29
ADDICTION AND THE MEANING OF DISEASE

Hanna Pickard

Is addiction a brain disease? Cards on the table: I don't know. My best guess is that in some cases it may be, in others not. But the fundamental aim of this chapter is not to argue for this admittedly weak conclusion. It is to do some conceptual spadework, so the ground is prepared for consideration of the question, whether the answer is ultimately yes, no, or, as I suspect, sometimes.

This conceptual spadework is often missing from addiction research. On the one hand, the question of whether addiction is a brain disease is not reliably distinguished from the question of whether labeling it thus has beneficial consequences. On the other, the question of what a disease is – let alone what a disease of the brain is – is rarely addressed. Consider any general question of the form: Is X a Y? How could we know the answer unless we know not only facts about X, but also what it is to be a Y? Yet few advocates or critics alike of the view that addiction is a brain disease have provided an account of what a brain disease is. Here, I hope to make a start.

The consequences of the brain disease label

Alan Leshner, director of the National Institute of Drug Abuse (NIDA) from 1994–2001, famously advocated for the brain disease model (BDMA) as an effective means of combating stigma (Leshner, 1997; Satel & Lilienfeld, 2014). Addiction is one of the most stigmatized of all psychiatric disorders (Barry et al., 2014); addiction stigma is associated with barriers to treatment (McGinty, Barry et al., 2018; McGinty et al., 2019) and public support for criminal justice rather than public health management strategies (Kennedy-Hendricks et al., 2017). Leshner hoped labeling addiction a brain disease would shift public perception from "bad person" to "chronic illness sufferer", thereby reducing stigma, improving funding for addiction neuroscience research and treatment, and eroding punitive attitudes and policies (Leshner, 1997, p. 45).

Leshner was right to recognize the importance of combating addiction stigma. But the effects of a label can only be established by empirical study. They are also likely to vary with culture and audience. To assess whether the brain disease label is effective at combating stigma, we should distinguish its possible effects on at least five groups: (1) the general public; (2) politicians and policy-makers; (3) family, friends, and colleagues of people addicted to drugs; (4) clinicians; and (5) addicted individuals themselves.

Consider first (1) the general public. Within the US, it is unclear how much public acceptance of the label has affected attitudes. Scientific support for the BDMA coincided with the "War on Drugs" in the 1980s and 1990s. This policy was politically advantageous because it appealed to a popular moralism about drugs that intersected with gender, class, and race (Courtwright, 2010). This dynamic is arguably still present: it is noteworthy that (2)

DOI: 10.4324/9781003032762-33

politicians and policy-makers became more willing to frame addiction as a public health rather than criminal justice problem as the opioid epidemic in the US impacted White communities (Hart, 2018, 2021). Meanwhile, increased public acceptance of biogenetic labels for mental disorders was not associated with decreased stigma between 1996 and 2006 (Pescosolido et al., 2010), and current rates of stigma toward people with mental disorders, including those with alcohol and opioid addiction, remain high (McGinty, Pescosolido et al., 2018; McGinty, Barry et al., 2018). Vignette studies complement these findings. Although biogenetic labels in general increase public support for services and treatment, they also increase social rejection and public perceptions of dangerousness, unpredictability, and difference (Pescosolido et al., 2010; Haslam & Kvaale, 2015; Mehta & Farina, 1997); a recent US national randomized study on the systematic effects of different labels for drug-related impairments confirm these findings for addiction (Kelly et al., 2020). On reflection, this is unsurprising. Many diseases are well-known to be highly stigmatizing, e.g., leprosy or HIV-positive status. Stigma may be associated with a disease label, rather than countered by it (Pickard, 2020b; Ross, 2020; see also Chapter 32, this volume).

By contrast, there is evidence that biogenetic labels in general, including the brain disease label in particular (Kelly et al., 2020), decrease attributions of individual blame, potentially enabling (3) friends, family, and colleagues of those addicted to drugs to maintain positive regard (Haslam & Kvaale, 2015; Mehta & Farina, 1997). Insofar as a disease label functions within a culture to exempt individuals from social expectations and responsibilities (Parsons, 1951), this is again unsurprising. Strikingly, however, biogenetic models appear associated with *less* empathy than psychosocial models in (4) clinicians (Haslam & Kvaale, 2015). This finding is consistent with research suggesting that personal narratives employing a psychosocial orientation may be more powerful than any label for combating stigma (Kennedy-Hendricks et al., 2016; McGinty, Pescosolido et al., 2018).

However, the effect of the brain disease label on (5) addicted individuals themselves appears negative. A sense of one's own agency for change may be crucial to successful treatment and recovery (Bandura, 1997; Lewis, 2015; Peele, 1985; Pickard, 2017; Satel & Lilienfeld, 2014; Wiens & Walker, 2015). Experimental manipulations of belief in the BDMA as compared with a psychosocial model in individuals with mild to moderate alcohol problems suggest it decreases a sense of agency, as measured by internal locus of control, self-efficacy, and an "entity view" of addiction as immutable and impervious to efforts to change, while failing to protect against internalized stigma and shame (Wiens & Walker, 2015). More generally, the brain disease label contributes to pessimism and helplessness about recovery (Haslam & Kvaale, 2015); indeed, a prospective study testing the relapse model of G. Alan Marlatt and Gordon (1985) found study participants' belief in the BDMA and absence of coping resources to be the strongest predictive factors of relapse (Miller et al., 1996). Acceptance of the BDMA also appears to impede problem recognition in problem drinkers (Morris et al., 2021). Disease labels divide people into two kinds: those who have the disease, and those who do not. This may motivate denial and other defense mechanisms so as to avoid the threat to self-identity and stigma associated with disease (Morris et al., 2021; Pickard, 2016, 2020a).

On balance, the brain disease *label* therefore appears to be at best a "mixed blessing" (Haslam & Kvaale, 2015), varying with audience in its effects. Evidence-based medicine is now orthodoxy; so, too, should be evidence-based public health messaging (McGinty & Barry, 2020). If we want to use labels to do good, we should find out what actually works, for which audience (Kelly et al., 2020).

Faith in the effectiveness of the BDMA to combat stigma is therefore a barrier to evidence-based public health messaging. But it is also a barrier to evaluating the evidence for or against the truth of the model, for it creates a dilemma: commit to the BDMA, or fail to combat stigma. This dilemma is false; there is no inconsistency in pairing agnosticism – or indeed, rejection – of the BDMA

with a commitment to combating stigma, decriminalizing drug possession, and fighting for access to treatment and harm reduction strategies. Nonetheless, the dilemma has been invidious. In an article notable in other respects for its judicious assessment of the BDMA, Markus Heilig and colleagues write: "A view of addiction as a disease is justified, because it is beneficial" (Heilig et al., 2021, p. 3). Even if the brain disease label was an unequivocal blessing, that would not establish that addiction is as the label represents it to be. Scientific theories are justified by evidence, not by their good or bad effects on minds and society. Whatever the consequences of the brain disease *label* – whether these ultimately prove to be positive, negative, or a bit of both – they have no bearing on the validity of the brain disease *model*.

Two concepts of disease

The concept of disease has its home within a family that includes injury, illness, disability, and disorder. Ordinary use tracks rough-and-ready distinctions between these. Injury typically involves damage caused by accident. Illness evokes the subjective experience of symptoms. Disability points to limitations of ability, and disorder points to divergence from "order" – suggesting deviations from psychological and behavioral norms. Medical science and practice may overlook these distinctions for the sake of conceptual unity; these conditions, like disease, are all taken to be inconsistent with *health* (Boorse, 1977). Nonetheless, medical use, like ordinary use, tracks some differences.

Within medicine, there are two competing models of disease that Dominic Murphy (2009) has called the *minimal* and *strong* model (cf., Kenneth Kendler [2012], who refers to the models as *soft* and *hard*). The minimal model dates to the seventeenth century and treats diseases as *syndromes*: collections of observable signs and experienced symptoms that co-occur and unfold over time in a characteristic way (Figure 29.1). Emil Kraepelin (1899) famously applied this model as the basis of differential diagnosis in psychiatry, e.g., between dementia praecox and other psychiatric conditions. Crucially, the minimal model makes no claims about underlying causes. Diseases are defined at the personal level, by appeal to characteristic courses of signs and symptoms.

The strong model emerged over the course of the nineteenth century. It treats diseases as pathological states and processes in the body that *cause* the collection of signs and symptoms constitutive of syndromes. This causal hypothesis is the crucial difference between the strong and minimal models (Figure 29.2). Paradigmatic examples that fit the strong model include Mendelian genetic defects or infectious diseases, where there is a single, clear pathological state or process that causes a collection of signs and symptoms: a gene whose function is

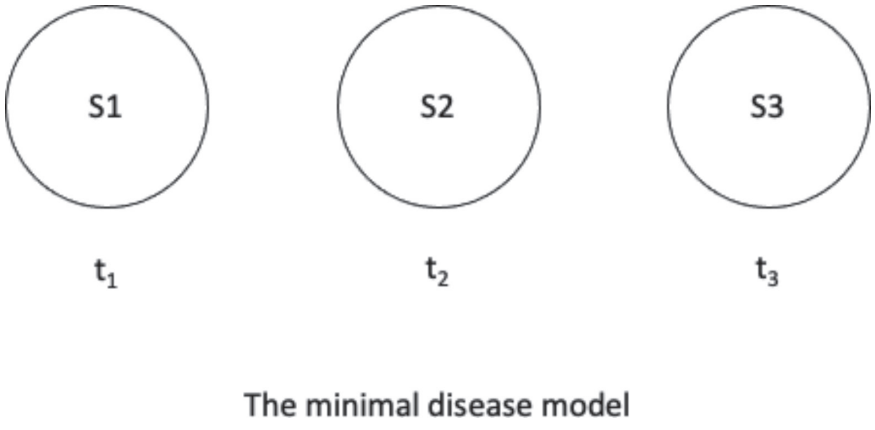

Figure 29.1 S1–S3 are signs and symptoms unfolding over a period of time t_1–t_3.

