CRÍTICA. Revista Hispanoamericana de Filosofía. Vol. 56, No. 167 (agosto 2024): 127–155

DOI: https://doi.org/10.22201/iifs.18704905e.2024.1530

THE SCIENCE AND MORAL PSYCHOLOGY OF ADDICTION: A CASE STUDY IN INTEGRATIVE PHILOSOPHY OF PSYCHIATRY

Quinn Hiroshi Gibson
Clemson University
Department of Philosophy and Religion
USA
qhgibson@berkeley.edu
https://orcid.org/0000-0002-9584-7049

SUMMARY: Though addiction is a complex empirical phenomenon, some of the most pressing questions about it concern how we should evaluate agents who are living with it. To that end, a fruitful methodology is to tease out from our best sciences consequences at the level of moral psychology. Taking account of epidemiology, behavioral science, animal studies and, chiefly, neuroscience, I argue for a view according to which addiction involves dysfunctional motivational states (which I call "hybrid intentions") as well as cognitive distortions. This argument can be made without needing to settle the traditional debate about whether addiction is a disease.

KEY WORDS: neuroscience, psychiatric disorder, responsibility, agency, disease

RESUMEN: Algunas de las preguntas más apremiantes acerca de la adicción tienen que ver con cómo debemos evaluar a los agentes que viven con ella. Para ello, una metodología fructífera implica extraer de nuestras mejores ciencias consecuencias a nivel de la psicología moral. Teniendo en cuenta la epidemiología, las ciencias del comportamiento, los estudios con animales y, principalmente, la neurociencia, defiendo una visión según la cual la adicción implica estados motivacionales disfuncionales (que yo llamo "hybrid intentions"), así como distorsiones cognitivas. Este argumento puede formularse sin necesidad de resolver el debate tradicional sobre si la adicción es una enfermedad.

PALABRAS CLAVE: neurociencia, desorden psiquiátrico, responsabilidad, agencia, enfermedad

1. *Addiction, Moral Psychology, and Philosophy of Psychiatry*

I intend this paper to be an exercise in what one might call integrative philosophy of psychiatry: philosophy of psychiatry which proceeds by trying to secure the best available empirical descriptions of disordered mental phenomena and integrating them into the ordinary personal and interpersonal practices of interpretation and evaluation which are central to the social world. This approach takes for granted that it is valuable to illuminate the ways in which what we consider to be disordered is and is not continuous with, and intelligible in light of, apparently "non-disordered" psychological and behavioral forms.[1] I hope this essay is witness to the virtues of this approach.[2]

This methodology is appropriate for the study of addiction because part of what makes addiction philosophically interesting is that our understanding of it straddles the scientific and manifest images of the world, i.e., the image of the world as presented to us by our best sciences, on the one hand, and the rich, normatively laden framework in terms of which we ordinarily understand one another and in which persons and their attitudes and reasons are central, on the other. (Sellars 1963) Science provides a detailed account of many aspects of drug addiction —the focus of this paper— but what matters to us as agents who inhabit a social world of which addiction is a part is how we should think, feel, and act towards those who suffer from it. It isn't *merely* that what interests us about addiction are its effects at the level of persons —e.g., on attitudes and behavior— but that addiction seems to be relevant for our *assessment* of persons, though it is unclear how. Addiction seems to lead people to act poorly —but how does it do that? It can also seem to involve an undermining, co-opting, or bypassing of the capacities relevant for moral responsibility, such as choice, judgment, and control. But each of these is controversial. *Which* of these does it do? *How* and *to what extent* is that relevant for assessing agents' behavior or their characters? There are lots of questions one can ask about addiction, but these are among the most pressing. It isn't so much that a purely scientific or aggressively reductionist picture of addiction would *leave something out* —though it might; even if we had such a picture, it would still be a *distinct question* how we should take it to bear on our attitudes and practices. Unless we are willing to give up such attitudes and practices altogether, I see no way around having to face this issue in some form.

It is common to simply not know how one should feel towards or what one should believe about someone living with addiction. Perhaps we are torn between pleas for excuse and hard love. The

[1] I have argued for this claim in detail elsewhere. See Gibson (2024).

[2] Despite the name, what I am calling "integrative" philosophy of psychiatry is only indirectly connected to work on what is sometimes called "the integration problem" in philosophy of psychiatry. (See, e.g., Gallagher (2022).) That problem concerns how variables at multiple levels hang together in the explanation of psychiatric disorders. Although I am here concerned in a certain sense with the relation between different levels, I am not primarily concerned with sketching a complete causal model.

DOI:https://doi.org/10.22201/iifs.18704905e.2024.1530

leading neuroscientific theories of addiction are compelling in part because, I shall claim, of how they can help us understand addiction at the level of moral psychology, which in turn can help us sort through the complex attitudes that we hold or might be drawn to. In this essay I will attempt to cash out these inter-level connections. I will argue that a close look at the sciences of addiction —especially the neuroscience of addiction— supports a picture on which both motivation and cognition are impaired. In particular, I will argue that addiction involves, in addition to cognitive distortions, *sui generis* motivational states, which I will call "hybrid intentions". Hybrid intentions are like typical intentions in that they are very closely connected to action, but they are unlike typical intentions by not being directly subject to volition. Hybrid intentions are thus motivational liabilities. They motivate action and seem to represent a practical judgment, but are not responsive to the agent's control in the same way that ordinary intentions are.

Section 2.1 is a broad-brush attempt to put some general constraints on theorizing about addiction. Here, the relevant constraining empirical findings are from epidemiology, behavioral science, and animal studies. Section 2.2 is a detailed engagement with the neuroscience of addiction in which I give an overview of two leading theories. I argue that both theories provide resources that we should help ourselves to and that, contrary to how they are typically understood, the two theories are not in explanatory competition and should instead be seen as complementary.

I then show how we can understand the moral psychological significance of the science by using it to refine theories of addiction originally proposed at that level. I begin with the idea that addiction is an impairment because of the distinctive power of *desires* in addiction. I argue that such a view needs to be modified carefully in light of neuroscience. These revisions point to a sketch which can begin to adequately capture the normative and valuational significance of addiction.

I close with a brief reflection on how the present investigation relates to the traditional question of whether addiction is a disease. I conclude that because my arguments do not depend on settling that question —especially as it is traditionally construed, as opposing disease to moral failing or choice— asking it is the wrong starting point if we want to cash out the significance of addiction in terms that matter to us.

2. *The Science(s) of Addiction*

2.1. Narrowing the Field

Debate about addiction is often organized around the following question: Is drug use in addiction a choice or is it compulsive? We can thus refer to proponents of the *compulsion* and *choice* (or *moral*) models of addiction. As we will see, it is not difficult to reject an uncompromising version of a compulsion model. Relatively uncontroversial and established science, which I review in this section, suffices. It is also important to reject what we might call an extreme pharmacological view according to which exposure, perhaps even brief exposure, to addictive drugs alone is sufficient to cause addiction. This view will also be rejected presently. At the same time, the assumption that neuroscience I review in the following section is relevant for understanding addiction is inconsistent with the opposite of an extreme pharmacological view. We need a view which acknowledges that there is something about the pharmacology and neurobiology of drug use that is distinctive and powerful, but which does not reduce to a cartoonish view of addiction as literal compulsion. The role of choice, which is native to the domain of moral psychology, I will have much to say about in section 3.4 once the other aspects of the picture have been outlined.

