice theory. The impairment of control is *not* a lack of control. The level of control according to RTA is complete (the same level of control covers a heroin addict as it does a 'normal' person). The impaired control is not explained in the same way as Berridge and Robinson in chapter 3 (or the diminished responsibility philosophical theories in chapter 4), but it is still nonetheless impaired control.

---

[16] There are two dimensions that rationality can be analyzed. The first dimension is relative to an individual's belief set. This is the dimension that RTA analyzes the decision-making process of the individual. The rationality of the individual can also be analyzed along a dimension relative to an 'appropriate' (perhaps intersubjectively shared) belief set (something like the 'right' belief set to hold). Along this dimension RTA analyzes the decision-making process of the individual in order to posit the impairment (by comparing the appropriate belief set with the actual individual's belief set). Further, these two dimensions of rationality can be used in order for RTA to be able to say an individual is *pathological*. If an individual is pursuing the consumption of crack cocaine every night for days in a row, then it is quite obvious that relative to the 'appropriate' belief set, this individual is behaving pathologically. They only get to say this, however, once these two dimensions of rationality are both held in mind and contrasted with each other.

Reinforcement, utility, tolerance, stock, time preference, and *wealth constraints* are the basic constituents of the RTA concept of addiction. Understanding these basic concepts is necessary in order to understand the mathematical model of RTA that utilizes these concepts.

Reinforcement is the effect that consumption has on the future utility (usefulness) of the addictive good (positive reinforcement increases utility and negative reinforcement decreases utility).

The utility of an addictive good for an individual is contingent on a relation between past consumption and present consumption (the type of relation that economists call an *adjacent complementarity*[17]). This idea operates behind what BM calls the basic idea of RTA: that "current consumption of [the addictive good] increases future consumption of [the addictive good]" (1988, 681). If current consumption increases future utility, then it is understandable why the consumer pursues using the addictive good in the future.

Calculating utility is derived from the relation between past and present use. The relation of adjacent complementarity between past and present consumption defines a necessary condition on whether a good is addictive or not.[18]

Tolerance is the aspect of consumption of a particular addictive good in which an increase in present consumption results in a decrease in future utility (i.e. more units of the addictive good must be consumed to maintain the same level of utility (BM, 1988, 682).

The concepts of reinforcement and tolerance are integral to each other and are what provide the physiological 'tension' to an addiction (if future use is simultaneously less satisfying

[17] Adjacent complementarity here means a relation between two temporally ordered events and the interdependence of the later event on the earlier event. In specific for addiction, consider two drug consumptions, happening apart from each other (maybe the two drug consumption experiences happen two nights in a row). The second drug consumption experience will be affected by the first drug consumption experience. The second night the person will be sensitized to the drug a bit more and of course tolerance will happen as well. This relation is known as an adjacent complementarity.

[18] As should be evident in the next paragraph on tolerance, the necessary reinforcement condition is not a sufficient condition for a good being addictive.

(tolerance) and more useful (reinforcement) then it is obvious why an addict would escalate their use). Reinforcement and tolerance do not develop together at a linear rate. Reinforcement might develop quicker than tolerance, or vice versa.[19,20]

The utility (u) of the addictive good for an individual at a specific time (t) equals a function of actual consumption at t (C_t) and the *stock* (put another way this is the *strength* of the addiction) that the individual has built up in the addictive good up to t (S_t). The relation of C_t to S_t is the complementarity I mentioned previously. The following formula illustrates this:

$$u_t = u(C_t, S_t) \qquad (1)$$ [21]

The stock an individual possesses in an addictive good is a function of reinforcement and tolerance effects relative to the individual's psychology (some individuals may be more resistant to the reinforcing effects of a drug than others). The level of stock that an individual has built up in the addictive good is derived from taking into account the rate of depreciation of the addictive good (Ω)[22] subtracted from one, multiplied by the stock of the addictive good the individual possesses at a time, t-1, plus the actual consumption at t-1. This is expressed as follows:

$$S_t = (1 - \Omega)\, S_{t-1} + c_{t-1} \qquad (2)$$

Also critical to an understanding of addiction is the individual's *time preference*. Time preference is an individual's predilection toward favoring, for instance, current utility versus future utility. BM write:

---

[19] Note that this formulation of the addiction concept applies equally well to beneficial or harmful addictions. BM distinguish beneficial and harmful addictions based upon whether or not the addiction has a positive or negative impact on total utility of the individual and/or earnings of the individual (1988, 684). The obvious candidate for a distinction between harmful and beneficial addictions (welfare of the individual) is not pursued by BM. According to Rogeberg, doing so would rule out their possibility for adopting the as-if stance (2003).

[20] There is also the possibility for the phenomenon of *reverse tolerance*. It is possible that by taking a substance, it is not necessary to take an equipotent dose in the future to achieve the same intensity of drug effects. Rather, it is possible to take *less* in the future than was required in the past. This occurs in substances that cause long-term biochemical changes (either in enzymatic pathways or receptor upregulation or some similarly isomorphic effect).

[21] This formula would be used in order to prove a good as addictive (if past consumption reinforced future consumption and tolerance effects were noted, then the good could be considered addictive).BM note that this point is *necessary* but far from *sufficient* to explain what addiction is (1988, 681). In addition, one would have to show the addictive good displays aspects of tolerance.

[22] Becker and Murphy call this number the "exogenous rate of disappearance of the physical and mental effects of past consumption of C" (1988, 677). This rate will be specific to each individual, since the particular psychological traits of the individual will influence the rate that the stock of the good diminishes (i.e. some people will remain more addicted to a substance based on past consumption than other people).

> Our analysis implies the common view that present-oriented individuals are potentially more addicted to harmful goods than future-oriented individuals [see formula 3 below for clarification -AR]. The reason for this is that an increase in past consumption leads to a smaller rise in full price when the future is more heavily discounted (1988, 682).

Finding the lifetime utility (U) of an addictive good for an individual at a certain time period (T) is just a matter of summing the utility across specific times across T adjusted by the individual's time preference (μ). This is done by taking the sum of 1 plus μ, raised by the difference of T minus t+1, multiplied by the utility of the addictive good that the individual possesses at a certain time, t.

$$U_T = (1 + \mu)^{T - (t+1)}(u(C_t, S_t)) \qquad (3)$$

If wealth was unrestricted, then individuals with time preferences biased toward the present would quickly escalate into using their addictive good to the exclusion of all other activities.[23] In the real world, however, addicts face constraints on their use. Wealth is one of the primary constraints. An addict under the RTA model balances current use with the level of wealth they possess, relative to plans for future use. Wealth is the constraint that BM chose to quantify in order to model the 'real-world' constraints on the addict.[24]

In this section I have provided a basic understanding of the concept/theoretical structure of RTA as well as the mathematical formulae underpinning RTA.[25] The following section examines the explanatory problem with RTA and I propose formal 'diagnostic conditions' for

---

[23] This 'unlimited-wealth-present-biased' addict would be reminiscent of experiments of a mouse with unlimited access to cocaine. The mouse (and the theoretical addict) would quickly drop all other activities (including food and sleep) and pursue the consumption of their addictive good to the point of death.