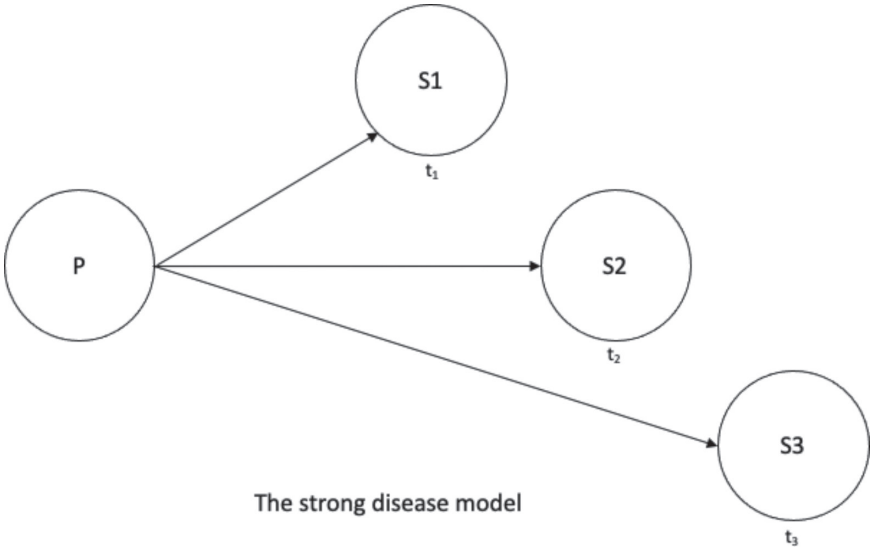

Figure 29.2 S1–S3 are signs and symptoms caused by underlying pathology P over a period of time t_1–t_3. Arrows represent causation.

disrupted or an invading organism (Kendler, 2012). Over the course of the twentieth century, with advances in biomedical and biogenetic science, our ordinary concept of disease has become increasingly geared toward the strong model. *Disease* is often conceptualized as *underlying pathology*, which manifests in personal-level signs and symptoms.

Within psychiatry, mental disorder as formally defined by DSM-5 (the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* of the American Psychiatric Association) aligns with the strong model: "A mental disorder is a syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behavior *that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning*" (APA, 2013, p. 20, my italics; I assume "reflects" implies "is caused by"). However, the appeal to an underlying dysfunctional process is theoretical only – at best a placeholder with respect to classification and clinical practice. Differential diagnosis is made in line with the minimal model: via collections of personal-level signs and symptoms. For example, although DSM-5 states that "an underlying change in brain circuits" (APA, 2013, p. 483) is an important characteristic of substance use disorders (SUD), this change is not a component of the core construct, which involves persistent substance use combined with (1) cravings and failures to limit use as intended as (2) drugs occupy increasing time and attention at the expense of other pursuits and despite (3) negative consequences, including social impairment and physical and psychological health risks. In other words, SUD is diagnosed by a characteristic course of personal-level signs and symptoms: mental states, actions, and their outcomes. An underlying dysfunctional process is neither necessary nor sufficient for diagnosis.

Addiction is not unusual in this respect. Despite increasing understanding of neurobiological mechanisms and polygenetic background associated with mental disorder in general, there has been little progress identifying disorder-specific underlying pathologies that act as a common cause for disorder-specific personal-level signs and symptoms. This has led some researchers, notably Denny Borsboom and colleagues, to abandon the strong model in favor of an alternative, network model (Borsboom, 2017; Borsboom et al., 2018). Network models posit direct causal interactions between personal-level signs and symptoms. Initially, a sign or symptom may be activated by a cause external to the network.

Once activated, it can then activate other signs or symptoms within the network. However, the mere fact that a network is activated is not indicative of disorder. Rather, what makes a network pathological is that it remains active *even when* the precipitating cause is removed. To "suffer from a disorder" is "to be trapped in the stable state of a self-sustaining symptom network" (Borsboom et al., 2018, p. 4) – a state that has come to have a life of its own. Network models of mental disorder therefore offer an elegant elaboration of the minimal model of disease as a *syndrome*. They explain the presence and unfolding of disorder-specific personal-level signs and symptoms not by appeal to underlying brain pathology, but by appeal to self-sustaining *causal connections* between the signs and symptoms themselves.

DSM-5 lists the disorder-specific diagnostic criteria independently of each other. But as anyone familiar with any kind of mental disorder knows, they are typically interconnected. In formalizing the causal connections between personal-level signs and symptoms, network models reveal a disorder's *narrative schema*. They make psychological sense of the syndrome. By way of example, consider addiction. As Lee Hogarth has emphasized, self-reported use of drugs as a means of coping with negative affect is strongly predictive of addiction severity (Hogarth, 2020; Hogarth & Field, 2020; see also Chapter 21, this volume). For individuals who use drugs to cope, early adverse events and exposure to repeated trauma may be the original cause of a disposition to negative affect. However, once drugs have become a means of coping, any negative affect – regardless of cause – will trigger a desire to use, thereby increasing the probability of consumption. Consumption, in turn, causes or compounds the negative consequences of addiction, including social impairment and physical and psychological health risks. These negative consequences may themselves cause negative affect, both because they are distressing and because the knowledge that they are caused by one's drug consumption may be in addition a source of shame (Flanagan, 2013). Hence, in some individuals, there may be a self-sustaining, psychologically intelligible, causal network between desire or craving, consumption, negative consequences, and negative affect (Figure 29.3).

By contrast with a network model of addiction, the BDMA is not a minimal but a strong disease model: a case of what Nancy Andreason has famously called the "broken brain" model of mental disorder (Andreason, 1984).

According to the original version of the BDMA, the various SUD diagnostic criteria are compressed into a single, core psychological concept: compulsive drug seeking and use (cf. Leshner, 1997). As Nick Heather (2017) has shown, there is no one agreed meaning or neurobiological mechanism associated with compulsion within addiction science. Compulsion

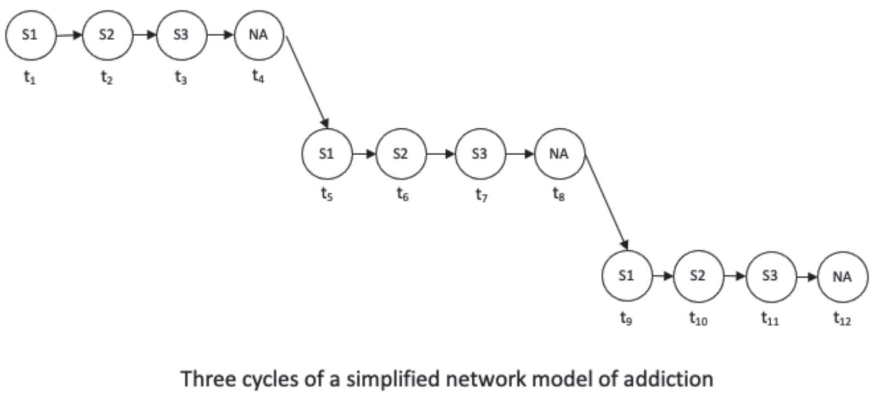

Figure 29.3 Three cycles of a simplified network model of addiction over a period of time t_1–t_{12}. S1 is desire or craving. S2 is drug consumption. S3 is negative consequences of drug consumption. NA is negative affect. Arrows represent causation.

has been identified with an array of automatic, habitual, motivational, and executive processes – some of which can work in tandem, while some cannot. Nonetheless, the explanatory function of an appeal to compulsion is clear: compulsion can explain why people continue to use drugs despite negative consequences that, given a basic folk psychological norm of rational self-interest, ought to disincentive use. This is what I have elsewhere called "the puzzle of addiction" that any theory of addiction must address: why do people keep using drugs, when the costs appear to outweigh benefits (Pickard, 2016, 2018, 2020a, 2020b; cf. Rachlin, 1997)? Compulsion offers a powerful and parsimonious answer to this puzzle. Simply put, the idea is that if people could stop using, they would. But they can't, which is why they don't. Once a person is addicted, drug seeking and use is no longer under voluntary control. According to the original version of the BDMA, this is because something is wrong with the brains of people who suffer from addiction. Underlying brain pathology *causes* what the model takes to be the core personal-level sign or symptom: compulsive drug seeking and use.

