Brain Change in Addiction as Learning, Not Disease

Marc Lewis, Ph.D.

During the past 30 years, the assumption that addiction is a disease or pathology has crystallized into the “brain disease model of addiction.” This trend was driven by the convergence of 12-step thinking with residential treatment approaches in the latter half of the 20th century, the explosion of neuroimaging technologies that began in the 1990s, and promotion by professional organizations and community groups. According to the brain disease model, addiction is a chronic disease brought about by changes in the brain systems that mediate the experience and anticipation of reward and in higher-order systems that underlie judgment and cognitive control. The proponents of the model propose that these changes are driven by exposure to drugs of abuse or alcohol, though links with behavioral addictions have also been explored.

The brain disease model is the most prevalent model of addiction in the western world. Particularly in the United States, it dominates professional and public discourse on prevention, treatment, research agendas, and policy issues. Because the disease model focuses on brain change, it has helped explain why persons with addictions find it difficult to change their thoughts and behaviors quickly or easily. Because it focuses on biologic factors rather than moral arguments, it has helped reduce the stigma faced by those with addictions and their families, at least in some respects. (See Table 1 for a broader discussion of stigma.) The brain disease model has also legitimized the role of doctors and other medical professionals in addiction treatment and driven research on new drugs to combat addiction, and it has been used to advocate for access to treatment and care rather than segregation and punishment.

These aims and outcomes are well intended, and they have been beneficial in some contexts, but the narrow focus of the disease model on the neurobiologic substrates of addiction has diverted attention (and research funding) from other models. Alternatives to the brain disease model often highlight the social and environmental factors that contribute to addiction, as well as the learning processes that translate these factors into negative outcomes. For example, it has been shown repeatedly that adverse experiences in childhood and adolescence increase the probability of later addiction. Also, exposure to physical, economic, or psychological trauma greatly increases susceptibility to addiction. Learning models propose that addiction, though obviously disadvantageous, is a natural, context-sensitive response to challenging environmental contingencies, not a disease. Yet the brain disease model construes addictive learning in terms of pathologic brain changes triggered mainly by substance abuse. Learning models also favor individual solutions for overcoming addiction, facilitated by cognitive modifications and personal agency. (See Table 2 for a discussion of empowerment.)

Learning models can include multiple levels of analysis: societal, social, psychological, and biologic. According to experts both inside and outside the medical
Especially when yet many people with addiction recoil from the disease label. It is true that some people with addiction feel consoled by the disease label. Proponents of the brain disease model of addiction have consistently claimed achievement over the pity bequeathed by the disease definition. Becoming better people as a result. Many would prefer respect for that partial success in achieving these humanitarian goals.

Yet the disease definition can replace one kind of stigma with another. The notion of a mental illness or disease can hurt more than help those with behavioral problems such as addiction, because it fuels discrimination and alienation of another sort. The disease designation can reinforce the belief that an inviolable or essentialist “badness” is built in and permanent, resulting in a sense that one is fundamentally different from “normal” people, with concomitant feelings of inferiority and shame. The label can also curtail attempts to improve one’s functioning without medical care. Biogenetic explanations carry the implication that people with addictions are not really trustworthy, now or in the future, because of a biologic proclivity they cannot control. Not only does this fuel one kind of stigmatization; it also helps rationalize a long-standing policy of withholding employment benefits and positions of authority from anyone who has ever been labeled an addict.

It is true that some people with addiction feel consolled by the disease label. In fact, psychiatric classifications have provided people who have diverse emotional and mental problems with a label and (sometimes) a hypothetical explanation for adversities that can otherwise seem indefinable, amorphous, and yet blameworthy. Distinct categories with concrete labels can help provide closure, context, and even a sense of belonging (to a particular group).

Yet many people with addiction recoil from the disease label. Especially when they are successful in galvanizing their willpower and rejigging their habits (i.e., recovering), they often find it confusing and debilitating to be told they are chronically ill. People with previous addictions (“recovered addicts”) usually want to feel that they have developed beyond their addiction and become better people as a result. Many would prefer respect for that achievement over the pity bequeathed by the disease definition.