[24] There are other constraints that could be added: familial/societal constraints, time constraints, intelligence constraints, etc (for the purpose of the mathematical formulae below wealth is the only constraint of this type added). I take it that a complete RTA framework would have to incorporate these constraints to properly predict behavior. Therefore, a final formula is necessary in order to situate RTA in a 'real-world' context: a formula that delineates the interaction between wealth (and the other mentioned constraints) and rising utility the individual possesses toward the addictive good. For sake of room here I will not go into this complex formula. It is enough to note that the individual has a 'real world' constraint (wealth) on the amount of consumption of the addictive good they can exercise (and embedded here is a normative dimension of what one *ought* to do given their preferences for using compared to their overall wealth level and the necessity to use wealth to maintain basic life functions: eating, sleeping, etc). The individual is constantly balancing current use with a plan for future use.

[25] This model is based on Becker and Murphy's model (adapted from Fehr and Zych's interpretation, 1998, 646).

the term ‘addict’ according to the theoretical structure of RTA. I use these diagnostic conditions to elucidate the problem with RTA and relapse by way of a thought experiment.

3.0 The Explanatory Problem in RTA

The main difficulty RTA faces is explaining the phenomenon of *relapse*. Relapse is particularly important for an addiction theorist to work on because working with addicts that are on the verge of relapse (or have relapsed before and are likely to again) comprises much of an addiction therapist’s work. It is, therefore, imperative for a theory of addiction to provide something substantive in way of therapeutic implications for relapse.

An addict is only able to relapses *after* they have decided they should quit using (and have quit). Imagine a person struggling with a heroin addiction for years, constantly battling the temptation to increase their regular dose, always chasing some fading image of a blissful high they remember from the early years of their habit. After enough help, either from friends or family or therapist, or even just a steeled reserve of their own mind, this person quits their habit. Days of sobriety turn into weeks and months of sobriety. The person begins to get their life and body in shape and recovered. They regain self-respect and respect from their peers and stabilize themselves financially. Enough time passes and the cravings to use again that once were so strong, so overwhelming, begin to subside. What was initially a strong phenomenological push to use might become just a latent desire, hardly noticeable.

There are three primary possibilities for the person in this situation. I will focus on the last two possibilities (as they are the possibilities that entail relapse). First, the person could continue to abstain from heroin (keeping their will in place) and not use the rest of their life, leading a normal life apart from a few rather bizarre memories of being on a lot of heroin laying numb somewhere in some room in some house and some vague memories of the pain of

intermittent withdrawals.

Second, this person could decide to drop the will to not use and form a will to use again (despite the memories of pain and suffering that the drug brought, the person desires so strongly the sense of drug euphoria that it seems 'worth it'). Here we have a relapse as a 'willing addict' that Harry Frankfurt discussed in 1971.

Third, and perhaps most tragically, the person could maintain their will to not use (maybe in the face of unexpected, mounting cravings) and despite their best efforts, a momentary lapse in judgment could result in their using again (and then of course the past habitual routine use begins once again). Here we have a relapse with Frankfurt's 'unwilling addict.'

What does RTA have to say about relapse? And, more importantly, what does RTA have to say about the two possibilities of relapse the addict faces? A *reason* can be formulated utilizing the RTA framework to explain why the addict relapsed in either case. This would be in the language of RTA and would have to do with the stock in the drug (heroin) that the individual possesses (i.e. how long ago they used, upregulation of opioid receptors in the individual's brain, lack of endogenous opioids from past abuse, etc), the time preference of the individual, etc. RTA is fine here at giving a reason for both types of users.

The problem RTA faces is making a distinction between the two types of relapse (i.e. relapse by a 'willing addict' and relapse by an 'unwilling addict'). Since there is no way to make a modification to the theoretical structure based on differences in volitional (willful) states of the organism, the RTA framework is only able to recognize these addicts on par as equivalents (on pain of assumption their past histories of consumption and their present behavior is isomorphic).

The problem, more succinctly, is that there are *irrational* aspects that contributed to the third possible user's relapse that were not present in the second possible user's relapse. The

irrationality of impulsive cravings is a strong deterrent to careful and calculated reasoning. Sometimes, despite having a good reason to not use and having made up one's mind to not use, the person will end up (tragically) using in the face of strong cravings. Quite simply, they cannot help themselves. It is not surprising RTA has difficulties explaining this possibility, because RTA also has commitments that the choices a drug user makes are done with *full responsibility* (so the concept of irrational forces of craving acting on them is difficult if not impossible to parse apart conceptually within the RTA framework).[26]

3.1 'Diagnostic' Conditions for the term 'addict'

In order to formalize the diagnostic conditions for the term 'addict' according to RTA, I propose conceptualizing the diagnostic conditions in terms of a multi-variable set. The variables are the same as the aspects analyzed in the mathematical formulae above.

{N, TP, U, W, X, S}

N: the name of the addict

TP: time preference

U: lifetime utility

W: wealth constraints

X: x constraints (where x is undesignated)[27]

S: stock

In order to provide the diagnostic conditions for a particular addict one would need to obtain the individual variables involved in the set above. The diagnostic conditions I have proposed here are basically a way to formalize the extent of the aspects of an addict that RTA focuses on within

[26] RTA, like the other addiction theories I am examining in this project, do not explicitly state their commitments to some degree of personal responsibility. It is left up to us to derive their commitment from their other statements.
[27] X could designate, for instance, social constraints (familial or workplace constraints). The point is that wealth constraints are not the only type of constraints worth taking into account.

their theory. Any aspect falling outside of the set above is not relevant for analysis according to the theoretical structure of RTA.

Now let us return to our two possibilities for relapse. I represent these two possibilities with two individuals (call them Sam and Rebecca). Sam has a will to use during his relapse (i.e. his use is in accord with his will) and Rebecca has a will to not use during her relapse (i.e. her use is *not* in accord with her will). According to this quick thought experiment, assume that Sam and Rebecca have identical drug use history and also have identical time preferences and wealth/social constraints. Fill in the variables for Sam and Rebecca and you will produce sets that are nearly identical (they differ only in the 'name' variable which designates the individual).

But, clearly the two types of possible addicts highlighted above (the 'willing addict' and the 'unwilling addict') are not the same kind of thing (or should not be construed as the same kind of thing). In fact, it would be reckless and negligent to construe these two possibilities in the same way in a therapeutic/clinical (or legal) setting. The first type of addict would not be expected to show much remorse upon using (since her actions are *ex hypothesi* in accord with her will) – until perhaps later when she has realized 'what she has gotten herself into' so to speak. The second type of addict with a will to not use would be expected to show great remorse and regret for her actions (after relapsing back into heroin use). A therapist would want to be ready to deal with the onset of depression and learned helplessness in this individual. **If RTA is unable to account for this distinction as it stands, then it is not a sufficient theory for explaining addictive behavior within a therapeutic context.**

4.0 Conclusion

RTA is a useful theory of addiction if the desideratum of the theory is restricted to prediction of consumption of addictive goods on a societal level (i.e. not relative to a singular

individual but based on group trends and probabilities). The theory runs into problems, however, when the framework and mathematical principles are applied to an individual in an attempt to *explain* what is going on in the individual's mind, specifically during and after the phenomenon of relapse (which is an aspect of addiction that is of primary interest for any therapeutic context).