The implications of this picture are profound. Human movement can be passive or active: there are things that happen to us and things we do (cf., Dretske, 1988). Some (but perhaps not all) of the things we do count as actions. Actions are caused by our beliefs, desires, goals, and intentions. They are voluntary and subject to choice and control. The things that happen to us are not actions. They are effects on us of some cause that lies outside our beliefs, desires, goals, and intentions. Prior to becoming addicted, drug seeking and use is a type of action. According to the original version of the BDMA, once addicted, this is no longer the case. Drug seeking and use is an effect of whatever is wrong with a person's brain – a passive sign or symptom of underlying brain pathology, analogous, perhaps, to how signs and symptoms of Parkinson's disease, such as tremor and slow movement, are caused by brain degeneration. This picture of addiction transforms actions into occurrences, and agents into patients.

Although I believe we should be agnostic about whether addiction is a brain disease, we should not be agnostic about whether it is aptly characterized by compulsion thus understood. It is not; this picture is not empirically credible. On the one hand, experimental studies suggest that, on the whole, drug consumption in human addiction is goal directed as opposed to automatic or habitual (Hogarth, 2020; see also Chapter 21, this volume). On the other hand, there is ever-increasing evidence stemming from many distinct sources – personal narrative to epidemiology to behavioral economics to preclinical animal models to human studies to effective clinical treatment and more – that people (and animals) with addiction respond to incentives and so retain choice and a degree of control over their use (Banks & Negus, 2017; Field et al., 2020; Heather, 2017; Heyman, 2009; see also Pickard, 2018, 2020a, 2020b). The effectiveness of contingency management (CM) treatment provides a particularly striking example (Higgins et al., 2004; Zajac et al., 2018). Based on behavioral conditioning principles, CM offers alternative rewards contingent upon drug abstinence – including vouchers, prizes, and most recently and successfully, employment (Silverman et al., 2016). People with addiction are not subject to irresistible desires over which they are powerless; they are making voluntary and value-based (if puzzling) decisions (Regier & Redish, 2015). Hence, if addiction is a brain disease at all, it is a disease not of compulsion but of *choice*.

This transition from compulsion to choice is increasingly orthodox among addiction researchers who advocate for the BDMA. For example, in the article by Markus Heilig and colleagues cited previously, they conclude: "There is a freedom of choice, yet there is a shift of prevailing choices that nevertheless can kill . . . Addiction is a brain disease in which a person's choice faculties become profoundly compromised" (Heilig et al., 2021, p. 6; cf., Berridge, 2017; Redish et al., 2008; see also Berridge, Chapter 6, this volume). An empirically credible strong disease model hypothesizes underlying brain pathology as the cause, not of compulsive drug seeking and use, but of the pattern of *drug choices* characteristic of addiction – however compromised or distorted (cf., Berridge, 2017) these may prove to be (Figure 29.4).

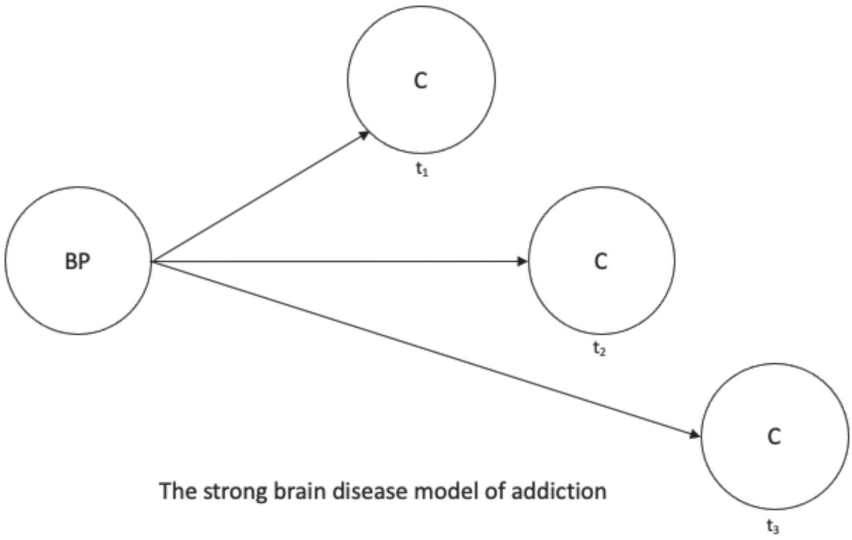

Figure 29.4 C is *either* compulsive drug seeking and use (original version) *or* a pattern of drug choices (new version) despite negative consequences over a period of time t_1–t_3. BP is brain pathology. Arrows represent causation.

Brains and pathology

Is the hypothesis that brain pathology causes the pattern of drug choices in addiction correct? To answer this question, it is essential to distinguish it from another – namely, whether brain science can illuminate this pattern. Brain science illuminates all human choice and behavior. Indeed, this is so whether or not it has a privileged place in explanation by comparison with other methods of study, including philosophy, psychology, the behavioral and social sciences, and the humanities. It is therefore not accurate that "viewing addiction as a brain disease simply states that neurobiology is an undeniable component of addiction" (Heilig et al., 2021, p. 5), on pain of the consequence that, since neurobiology is an undeniable component of all human choice and behavior, all human choice and behavior is a brain disease.

I labor this point because I believe it is common within addiction research to slide from the obviously true claim that brain science can illuminate addiction to the conclusion that addiction is a brain disease. This inference is unjustified even when scientific study of the brains of people who are addicted reveals either brain *changes* over time correlated with drug use, or brain *differences* between those who are and those who are not addicted. With respect to brain changes, as Marc Lewis (2015) and Maia Szalavitz (2016) have emphasized, all forms of learning (and, moreover, many kinds of psychiatric medication) produce brain changes; therefore, the fact that we find brain changes correlated with patterns of drug seeking and use (or, for that matter, with the ingestion of psychiatric medication) is not by itself evidence that these changes are pathological. With respect to brain differences, not only are they on the whole non-specific with respect to addiction and correlational as opposed to causal (Heilig et al., 2021), but it is a truism within philosophy of medicine that difference is not the same as disease; this point is foundational to theories of disability or disorder as forms of *diversity* (Barnes, 2016; Murphy, 2021). Statistical atypicality is neither necessary nor sufficient for pathology. Some pathology may be nearly universal relative to populations or times, e.g., atherosclerosis, tooth decay, infectious diseases during pandemics. Some atypicality is not pathological, e.g., athletic prowess, left-handedness, type O blood, low resting pulse rate, low levels of D2 receptor binding. This is so even if, as with resting pulse rate in relation to psychopathy (Raine, 2014),

or D2 receptor binding in relation to addiction (Martinez & Castillo, 2018), atypicality is associated with increased risk of disease; *risk* of disease is not the same as *disease* (Boorse, 1977). The question to ask, when we find some form of brain change or difference and wonder if it is a brain disease, is whether it should count as brain *dysfunction*. And to answer this question, we need an account of normal brain function.

We do not currently have an account of normal brain function. The theoretical and empirical challenges to developing one are surely not insurmountable, but they are nonetheless substantial. Three forms of challenge deserve mention here.

The first challenge is to provide an analysis of the concept of function as it applies to biology and related sciences, as opposed, e.g., to artifacts, which have a function by human design. Within philosophy, there are two classic analyses. The first is due to Larry Wright (1973, 1976). Functions are defined as effects that explain *why* something exists. Although this definition is highly general, it can be given an evolutionary interpretation (Millikan, 1984; Wakefield, 1992): the function of a part, mechanism, or trait is whatever effect explains its existence by contributing through natural selection to the survival and reproduction of the species over its evolutionary history. The second analysis is due to Robert Cummins (1975). Functions are defined as effects that explain *how* something contributes to specific capacities of the system containing it. Both analyses do scientific service, albeit with respect to different types of scientific explanation (Godfrey-Smith, 1993). To see this, consider the claim that the function of the myelin sheaths surrounding brains cells is to conduct signals. This may be an explanation of *why* the myelin is there. Or it may be an explanation of *how* the brain performs this task. Although the alignment with ordinary uses of *why* and *how* is not precise, I shall nonetheless term these *why-explanations* and *how-explanations*, respectively. Functions understood by appeal to why-explanations are an important component of evolutionary biology and psychology. Functions understood by appeal to how-explanations are an important component of neurobiology and cognitive science. However, these analyses can diverge; the function of a part, trait, or mechanism that explains why it exists may not be identical to its function within a system. For example, the explanation of why the visual cortex exists is, plausibly, to enable sight; but it is repurposed in congenitally blind people to serve higher-cognitive functions, including language and mathematical reasoning (Bedny, 2017). Is this dysfunctional? Yes, and no. Dysfunction depends on the analysis of function.

The second challenge is epistemological. With respect to either analysis, how do we avoid telling "just-so stories" (Kipling, 1902) and scientifically establish what something's normal function is? Just as pathological function is not simply atypicality, so too normal function is not simply typicality.