The brain disease model does not dismiss the importance of learning but views this learning as pathologic. Addictive behaviors are proposed to begin as impulsive bids for highly motivating rewards, consolidated through operant conditioning, by which animals work to receive rewards predicted by specific cues, and Pavlovian conditioning, by which animals respond automatically to the stimulus properties of cues themselves. Advances in cognitive psychology reveal that learning also involves planning, decision making, inhibitory control, and strings of cues that eventually lead to predicted rewards. The contemporary view from cognitive science has extended this understanding with models of “embodied cognition,” which propose that all cognitive activity (including learning) results from iterative, self-perpetuating interactions (i.e., feedback) between the animal and the environment. From this perspective, learning occurs when the animal’s neural capacities become entrained with an environmental context. Thus, learning is not just a response to stimuli but active engagement with meaningful aspects of the environment.

The brain disease model presents these levels of analysis should ideally be integrated for a comprehensive understanding of addiction. Unfortunately, however, the neural level of analysis is almost always ignored by nondisease models that emphasize learning. (Work by Szalavitz is a notable exception.) Rather than ignore (or dispute) evidence of brain change in addiction, the current learning model reinterprets such evidence. Psychological change, development, and indeed all learning involve brain change. It is therefore unnecessary and perhaps unreasonable for a learning model of addiction to dismiss neural findings. In this review, I examine addiction within a learning framework, informed by classic and contemporary cognitive principles, which can incorporate the brain changes seen in addiction without reference to pathology or disease. In doing so, I hope to connect neurobiologic and environmental accounts to make sense of addiction with a degree of depth and precision that could not be achieved by either one alone. I also interpret key neurocognitive findings from both learning and disease perspectives to highlight their parallels as well as their disparities (Table 3).
of these learning mechanisms are necessarily unique, the brain disease model of addiction views the progression of decreasing control as a reflection of pathologic brain changes.

Addiction neuroscience explores these brain changes. The shift from impulsive (operant, reward-driven) actions to compulsive (automatic, Pavlovian) associations is a case in point. When drug taking is found to be highly rewarding, the ventral striatum (including the nucleus accumbens) focuses attention on the desired goal, activates a behavioral sequence to achieve that goal, and produces a motivational urge to energize that behavior.35 Over time, however, as behavior becomes more compulsive and less impulsive (less reward-driven), activation increases in the dorsal striatum, the region most associated with automatic responses.10,33,36,37 This progression is thought to eradicate willpower,38 because conscious choice is no longer driving the behavior.

The neurotransmitter dopamine has often been the focus of neural models of addiction.36 But dopamine has many functions, both in the striatum and in the prefrontal cortex, depending partly on the receptor type absorbing it. For the purposes of this discussion, we can think of dopamine as activating synaptic activity and, over time, synaptic change, both in the ventral and dorsal striatum and in the prefrontal cortex (partly through its effect on glutamate transmission). The release of dopamine to these and other systems is triggered by the perception of cues paired with anticipated rewards (in the case of operant learning) or with automatic responses (in the case of Pavlovian conditioning). Yet dopamine metabolism also responds to the experience of rewards, increasing when rewards exceed expectations and decreasing when they fall short. Addiction neuroscientists highlight the long-lasting sensitization of the dopamine system to addictive rewards or the cues that predict them, resulting in craving and narrowed attention6-37 as well as the subsequent blunting of the dopamine system over time.1

Striatal systems engage in constant cross-talk with regions of the prefrontal cortex. Prefrontal activation (in the orbitofrontal cortex) determines the attractiveness of potential rewards and also (in the dorsolateral prefrontal cortex) the exercise of judgment and perspective shifting. In fact, disrupted activation of the lateral prefrontal cortex has been shown to increase delay discounting (i.e., the proportion of impulsive choices).39 A key finding in support of the brain disease model is that drug use reduces connectivity between the prefrontal cortex and striatum, and long-term addiction corresponds with reduced gray-matter density (synaptic loss) in several prefrontal and related regions. Such changes are hypothesized to underlie diminished capacities for judgment and self-control, or “impaired response inhibition,” in people with addictions.5,40

According to the brain disease model, the cognitive and neural changes characterizing ad-
diction are unique and pathologic. Some theories highlight distinct phases or stages: drug taking is driven by positive reinforcement at first, then by negative reinforcement (underpinned by reduced dopamine signaling and blunted receptor responses), and finally by the loss of prefrontal control.1,41 A closely related theory suggests that addictive urges are increasingly driven by the brain’s rebound from drug stimulation — an “antireward” effect resulting from an overactive stress-response system, dopamine blunting, and physical withdrawal symptoms.42 These theories emphasize repeated episodes of negative reinforcement (learning to avoid an aversive outcome) and positive reinforcement, plus changes in neurochemistry and circuitry.