The RTA theorist can maintain that the challenge of explanatory impotence misses the *point* of the theory. The theory just *is* a theory of group prediction for the consumption of addictive goods. My contention with this counter-objection is that despite (perhaps) originally being intended as simply a theory of prediction, RTA is cited in the literature as providing a possible basis for analysis of welfare and as an explanatory construct of what is actually happening in the individual's mind. In this respect, this essay has been an attempt to block any future attempts at extending the framework of RTA into a therapeutic context. More suitable theories of addiction for a therapeutic context are the diminished responsibility discussed in chapter 4. These theories provide an explanation of the many aspects of addiction that are required in order to approach a full understanding of this phenomenon.

In chapter 3, I discuss Berridge and Robinson's incentive salience theory of addiction. I examine their empirical claims and offer revisions to their conceptual commitment to addiction as *compulsion*.

CHAPTER 2

INCENTIVE SALIENCE THEORY OF ADDICTION

1.0 Introduction

In this chapter I examine a no responsibility account of addiction, Berridge and Robinson's neuropsychological theory of addiction as compulsion (Berridge, Robinson, 2000; Berridge, Robinson, 1993, Berridge, Robinson, 2004). Their account is committed to a no responsibility view[28] because of their commitment to addiction being a *compulsive* process (i.e. the addict *must* consume the drug once the desire for the drug hits them).

Section 1 outlines Berridge and Robinson's view of addiction. Their view is concerned with the processes in the brain responsible for the attribution of *incentive salience* to the representations consequent of environmental cues for drug consumption. The basic idea behind their theory is that as the person becomes habituated to drug use, cues in the environment that signal imminent drug consumption gain 'incentive salience' (i.e. they are recognized as more important) and focus the energy of the person upon the drug consumption. At a certain level, the level of 'wanting' (from increased attribution of incentive) becomes so high that the addict is compelled to consume the drug, despite perhaps having a desire to not do so (i.e. attempting to quit).

Section 2 draws the relevant normative implications that accompany a commitment to the no-responsibility aspect of their theory. It is my view that Berridge and Robinson's account of addiction does not require a commitment to compulsion as a *necessary* part of addiction, e.g. they use an inductive/empirical argument to support their conceptual claims of compulsion. However, if another conceptual claim (that addiction is being 'strongly urged') is equally

[28] As a reminder: No personal responsibility over one's actions means not being able to control oneself in a given situation; the prototypical version of this would be a *truly* compulsive drug user; a person who has no personal control over the drugs they consume and potentially no autonomy by potential harm of one's compulsive actions.

supported by the evidence they provide and furthermore the implementational mechanisms they describe apply equally consistent with this revised claim, then there appears to be nothing empirically to determine which claim to assent to. Thus, we turn to normative implications to determine whether we assent to their claim (of compulsion) or my claim (of 'strongly urged'). Their conception of addiction as compulsion has a negative normative implication that effectively deprives autonomy from the addict by constricting the possibilities she can imagine for herself in the recovery process. By modifying their claims of compulsion into a claim of 'strongly urged' their theory can become suitable for a therapeutic context and moreover preserves the rest of their theory (i.e. the physical/implementational aspects of their theory).

Before I explain BR's theory, in Section 0.1 I define 'human kind' terms in a way that will be helpful to frame my discussion of the normative implications of BR's conception of addiction (Hacking, 1991).

1.1 Human Kind Terms

The purpose of a human kind term is to describe a type of person in the world we live in. The therapeutic sciences (such as psychiatry) work with human kind terms. There are normative implications involved in the description of human kind terms that are not as relevant with other kind terms. The point is that the descriptions of human kind terms are much more sensitive to the effects that they have on the world (i.e. on the phenomena they are describing), because the things they are describing (persons) actively change according to the descriptions assigned to them. Ian Hacking writes of human kind terms:

> [T]he histories of human kinds are quite different from those of natural kinds, for as some classifications are formed and moulded they loop back, interact with, and alter the individuals and the types of behaviour to which they apply. At a deeper level, they help determine the very space of possibility of action (Ian Hacking, 1990).

Consider *addict* as an exemplar of a human kind. How we describe the term 'addict' has an effect on actual addicts. Addicts are sensitive to the descriptions assigned to them. Human kind terms like addicts are quite different compared to, for instance, an electron. An electron does not care how we describe it, because an electron is simply not a thing that is capable of caring. Persons care – electrons basically just spin around nuclei.

If our description of the human kind term 'addict' does not respect the *space of possibility* within the imagination of the addict, then it is my view that we have effectively blocked the likelihood for recovery in some addicts. Imagining the possibility for change is critical for a healthy recovery from an addiction. The possibility for recovery is within the 'space of possibility' for an addict only if they are a person with an imagination that is unrestricted and unconstrained. If an addict cannot imagine herself as free from the addiction, then it will be very hard, if not impossible, for the addict to *become* free from the addiction.

Any unwarranted conception of an addict that constrains the possibility for recovery and describes the addict as powerless is not a good conception (so long as we value individual autonomy and personal freedom within one's own 'space of possibility'). By unwarranted here I mean that there are other conceptual claims and descriptions of addict that are equally warranted given the scientific/empirical evidence provided. So, without providing more empirical evidence, there is no way to disambiguate between which is the right claim and which is the wrong claim, unless the attempt to disambiguate incorporates normative implications (like I have suggested).

2.0 Berridge and Robinson: The Incentive Sensitization Theory - Addiction as Compulsion

The Berridge and Robinson (BR hereafter) model of addiction construes addiction as a compulsion. The process of getting addicted is a process of increased attribution of incentive salience for drug cues and drug-related stimuli to the point where the person perceives these

cues/stimuli in their environment they have an internal state that provides a very high level of motivation (or 'wanting'[29] in BR's terminology) to consume their drug of choice. If the incentive attribution is high enough, they compulsively pursue that consumption (i.e. regardless of whatever other internal self-management of behavior is being attempted, BR, 1993, 249).

BR state that:

> [D]rug addiction is defined as a pattern of 'compulsive drug-taking behavior'. Drug taking does not in itself constitute addictive behavior. Only when the repeated self-administration of drugs leads to a pattern of compulsive drug-seeking and drug-taking behavior, which occurs at the expense of most other activities, is a person said to be addicted (1993, 248).

The aim of BR's theory is to explain: (1) the biopsychological processes that underpin *drug craving* and (2) why are they so persistent, and (3) how, even in the face of intense disliking and dysphoria in the later stages of an addiction, how it is possible that the drugs are still craved.