The third challenge is specific to providing an account of normal function with respect to the mind/brain. We can study the mind/brain from at least three levels of explanation: personal, cognitive, and neurobiological (Figure 29.5). The personal level is manifest in conscious experience and behavior, and captured in the folk psychological concepts, generalizations, expectations, and norms that we use to understand, explain, and predict our own and each other's mental states and actions. The cognitive level is computational (what problem is solved or what task is performed) and algorithmic (how the problem is solved or task is performed). To use the metaphor that has dominated our understanding for decades:

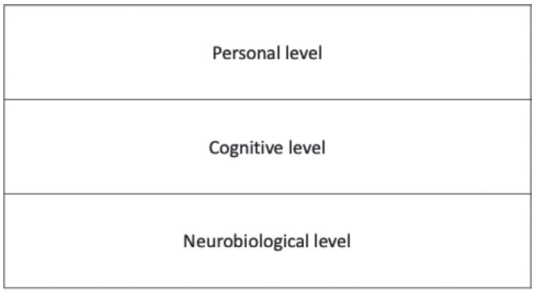

Figure 29.5 The personal level is folk psychological. The cognitive level is computational and algorithmic. The neurobiological level is the physical implementation of the system.

if the mind/brain is a computer, the cognitive level is the software. By contrast, the neurobiological level is the hardware: the physical system that implements the software. This characterization is crude, but sufficient to make the point of relevance here, which is that it is possible that the mind/brain is functioning normally at one level but not another. This means that, contra some advocates of the brain disease model of disorder (Andreason, 1984), we cannot slide between levels and infer that because there is something wrong at one level, there is also something wrong at another. Lower-level dysfunction may not manifest at the personal level; violations of folk psychological generalizations, expectations, and norms may occur in absence of lower-level dysfunction.

Jerry Wakefield (2018) offers a parable to illustrate this point. Imagine a gosling that has the misfortune of imprinting on a fox; the gosling follows the fox everywhere and is distressed at its absence, until it is, all too predictably, eaten. Nothing is wrong with the gosling's mind/brain at the neurobiological level; it is functioning as it should. But something arguably has gone wrong at the personal level. A gosling who chases after a fox is not *behaving* as goslings normally do: it violates the "folk" psychological and behavioral norms for geese.

Wakefield holds that something has also gone wrong at the cognitive level. This is because he takes the function of the imprinting mechanism to be the storage of a representation of the gosling's mother. However, he does not distinguish the computational and algorithmic dimensions of the cognitive level. We can see these dimensions as aligned, however inexactly, with functions understood by appeal to why-explanations and functions understood by appeal to how-explanations, respectively.

Consider first *why* the imprinting mechanism exists – the effect for which it was naturally selected over goose evolution. There is a well-known problem of indeterminacy in specifying functions thus understood (for a review, see Nicholas Shea [2018]). Wakefield simply assumes that the why-explanation function of the imprinting mechanism is to store a representation *of* the mother *because* the effect for which it has been naturally selected is a mother-directed pattern of behavior. But why specify this pattern in relation to the mother, as opposed to, e.g., a nurturing object? After all, it is only *because* mother geese are reliably nurturing that it would have been fitness-enhancing over goose evolution for goslings to imprint on them. This matters, because how the why-explanation function is specified affects what counts as dysfunction. If the function of the imprinting mechanism is to store a representation of *its mother*, then there is dysfunction when the gosling stores a representation of the fox. If it is to store a representation of *a nurturing object*, then there is dysfunction when it stores a representation of a fox (or indeed, a mother goose) that is non-nurturing, but not an animal – whatever the species – that is nurturing.

Consider next *how* the imprinting mechanism works. There are built-in parameters to what the mechanism takes as input: not only that the object imprinted upon is the first object perceived, but some specification, however minimal, of object size and movement. The fox fell within these parameters. Just as the gosling's mind/brain is functioning as it should at the neurobiological level, so, too, is it functioning as it should, considered in its algorithmic capacity; the imprinting mechanism received appropriate input and processed it correctly, leading to successful storage of a representation of an object – which just happened to be a non-nurturing fox. This is bad luck, not dysfunction.

Wakefield's claim that the gosling is dysfunctional at the cognitive level, therefore, confronts a dilemma. If we understand normal cognitive function by appeal to why-explanations, we must solve the problem of indeterminacy in a way which is scientifically credible, not based on "just-so stories". If we understand normal cognitive function by appeal to how-explanations, then it does not seem the gosling's imprinting mechanism is dysfunctional at all. To be sure, the gosling is not behaving normally; nor is its behavior conducive to its survival and reproduction. But this may be true and yet nothing is wrong with its brain – either at the neurobiological or at the cognitive level.

The moral of Wakefield's parable applied to addiction is *the mismatch model* (Nesse & Berridge, 1997; cf., Levy, 2013). The mismatch model explains addiction as a mismatch between the ancestral and current environment. The possibility of normal function at one level of explanation but dysfunction at another arises in part because natural selection occurs not in a vacuum but in a specific ancestral environment. We assume geese to have evolved in an environment where mothers are reliably nurturing and visible upon hatching. Otherwise, the imprinting mechanism would not exist in the form it does. But it takes some luck in the current environment for the mechanism to have the effect it was selected for in the ancestral environment; the conditions are not guaranteed.

Drug consumption is not new to our era; there is evidence of it since the start of civilization (Slingerland, 2021). Edward Hagen and Roger Sullivan (2018) point out that the majority of drugs are found in plants and, despite their potential toxicity, have many benefits, including not only medicinal properties – such as defending against pathogens and parasites – but also social ones, such as facilitating mating and bonding. They are therefore potentially fitness-enhancing natural rewards, akin to food and sex. Herbivores are known to ingest such plant toxins; Hagen and Sullivan's *neurotoxin regulation hypothesis* proposes we have similarly evolved to have a mechanism to reap the benefits of drugs while protecting against their potential toxicity, by careful regulation of intake, particularly during critical periods of individual or fetal brain development (Hagen et al., 2013).

Nonetheless, we have never before lived in an environment where drugs are so readily available, potent, and able to overcome natural toxin defenses, whether through, e.g., hypodermic injection or pills that are swallowed whole – both of which sidestep an evolved aversion to bitter taste, and the former of which in addition increases speed of delivery to the brain compared to oral ingestion. According to a mismatch model of addiction, even if nothing is wrong with the mind/brain at either a neurobiological or cognitive level, something certainly appears to be wrong at the personal level, as described by the core DSM-5 SUD construct understood as a syndrome. In continuing to use drugs despite negative consequences that ought to disincentive use, people with SUD are not behaving as we expect, for they violate the folk psychological norm of rational self-interest. The mismatch explanation for this violation is simply that, although we may have evolved to take drugs, we have not evolved to deal with drug availability, potency, and method of delivery, as found in our current environment. Hence, according to the mismatch model, the pattern of drug choices characteristic of addiction can be explained without recourse to brain pathology.

Note that a virtue of this model is the capacity to unify drug and behavioral addictions, which display a similar syndrome at the personal level despite the fact that behavioral addictions do not involve ingestion of substances that could directly cause brain pathology. For example, Don Ross (2020; see also Chapter 32, this volume) argues that gambling addiction is the result of a mismatch between our ancestral and current environment, which has been engineered to exploit normal learning mechanisms, e.g., high rates of responding to variable reinforcement schedules, to produce a pattern of behavior that makes the gambling industry money. We did not evolve in the company of slot machines (cf., Courtwright, 2019).

Advocates of the BDMA typically locate pathology at the neurobiological level. If there is indeed neurobiological dysfunction in individuals with addiction, and this dysfunction is the cause of their signs and symptoms, then addiction is a neurobiological brain disease, according to the strong disease model – but there is no reason to rule out the cognitive level from what is meant by "the brain". There can be bugs in software, just as there can be breakdowns in hardware (cf., Pickard, 2016; Segal, 2017). If there is indeed cognitive dysfunction in individuals with addiction, and this dysfunction is the cause of their signs and symptoms, then addiction is a cognitive brain disease, according to the strong disease model. The point is that the strong disease model postulates not only that the cause

of the signs and symptoms of addiction is found in the brain, *but that this cause is pathological*. However, we cannot infer underlying brain pathology merely from violations of personal-level generalizations, expectations, and norms such as, e.g., a diagnosis with SUD, as the syndrome is defined in DSM-5; brain pathology must be independently established. And for this, we need something we do not yet have: a theoretically and empirically well-evidenced account of the nature of normal (not just typical) neurobiological and/or cognitive brain function, by which to measure dysfunction.

"The cause"

Although the causal hypothesis from underlying brain pathology to personal-level signs and symptoms is essential to the strong disease model, it is worth noting that the picture emerging from the foregoing discussion suggests the radical possibility of *asymptomatic* addiction. Brain pathology could in theory be present in absence of *any and all* signs and symptoms of addiction (including, arguably, conscious craving and urges; see Berridge & Robinson, 1995). Less radically, it is also possible that brain pathology could be present yet not be *the cause* of the signs and symptoms. In which case, addiction would not be a brain disease, according to the strong model.