But are the neurocognitive processes that give rise to addiction actually pathologic, or are they constituents of normal learning with detrimental consequences? To help resolve this question, I examine four neurocognitive changes central to brain disease models. The first is the hypothesized shift from impulsive behavior mediated by the ventral striatum to compulsive responses mediated by the dorsal striatum.35 The second change, which also supports the presumption of involuntary behavior, is a reduction in functional and structural connectivity between the striatum and prefrontal cortex.1,5 The third change is increased and enduring sensitivity (i.e., sensitization) to cues predicting addictive rewards, underpinned by mesolimbic dopamine.37 The fourth change is a decrease in sensitivity, not only to alternative rewards but even to addictive rewards themselves.1 I argue that these four neurocognitive changes are not specific to addiction and do not indicate a disease process.

### Reinterpreting the Neurocognitive Data

### Role of Compulsive or Automatic Responses

According to the brain disease model, impulsive drug seeking and use are linked with activation of the ventral striatum or nucleus accumbens at first, but these behaviors become compulsive and automatic with activation of the dorsal striatum over time.35,43 Yet behavior generally becomes more automatic with practice, as novelty is replaced by familiarity, and dorsal striatal (including globus pallidus) involvement underlies this automatization even in a simple finger-tapping task.44 As Everitt and Robbins, acknowledged experts on the ventral-to-dorsal shift, state, “There is nothing aberrant or unusual about devolving behavioural control to a dorsal striatal S-R [stimulus–response or Pavlovian] ‘habit’ mechanism.”35 They assert that this shift is to be expected in

<table>
<thead>
<tr>
<th>Disease Model</th>
<th>Learning Model</th>
<th>Evidence for Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addiction is characterized by a shift from impulsive to compulsive processing, loss of free will, and a shift of activation to dorsal striatum.</td>
<td>All behavioral habits devolve to stimulus–response mechanisms; automatization is a normal outcome of learning.</td>
<td>Dorsal striatal activation or behavioral automatization is seen with practice of even simple (e.g., motor) tasks; for people with addiction, operant contingencies facilitate the choice to abstain from using drugs.</td>
</tr>
<tr>
<td>Functional connectivity between striatum and PFC is lost, with reduced synaptic density in specific PFC regions.</td>
<td>When planning and decision making are bypassed, PFC demand is reduced; extended plasticity is normal; underused synapses may be pruned.</td>
<td>Immediate or valued rewards lead to increased striatal activation and decreased dorsolateral PFC activation and cognitive control; synaptic density in the PFC has been shown to rebound with recovery.</td>
</tr>
<tr>
<td>Sensitization to drug cues is increased (and enduring), mediated by increased mesolimbic dopamine uptake.</td>
<td>Sensitization to valued rewards is normal; an ongoing need or desire leads to ongoing sensitization (e.g., love, attachment, wealth acquisition, religious practice).</td>
<td>Motivated goal pursuit leads to increased dopamine, cue sensitization, and learning; high emotional salience facilitates lasting synaptic alterations (e.g., after trauma).</td>
</tr>
<tr>
<td>Ongoing drug use leads to loss of receptor availability or sensitivity and reduced pleasure (dopaminergic blunting).</td>
<td>Adversity, trauma (with or without drug use), isolation, and overstimulation lead to reduced dopamine-receptor response or pleasure.</td>
<td>Loss of social status or trauma leads to reduced D2 or D3 receptor availability; high levels of mating behavior, eating, engagement with pornography, and Internet use lead to a hypodopaminergic system.</td>
</tr>
</tbody>
</table>

*PFC denotes prefrontal cortex.
many aspects of our lives, including eating and other habitual activities. “Automatisation of behaviour frees up cognitive processes,” Everitt and Robbins continue, which explains why we can talk, eat, and drive at the same time.