The major conceptual shift that BR propose in their theory compared to other past models[30] is the separation between the mechanisms underpinning 'wanting' and the mechanisms underpinning 'liking,' i.e. the hedonic or euphoric aspect of consciousness (2000, 94).[31] The impetus behind this is clinical evidence that an addict need not *like* a drug in order to *want* the

---

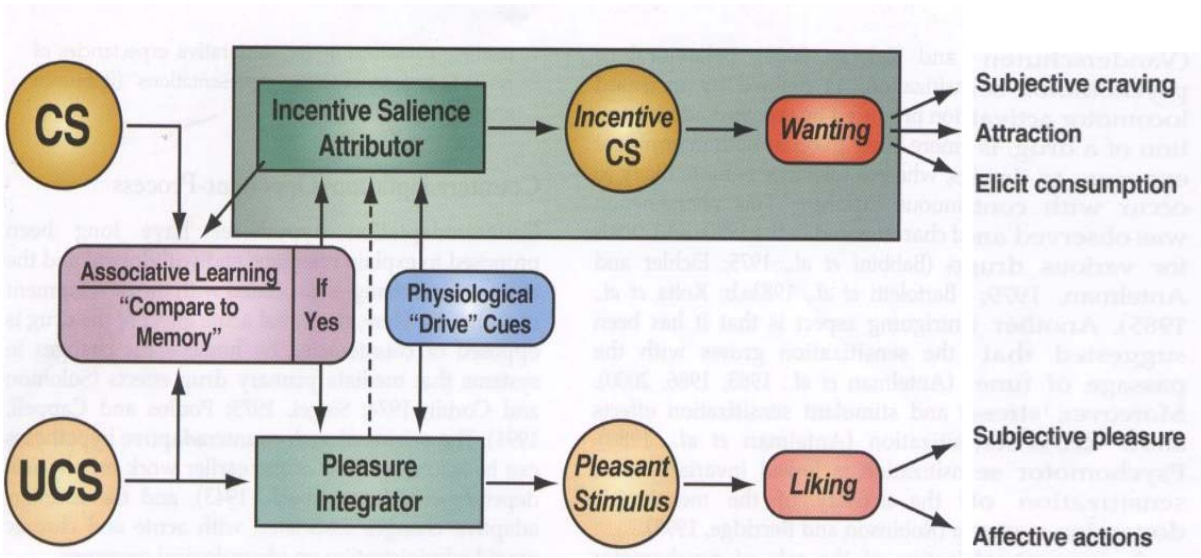
[29] Wanting according to BR is incentive attribution. The organism attributes incentive to something in the environment and as a result is urged toward focusing on that aspect (and depending on what it is either consumes or takes possession in some way). Since we are talking about drugs here: the 'wanting' activation results in drug-seeking and drug-consumption behavior.

[30] By reference to past models, I am talking mainly about positive and negative hedonic reinforcement models of addiction. On the positive reinforcement model, the drug user continues to use the drug because of the reward (drug euphoria) and eventually after enough use becomes accustomed to the drug and strongly craves it because she likes it so much (i.e. because the euphoria is so pleasurable). On the negative reinforcement model, the user may begin by using for the pleasurable effects, but once she is addicted, then she uses to keep away dysphoric withdrawal effects that are not pleasurable. The problem that BR found is that liking does not track using very well. In fact, in the later stages of addiction, the 'liking' aspects of drug consumption usually fall away and all that remains is the 'wanting' or desiring aspects (BR, 2000, 250-255).

[31] As you will see in Figure 2.1, the wanting mechanisms have both afferent and efferent connections to hedonic 'hot-spots' (or in other words the 'liking' mechanisms). It is understandable, therefore, that in normal persons activation of the wanting mechanisms is somewhat pleasurable. This is a correlated aspect to the wanting mechanism activation and not necessarily correlated. One of the groundbreaking points that Berridge and Robinson proved is that the liking mechanisms can be abused to such an extent (through drug use for instance) that they will not strongly activate when the wanting mechanisms activate. This phenomenon is seen clinically in long-term opiate users and their inability to feel hedonic reward for stimuli that would normally produce pleasure (even their regular heroin shots eventually cease to produce pleasure).

drug (in fact, it is common in the later stages of drug addiction that the drug elicits very low levels of euphoria upon consumption based on complex biochemical changes (downregulation of receptors and increased production of enzymes that function to excrete the drug from the brain)).[32]

Below is a schematic illustration of their model where they have illustrated the separation between 'wanting' and 'liking' (1993, 263): Fig. 2.1



Since the mechanisms that underpin the euphoric 'liking' effects of drug consumption are separate from the 'wanting' effects, it is possible that through downregulation (tolerance) and increased production of enzymes that are responsible for excreting the drug (also tolerance effects), the euphoric effects can all but subside (even shift into *dys*phoric effects), while the 'wanting' effects are maintained or even strengthened. This accounts for the situation an addict

[32] There is a possible worry here that one might entertain: are these neuropsychological theorists ignoring the distinction between first-order and second-order desires and might this distinction help clear up the difference between 'wanting' and 'liking'? A first order desire comes in the form of: Person A1 desires X1. A second-order desire comes in the form of: Person A1 desires to desire X1. I do not think the theorists are familiar with this philosophical distinction of first-order/second-order desire. However, I do not think this distinction will be useful here to explain the difference between 'wanting' and 'liking'. The activation of 'wanting' mechanisms more closely maps onto the concept of 'motivation' in the organism (i.e. provides an impetus to pursue either investigation of the object of incentive or consumption/possession). 'Liking' here refers to the release of euphoria/pleasure. The 'liking' structures are the structures (ventral tegmental, periaqueductal gray) in the human brain that are activated when intense pleasure is felt, for instance, during an orgasm (Bianchi-Demicheli, Ortigue, 2007).

might find herself in the later stages of her addiction where she no longer even likes the drug or gets much enjoyment out of it (because she has so massively downregulated the receptors and/or temporarily depleted the monoamines/peptides responsible for the euphoric effect).[33]

The 'wanting' mechanisms, according to BR, are realized in the mesotelencephalic dopamine systems (1993, 263). This portion of the brain is responsible for the attribution of incentive salience to the drug itself as well as related drug cues (imagine, for instance, the rush of desire a heroin addict feels when they open up a new bundle of heroin and begin loading the drug into their syringe).

Nobody is an addict the first time they try a drug. The conformational changes in the mesotelencephalic system take multiple exposures to the drug to begin to reach the level of increased incentive salience that would precipitate compulsive drug-seeking and drug-using behavior. The rate at which this happens depends on the individual's psychology, the drug used (different drugs will affect the dopamine system in the brain differently), the dosage used, the frequency of use, as well as environmental/social aspects the user finds herself in.

Over enough exposures, the mesotelencephalic 'wanting' dopamine system becomes sensitized to a range of cues involving the drug. When the user perceives one of these cues in their environment (or imagines the cue) the dopamine systems activate and an intense urge to consume the drug hits the user. Again, imagine the heroin addict and the state she is in when she is handed her new bundle of heroin (she desperately 'wants' to consume and will do whatever it takes to do so). According to BR's theory, the user is only an *addict* if it is not physically possible that she could resist by her own strength of will.

---

[33] There is an ambiguity in the use of our term liking. I am using BR's notion of 'liking' which has to do with hedonic reward/release of euphoria. There is a sense in which the heroin addict $likes_1$ their heroin despite not getting any euphoria from it anymore (i.e. they $like_1$ using the drug better than not). However, this is not the sense of 'liking' that BR intend. This $like_1$ has more to do with *preference* than hedonic reward.