Despite the reservations of the last section, I assume that there are (at least) two stand-out candidates for pathology in the brains of those who are addicted: (1) brain damage caused by drug neurotoxicity, e.g., the extensive gray matter loss found in advanced alcohol use disorder; and (2) neuroadaptations in the mesocorticolimbic dopamine system caused by the direct effect of drugs on extracellular dopamine levels, e.g., excessive neural sensitization or hyper-reactivity to drug cues (Berridge, 2017; Robinson & Berridge, 1993; Samaha et al., 2021; see also Chapter 6, this volume), potentially in combination with blunted striatal activations or hypo-reactivity to non-drug reward cues (Ihssen et al., 2011; Leyton & Vezina, 2013, 2014; cf., Rachlin, 1997). Although some theorists have dismissed brain damage as irrelevant to establishing the BDMA (Ross, 2020; Watson, 1999), I see no reason for this view *if* the damage is indeed the cause of the signs and symptoms of addiction: surely brain damage is brain pathology if anything is. With respect to neuroadaptations, the crucial questions are how excessive the hyper-reactivity to drug cues and hypo-reactivity to non-drug alternative reward cues is, and whether there is reason to view this excess as crossing a threshold from atypical function into dysfunction. I take no view on the threshold question here; but more research is needed to establish the extent of hyper- and hypo-reactivity. Increased neural activation to reward cues is found across drug and behavioral addictions (Noori et al., 2016; Starcke et al., 2018). There are functional magnetic resonance imaging (fMRI) studies comparing response to cocaine, sexual, neutral, and aversive backward-masked cues in cocaine-addicted subjects (Childress et al., 2008); response to choice of drug, food, and neutral images in active and abstinent cocaine-addicted subjects alongside non-users (Moeller et al., 2018a); response to alcohol and neutral images in light and heavy drinkers (Vollstädt-Klein et al., 2010); response to alcohol, food, neutral, and aversive images in light and heavy drinkers (Ihssen et al., 2011); and response to alcohol images and anticipated monetary reward in healthy controls and abstinent alcohol-addicted subjects (Wrase et al., 2007). There are also positron emission tomography (PET) studies comparing dopamine response to cocaine-cue and neutral videos in cocaine-addicted subjects (Volkow et al., 2006); and dopamine response to choice of drug, pleasant, unpleasant, and aversive images in methamphetamine-addicted subjects and non-users (Moeller et al., 2018b). But there are not yet, to my knowledge, PET studies either directly comparing dopamine response to drug cues versus dopamine response to meaningful non-drug reward cues in individuals with drug addiction, or directly comparing dopamine response to drug cues in individuals with drug addiction to dopamine response to meaningful non-drug reward cues in non-users. Nonetheless, on the assumption that brain damage and/or neuroadaptations constitute brain pathology, the

question is whether they cause the pattern of drug choices characteristic of addiction.

The answer depends in part on what we mean by and how we operationalize "the cause". Within philosophy, there are various, competing analyses of causation, including regularity theories, counterfactual theories, structural equation accounts, contrastive accounts, and more – some of which return different verdicts on test cases, such that what counts as the cause of an effect according to one analysis will not so count according to another (Paul & Hall, 2013). Nonetheless, within experimental science, there is a relatively standard method used to establish causation: intervention. Causes are difference-makers. Suppose we know that there is a correlation between two variables, and we hypothesize that one is the cause of the other. To test this hypothesis, we can design an experiment that, holding all else equal, manipulates the variable hypothesized to be the cause; we then observe whether this manipulation has an effect on the other variable. If so, that is prima facie evidence for the causal hypothesis. If not, that is prima facie evidence against it. Effect size can be taken to indicate causal significance. Some "natural experiments" have a similar structure. Interventionist models of causation use this process to provide a non-reductive analysis of causation. Given that a range of background conditions are met (e.g., the variables are independent, there is no common cause of both variables, the intervention does not influence the variable hypothesized as effect otherwise than by influencing the variable hypothesized as cause), something is a cause of an effect if and only if there is a possible intervention on it that changes the effect (Woodward, 2003).

Suppose we ask whether brain damage and/or neuroadaptations cause the signs and symptoms of addiction. The simplest way to test this hypothesis would be to take two sample groups matched for brain pathology, and, holding all else equal, fix the brains of one group while leaving the other unchanged and observe the effect.

We cannot run this experiment because we cannot just "fix" the brain. However, we can nonetheless test whether *other* variables cause the pattern of drug choices characteristic of addiction. Ken Kendler (2012) argues that alcohol use disorder does not fit a strong disease model because of the multiplicity of causes, none of which is privileged, that occur across many levels of explanation, including, e.g., genetic risk, molecular genetic variants, executive deficits, personality traits, drinking expectancies and reasons for drinking, childhood sexual abuse, peer substance use, cultural norms, and alcohol availability and cost. To this list we might add further specific factors, including, but by no means limited to: self-reported drug consumption to cope with negative affect (as mentioned previously; Hogarth, 2020; Hogarth & Field, 2020; see also Chapter 21, this volume), diminished insight and self-awareness (Goldstein et al., 2009), denial (Pickard, 2016), an "addict" self-identity (Pickard, 2020a), and, importantly, lack of alternative reinforcers (Banks & Negus, 2017). There is a robust, inverse relationship between drug use and drug-free reinforcement (Acuff et al., 2019; see also Chapter 39, this volume). Forced choice studies with rodents (Ahmed, 2010; Ahmed et al., 2013) and community reinforcement (CR) and contingency management (CM) treatment with humans (Silverman et al., 2016; Stitzer et al., 2011; Zajac et al., 2018) complement this finding to converge on the same hypothesis. If we manipulate the environment so that it supports a richer array of goods, many individuals with addiction will forego drugs. Indeed, a recent rodent study found that almost 100% chose social over drug reward irrespective of sex, drug class, training conditions, size of dose, length of abstinence since last dose and "addiction score" based on a DSM-style model adapted to rodents (Fredriksson et al., 2021; Venniro et al., 2018). This suggests that lack of alternative reinforcers is a cause of addiction.

Advocates of the BDMA may respond to these findings by pointing out that although the effect is large, it is not without exception. Environmental determinants, such as relative magnitude, effort or cost, and delay of delivery, influence choice for either drug rewards or alternative reinforcers in

preclinical and human studies (Banks & Negus, 2017; Venniro et al. 2021). Nonetheless, a small minority of animals continue to choose drugs in contexts where, controlling for such determinants, the majority switch to alternative reinforcers (Ahmed, 2010). Similarly, approximately 10–20% of humans do not spontaneously recover (Heyman, 2009); and CR and CM do not work for everyone (Venniro et al., 2020). Perhaps, then, brain pathology explains why some individuals do not forego drugs for alternative reinforcers in contexts where the majority do (Ahmed, 2010; Heather, 2017; Lüscher et al., 2020; Segal, 2017).

In theory, if not in practice, we could experimentally test this hypothesis. Take two sample groups of individuals with addiction matched for presence of brain pathology and lack of alternative reinforcers, and one control group. Fix the brains of one sample group while leaving their environment unchanged; fix the environment for the other sample group while leaving their brains unchanged; compare to the baseline provided by the control group and observe which intervention has the greater effect on the pattern of drug choices. This, of course, is a simplified example of the kind of complicated experiment routinely designed by scientists to determine causation under real-world constraints.

I do not know what we would find if we could do this experiment, but we should be open to the possibility that, tested against alternative reinforcers (and/or other possible variables), brain damage and/or neuroadaptations could be more significant in some cases, in line with a strong disease model. Would this vindicate the BDMA? On the assumption that brain damage and/or neuroadaptations constitute pathology, the answer is yes, *but only in those cases*. This caveat is important: as David Epstein (2020) argues, we should not assume that addiction is homogenous rather than heterogenous (cf., Pickard, 2018, 2020a, 2020b). Instead of a "winner-takes-all" approach to theory building, we should care about determining the relative size of the effects of *all* possible causes for *all* individuals who suffer. As Epstein remarks, "the questions are when, for whom, and to what extent" (Epstein, 2020, p. 2). I agree, but I would add to these the question of allocation of resources. Addiction neuroscience and the search for pharmacological treatment receive significant funding and attention compared with psychological, behavioral, social, and economic interventions – which we know help many people recover. Given that the existence and/or causal significance of brain pathology relative to other possible causes of drug choices in addiction has not yet been fully demonstrated, not only should we be agnostic about the BDMA and embrace the possibility of heterogeneity in the causation and explanation of addiction, but alternative areas of research and interventions should be given their due (Pickard, 2020b; Room, 2021).

Alternative concepts of disease and the fight against stigma

The BDMA is a strong disease model. However, there are alternative conceptions of disease (Murphy, 2021). Some theorists analyze disease by appeal to *values*, as any mental or physical condition that adversely affects wellbeing (e.g. Goosens, 1980; cf., Boorse, 1977); others locate it in relation to *medical practice*, as any mental or physical condition treated by doctors and allied health care professionals (e.g., Engelhardt, 1975; cf., Boorse, 1977; Kendell, 1975). These conceptions are far from ordinary usage, as they entail that conditions as diverse as aging, ugliness, hunger, obesity, pregnancy, and natural suffering – to name but a few – all count as disease. But ordinary usage need not be the final word. Don Ross (2020) argues that disease is a context-sensitive, unstructured concept that warrants a broad-church approach to its analysis (cf., Heilig et al., 2021), notwithstanding the tendencies of philosophers to want to serve as "conceptual janitors" (Ross, 2020, p. 4). Nonetheless, he suggests that if any discipline can lay claim to the concept, it is epidemiology, which uses it to delineate conditions that undermine public health and so warrant public health interventions.

I have no interest in being a conceptual janitor. If we stipulate that, by "disease", we simply mean a mental or physical condition that adversely affects wellbeing and is part of medical practice and public health, then addiction is a disease thus defined. But in so doing, we should not lose sight of the following three points.