Consider the extreme pharmacological view. The influential pharmacologist Avram Goldstein seems to endorse such a view in what follows:

> If we arrange matters so that when an animal presses a lever, it gets a shot of heroin into a vein, that animal will press the lever repeatedly, to the exclusion of other activities (food, sex, etc.); it will become a heroin addict. A rat addicted to heroin is not rebelling against society, is not a victim of socioeconomic circumstances, is not a product of a dysfunctional family, and is not a criminal.[3]

Though this is a particularly strident expression of such a view, what Goldstein is saying is familiar to many who were exposed to public-facing anti-drug use policy statements from the ‘80s and ‘90s containing phrases such “[this drug is] so addictive once is enough”, and so on.

[3] The quotation is from “Neurobiology of Heroin Addiction and of Methadone Treatment”, available at http://www.aatod.org/media/archived-aatod-news/neurobiology-of-heroin-addiction-and-of-methadone-treatment/ (retrieved November 15, 2023).

We can safely reject this view. Most people who use addictive drugs do not become addicted. Approximately 40% of Americans admit to having used an *illegal* drug in their lifetimes, but even the highest estimates put the addiction rate —understood to include addiction to prescription pain medications and alcohol— somewhere between 8% and 10%. Further, most people who do abuse drugs at some point in their lives manage to stop. The U.S. National Institute of Mental Health (NIMH) conducted a landmark study between 1980 and 1985 to measure the prevalence of psychiatric disorders amongst the U.S. population according to the then-current DSM III criteria, *the Epidemiologic Catchment Area Study* (US DHS 1994). One of the study's most striking findings is that more than half of those who previously met the criteria for drug abuse or dependence reported no symptoms at all by age 24; by age 37, almost 75% are symptom-free.

Another consideration which appears to be inconsistent with the extreme pharmacological view comes from questioning the experimental paradigm used in earlier animal studies on addiction. Most notable among researchers pursuing this line is Bruce Alexander:

> We compared the drug intake of rats housed in a reasonably normal environment 24 hours a day with rats kept in isolation in the solitary confinement cages [...]. This required building a great big plywood box on the floor of our laboratory, filling it with things that rats like [...]. The rats loved it and we loved it too, so we called it "Rat Park". (2010, p. 3)

Alexander and his colleagues found that compared with animals housed in Skinner boxes, the rats in Rat Park took morphine at dramatically lower levels. They also found that this held for rats bred to have a metabolic dependence on morphine at birth (Alexander et al. 1981). This suggests that the addiction of the caged rats to morphine is not caused by exposure to morphine, which they have in common with the rats in Rat Park, but by something else.[4]

We can also reject the extreme compulsion view. That view says that persons living with addiction possess very little or no control over their drug-taking. Though such a view is indeed extreme, it is by no means fringe. For instance, Nora Volkow, current head of the National Institute on Drug Abuse (NIDA), a $1 billion U.S. federal agency, appears to take such a view, calling addiction a "fundamental" disruption to "self-control" (2015). Moreover, when researchers

[4] These findings are robust. See also Ahmed (2010) and Zernig et al. (2013).

and clinicians characterize addiction as a disease (a controversial but by no means fringe position; see section 4 for further discussion), they often intend to invoke compulsion as the hallmark of such a disease. As Heyman and Mims report, "when addiction specialists say that addiction is a disease, they mean that drug use has become involuntary" (2016, p. 386).

Notice that taken together, the extreme compulsion view and the extreme pharmacological view are especially implausible. They imply that *mere* exposure to drugs can cause a fundamental disruption to one's capacity for self-control. The view in this form is refuted by ordinary experience.

The extreme compulsion view on its own would appear to be refuted by the fact that persons with addictions can modulate their drug-taking behavior in response to incentives. For instance, when physicians in treatment for addiction are randomly tested and threatened with job loss if they relapse, abstinence rates remain as high as 80–90% (Ganley et al. 2005; Bohigian et al. 2005). Smaller incentives also show impressive effects. Further, if those who are incentivized to remain abstinent with vouchers that they can exchange for modest goods are compared to controls who receive only counseling, those in the voucher program consistently do better (Higgins et al. 1991; 1994; 1995). The most natural interpretation of this is that subjects are exercising self-control in these settings.[5]

The epidemiological and environmental considerations adduced against the extreme pharmacological view also have force against the extreme compulsion view. If addiction involves a "fundamental" disruption to self-control, why do most people with addictions "mature out"? And how is it that rats who have a morphine "addiction" are able, given the right environment, to refrain from taking the drug when it is readily available? One possible response is that somehow all of these creatures just fail to have true addictions. But this threatens to cause us to lose an independent handle on the phenomenon. It requires thinking that addiction must be *more* than, e.g., morphine dependence from birth, or that it is not reliably tracked by the behavioral criteria used in the Epidemiologic Catchment Area Study. That may be true, but until the reply is buttressed with an independent positive characterization of addiction such that those who mature out

[5] See Fingarette (1988, p. 38) for a helpful summary of results like these pertaining to alcoholism.

are not true addicts, it seems to involve helping itself to what is under dispute.[6]

2.2. Neuroscience

Let us now consider the neuroscience of addiction. The goal in this section is to draw valuable lessons that can be applied to the moral psychology of addiction. Addictive drugs all increase dopaminergic activity[7] and three of the brain's main dopaminergic pathways, the mesolimbic and mesocortical pathways, and the nigrostriatal pathway, are widely thought to be implicated in addiction. Crucially, it is also well-established that the mesocorticolimbic system not only responds to rewards, but to *cues* that reliably predict rewards. (Schultz et al. 1992; Shultz et al. 1997). In recent years, two major theories have been advanced regarding the role of dopamine neurotransmission in addiction: (1) the prediction error theory and (2) the incentive sensitization theory. These theories are often presented as competing (Hu 2016, pp. 300–301; Berridge 2007).[8] Against this, I will argue that they should be taken to be complementary.

2.2.1. The Prediction Error Theory

The prediction error theory says that addictive drugs interfere with learning by causing drug-taking to acquire, effectively, ever-increasing value in an animal's representational systems. Quite generally, animals can learn to update the value of rewards (Unconditioned Stimuli, US) associated with cues (Conditional Stimuli, CS) by keeping track of the discrepancy between expected value and actual value. When a cue turns out to signal greater than expected value —when there is a prediction-error— dopamine is released, strengthening the CS-US association. With natural rewards, animals eventually learn to predict their values accurately, and dopamine release at reward delivery is attenuated. However, drugs of abuse are themselves causally responsible for increasing levels of intercellular dopamine. So, un-

[6] One might also worry that such a response is dangerous in a context where the class of the most severe cases (by behavioral standards) is dominated by subjects with high rates of psychiatric comorbidity (Pickard and Pearce 2013). Clinically speaking, those subjects might be typical of persons with addictions, but it would be misleading to think that they accurately represent the phenomenon generally.

[7] Different drugs achieve this by different mechanisms, but all are fairly well understood. See, e.g., Sulzer (2011).

[8] Though see Redish, Jensen, and Johnson (2008, p. 416) for an acknowledgment the leading theories of addiction "are not incompatible with each other".

like natural rewards, drug-taking behavior is accompanied by non-diminishing dopaminergic activity at reward delivery. The prediction error theory says that the system is thus unable to accurately *represent* the value of drug-taking because the mechanism that is supposed to predict its value is being directly manipulated by the action of the drug to return the result that it is always greater than expected (Redish 2004).

2.2.2. The Incentive Sensitization Theory

According to the incentive sensitization theory, "the dopamine signals are not learning signals, in the sense that they do not give rise to [representational states] at all. Instead, they give rise to desires directly —or, more accurately, to a sensitivity to experience desires when cued with appropriate stimuli" (Holton and Berridge 2013, p. 247). On this view, the role of dopamine is to amplify the motivational power of already existing relationships between cues and rewards.