In other words, according to BR's model of addiction, a person is either an addict or they are not.[34] If the incentive salience has reached a high enough degree that it provides such a strong motivation to elicit drug consumption behavior in a compulsive manner, then the person is an addict. Addicted drug use does rest on a continuum with casual/social use. The degree property of interest is the degree that the incentive salience structures are activated. They become stronger as use continues (all things being equal) and over time reach the threshold point where drug use becomes addiction (BR set this point at whatever point drug consumption compulsion occurs at). On one side we have social drug users that are experimenting with different drugs, some for their first times. As you follow the continuum, the drug use becomes more serious and perhaps the social aspect falls away (and we have isolated individuals using alone). As drug use becomes more serious and regular, the dopamine systems in the brains of the users are conforming to attribute increased salience to a select number of representations that cue drug consumption. Below is a diagram that illustrates the continuum:

Fig. 2.2

low incentive salience …increase in 'wanting'… high incentive salience

→

casual/social drug use addiction

At a certain point the increased salience reaches the level of making the individual 'want' the drug akin to how a 'normal' person would 'want' some food if they hadn't ate for 8 hours. The craving for the drug reaches such a high intensity of 'wanting' because the brain is firing off signals that the consumption of the drug is of supreme importance (high incentive). Finally, we reach a point on the continuum where the 'wanting' is so intense that the drug user cannot help but consume the drug once the dopamine system has sensitized itself to the related drug cues. No

[34] I will remain agnostic here on semantic worries of vagueness (i.e. whether there is an area of indeterminacy – a 'penumbra' between addict/not-addict.

matter what strength of will or volition the person has against using, she will use (she is helpless against the intense 'wanting' state inside of her). Without external management (for instance, from friends, family, or professional therapists) the person will continue to use the drug. She lacks internal self-management for her use. She is a compulsive drug user and has reached the point of the continuum that qualifies her as an 'addict' according to BR's model of addiction.[35] This conceptual point (the necessity of compulsion for addiction) will be the focus of the next section.

The mesotelencephalic dopamine system is quite flexible (plastic) to change, but once it adapted and has started to attribute increase incentive salience to a certain number of representations of the environment (i.e. drug cues), its ability to revert back to its previous state is greatly diminished. This explains why it is the case that an addict who is years away from their abuse can still quite easily relapse back into prior use patterns. The incentive salience is still extremely high for certain cues that would lead to drug consumption behavior, despite not being activated for years. It is not surprising that drugs promote change in the mesotelencephalic dopamine system, since they can be extremely powerful pharmacologically (compared to 'normal' experiential states of consciousness). This system is quite plastic in the range of cues to increase incentive to, but once it has changed, then 'wiping clean' is usually not an easy option. In common parlance, 'once an addict, always an addict.'[36]

---

[35] There is a problem here in BR's conception of addiction that I do not have space for, but will briefly cite here. The problem is a difference in *extension* of the term addict than would be appropriate in a therapeutic context. If only compulsive drug users are addicts, then that means a person falling on the continuum *just* before compulsivity would not be considered an addict. Yet, this person still strikes me as falling into the extension of addict insofar as they would display classic problems of addiction and faces the same internal battles of craving and weakness of will. Ought we to accept BR's term that departs so drastically from the 'common' extension of addict knowing that in doing so our acceptance might have a reciprocal effect on our common-use term addict? I do not have space here to pursue this, but I will end by saying that I think this indicates BR has cut the extension of addict too tight (they have identified the extension of addict with the extension of compulsive user, which I think may rightly be seen as a non-coextensive subset of addict).

[36] According to BR's theory, we might say: 'once an addict, an addict until sensitization in mesotelencephalic dopamine systems is desensitized.' Well, I admit, it doesn't have *quite* the same ring to it.

2.1 Incentive Salience and Therapy

What are the therapeutic approaches that are suggested by BR's theory? Of a growing popular method of drug addiction treatment they write:

> A recent trend in the psychotherapeutic treatment of addiction is based on the recognition that drug-conditioned stimuli are very potent in eliciting craving and precipitating relapse. There have been attempts, therefore, to 'extinguish' conditioned responses to such stimuli. Indeed, the repeated presentation of drug-related stimuli, in a laboratory setting, results in a progressive decline in drug craving elicited by drug-related stimuli. It is interesting, however, that some of the autonomic responses to such stimuli are more resistant to extinction than the subjective effects, and non-specific changes in mood state (especially anger) can rapidly reinstate conditioned stimulus-induced drug craving (1993, 271).

This is the obvious extension of their theory into drug-therapy: if drug addiction is caused by increased attribution of salience that happens within the mesotelencephalic dopamine systems, then by desensitizing this system to the relevant drug-related cues it is (theoretically) possible to decrease the intensity/salience of the 'wanting' in the drug addict.

Pharmacologically speaking, they write:

> The Incentive-Sensitization Theory predicts neither of these approaches will be very successful in eliminating addictive behavior, because neither target the fundamental neuroadaptations underlying sensitization (1993, 271).

3.0 Is BR's Theory Appropriate in a Therapeutic Context?

1) Are the normative implications of the description of an addict as 'a person that is *compelled* to use drugs' (i.e. the harm caused by the 'looping effect' of the description on the addict) strong enough to justify revision of the description?
2) And if so, can we revise the commitment to compulsion while still preserving the rest of the theoretical structure?

I will explain and answer these two questions in turn. First, I will unpack the relevant normative

implications of a commitment to no responsibility as a result of compulsion. Remember Hacking's looping effect: whenever you make a description about a human kind (i.e. whenever you describe a person), you have an effect on the person you are describing because that person is capable of coming into contact with your description and changing their behavior as a result of the description (in common parlance, 'if you tell somebody they are something enough times, eventually they will become it').

I contend here that BR are on dangerous ground with their commitment to a view that describes humans as being powerless against the powerful drug urges they have. In their commitment to a description of the addict as having no responsibility[37,38], BR have effectively constrained the 'space of possibility' in the imagination of the addict. If you tell a person they are some type of thing long enough, they eventually will begin to conform their behavior based on the description you have assigned to them (Hacking's 'looping effect'). We should be sensitive to this (potential) effect in addict's because it is imagination that is a critical faculty in the process of recovery (if a patient cannot imagine a better life and see themselves free of the 'need' to use the drug, then it will be quite difficult to help them get to that point).

It is not enough, of course, to simply note that there are negative normative implications to their description of addict and therefore the theory should be revised. It may very well be the case that a description of a term carries negative normative implications. However, in order for this description to be warranted there must be no other description available to the theorist that is

---

[37] Remember, the set of people who are compelled to use drugs just *is* the set of addicts according to BR. There are no addicts that fall outside of the continuum of compelled users (refer to Fig. 2.2)

[38] Berridge and Robinson do not explicitly discuss personal responsibility. The concept of personal responsibility is a philosophical concept and, as such, does not get explicitly covered by these psychology theorists. However, they do have *implicit* commitments to some degree of personal responsibility based on their other explicit conceptual commitments. If they hold that addiction is drug compulsion, then they have effectively committed themselves (implicitly) to a no responsibility view of addiction. I remain agnostic here still whether this commits them to a view of no *moral* responsibility (I do not believe there is a fact of the matter whether moral responsibility closely tracks personal responsibility. This will be left up to the details of one's theory of moral responsibility.).

equally well supported by the available empirical evidence. If there is another description that is equally well supported by the empirical evidence and does not bear the negative normative implications of the former description, then it ought to be endorsed if the description is to be utilized in a therapeutic context.