First, a theoretical point: conceptual stipulation does not answer any questions about what causes the pattern of drug choices characteristic of addiction and whether or not some of these causes some of the time involve brain pathology.

Second, a point about the need to provide medical treatment and support public health initiatives, as recognized by Leshner (1997): medicine and public health aim to alleviate suffering and promote health and wellbeing. They therefore target not only disease, however defined, but also injury, illness, disability, and disorder, as well as physical, psychological, social, and economic determinants of these, and entirely natural, biological processes, such as, e.g., stress, sex, pregnancy, and aging, where individuals benefit from treatment and/or various forms of support and intervention. As treatment, support, and interventions are in fact offered for a host of conditions that are not diseases, it is equally appropriate to offer them for addiction, no matter its disease status. This includes access to pharmacotherapy, such as, e.g., nicotine patches for cigarette addiction, and methadone, buprenorphine, and injectable heroin in refractory cases (Strang et al., 2015) for opioid addiction. Treatments that target the brain can be effective irrespective of whether addiction is or is not a brain disease (Papineau, 1994).

Third, a point about the moral atrocity of state punishment as a strategy for regulating individual drug consumption, also raised by Leshner (1997): the most powerful arguments for the decriminalization of drug possession – which functions as a proxy for drug consumption – in no way depend on the disease status of addiction; indeed, they are equally forceful in relation to *all* drug use, addicted or not. The War on Drugs has failed as a regulative strategy, at the expense of education, public health, and alternative regulative strategies, while imposing terrible costs on individuals and families (Earp et al., 2021). It has also weaponized the criminal justice system against Black and ethnic minority communities (Alexander, 2010; Hart, 2021). In addition, there is a foundational legal argument for *decriminalization*, which is that the current status quo of the *criminalization* of drug possession violates a core principle of criminal justice (Husak, 2002). Criminalization is the state's most coercive regulatory tool, and it is widely considered to be legitimate only with respect to behavior that is morally wrong and constitutes a serious harm to others. But there is nothing *intrinsically* morally wrong or harmful to others about using drugs (Husak, 2002; Pickard, 2020a, 2020b). Decriminalization is therefore necessary to rectify the injustices of the criminalization and punishment of drug possession perpetrated by the state against all people who use drugs – regardless of the disease status of addiction.

Many have put faith in the BDMA as an antidote to stigma. At bottom, the idea guiding this faith is that, if addiction is a brain disease, then people who are addicted cannot help using drugs, and so are not responsible for doing so. One problem with this idea is that it is less compelling once it is acknowledged that, if addiction is a disease at all, then it is a disease whose signs and symptoms involve choice. But a deeper problem is that it does not reject, but rather accepts, moralism about drugs, thereby buying into addiction stigma. For the idea assumes that people who use drugs and struggle with addiction need an excuse – but for what, exactly? We need excuses when otherwise we would be responsible for doing something morally wrong. By contrast, if we reject moralism about drugs, then no one who struggles with addiction needs an excuse in the first instance, for people are not doing anything morally wrong simply in virtue of using drugs – whether they are addicted or not. My best guess is not only that addiction is at most sometimes a brain disease. It is also that the best way to fight addiction stigma is to stop resting our hopes on the brain disease label and instead to fight moralism about drugs and moralistic drug policies directly.

Acknowledgments

I would like to thank Peter Achinstein, Serge H. Ahmed, Marina Bedny, Kent C. Berridge, David Epstein, Matt Field, Nick Heather, Lee Hogarth, Diana Martinez, Laurie Paul, Ian Phillips, and especially Yavin Shaham, for very helpful comments and discussion.

References

Acuff, S. F., Dennhardt, A. A., Correia, C. J., & Murphy, J. G. (2019). Measurement of substance-free reinforcement in addiction: A systematic review. *Clinical Psychology Review*, *70*, 79–90. https://doi.org/10.1016/j.cpr.2019.04.003

Ahmed, S. H. (2010). Validation crisis in animal models of drug addiction: Beyond non-disordered drug use toward drug addiction. *Neuroscience & Biobehavioral Reviews*, *35*(2), 172–184. https://doi.org/10.1016/j.neubiorev.2010.04.005

Ahmed, S. H., Lenoir, M., & Guillem, C. (2013). Neurobiology of addiction versus drug use driven by lack of choice. *Current Opinion in Neurobiology*, *23*(4), 581–587. https;//doi.org/10.1016/j.conb.2013.01.028

Alexander, M. (2010). *The new Jim Crow: Mass incarceration in the age of colorblindness*. The New Press.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Author.

Andreason, N. C. (1984). *The broken brain: The biological revolution in psychiatry*. Harper & Row Publishers.

Bandura, A. (1997). *Social learning theory*. Prentice-Hall.

Banks, M. L., & Negus, S. S. (2017). Insights from preclinical choice models on treating drug addiction. *Trends in Pharmacological Sciences*, *38*(2), 181–194. https://doi.org/10.1016/j.tips.2016.11.002

Barnes, E. (2016). *The minority body*. Oxford University Press.

Barry, C. L., McGinty, E. E., Pescosolido, B. A., & Goldman, H. H. (2014). Stigma, discrimination, treatment effectiveness, and policy: Public views about drug addiction and mental illness. *Psychiatric Services (Washington, D.C.)*, *65*(10), 1269–1272. https://doi.org/10.1176/appi.ps.201400140

Bedny, M. (2017). Evidence from blindness for a cognitively pluripotent cortex. *Trends in Cognitive Sciences*, *21*(9), 637–648. https://doi.org/10.1016/j.tics.2017.06.003

Berridge, K. C. (2017). Is addiction a brain disease? *Neuroethics*, *10*, 29–33. https://doi.org/10.1007/s12152-016-9286-3

Berridge, K. C., & Robinson, T. E. (1995). The mind of an addicted brain: Neural sensitization of wanting versus liking. *Current Directions in Psychological Science, 4*(3), 71–76. https://doi.org/10.1111/1467-8721.ep10772316

Boorse, C. (1977). Health as a theoretical concept. *Philosophy of Science*, *44*(4), 542–573. https://doi.org/10.1086/288768

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13. https://doi.org/10.1002/wps.20375

Borsboom, D., Cramer, A., & Kalis, A. (2018). Brain disorders? Not really . . . Why network structures block reductionism in psychopathology research. *Behavioral & Brain Sciences*, 1–54. https://doi.org/10.1017/S0140525X17002266

Childress, A., Ehrman, R., Wang, Z., Li, Y., Sciortino, N., Hakun, J., Jens, W., Suh, J., Listerud, J., Marquez, K., Franklin, T., Langleben, D., Detre, J., & O'Brien, C. (2008). Prelude to passion: Limbic activation by "unseen" drug and sexual cues. *PLoS ONE*, *3*(1), e1506. http://www.f1000biology.com/article/id/1101048

Courtwright, D. T. (2010). The NIDA brain disease paradigm: History, resistance and spinoffs. *BioSocieties*, *5*, 137–147. https://doi.org/10.1057/biosoc.2009

Courtwright, D. T. (2019). *The age of addiction: How bad habits became big business*. The Belknap Press.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy*, *72*, 741–765. https://doi.org/10.2307/2024640

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. MIT Press.

Earp, B. D., Lewis, J., Hart, C. L., & with Bioethicists and Allied Professionals for Drug Policy Reform. (2021). Racial justice requires ending the War on Drugs. *American Journal of Bioethics*, 1–29. Advance online publication. https://doi.org/10.1080/15265161.2020.1861364

Engelhardt, H. T., Jr. (1975). The concepts of health and disease. In H. T. Engelhardt & S. F. Spicker (Eds.), *Evaluation and explanation in the biomedical sciences* (pp. 125–142). Reidel.

Epstein, D. H. (2020). Let's agree to agree: A comment on Hogarth (2020), with a plea for not-so-competing theories of addiction. *Neuropsychopharmacology*, *45*, 715–716. https://doi.org/10.1038/s41386-020-0618-y

Field, M., Heather, N., Murphy, J. G., Stafford, T., Tucker, J. A., & Witkiewitz, K. (2020). Recovery from addiction: Behavioral economics and value-based decision making. *Psychology of Addictive Behaviors*, *34*(1), 182–193. https://doi.org/10.1037/adb0000518

Flanagan, O. (2013). The shame of addiction. *Frontiers in Psychiatry*, *4*, 120. https://doi.org/10.3389/fpsyt.2013.00120

Fredriksson, I., Venniro, M., Reiner, D. J., Chow, J. J., Bossert, J. M., & Shaham, Y. (2021). Animal models of drug relapse and craving after voluntary abstinence: A review. *Pharmacological Review*, *73*(3), 1050–1083. https://doi.org/10.1124/pharmrev.120.000191

Godfrey-Smith, P. (1993). Functions: Consensus without unity. *Pacific Philosophical Quarterly*, *73*(3), 196–208. https://doi.org/10.1111/j.1468-0114.1993.tb00358.x

Goldstein, R. Z., Craig, A. D., Bechara, A., Garavan, H., Childress, A. R., Paulus, M. P., & Volkow, N. D. (2009). The neurocircuitry of impaired insight in drug addiction. *Trends in Cognitive Sciences*, *13*(9), 372–380. https://doi.org/10.1016/j.tics.2009.06.004

Goosens, W. (1980). Values, health, and medicine. *Philosophy of Science*, *47*(1), 100–115. https://doi.org/10.1086/288912

Hagen, E. H., Roulette, C. J., & Sullivan, R. J. (2013). Explaining human recreational use of 'pesticides': The neurotoxin regulation model of substance use vs. the hijack model and implications for age and sex differences in drug consumption. *Frontiers in Psychiatry*, *4*, 142. https://doi.org/10.3389/fpsyt.2013.00142

Hagen, E. H., & Sullivan, R. (2018). The evolutionary significance of drug toxicity over reward. In H. Pickard & S. H. Ahmed (Eds.), *The philosophy and science of addiction* (pp. 102–109). Routledge.