This can be investigated as follows. If rats are trained to associate both a sound and an activity (say, lever-pressing) with a sugar reward, they will learn to associate the sound and the lever-pressing with one another. That is, hearing the sound will tend to trigger the lever-pressing. Using this paradigm, it is possible to test for whether a given manipulation modulates the strength with which the rats "want" the reward, by measuring the effect the manipulation has on how much the rats press the lever in the presence of the sound. Rats given amphetamine (which increases dopamine neurotransmission) during the test will press the lever up to three times as much upon hearing the sound, apparently "wanting" the reward more strongly (Wyvell and Berridge 2000). Strikingly, nearly identical results were found when the rats were *not* given amphetamine during the trial, but instead were given an escalating regime of amphetamines that ended weeks before the trial, "sensitizing" their brains (Wyvell and Berridge 2000). In both experiments, there was no significant deviation from the rats' baseline wanting of the reward, or to their lever pressing, when given an unrelated stimulus and, crucially, there was no change in the degree to which the rats "liked" the reward.

If this theory is correct, drug use lays down intrinsic dispositional desires to take the drug, and persons with addictions are susceptible to having occurrent desires to take drugs triggered (perhaps via boosting phasic glutamate (Holton and Berridge 2011, p. 259)) by stimuli associated with drug-taking. Crucially, because the cues give rise to occurrent desires directly, these desires will be triggered even when the object of the desire is not taken to be good and even when it is not anticipated that it will be liked.

2.2.3. How are the Theories Related?

Are these two theories competing theories? One way of denying that they are in competition is to deny that they are *distinct*, i.e., to reduce the constructs of one to the constructs of the other. Indeed, Redish claims that the postulates of the incentive sensitization theory correspond to variables that are already in the prediction error model:

> Robinson and Berridge's concept of incentive salience has a direct correspondence to variables in [the prediction error model]: the value of a state reachable by an action. If agent is in state S0 and can achieve state S1 via action [x] and if state S1 has a much greater value than state S0, then [x] can be said to be a pathway with great incentive salience. (2004, p. 1946)

This strategy for integrating the two theories entails that incentive sensitization —understood now to mean that the states where the drug reward is delivered are assigned values high enough that they "win out" over other states— comes about because the pharmacological action of the drug "stamps in"[9] the association between US (drug) and CS, i.e., that the role of dopamine is primarily to be found in the *induction* of the learned association between US and CS.

However, Berridge and colleagues may point out that there is no difference in terms of *induction* of a learned association between the rats given amphetamine and the rats not given amphetamine in the experiments described above. All of the relevant induction occurred before the trial, and both groups of rats received the same training. So, there is no reason to suppose that the pharmacological action of the drug is responsible for a difference in induction of a learned association. This point is presumably only buttressed by Berridge's (2007) persuasive arguments that dopamine is neither necessary nor sufficient for learning.

If this response is successful it amounts to a defense of the independence of the variables in the two models and the claim that incentive sensitization is a phenomenon primarily of the *expression* of a learned association, not its induction. I believe this is the correct conclusion. But this doesn't show that addiction is accounted for by

[9] See Berridge (2007) for an extensive discussion of this possibility.

the phenomenon of incentive sensitization alone. Consider the following difficulty for the incentive sensitization account. Berridge and colleagues are clear that the kind of sensitization invoked by their view is the standard variety that is opposed to tolerance. So, what is the relevant drug effect which is increased with repeated use? The effect that the drug has on "wanting" of the drug itself. In the experiments cited above it was desire for natural reward (sugar pellets) that was being modulated by amphetamine or by amphetamine-driven sensitization. What happens when the object of wanting is the drug itself?

If the drug is not wanted at all, no drug effect that boosts antecedent wanting, and no increase in that drug effect, could account for intrinsic dispositional desires of increasing strength. Of course, it isn't particularly controversial to suppose that the drug *is* wanted to a non-zero degree, but this is only because the drug has been encountered and learning has begun to take place.[10] The incentive sensitization theory says that addiction is not a pathology of learning, but once learning is in the picture, it becomes an open question whether it is of the pathological variety invoked by the prediction error theory.

One positive reason for thinking it is is provided by recalling that cue-reward relationships are not only enhanced by a *presently* higher level of dopaminergic activity but are also enhanced by exposure to prior regimen of dopaminergic drugs. It is, to my knowledge, a largely unaddressed challenge for the incentive sensitization theory why the relevant kind of sensitization wouldn't cause those with a history of use of dopaminergic drugs to experience strengthening of *all* conditioned responses.[11] This is easier to account for if there is something distinctive about drugs which makes it difficult to accurately represent their value.

I take these reasons favoring the complementarity of the prediction error model and the incentive sensitization model to be highly suggestive, and it is very plausible that drug addiction involves both representational dysfunction and motivational dysfunction. If this is right, the incentive sensitization theory provides an account of one aspect of addiction, viz., the strength and persistence of desires to take drugs independently of how much they are liked; the prediction error theory provides an account of another, viz., the difficulty involved in accurately representing the value of drug-taking, which helps to explain why drug-enhanced wanting is targeted specifically onto drugs.

[10] As Robinson and Berridge acknowledge, "[l]earning specifies the object of desire" (2008, p. 3138).

[11] Though see Berridge (2007), fn. 6, where he notes that there is occasionally "spillover", e.g., some people with cocaine addiction have compulsive sexuality.

In what follows I will try to flesh out some of the consequences of this combined picture. Nevertheless, it will be clear at each point which aspects of the combined theory are under discussion so that, even if the combined theory should turn out to be false, the consequences of each part can still be tracked. In any case, I take the truth of *either* theory to be good reason to think that the very action of drugs and/or their effects on the brain are causally implicated in addiction, even if addiction is modulated by other contributing causes.

3. *Moral Psychology*

What is the best way to understand the significance of these findings at the psychological level? I suggest that we can make progress here by considering the highly natural idea that addiction interferes with an agent's ability to act on her considered judgment by subjecting her to especially strong or persistent *desires*. On this telling, addicted persons can unproblematically appreciate the importance of reforming their behavior, but they fail (when they do) because they are led astray by unruly appetites which have been installed in them by their drug use. In broad terms, this is the sort of view defended by Wallace (1999a), Watson (2004a), Dill and Holton (2014), and Burdman (2024). Such views offer a sketch of how addiction impairs agency, commitment, volition, and other things that are of personal and social significance and do so in terms that are familiar from ordinary psychological explanation. I will argue that the best way to translate the neuroscience we have just seen to the psychological level is to attend carefully to the details of how the basic states in such a framework —beliefs, desires, and volitional states— operate and interact in addiction.

3.1. Addiction as a "Defect of the Will"

Let us begin by taking Wallace's view as representative. His view is meant to capture the sense in which addiction is what he calls a "defect of the will". Wallace argues that motivational states come in two varieties, viz., those that are merely given to us, and those that are the result of primitive episodes of agency such as choice, decision, or the formation of an intention. The latter show deliberative and volitional capacities at work, whereas the former come unbidden and can be obstacles to acting in accordance with considered judgment. When we add to this picture the ordinary notion of belief, we get a tripartite moral psychology.[12] As we will see, in addiction all three elements are present but distorted.