If a description is capable of doing this kind of harm in a therapeutic setting, then I contend that it should not be utilized (so long as there is no overwhelming scientific warrant for its usage, i.e. it is the only available description with not competing equally well supported descriptions). We should be careful here. I am not saying that we should ignore the truth of our empirical findings if they have negative normative implications. This would be a mistake and would only hinder our understanding of the phenomenon (and even if we did this type of move with altruistic normative implications in mind, we would most likely end up confusing and hindering therapy despite our best intentions). I am saying here that if there is a conceptual claim that is equally well supported compared to a similar conceptual claim (but the latter claim lacks the negative normative implications of the former), then all else being equal, we should favor the latter claim if our theory is to be utilized in a therapeutic context.[39]

It is my view that we should not assent to the claim that addicts are compulsive drug users (i.e. powerless) if there is an alternate claim that is equally well supported by the empirical evidence used to inductively support the former claim that preserves a small degree of personal responsibility. There is such an alternate claim (my claim that addicts are persons that are 'strongly urged,' yet partially in control, i.e. diminished responsibility) and since it does not suffer the normative implications of the compulsion commitment, we should favor my revision over their original formulation.

An analogy here is appropriate. Consider a patient that is being informed he is in stage

[39] Berridge and Robinson have made clear (1993, 2000) that their theory is intended to benefit therapy.

four prostate cancer and has very little time to live. This news definitely constrains the space of possibility in the patient's mind. The news, however, despite constraining the possibilities the patient can imagine for himself, is inductively warranted given the tests and analyses at the hospital. There is no equally supported claim given the evidence that the hospital tests provide. Therefore, there is no reason to turn to normative implications to disambiguate between claims of cancer or something else (because there is no 'something else' up for grabs here). In this case, it is just a fact that the person has prostate cancer and has very little time to live.[40]

There is a relevant dissimilarity, however, to the case of the addict. There are two equally well supported claims 'up for grab' here: addiction as compulsion or addiction as being 'strongly urged.' The empirical evidence does not disambiguate between the claims, therefore, normative implications must be considered (since the theory of addiction is concerned with therapy). There is an alternative option here: make no conceptual claims about the empirical evidence and wait until further evidence is created. I do not think, however, that Berridge and Robinson want to do this (in fact, their going ahead and asserting the claims about compulsion make it clear that they do not want to wait on an issue of indeterminacy).

Conceptually, the two properties of a drug user (property of being a compulsive drug user and the property of being a strongly urged drug user) are quite distinct. In the case of compulsive use, the drug user has *no option for internal self-management*. The drug user will use the drugs (barring no insurmountable external force acting on their desire to use). In the case of being strongly urged to use the drug, it is possible that the drug user can resist using based on internal processes (strength of will).

If its true that BR's commitment to compulsion as a necessary part of their theory of

[40] Of course, this does not mean that people involved with informing the patient should not be sensitive to the patient's well-being. Quite the opposite, it is critical that the doctor informs the patient about his condition with care and respect.

addiction is actually *not* necessary, then is it possible to suggest a revision (conceptually) to their theory while still preserving the main tenants of their theory? I suggest that Berridge and Robinson define the extension of 'addict' as the set of drug users that are 'strongly urged' to use drugs based on the increased attribution of incentive salience to the drug and drug cues. Backing their claim down from compulsiveness to 'strongly urged' still preserves the main tenants of their theory (incentive salience underpinning the 'wanting' mechanisms), yet allows for a more tempered conceptual claim that circumvents the normative problems of labeling a person as a thing that has no control over itself.

This revision effectively converts their theory from a no responsibility account of addiction into a diminished responsibility account of addiction. The revision also widens the extension of the term (according to Fig 2.2 the spectrum of addict would occur further toward the middle than before). I take this to be the upshot of my revision to their theory: there is no reason to commit themselves to a no responsibility account of addiction when a diminished responsibility account of addiction is equally (or more) amenable to the data the have presented in their 2000 article. This revision also suggests the possibility of joining together the therapeutic practices they suggest (that I reviewed in section 1.1) with a more traditional 'talking' therapy that would be useful to strengthen the addict's coping strategies (if they are likely to fall in the set of addicts that are 'strongly urged').

4.0 Conclusion

In this chapter I have outlined and explained the Berridge and Robinson view of addiction (the 'incentive salience' theory of addiction as compulsion). I have chosen this theory for more than just its usefulness and what strikes me as its correctness given the evidence that BR present in their 2000 (and to some extent their 1993). I chose this theory because it is one of

the only operational and *prima facie* consistent/useful theories of addiction that construe the addict as having no responsibility for his/her actions. Compulsion, in the sense that BR use the term, quite clearly means a lack of control over one's actions in regards to drug consumption and incentive salience attribution.

In the next chapter, I intend to review various philosophical accounts of partially diminished responsibility in addiction. I then will show how these philosophical accounts of addiction are consistent with the neuroadaptationist model provided above by Berridge and Robinson, bearing my conceptual revisions in mind.

CHAPTER 3

DIMINISHED RESPONSIBILITY THEORIES

1.0 Introduction

The last two chapters have been an effort to tease out problems within the full responsibility theory (RTA) and the no responsibility theory (BR's incentive salience theory). I have identified explanatory problems with RTA (the framework is insufficient for explaining aspects of addiction, particularly the phenomenon of craving and of relapse). I cited normative problems with BR's incentive salience theory, particularly their commitment to construing addiction as compulsion. I then suggested revision to their theory that changed their conceptual claims from compulsion to 'strongly urged.'

If the viable full responsibility theory and the viable no responsibility theory both suffered problems, then where is there to turn to for the correct view of addiction and responsibility? Diminished responsibility theories are the last option and they are, in my view, not committing themselves to any explanatory or normative problems like the above two theories.

In section 1.0, I examine various different views (all by philosophers) that share some overlap (and some important incongruence). I start with Harry Frankfurt (1971) and then examine R. Jay Wallace's claim about addictive desires (2006) and George Graham's view of addiction as impairment in self-responsibility (forthcoming). In section 2.0 I discuss the possibility for combining the views of some of the above diminished responsibility theorists with BR's incentive salience theory.

2.0 Diminished Responsibility Theories

Frankfurt is the historical (and conceptual) starting point of the diminished responsibility theorists. In 1971 he formulated a distinction between a "willing addict" and an "unwilling addict" (8-14). This distinction is a conceptual implication of his ontology of 'wanting and desiring.' His claim (a conceptual point) is that a person can desire to do something, for instance consume a drug, while simultaneously having a volition (which is a higher-order conceptual state on his view) to *not* want or do that activity. If the desire is satisfied (i.e. the drug is consumed), despite the volition to not consume, then we have an example of his unwilling addict. The unwilling addict is, as he puts it, "helplessly violated by his own desires" (1971, 12). This provides an obvious tension in the decision-making process of the individual.

