The American Journal of Bioethics, 7(1): 8–11, 2007

ISSN: 1526-5161 print / 1536-0075 online
DOI: 10.1080/15265160601063969

Target Article

The Neurobiology of Addiction: Implications for Voluntary Control of Behavior

Steven E. Hyman, Harvard University

There continues to be a debate on whether addiction is best understood as a brain disease or a moral condition. This debate, which may influence both the stigma attached to addiction and access to treatment, is often motivated by the question of whether and to what extent we can justly hold addicted individuals responsible for their actions. In fact, there is substantial evidence for a disease model, but the disease model per se does not resolve the question of voluntary control. Recent research at the intersection of neuroscience and psychology suggests that addicted individuals have substantial impairments in cognitive control of behavior, but this "loss of control" is not complete or simple. Possible mechanisms and implications are briefly reviewed.

Cognitive and social neuroscience and studies of the pathophysiologic processes underlying neuropsychiatric disorders have begun to probe the mechanisms by which human beings regulate their behavior in conformity with social conventions and in pursuit of chosen goals—and the circumstances under which such "cognitive control" may be eroded (Miller and Cohen 2001; Montague et al. 2004; Miller and D'Esposito 2005). The resulting ideas call into question folk psychology views on the voluntary control of behavior, that is, for the most part, we regulate our actions based on conscious "reasons." Even in health, critical processes that intervene between sensory inputs to the brain and the execution of actions, including processes that permit "top-down" or "cognitive" control of behavior, do not appear to depend on conscious exertion of will (Wenger 2002). Challenges to folk psychology views of the voluntary control of behavior may be highlighted most vividly, however, by conditions such as addiction, in which the core symptoms reflect a failure of the underlying processes (Montague et al. 2004; Hyman 2005; Kalivas and Volkow 2005), which I refer to as *cognitive control*.

The major justification for demarcating neuroethics from the broader field of bioethics derives from the special status of the brain (Roskies 2002), which is the causal underpinning of our conscious mental lives and of our behavior. This is not a reductionist claim. The structure and function of the brain is influenced not only by "bottom-up" factors such as genes, but also by top-down factors such "lived experience" and context. Moreover, neuroscience does not obviate the need for social and psychological level explanations intervening between the levels of cells, synapses, and circuits and that of ethical judgments. Indeed, modern cognitive and social neuroscience (Cacioppo et al. 2002; Gazzaniga 2004) are, in no small measure, attempts to mediate between understandings of the functioning of neural networks in one regard and of sensation, thought, and action in another. What neuroscience contributes to ethical discourse is mechanistic insight that constrains our interpretations of psychological observations and that suggests new explanatory frameworks for thought and behavior. Neuroscience should make it possible to ask how the nature of our brains shapes and constrains what we call *rationality*, and therefore, ethical principals themselves, and it should permit us to probe deeply into the nature of reason, emotion, and the control of behavior (Churchland 2006). Having recently reviewed the neurobiology of addiction for clinicians (Hyman 2005) and for neuroscientists (Hyman et al. 2006), I would like to examine the implications of emerging ideas about reward, cognitive control, and the pathophysiology of addiction for insights into the voluntary control behavior.

The question of whether and to what extent an addicted individual is responsible for his or her actions remains a matter of unsettled debate. One proxy (albeit imperfect) for this question is disagreement as to whether addiction is best conceptualized as a brain disease (Leshner 1997; McLellan et al. 2000), as a moral condition (Satel 1999), or as some combination of the two (Morse 2004b). Those who argue for the disease model not only believe it is justified by empirical data, but also see virtue in the possibility that a disease model decreases the stigmatization of addicted people and increases their access to medical treatments. Those who argue that addiction is best conceptualized as a moral condition are struck by the observation that drug seeking and drug taking involve a series of voluntary acts that often require planning and flexible responses to changing conditions—not simply impulsive or robotic acts. They worry that medicalization

Address correspondence to Steven E. Hyman, Office of the Provost, Massachusetts Hall, Harvard University, Cambridge, MA 02138. E-mail: seh@harvard.edu

will lead addicted people to fatalism about their condition and to excuses for their actions rather than full engagement with treatment and rehabilitation and an effort to conform to basic societal expectations.