Not only is normal behavior partly automatic, but also addictive behavior, even in its later stages, remains partly operant (reward-driven). Supporting evidence comes from numerous studies in which the reward value of the addictive goal (e.g., the amount of drug offered) shifts in relation to the reward value of an alternative goal (e.g., money). In fact, these studies show that the probability of abstaining is proportional to the relative reward value of the two choices; this sensitivity to environmental contingencies is the hallmark of operant learning. Contingency management programs, based on these principles, have shown a consistent effect in the reduction of drug use. The ventral striatum continues to be involved in reward seeking in later-stage addiction, even when the dorsal striatum dominates behavior control. In sum, a combination of deliberate and automatic neurobehavioral mechanisms characterizes both addiction and “normal” habitual behavior.

**LOSS OF PREFRONTAL CONNECTIVITY AND SYNAPTIC PRUNING**

Evidence of a functional and (in some studies) structural disconnection between the prefrontal cortex and striatum has been pivotal for defining addiction as a brain disease. Unfortunately, these findings come from cross-sectional, not longitudinal, research, so some cortical differences must precede rather than follow addiction, as acknowledged by the researchers. Yet even cortical changes that arise from (or with) addictive drug use need not be considered pathologic.

When skills become streamlined with practice, they no longer engage conscious, reflective, or effortful control. In fact, higher-order cognition is unnecessary once behavior becomes habitual, as any professional musician or athlete can demonstrate. Also, rewards perceived as both immediate and valuable often bypass cognitive control, as seen in the reduction of planning, decision making, and concomitant prefrontal involvement when it comes to sex, gambling, and eating fast food. Research points to an inverse correlation between striatal activation and dorso-lateral prefrontal engagement, both in delay discounting and more generally in effortful reward seeking. But would this loss of functional connectivity normally lead to structural changes? Indeed, the elimination, or “pruning,” of underused synapses is considered a key mechanism of learning. Massive cortical pruning has traditionally been associated with adolescence, when most addictions develop. However, since pruning makes the brain more efficient when new skills are practiced and consolidated, it is now thought to underpin learning over the lifespan.

Synaptic density in certain prefrontal regions decreases with the duration of drug use, but a contrasting increase in synaptic density (in similar but not identical regions) correlates with the number of weeks of abstinence. In studies using functional magnetic resonance imaging (fMRI), “cocaine-dependent” participants who became abstinent no longer differed from controls with respect to the activation of inhibitory control networks in the prefrontal cortex or the performance of motor-inhibition tasks. Thus, reductions in prefrontal involvement and synaptic density appear to be restricted to the period of habitual drug use, which may be followed by a period of synaptic growth when a new skill — abstinence — is learned. This two-way street in frontal neuroplasticity is consistent with evidence that most people with addiction recover, and most of those who recover do so without treatment. This finding would seem to be impossible if prefrontal changes were permanent and therefore pathologic.

**SENSITIZATION TO CUES**

People with drug addiction are highly sensitive to drug-related cues, even after they quit using drugs. To account for this sensitization, the brain disease model points to a sharp rise in mesolimbic (reward-related) dopamine uptake. The motivational drive provided by mesolimbic dopamine is essential for survival, because it ensures that we prioritize eating, social relationships, and procreation. Addiction neuroscientists acknowledge that the levels of cue-triggered dopamine seen in addiction can parallel those related to “natural” rewards. Indeed, romantic relationships depend on motivational dopamine uptake, and desire after romantic rejection matches the craving for cocaine. Motivated pursuits (natural or otherwise), including shopping, sports, religious practice, wealth acquisition, gambling, binge eat-
ing, romantic love, and pornography, correspond with cue sensitization and increased activation of striatal dopamine. Even a simple increase in reward availability on a computer screen is sufficient to increase mesolimbic dopamine, with a concomitant increase in effort.

Proponents of the brain disease model emphasize that cue sensitivity in addiction is not only extreme but also prolonged, whereas cue sensitivity returns to normal levels in relation to natural reinforcers, once the need has been met. This prolonged sensitization is seen as the cause of relapse. Yet prolonged sensitization also results from normal learning of emotionally salient associations, through synaptic alterations in regions that process emotion, such as the amygdala. Stimuli associated with past triumphs or traumas or even a once-loved song will reliably trigger strong feelings. Because these cues refer to still-meaningful experiences, dopamine uptake remains adaptive (rather than pathologic) for ongoing behavioral adaptations.