In Wallace's formulation, what makes drug addiction practically challenging is that it involves particularly strong and persistent motivational states of the merely given variety. Wallace says:

> On the [...] account I have offered, [desires to take drugs] involve the intense focusing of one's attention onto the anticipated pleasures of (say) drug consumption [...]. But someone subject to such a quasi-perceptual state will presumably find it difficult to think clearly about the overall balance of reasons bearing on the decision to consume or abstain from consuming the drug. Let us suppose [the balance of reasons favors abstaining.] Adding such a desire to the mix [...] would make it much harder for the agent to reach this conclusion and to keep it firmly in view. The focusing of one's attention onto the pleasures of consumption [distorts judgment], and this distorting effect can be considered an impairment of the agent's capacities for practical rationality. (1999a, pp. 645–646)

The neuroscience suggests that Wallace is correct to emphasize the way in which addictive motivations are persistent and operate primarily by grabbing attention.[13] But the same neuroscience suggests that he is overemphasizing, or perhaps mislocating, the role of pleasure. Indeed, Wallace explicitly links pleasure and salience by adopting a "phenomenological" conception of desire, according to which at least the desires operative in addiction represent, in a quasi-perceptual mode, the action of drug-taking *as pleasant* (1999a, pp. 641–643).

But salience can be independent of pleasure or anticipated pleasure. People with addictions often report wanting to use drugs even though they know it won't be pleasant. Insofar as the operative motivational state in addiction is a cued intrinsic desire for drug taking of the variety posited by the incentive sensitization account, this is not surprising. On that model, not only needn't the action of drug-taking be represented as pleasant, its satisfaction also needn't ultimately *be* pleasant, as the distinction between liking and wanting suggests.

Some theorists also think that there may be a distinctive role for hedonic dysregulation in some phases of addiction and not others. As Meyer et al. put it: "[H]edonic allostasis may maintain drug-taking behaviors during early withdrawal or attempts to reduce drug

[12] An exhaustive defense of a tripartite moral psychology is beyond the scope of the present inquiry, but in broad strokes I endorse the arguments given in Wallace (1999b).

[13] Watson (2004a) also correctly emphasizes this, as does Burdman (2024).

use, whereas incentive sensitization may promote relapse even after physiological and emotional withdrawal symptoms have subsided" (2016, p. 478). This could mean that pleasure and relief from distress are involved in relapse in early abstinence but not in later abstinence.

In any case, the role of pleasure and relief from distress —assuming that these are two sides of the same conceptual coin— is not as straightforward as Wallace supposes. Where desire to take drugs does not represent that course of action as pleasant we need to know: What is the mode of presentation characteristic of such a desire?

3.2. Hybrid Intentions

An additional necessary refinement to Wallace's view points the way. Wallace is right to contrast desires which are simply given to us with those motivational states which are up to us, and he naturally thinks that, on this dimension of comparison, addictive desires should be grouped with the former. However, a leading and natural interpretation of the incentive sensitization theory has it that addictive desires are like *the latter* in that they lead directly to action, even if they are unlike them in not being up to us:

> [A]ddictive desire does not typically function like [an ordinary desire]. It does not serve as an input to deliberation, something to be weighed, along with other competing desires, in deciding what to do. Instead, addictive desire functions as something more like an intention: as something that, unless checked, will lead, in a rather direct way, to action. (Holton and Berridge 2011, p. 241)

Wallace understands the contrast between the two kinds of motivation to be a contrast between what is merely "hydraulic" —forces to which we are largely passive bystanders— and what are intimate products of our wills. Paradigmatically, intentions are like the latter, but if addictive motivations operate as suggested here, the operative states in addiction might not be desires at all, but instead deviant intentions which merely assail us.

One could attempt to make a functional distinction between the intention-like states we find in addiction and true intentions to preserve the tidiness of Wallace's distinction. Perhaps, for instance, it is a functional requirement on intentions that they be linked in the right way to the exercise of deliberative and evaluative capacities.

This move is attractive, but it rests on a refusal to acknowledge deviant members of a functional type which would be very costly in the present context. Consider what a traditional functionalist might have to say about, e.g., a defective heart. If we consider only what the heart *in fact does* to determine what functional type it belongs to, we might be forced to say it is not a heart at all, but this doesn't seem like the right thing to say. Teleofunctionalists in the philosophy of biology (Neander 1991; 1995) capture this by saying that a defective heart counts as a heart even though it cannot pump blood because it is an entity of a type whose function is to pump blood. This fact about function, in turn, is determined by evolutionary history. It was *to pump blood* for which entities of the heart type were selected. This is as true of the defective heart as it is of the efficient heart.

Addiction seems to involve dysfunction at multiple levels. If we adopt teleofunctionalism, these dysfunctions need not be type-disqualifying. This is attractive if we want to avoid a proliferation of new functional types up and down our analysis of addition. Moreover, it offers an attractive way of thinking about the deviant intentions which the incentive sensitization theory posits. If intentions are generated by certain cognitive mechanisms, and those mechanisms have the right kind of evolutionary history, the product of those mechanisms will count as intentions, on teleofunctionalist grounds, even if they don't have the complete functional profile of typical intentions.

If this is correct, then the intentions characteristic of addiction straddle the standard boundaries of tripartite moral psychology. I shall therefore call them *hybrid intentions*. They are intimately related to agency on the front end, so to speak, even if the connection with volition that is typical of intention on the back end is not present. The result is a state which drives action in accordance with one's internal states, but which is not responsive to the will.[14]

To further flesh out the difference between paradigmatic intentions and hybrid intentions, consider some of the features that are standardly attributed to intentions. Intentions preserve the motivational force of a choice or decision for later, resisting, at least to a fairly high degree, revision. Bratman (1987, p. 16) calls this feature of intentions *stability*. Typical intentions are also what Bratman (1987, p. 16) calls *controlling*. Holton, echoing Bratman, says, "The agent forms the intention at one time either by making an explicit conscious decision to perform the action or by some less deliberate, more automatic, process. Then, unless it is revised, the intention will

[14] An interesting possibility here is to think of the perception of affordances, including cues, as issuing in the prepotent formation of a kind of proto-intention which, if not inhibited, leads to a motivationally efficacious state. On such a view, cue-based motivations for drugs become intentions when they are not inhibited, and they are not easy to inhibit because of the strength of the drug-cue relationship.

lead the agent to perform the action *directly*" (2009, p. 2; emphasis in original).

As Holton says, it is not necessary that intentions be arrived at through conscious deliberation. The more automatic processes that can lead to intentions might include unconscious deliberation (if one believes in such a thing) or might involve the output of certain cognitive modules which are relevant for action and evaluation. It is important, however, that they are responsive to reasons. Perhaps practical judgment can be made unconsciously, or can partly arise from subpersonal processes. But if one were to *revise* one's judgment, one's intentions would (ideally) change accordingly; and *without* revising such a judgment, intentions, it seems, cannot change. One could say, borrowing a term from Scanlon (1998), that intentions are "judgment-sensitive" attitudes.

So, if there were to be a state that *didn't* reflect judgment[15] but was nevertheless stable and controlling, it could be a liability with respect to acting well. It would be the sort of state whose formation is not subject to volitional control, as typical intentions are, but which possesses a high degree of motivational efficacy which persists over time. It would be a (defeasibly) action-controlling state over which its subject would have, in turn, limited control.

The sort of motivational states invoked by the incentive sensitization model appear to be states of just this kind. They come on the scene in direct response to a cue and so they are not the result of deliberation, conscious or unconscious. Moreover, they are also unlike typical intentions which arise from sub-personal mechanisms (if such there be) because they can't be overthrown with a revision of judgment. All the same, the connection between them and action is very close. Unlike desires, they don't serve as *inputs* to deliberation. Rather than partially contributing to how practical questions are answered, they represent such questions as having already been decided. Their motivational force is also persistent and, because the process by which drug-cue relationships are established works by mere association, almost anything that can be associated with a drug can operate as a hybrid- intention-triggering cue.

[15] This is a central feature of hybrid intentions which has been recognized by other theorists, even if they type the state differently. For instance, Dill and Holton take the motivational states triggered by drug cues, according to the incentive sensitization theory, to be deviant desires, not deviant intentions. But they make clear that they are deviant by being judgment-insensitive: "While incentive salience desires are by nature insensitive to our judgments about what is good, not all desires share this feature" (2014, p. 4).