Hart, C. L. (2018). Don't be fooled by the euphemistic language attesting to a gentler War on Drugs. In H. Pickard & S. H. Ahmed (Eds.), *The philosophy and science of addiction* (pp. 511–517). Routledge.

Hart, C. L. (2021). *Drug use for grown-ups.* Penguin Press.

Haslam, N., & Kvaale, E. P. (2015). Biogenetic explanations of mental disorder: The mixed-blessings model. *Current Directions in Psychological Science*, *24*(5), 399–404. https://doi.org/10.1177/0963721415588082

Heather, N. (2017). Is the concept of compulsion useful in the explanation or description of addictive behaviour and experience? *Addictive Behaviors Reports*, *6*, 15–38. https://doi.org/10.1016/j.abrep.2017.05.002

Heilig, M., MacKillop, J., Martinez, D., Rehm, J., Leggio, L., & Louk, J. M. J. (2021). Addiction as a brain disease revised: Why it still matters, and the need for consilience. *Neuropsychopharmacology*, *46*, 1715–1723. https://doi.org/10.1038/s41386-020-00950-y

Heyman, G. M. (2009). *Addiction: A disorder of choice.* Harvard University Press.

Higgins, S. T., Heil, S. H., & Lussier, J. P. (2004). Clinical implications of reinforcement as a determinant of substance use disorders. *Annual Review of Psychology*, *55*, 443–461.

Hogarth, L. (2020). Addiction is driven by excessive goal-directed drug choice under negative affect: Translational critique of habit and compulsion theory. *Neuropsychopharmacology*, *45*, 720–735. https://doi.org/10.1038/s41386-020-0600-8

Hogarth, L., & Field, M. (2020). Relative expected value of drugs versus competing rewards underpins vulnerability to and recovery from addiction. *Behavioural Brain Research*, *394*, 112815. https://doi.org/10.1016/j.bbr.2020.112815

Husak, D. (2002). *Legalize this! The case for decriminalizing drugs.* Verso.

Ihssen, N., Cox, W. M., Wiggett, A., Fadardi, J. S., & Linden, D. E. (2011). Differentiating heavy from light drinkers by neural responses to visual alcohol cues and other motivational stimuli. *Cerebral cortex (New York, N.Y: 1991)*, *21*(6), 1408–1415. https://doi.org/10.1093/cercor/bhq220

Kelly, J. F., Greene, M. C., & Abry, A. (2020). A US national randomized study to guide how best to reduce stigma when describing drug-related impairment in practice and policy. *Addiction, 116,* 1757–1767. doi:10.1111/add.15333

Kendell, R. E. (1975).The concept of disease and its implications for psychiatry. *British Journal of Psychiatry*, *127*, 305–315.

Kendler, K. S. (2012). Levels of explanation in psychiatric and substance use disorders: Implications for the development of an etiologically based nosology. *Molecular Psychiatry*, *17*(1), 11–21. https://doi.org/10.1038/mp.2011.70

Kennedy-Hendricks, A., Barry, C. L., Gollust, S. E., Ensminger, M. E., Chisolm, M. S., & McGinty, E. E. (2017). Social stigma toward persons with prescription opioid use disorder: Associations with public support for punitive and public health-oriented policies. *Psychiatric Services (Washington, D.C.)*, *68*(5), 462–469. https://doi.org/10.1176/appi.ps.201600056

Kennedy-Hendricks, A., McGinty, E. E., & Barry, C. L. (2016). Effects of competing narratives on public perceptions of opioid pain reliever addiction during pregnancy. *Journal of Health Politics, Policy and Law*, *41*(5), 873–916. https://doi.org/10.1215/03616878-3632230

Kipling, R. (1902). *Just so stories for little children*. Macmillan.

Kraepelin, E. (1899 [1990]). *Psychiatry: A textbook for students and physicians* (6th ed., reprinted). Science History Publications.

Leshner, A. I. (1997). Addiction is a brain disease, and it matters. *Science*, *278*(5335), 45–47. https://doi.org/10.1126/science.278.5335.45

Levy, N. (2013). Addiction is not a brain disease (and it matters). *Frontiers in Psychiatry*, *4*, 24. https://doi.org/10.3389/fpsyt.2013.00024

Lewis, M. (2015). *The biology of desire: Why addiction is not a disease.* Public Affairs.

Leyton, M., & Vezina, P. (2013). Striatal ups and downs: their roles in vulnerability to addictions in humans. *Neuroscience and Biobehavioral Reviews*, *37*(9 Pt A), 1999–2014. https://doi.org/10.1016/j.neubiorev.2013.01.018

Leyton, M., & Vezina, P. (2014). Dopamine ups and downs in vulnerability to addictions: a neurodevelopmental model. *Trends in Pharmacological Sciences*, *35*(6), 268–276. https://doi.org/10.1016/j.tips.2014.04.002

Lüscher, C., Robbins, T. W., & Everitt, B. J. (2020). The transition to compulsion in addiction. *Nature Review Neuroscience*, *21*, 247–263. https://doi.org/10.1038/s41583-020-0289-z

Marlatt, G. A., & Gordon, J. R. (Eds.) (1985). *Relapse prevention.* Guilford.

Martinez, D., & Castillo, F. (2018). Imaging dopamine signalling in addiction. In H. Pickard & S. H. Ahmed (Eds.), *The Routledge handbook of philosophy and science of addiction* (pp. 380–391). Routledge.

McGinty, E. E., & Barry, C. L. (2020). Stigma reduction to combat the addiction crisis – Developing an evidence base. *New England Journal of Medicine*, *382*(14), 1291–1292. https://doi.org/10.1056/NEJMp2000227

McGinty, E. E., Barry, C. L., Stone, E. M., Niederdeppe, J., Kennedy-Hendricks, A., Linden, S., & Sherman, S. G. (2018). Public support for safe consumption sites and syringe services programs to combat the opioid

epidemic. *Preventive Medicine*, *111*, 73–77. https://doi.org/10.1016/j.ypmed.2018.02.026

McGinty, E. E., Pescosolido, B., Kennedy-Hendricks, A., & Barry, C. L. (2018). Communication strategies to counter stigma and improve mental illness and substance use disorder policy. *Psychiatric Services (Washington, D.C.)*, *69*(2), 136–146. https://doi.org/10.1176/appi.ps.201700076

McGinty, E. E., Stone, E. M., Kennedy-Hendricks, A., & Barry, C. L. (2019). Stigmatizing language in news media coverage of the opioid epidemic: Implications for public health. *Preventive Medicine*, *124*, 110–114. https://doi.org/10.1016/j.ypmed.2019.03.018

Mehta, S., & Farina, A. (1997). Is being "sick" really better? Effect of the disease view of mental disorder on stigma. *Journal of Social & Clinical Psychology*, *16*(4), 405–419. https://doi.org/10.1521/jscp.1997.16.4.405

Miller, W. R., Westerberg, V. S., Harris, R. J., & Tonigan, J. S. (1996). What predicts relapse? Prospective testing of antecedent models. *Addiction*, *91*(suppl.), S155–S171. https://doi.org/10.1046/j.1360-0443.91.12s1.7.x

Millikan, R. G. (1984). *Language, thought, and other biological categories*. MIT Press.

Moeller, S. J., Okita, K., Robertson, C. L., Ballard, M. E., Konova, A. B., Goldstein, R. Z., Mandelkern, M. A., & London, E. D. (2018b). Low striatal dopamine D2-type receptor availability is linked to simulated drug choice in methamphetamine users. *Neuropsychopharmacology*, *43*(4), 751–760. https://doi.org/10.1038/npp.2017.138

Moeller, S. J., Zilverstand, A., Konova, A. B., Kundu, P., Parvaz, M. A., Preston-Campbell, R., Bachi, K., Alia-Klein, N., & Goldstein, R. Z. (2018a). Neural correlates of drug-biased choice in currently-using and abstinent individuals with cocaine use disorder. *Biological Psychiatry Cognitive Neuroscience Neuroimaging*, *3*(5), 485–494. https://doi.org/10.1016/j.bpsc.2017.11.001

Morris, J., Albery, I. P., Moss, A. C., & Heather, N. (2021). Promoting problem recognition amongst harmful drinkers: A conceptual model for problem framing factors. In D. Frings & I. P. Albery (Eds.), *The handbook of alcohol use: From synapse to society* (pp. 221–226). Academic Press.