Perhaps the most parsimonious explanation for enduring cue sensitivity is that, in addiction, goal seeking remains unfulfilled. The drug or activity that was pursued to satisfy emotional needs may have lost its effect because of a short duration of action, chemical tolerance, or habituation. The value of addictive rewards is always determined by context, including both the strength of aversive feelings and the effectiveness of drugs, for example, in quelling them. Unresolved needs can make drug taking relevant indefinitely.

Desensitization to Drug-Related and Natural Rewards

In parallel with cue sensitization and increased levels of dopamine release, there is an apparently paradoxical decrease in sensitivity to alternative rewards and even to drugs themselves. This reward desensitization is thought to contribute to increasing drug consumption. Brain disease models ascribe this blunting to the down-regulation (reduced availability or responsiveness) of dopamine receptors (e.g., D2 and D3 receptors), a pathologic process that may be manifested as tolerance or withdrawal effects. Yet many studies of addiction use psychostimulants (e.g., cocaine and methamphetamine), seriously confounding this observation. The buildup (e.g., delayed reuptake) of dopamine resulting from psychostimulants may directly trigger a chemical rebound effect, independent of addictive learning.

But even if addictive learning results in dopaminergic blunting, it need not denote pathologic brain change. Poverty, trauma, and diminished social status reduce the availability of the D2 and D3 dopamine receptors in humans and nonhuman primates. In fact, a reduction in D2 or D3 receptor availability has been shown to correspond with reduced social dominance or isolation, driving drug or alcohol use as a means of countering anxiety or distress. As noted above, early adversity and trauma are reliable predictors of subsequent drug use. However, social adversity may also result from drug use itself. Society responds to illicit drug use by excluding or punishing users, which in turn leads to broken relationships and erosion of self-esteem. Thus, social and psychological hardships may result in dopaminergic blunting, which then encourages addictive activities, amplifying these hardships.

Dopaminergic blunting can also result from nondrug rewards. Mating behavior in rats reduces dopamine output in mesolimbic dopamine circuitry, leading to “a hypodopaminergic system,” and identical changes result from prolonged exposure to opiates. In addition, obesity has been linked to reduced dopamine receptivity, with the hypothetical explanation that dopaminergic blunting leads to increased food consumption. Exposure to other potentially habit-forming pleasurable activities also leads to dopaminergic blunting, as shown with pornography use and extensive Internet use. Thus, it seems that dopaminergic blunting can result from frequent activation of the mesolimbic dopamine system by any repetitive reward-seeking behavior rather than by drug exposure itself. Kent Berridge, a renowned addiction neuroscientist, views dopaminergic suppression as a temporary effect of overstimulation, which may result from drug addiction but does not cause it.

Addiction as Organism–Environment Entrainment

Most alternatives to the brain disease model of addiction share the view that explanations of addiction should include societal, social, and familial factors that predict drug misuse. The brain disease model has acknowledged these factors,
but its emphasis on brain pathology sidelines their causal status and their relevance to prevention and treatment efforts. Yet viewing addiction solely as the product of environmental forces tends to ignore the properties of the organism, its nervous system, and its response proclivities. A comprehensive, balanced model of addiction needs to recognize that the organism and its environment are connected at every level, from perception to cognition to behavior, and interact continuously as an open system.

I have presented arguments and evidence that automatization, reduced neural flexibility, enduring cue sensitization, and reward desensitization are normal features of learning highly motivating, repetitive, and habitual behavioral patterns. Thus, I dispute the idea that addiction is pathologic. Nevertheless, there is considerable potential for reconciliation between aspects of the brain disease model and an environmental model of addiction, given that both view a rigidified behavioral pattern as learned, and learned deeply. Classic learning models have limited value for this synthesis, since they view the learner as an independent agent responding to a static environment. In contrast, principles of embodied cognition construe learning as a process of reciprocal adjustments between the activities of the organism and meaningful features of the environment. What is meaningful is assumed to be constrained by biologic antecedents and emerging biologic sensitivities, as well as the stimulus properties of the animal’s environment (i.e., features of the environment that afford or invite specific actions, known as affordances).