As with typical intentions, the presence of a hybrid intention does not, of course, *guarantee* that the corresponding action will be performed. However, it is much more difficult to overthrow a state which is not under volitional control than it is to overthrow one which is. If a state is not judgment-sensitive in the first place, it may be impervious to changes (or retrenchments) of judgment, no matter how clear or forceful.

If this is how hybrid intentions function, and hybrid intentions are central to motivation in addiction, it would be a mistake to think that trying to act well in the face of such motivations can be understood on the model of struggling against a desire. Watson seems to be getting at something similar when he says:

> We will do well, I think, to abandon [a view of addiction according to which we center] the power of addictive desire to *defeat* our best efforts and, instead, to understand the relevant notion of compulsion in terms of the tendency of certain incentives to *impair our capacity to make those efforts*. We are not so much overpowered by brute force as seduced. (2004a, p. 71; my emphasis.)

Burdman (2024) also emphasizes that we should not think of the addicted person as struggling "against a force she cannot oppose" (p. 58). But it is unclear that we can make the required shift to the recommended picture without recognizing the way in which intentions in addiction are hybrid states. Being seduced is, I take it, to be understood in terms of the fact that intentions in addiction constitute part of the agent's practical standpoint, the point of view from which action springs, as typical intentions do. Yet, they do so without being subject to the will. That is, typically, volitional states like intentions and choices are closer, logically and causally, to action than desires. Hybrid intentions share this property. By contrast, desires of the merely given variety are not themselves deliberable and are not, except where they are instrumental, the outputs of deliberation. Hybrid intentions share this property. One could say that hybrid intentions (though it is admittedly anti-Anscombean to think of them as mental states) correspond to all three of Anscombe's forms or guises of intention: they concern the future, they pick out that for the sake of which the action is done, and they make action done accordingly intentional action (2000). They are therefore action-theoretically thicker than mere desires, though because they are judgment-insensitive, they can lack a hallmark of typical intentions, viz., "an all-out, unconditional judgement that the action is desirable" (Davidson 1980, p. 99).

I take this to suggest a somewhat more sophisticated picture of how addiction is a volitional impairment than the one with which we began. Merely being subject to states with this anomalous functional profile is a volitional vulnerability simply because one's overall inventory of non-deliberable states is thereby expanded, and expanded by the addition of states which bear a direct connection with action. Perhaps more importantly, insofar as an agent's action-facing states constitute her practical view on the world, that practical orientation is compromised when it is infiltrated by states which are not responsive to her judgments or valuations.

3.3. Unstable Representations and Cognitive Distortion

Wallace assumes that representational states in addiction are not dysfunctional. But the prediction error theory suggests that this is not so. It implies that people with addictions cannot accurately represent the value of drug-taking, i.e., that the very representation of the value of drug-taking is manipulated by the pharmacological action of drugs to be ever-increasing.

Since this means that some of the cognitive materials on which practical deliberation takes place are themselves distorted, we should expect this to have downstream effects in action and to have effects elsewhere in the agent's practical decision network, manifesting in both a biasing towards the action of drug-taking and the spreading of cognitive distortion to states which are in an instrumental relation to that action.

We should also expect such distortion to be a springboard for various other cognitive biases. We know that biases such as the confirmation bias recruit cognition in defense of representational states. Pickard (2016) highlights role of denial in addiction, but she also makes clear that denial is only one manifestation of the face of addiction as a "disorder of cognition" (p. 278). Other aspects include: "information-processing biases, motivational influences on belief formation and self-deception, and cognitive deficits with respect to insight and self-awareness"[16] (p. 277). All of these processes are directly related to the artificially positive representations which the prediction

[16] The connection with self-deception here is especially interesting. According to my own view (Gibson 2020), it is sufficient for one to be self-deceived that one have an externally defeated belief, the evidence against which is not appreciated due to a desire to continue believing as one does, even if the belief is originally installed by a biased sub-personal mechanism. Further discussion would be needed to show that cognitive distortions in addiction fit this paradigm, but at first blush, they appear to.

error theory predicts we should find that those with addictions have towards drug-taking.

Other views in the philosophical literature also model addiction at the level of belief or cognition. For instance, Levy (2014) understands addiction as involving oscillations between all-things-considered judgments for and against drug-taking caused by competition between top-down and bottom-up processing. Sripada (2022) understands addiction to involve unreliable control over automatic thoughts which contribute to a distorted understanding of the true value of drug-taking. Levy's model makes explicit appeal to the prediction error theory in explaining why the oscillation occurs: minimizing prediction error in the presence of drug cues leads to a model in which drug-taking is highly valued; but minimizing prediction error under other circumstances leads to the judgment that abstention is more valuable. Sirpada's does not, but insofar as his view posits an inability to form a stable judgment of the value of drug-taking, seems consonant with it.

One of the chief advantages of the approach that I am taking here —and one of the key payoffs of arguing, however tentatively, for the complementarity of the prediction error theory and the incentive sensitization theory— is that I can recognize the role played by prediction-error-caused cognitive distortions, while also assigning a central role to what seems to be an ineliminable and distinct aspect of addiction, viz., volitional and motivational dysfunction. If this is correct, competing views that do not recognize the role of hybrid intentions are leaving out a key piece of the picture.

3.4. Choice, Agency, and Responsibility in Addiction

Acknowledging all of these ways in which addiction affects the ordinary economy of judgment, motivation, and volition is consistent with the claim that those who have addictions choose to take drugs. But this is a rather superficial truth. The impairment of addiction does not show itself in how it bypasses choice, it shows itself in how the capacity for choice is re-directed. If choice and other deliverances of the will are the bottlenecks through which practical agency ultimately reaches action generally, this is what we should expect.

Nevertheless, we should reject certain *choice-based models* of addiction, such as Heyman's (2009). Heyman's view is sophisticated, but in sum he thinks that addiction can be modeled using the tools of rational choice theory and behavioral economics. According to Heyman, from a local-choice perspective, drug-taking will often be the rationally superior choice even if, from a global-choice perspective, abstinence or reduction is clearly warranted. Addiction, then, involves making locally rational but globally irrational choices.

It is true that there are often clear temporal and probabilistic disparities between the value of drug-taking when assessed from the global- and local-choice perspectives. As Heyman says,

> From the perspective of current choice, specious rewards [like drugs] have high value (because of immediate benefits and hidden costs), whereas from the perspective of global choice, their effective value is their true value because the costs have as much weight as the benefits. (2009, p. 146)

Some natural rewards (such as fatty and sugary foods) are specious in this sense and are familiar as such. But drugs have several properties which natural rewards lack which manipulate the local choice architecture and bias towards local-choice perspective-taking. For example, drugs undermine the value of competing choices by being "behaviorally toxic" (Heyman 2009, p. 145). Typically, when we attend to a particular activity, the value of doing other things will increase, perhaps by becoming more pressing, or as alternatives become more interesting. Drugs often don't work like that. Instead, they continue to grab attention and the effects of withdrawal and intoxication make other things less rewarding. This tilts the comparative judgment between locally available options in favor of drug-taking.

Heyman's analysis illuminates the ways in which drugs manipulate choice architecture and set up a framework within which drug-taking choices are locally rational. However, the framing of addiction fundamentally through the lens of choice can start to look misguided once the neuroscience is under our belts. We needn't dispute that there are intentional states we can attribute to agents in virtue of which their behavior would appear to be locally rational. But if the ultimate source of such states is direct manipulation by the pharmacological action of drugs and/or resulting neuroplasticity, it is unclear what the *significance* of the availability of this interpretation is supposed to be.