Frankfurt also posits the possibility for a 'willing addict' – a person that is addicted to a drug or activity and also has higher-order volitions that line up with the use of the drug or activity (i.e. they agree on a volitional level to continue using). His primary point in his 1971 is that the conditions for satisfaction of a person's volitions do not *always* obtain because of the causal force of other lower-order states (he calls them desires, but recent philosophers have called them impulses or cravings, Graham, forthcoming).

Wallace follows Frankfurt, conceptually and historically. He is also a diminished responsibility theorist. He believes that addicts are subject to desires that are particularly unresponsive to deliberation (2006, 165). The particular desires that assail an addict form a subset of desires: he calls them a-desires (167). An a-desire provides a particularly strong phenomenological force in the agent's consciousness that focuses attention and cognitive resources on a narrower range of thoughts and beliefs than would typically be available to the

agent for reasoning and planning about how they will act next.[41] The result is a diminished capacity for exerting volitional control over one's own behavior (because of the strength of the a-desires).

Imagine a physically dependent injecting heroin addict that has not had a shot in 8 hours and is going into the initial stages of withdrawal. A dysphoric feeling manifests within her consciousness, accompanied with trembling, limb pain, and emotional dissonance. She begins to do everything she can to acquire a dose of heroin and *any* syringe to inject with, going so far as to ignore whether or not the syringe she is using has been used by her or another one of her friends. All she can think about is getting some heroin into her bloodstream as fast as possible. She finally acquires the heroin from her dealer as her withdrawal symptoms become more intense and immediately shoots up, feeling a profound sense of acquiescence into the heroin high and a dissipation of all the frenetic energy that was building as a result of the impulsive desire to inject heroin. She has placated the 'a-desire' and her dysphoria from the need for the heroin has been replaced with a deep and (falsely) abiding sense of euphoria from the heroin. The a-desire, however, will return in 7-8 hours with a reinforced intensity. This is the nature of addiction, and thus, the downward spiral.

The locus of philosophical interest of a-desires, for Wallace, is their *resilience* (2006, 168). A-desires are resilient in the face of rational deliberation, *persisting* in the conscious experience of the agent, exerting an abiding effect on the rationale of the agent despite even the agent's wishes that they would go away. This resilience contributes to their efficacy. Imagine a therapy that orders the addict, to simply tell oneself not to use whenever the craving strikes. In

[41] Wallace appears to hold the phenomenology of A-desires as central to their characterization, but it does not appear as if one must necessarily hold his view to buy into the larger picture (that something like A-desires or addictive impulses have an effect on the agent's reasoning capabilities, temporarily restricting certain 'epistemic routes' the agent can take in the reasoning process. The argument supporting abandoning the phenomenological claim of Wallace's would be an empirical one, citing evidence that not all addicts suffer from these terribly powerful cravings (A-desires using his terminology). Sometimes (the argument would go) an addict does not experience the distinct phenomenological cravings that Wallace seems to think constitutes addiction. I think for our purposes here it does not matter about deciding whether or not phenomenology must go.

Wallace's view, and that of other diminished responsibility theorists, this is a doomed therapeutic strategy. It is precisely this point of commanding oneself against these desires where the deficit of addiction rests.

In virtue of the causal connections a-desires have to conceptual reasoning, they influence and restricts the 'scope' of the agent's reasoning processes, focusing the agent on solving the 'problem' the a-desire presents (how to placate the a-desire, in our example the 'solution' or satisfaction conditions of the desire is to shoot up some heroin).

In prototypical, non-pathological cases of desire the scope of the agent's reasoning is narrowed, in virtue of having the desire, usually to her gain (the reasoning process is streamlined and choice is made more efficiently; e.g. she desires the apple and in virtue of having this desire it makes her search for and choice for the apple more easily made). Under Wallace's analysis of addiction, however, the pathology obtains from the intensity and resilience of the a-desire and the resultant impairment this has on the agent's reasoning process (*ex hypothesi*, to her deficit, since we are dealing with activities like injecting heroin, which are activities that most reasonable persons can agree do not support a good life).

George Graham also discusses the idea of an addict with an impaired faculty for self-control in his forthcoming book, The Disordered Mind. For Graham, the addict suffers an impairment in her ability to exert self-control over her own impulses (very close to Wallace's a-desires, albeit (purposely) missing the phenomenological component of Wallace's view[42]). The impairment in self-control is not the full extent of addiction for Graham. He also thinks addicts

[42] Graham does not think that the phenomenological component is a necessary component in a theory of addiction. In the beginning of breaking an addiction, the impulses and cravings are very strong (along the lines of what Wallace is describing). After enough time, however, the impulses weaken until eventually they may not even be felt anymore. Just because the impulses are not being consciously felt, does not mean that they are not having an effect on the reasoning process of the individual. It is possible that the impulses have such an effect so as to override a volition to quit (such as in the case of relapse in an 'unwilling addict'). The reasoning behind Graham's objection to Wallace's claim of the necessity of a phenomenological aspect is to support a theory that describes and helps prepare for the possibility of relapse (personal communication). Robert West discusses this in his 2002 book. He mentions that there is enough clinical evidence that addicts may relapse independent of withdrawal discomfort or of the anticipation of pleasure. The felt component does not seem necessary to an explanation of relapse.

suffer problems in viewing themselves as persons with historically bounded narratives and self-identities that extend past the immediate moment.[43] He has said, "The failure of addiction is a failure in treating oneself as an extended agent – extended in time" (personal communication). Under the traumatic push of an addiction a person, in my experience, tends to collapse their timelines into the immediate moment, forgetting past achievements and future prospects. So, a complete philosophical view of addiction will have something important to say about self-worth and self-identity.

The basic idea that diminished responsibility theorists agree upon is that an addict is subject to forces that are at times out of their control. These forces can provide a particularly strong push toward deciding, choosing, or acting a certain way. It is important to note that none of the diminished responsibility theorists think that addicts work on a brute mechanistic level (i.e. like an automaton). Addicts are still rational (and partially reasonable) agents. They still have the cognitive faculties that a 'normal' person has; however, these faculties are impaired/impinged upon. The disorder results in a diminishment in responsibility (but not the lack of *any* responsibility). The diminishment in the faculty of reasoning is the justification for the diminishment in responsibility.

In the next section, I turn back to the incentive salience theory and show how that view is consistent with these previous philosophical remarks.

3.0 A Synthesis of Diminished Responsibility Theories and the Incentive Salience Theory

The evidence that Berridge and Robinson use to support their conceptual claims about addiction are consistent with the claims of the previously cited diminished responsibility theorists above. As a reminder of BR's theory of addiction: they believe that the plasticity of the

[43] In this way some of Graham's work on addiction relates to Valerie Hardcastle's work on 'narratives' on self-identity.

mesotelencephalic dopamine system (that underpins the ‘wanting’ mechanism in the user’s mind) is the aspect that is sensitized over time to repeated exposure to the drug and related drug cues. After enough exposures the mesotelencephalic system begins to increase attribution of incentive salience to the drug and related drug cues (by strengthening the dopamine pathways that activate with the perception of the drug and related drug cues).

How is this view consistent with the diminished responsibility views cited above? According to Wallace’s view, addiction is the inability of the drug user to adequately control powerful a-desires or urges to use the drug. Note, this does not construe the addict as completely helpless against the a-desire (the urge or ‘wanting’ of the drug), but rather construes the addict as being strongly disposed toward drug consumption (when the a-desire hits phenomenologically).