Murphy, D. (2009). Psychiatry and the concept of disease as pathology. In M. Broome & L. Bortolotti (Eds.), *Psychiatry as cognitive neuroscience* (pp. 103–117). Oxford University Press.

Murphy, D. (2021). Concepts of disease and health. In Edward N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2021 ed.). Stanford University. URL = https://plato.stanford.edu/archives/spr2021/entries/health-disease

Nesse, R. M., & Berridge, K. C. (1997). Psychoactive drug use in evolutionary perspective. *Science*, *278*(5335), 63–66. https://doi.org/10.1126/science.278.5335.63

Noori, H. R., Linan, A. C., & Spanagel, R. (2016). Largely overlapping neuronal substrates of reactivity to drug, gambling, food and sexual cues: A comprehensive meta-analysis. *European Neuropsychopharmacology*, *26*, 1419–1430. http://doi.org/10.1016/j.euroneuro.2016.06.013

Papineau, D. (1994). Mental disorder, illness and biological dysfunction. In A. Phillips Griffiths (Ed.), *Philosophy, Psychology, and Psychiatry* (pp. 73–82). Cambridge University Press.

Parsons, T. (1951). *The social system*. Free Press.

Paul, L. A., & Hall, N. (2013). *Causation: A user's guide*. Oxford University Press.

Peele, S. (1985). *The meaning of addiction: Compulsive experience and its interpretation*. DC Heath.

Pescosolido, B. A., Martin, J. K., Long, J. S., Medina, T. R., Phelan, J. C., & Link, B. G. (2010). "A disease like any other"? A decade of change in public reactions to schizophrenia, depression, and alcohol dependence. *American Journal of Psychiatry*, *167*(11), 1321–1330. https://doi.org/10.1176/appi.ajp.2010.09121743

Pickard, H. (2016). Denial in addiction. *Mind & Language*, *31*(3), 277–299. https://doi.org/10.1111/mila.12106

Pickard, H. (2017). Responsibility without blame for addiction. *Neuroethics*, *10*, 169–180. https://doi.org/10.1007/s12152-016-9295-2

Pickard, H. (2018). The puzzle of addiction. In H. Pickard & S. H. Ahmed (Eds.), *The Routledge handbook of philosophy and science of addiction* (pp. 9–22). Routledge.

Pickard, H. (2020a). Addiction and the self. *Noûs*, 1– 25. https://doi.org/10.1111/nous.12328

Pickard, H. (2020b). What we're not talking about when we talk about addiction. *Hastings Center Report*, *50*(4), 37–46. https://doi.org/10.1002/hast.1172

Rachlin, H. (1997). Four teleological theories of addiction. *Psychonomic Bulletin & Review*, *4*(4), 462–473. https://doi.org/10.3758/BF03214335

Raine, A. (2014). *The Anatomy of violence*. Vintage Books.

Redish, A. D., Jensen, S., & Johnson, A. (2008). A unified framework for addiction: Vulnerabilities in the decision process. *Behavioral & Brain Sciences*, *31*(4), 415–487. https://doi.org/10.1017/S0140525X0800472X

Regier, P. S., & Redish, A. D. (2015). Contingency management and deliberative decision-making processes. *Frontiers in Psychiatry*, *6*, 76. https://doi.org/10.3389/fpsyt.2015.00076

Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving: An incentive-sensitization theory of addiction. *Brain Research Reviews*, *18*(3), 247–291. https://doi.org/10.1016/0165-0173(93)90013-p

Room, R. (2021). No level has primacy in what is called addiction: "Addiction is a social disease" would be just as tenable. *Neuropsychopharmacology*, 1. https://doi.org/10.1038/s41386-021-01015-4

Ross, D. (2020). Addiction is socially engineered exploitation of natural biological vulnerability. *Behavioural Brain Research*, *386*, 112598. https://doi.org/10.1016/j.bbr.2020.112598

Samaha, A.-N., Khoo, S. Y. -S., Ferrario, C. R., & Robinson, T. E. (2021). Dopamine 'ups and downs' in

addiction revisited. *Trends in Neurosciences, 44*(7), 516–526. https://doi.org/10.1016/j.tins.2021.03.003

Satel, S., & Lilienfeld, S. O. (2014). Addiction and the brain-disease fallacy. *Frontiers in Psychiatry*, *4*, 141. https://doi.org/10.3389/fpsyt.2013.00141

Segal, G. (2017). Ambiguous terms and false dichotomies. In N. Heather & G. Segal (Eds.), *Addiction and choice: Rethinking the relationship* (pp. 449–462). Oxford University Press.

Shea, N. (2018). *Representation in cognitive science*. Oxford University Press.

Silverman, K., Holtyn, R., & Morrison, R. (2016). The therapeutic utility of employment in treating drug addiction: Science to application. *Translational Issues in Psychological Science*, *2*(2), 203–212. https://doi.org/10.1037/tps0000061

Slingerland, E. (2021). *Drunk: How we sipped, danced, and stumbled our way to civilization*. Little, Brown Spark.

Starcke, K., Antonis, S., Trotzke, P., & Brand, M. (2018). Cue-reactivity in behavioral addictions: A meta-analysis and methodological considersations. *Journal of Behavioral Addictions*, *7*(2), 227–238. https://doi.org/10.1556/2006.7.2018.39

Stitzer, M. L., Jones, H. E., Tuten, M., & Wong, C. (2011). Community reinforcement approach and contingency management interventions for substance abuse. In W. M. Cox & E. Klinger (Eds.), *Handbook of motivational counseling: Goal-based approaches to assessment and intervention with addiction and other problems* (pp. 549–569). John Wiley & Sons. https://doi.org/10.1002/9780470979952.ch23

Strang, J., Groshkova, T., Uchtenhagen, A., van den Brink, W., Haasen, C., Schechter, M., Lintzeris, N., Bell, J., Pirona, A., Oviedo-Joekes, E., Simon, R., & Metrebian, N. (2015). Heroin on trial: Systematic review and meta-analysis of randomised trials of diamorphine-prescribing as treatment for refractory heroin addiction. *British Journal of Psychiatry*, *207*(1), 5–14. https://doi.org/10.1192/bjp.bp.114.149195

Szalavitz, M. (2016). *Unbroken brain*. New Picador Press.

Venniro, M., Banks, M. L., Heilig, M., Epstein, D. H., & Shaham, Y. (2020). Improving translation of animal models of addiction and relapse by reverse translation. *Nature Reviews Neuroscience*, *11*, 625–643. https://doi.org/10.1038/s41583-020-0378-z

Venniro, M., Zhang, M., Caprioli, D., Hoots, J. K., Golden, S. A., Heins, C., Morales, M., Epstein, D. H., & Shaham, Y. (2018). Volitional social interaction prevents drug addiction in rat models. *Nature Neuroscience*, *21*(11), 1520–1529. https://doi.org/10.1038/s41593-018-0246-6

Vollstädt-Klein S, Wichert S, Rabinstein J, Bühler M, Klein O, Ende G, Hermann D, & Mann K. (2020). Initial, habitual and compulsive alcohol use is characterized by a shift of cue processing from ventral to dorsal striatum. *Addiction, 105*(10), 1741–1749. doi: 10.1111/j.1360-0443.2010.03022.x.

Volkow, N. D., Wang, G.-J., Telang, F., Fowler, J. S., Logan, J., Childress, A.-R., Jayne, M., Ma, Y., & Wong, C. (2006). Cocaine cues and dopamine in dorsal striatum: Mechanism of craving in cocaine addiction. *Journal of Neuroscience*, *26*(24), 6583–6588. https://doi.org/10.1523/JNEUROSCI.1544-06.2006

Wakefield, J. C. (1992). The concept of mental disorder. On the boundary between biological facts and social values. *American Psychologist*, *47*(3), 373–388. https://doi.org/10.1037//0003-066x.47.3.373

Wakefield, J. C. (2018). The harmful dysfunction analysis of addiction: Normal brains and abnormal states of mind. In H. Pickard & S. H. Ahmed (Eds.), *The Routledge handbook of philosophy and science of addiction* (pp. 90–101). Routledge.

Watson, G. (1999). Disordered appetites. In J. Elster (Ed.), *Addiction: Entries and exits* (pp. 3–28). Russell Sage Foundation.

Wiens, T. K., & Walker, L. J. (2015). The chronic disease concept of addiction: Helpful or harmful? *Addiction Research & Theory*, *23*(4), 309–321. https://doi.org/10.3109/16066359.2014.987760

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.

Wrase, J., Schlagenhauf, F., Kienast, T., Wüstenberg, T., Bermpohl, Kahnt, T., Beck, A., Ströhle, A., Juckel, G., Knutson, B., & Heinz, A. (2007). Dysfunction of reward processing correlates with alcohol craving in detoxified alcoholics. *NeuroImage, 2*(1), 787–794. https://doi.org/10.1016/j.neuroimage.2006.11.043

Wright, L. (1973). Functions. *Philosophical Review*, *82*, 139–168.

Wright, L. (1976). *Teleological explanations*. University of California Press.

Zajac, K., Alessi, S. M., & Petry, N. M. (2018). Contingency management approaches. In H. Pickard & S. H. Ahmed (Eds.), *The Routledge handbook of philosophy and science of addiction* (pp. 455–463). Routledge.