If pharmacology and neurobiology are at the root of the observed behavior, asking whether such behavior can be given a rational interpretation is not altogether probative. Doing so simply involves taking for granted the valuational and motivational states which only arise because of drug-induced neuroplasticity and asking what is rational *relative to them.* But such a perspective overlooks the way in which addiction involves disruptions to the very states and processes against which assessments of local rationality take place.[17]

In contexts where an agent is in full command of her cognitive and volitional capacities and is fully informed about the possible costs of acting in a certain way, choices bear an intimate connection with moral responsibility. Paradigmatic choices are made on the basis of the agent's assessment of the balance of reasons. As such, inability to successfully justify a choice can occasion a number of different responses and forms of negative moral appraisal.

But the connection between choice and the appropriateness of these responses is unclear in addiction. Choices to take drugs can either fail to reflect the agent's assessment of the balance of reasons (if action follows directly from a hybrid intention) or fail to reflect a *competent* assessment of those reasons (if based on distorted cognition). Choices paradigmatically render agents accountable,[18] in the sense of being the appropriate targets of reactive attitudes and other sanctions, but this is unclear when the choices have the properties that they have in addiction.

Nevertheless, we should notice that two kinds of assessment still seem to have relatively unproblematic purchase. First, the states that impair cognition and volition in addiction do not operate in isolation from other things the agent believes, wants, or is committed to. The fact that addiction does not involve literal compulsion is enough to show this. Those with addictions respond positively to therapeutic support, to well-crafted choice architecture, to the cultivation of self-insight, and perhaps more crucially, to being held accountable to a community (Pickard and Pearce 2013). All of this only seems appropriate if we see addiction broadly within the framework of agency. The fact that recovery is possible under such conditions suggests that failure to do so soon enough or effectively enough could render an agent negatively assessable[19] if enough supportive background conditions are satisfied.

[17] As such, I am not objecting to choice views *per se*, but only to views which take the status of choice to be relatively unproblematic in the context of addiction.

[18] I use this term here to refer to an agent's being responsible in the sense of being the appropriate target of sanctions and reactive attitudes. This, I take it, is roughly how Watson (2004b) uses the term. Though, see Shoemaker 2011 for a distinction between "accountability" and "answerability", where it is the latter which corresponds to being the appropriate target of sanctions and reactive attitudes and the former captures a distinctive form of response to the violation of relationship-constituting norms. See, in turn, Smith 2012 for an argument against that distinction.

[19] Though there are pragmatic questions about which forms of assessment are most compatible with ultimate success in recovery (Pickard 2017). Whether this

Second, the action-facing states which are triggered by drug cues and the (distorted) assessment of the value of drug-taking are still *states of the agent*. Thus, if instead of asking whether it would be fair to blame an addict for her conduct, we engage in assessment of her character —roughly, one's more-or-less stable set of cognitive, behavioral, and affective dispositions— we may find an appropriate target for moral assessment. The states in question are states for which the agent is not fully responsible, but in general that is not a condition on being an object of aretaic assessment (Watson 2004b), as evidenced by the banal fact that no one is wholly self-created (Wolf 1987) and that we nevertheless (rightly, it seems) take the ways people are to have great interpersonal significance (Smith 2005).

So, while the psychological states and processes characteristic of addiction may not be such that the agent is fully accountable for them and their behavioral upshots, they may nevertheless be *attributable* to the agent (Watson 2004b). They are states *of* the agent and because of their stability and (in severe cases) relative centrality to the character of the agent, ground certain judgments about that character.

This is relevant not only because it allows a certain kind of assessment to get a foothold. It is also only *because* those states are attributable to the agent that they can come under her indirect control and can be tempered, modified, inhibited, or displaced given the other traits of character and states of mind that constitute her practical self. Indeed, the degree to which the subject ultimately succeeds in taking control of her life depends on the rest of her character being brought to bear in this way. It is thus no surprise that treating those who are addicted as responsible —that is, as agents who face obstacles to their agency but who nevertheless have the power to reform— is a promising therapeutic avenue for recovery (Pickard and Pearce 2013).

Although there are a great many complications about addiction and moral responsibility that must remain off the page, I take myself to have pointed the way to the following:

I. The pharmacology of addictive drugs and the neurobiology of addiction support the claim that there are both motivational and representational aspects to drug addiction. Epidemiology, behavioral science, and animal studies show that addiction is modulated by environmental factors and often spontaneously resolves.

can be substituted for the question of which forms are most appropriate without qualification, I leave open.

II. An analysis of addiction in terms of recalcitrant desires alone is inadequate because (a) the operative motivational states in addiction share too many interesting properties with intentions and (b) such an analysis leaves out the representational aspect of addiction.

III. The motivational and representational dysfunctions characteristic of addiction are consistent with drug-taking being grounded in choice but the usual connection between choice and responsibility is rendered unclear; though agency is involved in drug-taking behavior, it is subject to gradable impairment which should temper judgments of accountability.

IV. Addictive dysfunction nevertheless manifests in traits of character which are morally assessable.

V. Because addiction is consistent with agency, recovery is predicted by treating those with addictions *like agents* and by other scaffolds to self-control, e.g., self-insight, accountability, choice architecture, incentives, and other forms of support.

VI. Failure at recovery under such conditions of support is potentially negatively assessable.

I take these conclusions to be significant, in part, because of the way they relate agency in addiction to non-disordered agency. Agency in addiction isn't just agency in the face of unruly appetites. But it is also not *merely* being pushed around mechanically by states which bear no relation to the states recognized by ordinary psychological explanation. With the major signposts on the terrain, we can begin in earnest to tease out the ultimate significance of addiction in context.

4. *Disease?*

How do these conclusions relate to whether addiction *is* a disease? This question is complicated by having two senses. The first concerns whether addiction is a disease in the sense investigated by philosophers of medicine and biology. The second concerns whether addiction is a disease as opposed to a set of moralized choices. In this second sense, whether addiction is a disease is tantamount, as we have already seen, to whether addiction involves compulsion. As Pickard (2017, p. 170) says, "The moral model of addiction has two distinctive features. First, it views drug use as a choice, even for those with addictions. Second, it adopts a critical moral stance against this choice.” The disease model, in this second sense, is meant to offer an alternative which avoids the critical moral stance of the moral model, but, as we have already seen, it is typically understood to entail that drug-taking in addiction is involuntary.

So understood, the disease model is equivalent to a compulsion model and is false. But more importantly, the opposition between the disease model and the moral model is artificial and non-exhaustive. Obviously, and as Pickard points out, one could deny the moral model by denying that one should take a critical stance towards addiction (2017). More fundamentally, there is simply no logical connection between (a) whether something is partially constituted by choices (whether or not those choices are the objects of disapprobation) and (b) being a disease (in the technical or non-technical sense). Choice is implicated in complex ways in the etiology and symptomatology of countless canonical diseases, as component cause in chronic lifestyle diseases such as heart disease and some cancers, and constitutively as in phobias, depressive disorders, and anxiety disorders. Such conditions may even be consequently stigmatized. But the fact that it is possible to (often in bad faith) stigmatize behavior that can be construed as choice-consistent should not drive theorizing.[20]

One could attempt to claim that the psychiatric conditions I have listed do not involve choices, but rather some form of compulsion. Perhaps if they were truly choices, people could “simply choose” *not* to repeat a ritual in response to intrusive thoughts, or not to stay in bed all day when depressed, and so on. Setting aside the general probity of this test, in order for it to be relevant to whether addiction is a disease, one must simply assume that choice and disease mutually exclude, which is the thing presently at issue and for which I see no *a priori* grounds. Moreover, as I hope to have shown in section 3.4, a superficial analysis in terms of choice is consistent with quite serious underlying dysfunction, so it is highly implausible that the distinction between choice and compulsion can bear the weight that it being asked to bear here.