The a-desire Wallace is discussing can map on very well, physically speaking, to activation of the mesotelencephalic dopamine system. If we are looking for a ‘physical realizer’ of this a-desire, then I think we need not look further beyond this dopamine system. It is also consistent with Wallace’s view that the system can be strengthened as the addiction grows stronger from repeated administration (remember the reinforcement effects discussed above in the heroin addict). Also consistent with Wallace’s framework is Berridge and Robinson’s commitment to the possibility that the pleasure an addict receives from the drug use need not be in direct correlation to the strength of the ‘wanting’ activation (or in Wallace’s words the pleasure from the drug need not correlate with the intensity of the a-desire). It is quite common actually that an addict receives very little pleasure in the darkest moments of their addiction when the urges and impulses to use are at their strongest (heroin addicts speak, clinically, of their dose as making them ‘normal’ but not happy or euphoric like it used to).

Likewise, BR’s neuropsychological view of addiction is consistent with the claims

George Graham makes about addiction. Graham discusses addiction as impairment to the faculty of self-control. This is precisely the aspect that diminishes as the mesotelencephalic dopamine system is strengthened and the 'wanting' mechanisms exert a stronger effect on the addicts decision-making process.

We can understand the claims that BR make about addiction as providing a view of the 'implementational' level of addiction (the physical level). If we bear in mind the suggested revision I propose in chapter 3, then we can convert BR's theory of addiction from a no responsibility theory of addiction into a diminished responsibility theory of addiction that focuses on the physical/implementational level in the brain.

4.0 Conclusion

In this thesis, I have investigated the different levels of personal responsibility that an addiction theorist can commit him or herself to. I have cited popular full responsibility and no responsibility theories of addiction and found problems with both types. After a suggested revision I show how it is possible to convert Berridge and Robinson's view into a diminished responsibility view of addiction and how that can be consistent with other insightful remarks made by philosophical addiction theorists (Graham and Wallace).

From here, I would like to see strengthened the implementational level that Berridge and Robinson have begun to define for addiction. I would like to see research done on the peptidergic system that underpins 'liking' (the euphoric effects of drug consumption) and their interactional effects on the 'wanting' system. I would also like to see higher-order (so called 'top-down' modulation) to the 'wanting' system. I suspect these higher-order processes may be subject to their own non-rational influences much like the 'wanting' system is subject to non-

rational influences by sensitization.[44]

In this thesis I have tried to show how diminished responsibility theories of addiction are the most appropriate theories for a therapeutic context. The full responsibility and no responsibility theories I analyzed both showed problems when applied to a therapeutic context. However, by amending conceptual commitments in Berridge and Robinson's theory I have effectively converted their claims into a diminished responsibility theory, which in the previous section I have attempted to show how it is consistent with philosophical diminished responsibility views of addiction. I have argued that my conceptualization of their theory is equally supported by the evidence they give and because it does not suffer normative problems that their conceptualization suffers *and* is consistent with other useful philosophical accounts of addiction, is therefore the better theory and ought to be revised according to my recommendations.

[44] These higher-order modulations are probably associated with trace amine receptors (that are not fully understood or fully discovered yet) and release of minute quantities of tryptamine serotonergic compounds such as 5-methoxy-dimethyltryptamine and n,n-dimethyltryptamine (Borowsky, et al., 2001; Wallach, 2008). The work in this area, however, largely remains to be done. We will just have to wait to see how trace amine receptor agonists such as DMT will be incorporated into a larger hybrid therapeutic approach to treating addiction.

WORKS CITED

Ainslie, G. (1992). Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person. Cambridge. Cambridge UP.

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. American Psychiatric Press, Washington, DC.

Becker, Gary and Murphy, Kevin (1988). "A Theory of Rational Addiction." *J Political Economy*, 96 (4): 675-700.

Becker, Gary, and Grossman, Michael, and Murphy, Kevin (1990). "An Empirical Analysis of Cigarette Addiction." *NBER Working Paper #3322*.

Berridge, Kent and Robinson (1993). "The neural basis of drug craving: an incentive-sensitization theory of addiction." 247-291.

Berridge, Kent and Robinson (2000). The psychology and neurobiology of addiction: an incentive–sensitization view." *Addiction* (2000) 95 (Supplement 2), S91–S117.

Berridge, Kent and Robinson (2000). "Incentive-sensitization and drug wanting" *Psychopharmacology* (2004) 171:352–353.

Bianchi-Demicheli, Francesco; Ortigue, Stephanie (2007). "Toward an understanding of the cerebral substrates of woman's orgasm." *j neuropsychologia*.

Borowsky B, Adham N, Jones KA, Raddatz R, Artymyshyn R, Ogozalek KL, Durkin MM, Lakhlani PP, Bonini JA, Pathirana S, Boyle N, Pu X, Kouranova E, Lichtblau H, Ochoa FY, Branchek TA, Gerald C (2001). "Trace amines: identification of a family of mammalian G protein-coupled receptors". *Proc. Natl. Acad. Sci.* **98** (16): 8966–71.

Cook, Phillip and Tauchen, George (1982). "The Effect of Liquor Taxes on Heavy Drinking. *Bell Journal of Economics*. 13: 379-390.

Fehr, Ernst and Zych, Peter (1998). “Do Addicts Behave Rationally?” *Scandinavian J Econ*, 100 (3): 643-662.

Ferguson, Brian (2000). “Interpreting the Rational Addiction Model.” *Health Econ*, 9: 587-598.

Frankfurt, Harry (1971). “Freedom of the Will and the Concept of a Person.” *J of Phil*, 68 (1): 5-20.

Graham, George (2008). The Disordered Mind. MS.

Gruber, Jonathon and Köszegi, Botond (2001). “Is Addiction “Rational”? Theory and Evidence.” *Quarterly J Econ*, 116 (4): 1261-1303.

Hacking, Ian (1991). “On Boyd.” *Philosophical Studies* 61: 149-154.

Kauer, Julie; and Malenka, Robert (2007). “Synaptic Plasticity and Addiction.” *Nature Reviews Neurosci*., in press.

K.R. Alper, H.S. Lotsof, C.D. Kaplan (2008). "The Ibogaine Medical Subculture". *J Ethnopharmacology*. **115**: 9–24.

Peciña, Susana and Berridge, Kent (2000). “Opioid site in nucleus accumbens shell mediates eating and hedonic 'liking' for food: map based on microinjection Fos plumes.” *Brain Res*. 863(1-2):71-86.

Peciña, Susana and Berridge, Kent (2005). “Hedonic Hot Spot in Nucleus Accumbens Shell: Where Do μ-Opioids Cause Increased Hedonic Impact of Sweetness?” *J Neuroscience*, 25(50): 11777-11786.

Wallace, R. Jay (2006). Normativity and the Will. Clarendon P: Oxford.

Wallach J (2008). "Endogenous hallucinogens as ligands of the trace amine receptors: A possible role in sensory perception". *Med Hypotheses*. in print.

West, Robert (2002). Theory of Addiction. Blackwell P.