For a young human, the range of potentially meaningful environmental features can be vast, at least until social, familial, and psychological setbacks narrow it down to a small subset of suboptimal rewards. For example, many children grow up with an unpredictable, disengaged, or violent parent. As adolescents, they may face disruptions in education, employment, or relationships as a result of financial or other disadvantages. These persons tend to find increased meaning in drugs that reduce stress or promote feelings of security and well-being, especially because these effects can be attained without mediation by other people. As drug use progresses and becomes a more consistent focus of attention and behavior, the properties of the individual and of the environment tend to become synchronized through mutual adjustments. Behavioral outcomes continue to shape a social environment that progressively narrows behavioral options. For example, the social environment may become increasingly limited to people who can supply drugs (dealers or doctors), people with whom to take drugs, and “friends” who remain apathetic and disengaged. Behavioral proclivities will change accordingly. Besides the increasing habit strength of drug pursuit itself, there is likely to be increased lying to avoid rejection or punishment, as well as disengagement from romantic partners and family members, further limiting the chance to feel connected and protected. These changes would be mediated by cognitive modifications — changes in attentional foci, belief systems, identity, and self-esteem — as well as by immature habits of emotion regulation (e.g., suppression or denial) more generally.

But how might this addiction spiral get started? The embodied-cognition view encourages us to look for biologic and environmental vulnerabilities that amplify and reinforce each other. The goal here is not to list organismic (e.g., genetic) and environmental risk factors and add them together, but instead to track the interaction of factors that reciprocally influence each other. I suggest that the addiction spiral gets started with early psychosocial adversity. First, we already know that early adversity and trauma are strong predictors of later addiction. Second, developmental psychologists have shown that early trauma (physical, emotional, or sexual) leaves enduring effects on nervous system function, such as sympathetic or parasympathetic overattunement (causing hyperreactivity or hypo-reactivity), oversensitivity to threat based on accelerated amygdala development, and hippocampal damage resulting from excessive cortisol levels. Third, in animal models, researchers have pinpointed epigenetic changes (e.g., methylation of a gene that tunes the glucocorticoid feedback loop) that take place in utero or the first year of life in response to inadequate nurturing. But these neuropsychological insults do not emerge in a vacuum. Both trauma and “stress methylation” can begin with overstressed parents and even grandparents in families challenged by unemployment, marital discord, histories of abuse, or alienation from the community, affecting the stress response in childhood and throughout life. From these beginnings, a narrowing spiral of
ineffective coregulation emerges between developing children and their caregivers, leading eventually to entrainment between drug seeking and its environmental concomitants. From the Rat Park studies of the 1970s and 1980s, in which even addicted rats avoided ingesting morphine when allowed to socialize and play, to contemporary evidence of the adverse consequences of socioeconomic fragmentation, Bruce Alexander has shown that addiction emerges universally as a response to the disruption of normal social interactions. Therefore, models of addiction predicated on embodied cognition should focus on environments in which social stressors affect early neuropsychological development, as a gateway to ongoing reciprocal adjustments between disadvantageous organismic adaptations and narrowing environmental opportunities.

In summary, the embodied-cognition framework can help model the interaction between neurobiologic and social-environmental contributors to addiction. Addictive activities are determined neither solely by brain changes nor solely by social conditions. Although they indeed result from and contribute to brain changes, addictive activities also feed back to the social environment, further narrowing what are often already limited opportunities for well-being, which in turn further narrows cognitive and neural flexibility. It follows that the narrowing seen in addiction takes place within the behavioral repertoire, the social surround, and the brain — all at the same time. It also follows that growth beyond addiction can be facilitated by improved social support, extended behavioral opportunities, targeted pharmacologic interventions, or some combination of these strategies.

Disclosure forms provided by the author are available with the full text of this article at NEJM.org.

I thank Shaun Shelly (University of Pretoria, Department of Family Medicine) for providing information and references supporting the arguments reviewed here, as well as feedback on previous versions of the manuscript.

REFERENCES

26. Dutra L, Statopoulos G, Basden SL,


Brain Change in Addiction as Learning, Not Disease


Copyright © 2018 Massachusetts Medical Society.