To settle whether *that* underlying dysfunction makes for disease, we need a philosophical theory of disease, and there is an extensive literature attempting to settle that question. I have spoken of “dysfunction” freely throughout this essay, and part of the long-standing

[20] Pickard (2017; 2022) makes this point forcefully, arguing against those who explicitly say that they champion the disease model or avoid a choice model because they think doing so is the only way to reduce stigma. Evidently, this amounts to an admission that theorizing is being held hostage to public sentiment.

DOI:https://doi.org/10.22201/iifs.18704905e.2024.1530

philosophical debate about the nature of disease concerns whether the notion of (dys)function is normative, and whether it is sufficient for disease. I happen to think that it is irreducibly normative and that is not sufficient for disease. In other words, though I can't settle the issue presently, I am convinced that naturalism about disease (in the style of Boorse 1977) is false.[21] But more to the point, *we don't need to settle this here*: I take everything I have said thus far to be consistent with the truth or falsity of naturalism about disease, as well as with the truth or falsity of any other philosophical theory of disease.

Once we see that there is no meaningful connection between addiction's status as a (non)disease and *what we care about*, one is tempted into (at least a local form of) *eliminativism* about disease.[22] As Ereshefsky (2009) says, instead of speaking of "health" and "disease", it is often profitable to speak in terms of "state descriptions" —descriptions of "physiological or psychological states" and "normative claims"— "judgments concerning whether we value or disvalue [such] states" (p. 225). Indeed, I have attempted to offer a sketch of this sort in this paper. We considered low-level state descriptions from neuroscience and raised them up to the psychological level in order to discover what effects they might have on our judgments about agency and responsibility. A full exploration of eliminativism will have to be a task for another day, but I hope to have shown, among other things, that a methodology focused on teasing out the normative consequences of the best available state description of mental disorders is a fruitful one for philosophy of psychiatry.

[21] I am persuaded by Kingma (2007) and Stegenga (2015 and 2017), among others.

[22] As an anonymous reviewer points out, at this juncture one might instead adopt Pickard's (2022) agnosticism about addiction's status as a disease. I agree with Pickard that the present state of knowledge is indecisive with respect to (what she takes to be) the key question of whether brain *dysfunction* is the cause of the behavioral syndrome of addiction. Nevertheless, it might not matter much for what we care about (assuming I have been right about that thus far) how that question turns out. Eliminativists about disease needn't deny that somewhat of a clear notion of disease can be made out and applied to certain cases. Fundamentally, what they deny is that there is any importance to the results of doing so (or not doing so).

REFERENCES

Ahmed, Serge H., 2010, "Validation Crisis in Animal Models of Drug Addiction: Beyond Non-Disordered Drug Use toward Drug Addiction", *Neuroscience Biobehavioral Review*, vol. 35, no. 2, pp. 172–184. https://doi.org/10.1016/j.neubiorev.2010.04.005

Alexander, Bruce K., 2010, "Addiction: The View from Rat Park". http://www.brucekalexander.com/articles-speeches/rat-park/148-addiction-the-view-from-rat-park

Alexander, Bruce K., Barry L. Beyerstein, Patricia F. Hadaway, and Robert B. Coambs, 1981, "Effect of Early and Later Colony Housing on Oral Ingestion of Morphine in Rats", *Pharmacology, Biochemistry, and Behavior*, vol. 15, no. 4, pp. 571–576.

Anscombe, G. Elizabeth M., 2000, *Intention*, 2nd edition, Harvard University Press, Cambridge, Mass.

Berridge Kent C., 2012, "From Prediction Error to Incentive Salience: Mesolimbic Computation of Reward Motivation", *The European Journal of Neuroscience*, vol. 35, no. 7, pp. 1124–1143. https://doi.org/10.1111/j.1460-9568.2012.07990.x

Berridge Kent C., 2007, "The Debate over Dopamine's Role in Reward: the Case for Incentive Salience", *Psychopharmacology* (Berl.), vol. 191, no. 3, pp. 391–431. https://doi.org/10.1007/s00213-006-0578-x

Berridge, Kent C., Terry E. Robinson, and J. Wayne Aldridge, 2009, "Dissecting Components of Reward: 'Liking', 'Wanting', and Learning", *Current Opinion in Pharmacology*, vol. 9, no. 1, pp. 65–73.

Bohigian, George M., Robert Bondurant, and Jack Croughan, 2005, "The Impaired and Disruptive Physician: The Missouri Physicians' Health Program—an Update", *Journal of the Addictive Diseases*, vol. 24, no. 1, pp. 13–23.

Boorse, Christopher, 1977, "Health as a Theoretical Concept", *Philosophy of Science*, vol. 44, no. 4, pp. 542–573.

Bratman, Michael, 1987, *Intention, Plans, and Practical Reason*, Harvard University Press, Cambridge, Mass.

Burdman, Federico, 2024, "Recalcitrant Desires in Addiction", in David Shoemaker, Santiago Amaya and Manuel Vargas (eds.), *Oxford Studies in Agency and Responsibility*, vol. 8, Oxford University Press, Oxford, pp. 57–79.

Davidson, Donald, 1980, "Intending", in *Essays on Actions and Events*, Oxford University Press, Oxford, pp. 83–102.

Dill, Brendan and Richard Holton, 2014, "The Addict in Us All", *Frontiers in Psychiatry*, vol. 5, no. 139, pp. 1–20.

Ereshefsky, Marc, 2009, "Defining 'Health' and 'Disease' ", *Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 40, no. 3, pp. 221–227.

Fingarette, Herbert, 1988, *Heavy Drinking: The Myth of Alcoholism as a Disease*, University of California Press, Berkeley.

Gallagher, Shaun, 2022, "Integration and Causality in Enactive Approaches to Psychiatry", *Frontiers in Psychiatry*, vol. 13. https://doi.org/10.3389/fpsyt.2022.870122

Ganley, Oswald H., Warren J. Pendergast, Michael W. Wilkerson, and Daniel E. Mattingly, 2005, "Outcome Study of Substance Impaired Physicians and Physician Assistants under Contract with North Carolina Physicians Health Program for the Period 1995–2000", *Journal of Addictive Diseases*, vol. 24, pp. 1–12.

Gibson, Quinn Hiroshi, 2024, "Philosophy's Role in Theorizing Psychopathology", *Philosophy, Psychiatry and Psychology*, vol. 31, no. 1, pp. 1–12.

Gibson, Quinn Hiroshi, 2020, "Self-Deception as Omission", *Philosophical Psychology*, vol. 33, no. 5, pp. 657–678. https://doi.org/10.1080/09515089.2020.1751100

Heyman, Gene, 2009, *Addiction: A Disorder of Choice*, Harvard University Press, Cambridge, Mass. and London, England.

Heyman, Gene and Verna Mims, 2016, "What Addicts Can Teach Us about Addiction: A Natural History Approach", in Nick Heather and Gabriel Segal (eds.), *Addiction and Choice: Rethinking the Relationship*, Oxford University Press, Oxford, pp. 385–408.

Higgins, Stephen T., Alan J. Budney, Warren K. Bickel, Gary J. Badger, Florian E. Foerg, and Doris Ogden, 1995, "Outpatient Behavioral Treatment for Cocaine Dependence: One-Year Outcome", *Experimental and Clinical Psychopharmacology*, vol. 3, pp. 205–212.

Higgins, Stephen T., Alan J. Budney, Warren K. Bickel, Florian E. Foerg, Robert Donham, and Gary J. Badger, 1994, "Incentives Improve Outcome in Outpatient Behavioral Treatment of Cocaine Dependence", *Archives of General Psychiatry*, vol. 51, pp. 568–576.

Higgins, Stephen T., Dawn D. Delaney, Alan J. Budney, and Warren K. Bickel, 1991, "A Behavioral Approach to Achieving Initial Cocaine Abstinence", *American Journal of Psychiatry*, vol. 148, no. 9, pp. 1218–1224.

Holton, Richard, 2009, *Willing, Wanting, Waiting*, Oxford University Press, Oxford.

Holton, Richard, and Kent Berridge, 2011, "Addiction between Compulsion and Choice", in Neil Levy (ed.), *Addiction and Self-Control*, Oxford University Press, Oxford.

Hu, Hailan, 2016, "Reward and Aversion", *Annual Review of Neuroscience*, vol. 3, pp. 297–324.

Kingma, Elselijn, 2007, "What Is It to Be Healthy?", *Analysis*, vol. 67, no. 2, pp. 128–133.

Levy, Neil, 2014, "Addiction as a Disorder of Belief", *Biology and Philosophy*, vol. 29, no. 3, pp. 337–355. https://doi.org/10.1007/s10539-014-9434-2

Meyer, Paul J., Christopher P. King, and Carrie R. Ferrario, 2016, "Motivational Processes Underlying Substance Abuse Disorder", *Current Topics in Behavioral Neurosciences*, vol. 27, pp. 473–506. https://doi.org/10.1007/7854_2015_391

Neander, Karen, 1995, "Malfunctioning and Misrepresenting", *Philosophical Studies*, vol. 79, pp. 109–141.

Neander, Karen, 1991, "Functions as Selected Effects", *Philosophy of Science*, vol. 58, pp. 168–184.

Pickard, Hanna, 2022, "Is Addiction a Brain Disease? A Plea for Agnosticism and Heterogeneity", *Psychopharmacology*, vol. 239, pp. 993–1007.

Pickard, Hanna, 2017, "Responsibility without Blame for Addiction", *Neuroethics*, vol. 10, no. 1, pp. 169–180. https://doi.org/10.1007/s12152-016-9295-2

Pickard, Hanna, 2016, "Denial in Addiction", *Mind & Language*, vol. 31, no. 3, pp. 277–299.

Pickard, Hanna and Steve Pearce, 2013, "Addiction in Context: Philosophical Lessons from the Personality Disorder Clinic", in Neil Levy (ed.), *Addiction and Self-Control*, Oxford University Press, Oxford, pp. 165–189.

Redish A. David, 2004, "Addiction as a Computational Process Gone Awry", *Science*, vol. 206, no. 5703, pp. 1944–1947.

Redish, A. David, Steve Jensen, and Adam Johnson, 2008, "A Unified Framework for Addiction: Vulnerabilities in the Decision Process", *Behavioral and Brain Sciences*, vol. 31, no. 4, pp. 415–487.

Robinson Terry E., and Kent C. Berridge, 2008, "The Incentive Sensitization Theory of Addiction: Some Current Issues", *Philosophical Transactions of the Royal Society*, vol. 363, no. 1507, pp. 3137–3146.

Robinson, Terry E., and Kent C. Berridge, 1993, "The Neural Basis of Drug Craving: An Incentive-Sensitization Theory of Addiction", *Brain Research Reviews*, vol. 18, no. 3, pp. 247–291. https://doi.org/10.1016/0165-0173(93)90013-P

Scanlon, Thomas Michael, 1998, *What We Owe to Each Other*, Harvard University Press, Cambridge, Mass., and London, England.

Schultz Wolfram, Peter Dayan, and P. Read Montague, 1997, "A Neural Substrate of Prediction and Reward", *Science*, vol. 275, no. 5306, pp. 1593–1599.

Schultz Wolfram, Paul Apicella, Eugenio Scarnati, and Tomas Ljungberg, 1992, "Neuronal Activity in Monkey Ventral Striatum Related to the Expectation of Reward", *Journal of Neuroscience*, vol. 12, no. 12, pp. 4595–4610.

Sellars, Wilfrid, 1963, "Philosophy and the Scientific Image of Man", in *Science, Perception, and Reality*, Ridgeview Publishing Company, Cal., pp. 1–40.

Shoemaker, David, 2011, "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility", *Ethics*, vol. 121, no. 3, pp. 602–632.

Smith, Angela M., 2012, "Attributability, Answerability, and Accountability: In Defense of a Unified Account", *Ethics*, vol. 122, no. 3, pp. 575–589.

Smith, Angela M., 2005, "Responsibility for Attitudes: Activity and Passivity in Mental Life", *Ethics*, vol. 115, no. 2, pp. 236–271.

Sripada Chandra, 2022, "Impaired Control in Addiction Involves Cognitive Distortions and Unreliable Self-Control, not Compulsive Desires and Overwhelmed Self-Control", *Behavioural Brain Research*, vol. 418, 113639.
https://doi.org/10.1016/j.bbr.2021.113639

Stegenga, Jacob, 2017, *Medical Nihilism*, Oxford University Press, Oxford.

Stegenga, Jacob, 2015, "Effectiveness of Medical Interventions", *Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 54, pp. 34–44.

Sulzer, David, 2011, "How Addictive Drugs Disrupt Presynaptic Dopamine Neurotransmission", *Neuron*, vol. 69, no. 4, pp. 628–649.
https://doi.org/10.1016/j.neuron.2011.02.010

United States Department of Health and Human Services. National Institutes of Mental Health, 1994, "Epidemiological Catchment Area Study, 1980–1985: [United States] (ICPSR 6153)".
https://doi.org/10.3886/ICPSR06153.v1

Volkow, Nora, 2015, "Addiction is a Disease of Free Will", William C. Menninger Memorial Convocation Lecture, 168th Annual Meeting of the American Psychiatric Association, Toronto, Canada.

Wallace, R. Jay, 1999a, "Addiction as a Defect of the Will: Some Philosophical Reflections", *Law and Philosophy*, vol. 18, no. 6, pp. 621–654.

Wallace, R. Jay, 1999b, "Three Conceptions of Rational Agency", *Ethical Theory and Moral Practice*, vol. 2, no. 3, pp. 217–242.

Watson, Gary, 2004a, "Disordered Appetites: Addiction, Compulsion, and Dependence", in *Agency and Answerability: Selected Essays*, Oxford University Press, Oxford, pp. 59–87.

Watson, Gary, 2004b, "Two Faces of Responsibility", in *Agency and Answerability: Selected Essays*, Oxford University Press, Oxford, pp. 260–288.

Wolf, Susan, 1987, "Sanity and the Metaphysics of Responsibility", in Ferdinand David Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, Cambridge University Press, Cambridge, pp. 46–62.

Wyvell, Cindy L., and Kent C. Berridge, 2001, "Incentive Sensitization by Previous Amphetamine Exposure: Increased Cue-Triggered 'Wanting' for Sucrose Reward", *Journal of Neuroscience*, vol. 21, pp. 7831–7840.

Wyvell, Cindy L., and Kent C. Berridge, 2000, "Intra-Accumbens Amphetamine Increases the Conditioned Incentive Salience of Sucrose Reward: Enhancement of Reward 'Wanting' without Enhanced 'Liking' or Response Reinforcement", *Journal of Neuroscience*, vol. 20, no. 21, pp. 8122–8130.

Zernig Gerald, Kai K. Kummer, and Janine M. Prast, 2013, "Dyadic Social Interaction as an Alternative Reward to Cocaine", *Frontiers in Psychiatry*, vol. 4, no. 100. https://doi.org/10.3389/fpsyt.2013.00100

Received: July 30, 2023; revised: October 13, 2023; accepted: October 31, 2023.