Current definitions of addiction come from medical texts and thus, not surprisingly, favor a disease model. Indeed, addiction looks very much like a disease (admittedly definitions of "disease" remain somewhat fuzzy). Addiction has known risk factors (family history, male sex) and a typical course and outcome: often a chronic course punctuated by periods of abstinence followed by relapse (McLellan et al. 2000; Hser et al. 2001). True, the precise alterations in physiology that account for the symptoms and course are not yet known with certainty, but there is little doubt in the scientific community that such mechanisms will be found (Chao and Nestler 2004). Similarly the search for the precise genetic variants that confer familial risk is in its early days, but existing data from family, twin, and adoption studies convincingly argue that genes play a central role in vulnerability (Goldman et al. 2005).

What is more interesting is that modern definitions of addiction focus squarely on the issue of voluntary control. The current medical consensus is that the cardinal feature of addiction is compulsive drug use despite significant negative consequences (American Psychiatric Association 1994). The term *compulsion* is imprecise, but at a minimum implies diminished ability to control drug use, even in the face of factors (e.g., illness, failure in life roles, loss of job, arrest) that should motivate cessation of drug use in a rational agent willing and able to exert control over behavior. The focus on "loss of control" is not derived primarily from a theory, but from extensive observation of the behavior of addicted individuals (Tiffany 1990; O'Brien et al. 1998) and indeed recognition of the failure of previous definitions to capture clinical realities. The current focus on compulsive use as the defining features of addiction superseded previous views that focused on dependence and withdrawal. These previous views implied that addicted individuals take drugs to seek pleasure and avoid aversive withdrawal symptoms. Although the avoidance of withdrawal might create strong motivation to take drugs, this view does not imply a loss of voluntary control. This previous view failed on several counts. First, some highly addictive drugs such as cocaine and amphetamine may produce mild withdrawal symptoms and lack a physical withdrawal syndrome entirely. Moreover, the previous view does not explain the stubborn persistence of relapse risk long after detoxification, long after the last withdrawal symptom, if any, has passed, and despite incentives to avoid a resumption of drug use (Hyman 2005).

Before discussing my views of the neural basis of addiction, I should stipulate that the science is in its early stages and that there is not yet a fully convincing theory of how addiction results from the interaction of risk factors, drugs, and the brain. Moreover, there are still disagreements at the theoretical level of what the existing data signifies for the mechanisms of addiction. (Compare, for example, Koob and Le Moal 2005; Robinson and Berridge 2003; and Hyman 2005). This state of affairs invites skepticism from those wary of a disease model (Satel 1999). Nonetheless, we cannot select models of human behavior based on desired social implications, but must rely on the scientific evidence we have. Despite somewhat different views of mechanism, all current mainstream formulations agree that addiction diminishes voluntary behavioral control. At the same time, none of the current views conceives of the addicted person to be devoid of all voluntary control and thus absolved of all responsibility for self-control.

Short of being harshly coerced, severely psychotic, or significantly demented, what can it mean to say that a person cannot control his or her actions? An alcoholic must obtain money, go to the liquor store or otherwise obtain alcohol (perhaps carefully hidden from a spouse) and consume drinks. A heroin user may have to go to great lengths to obtain the drug, perhaps committing one of more crimes, before beginning the ritual that ends in self-injection. How can these extended chains of apparently voluntary acts be the result of compulsion? In my view, addictive drugs tap into and, in vulnerable individuals, usurp powerful mechanisms by which survival-relevant goals shape behavior (Hyman 2005; Hyman et al. 2006).

Diverse organisms, including humans, pursue goals with positive survival value such as food, safety, and opportunities for mating; such goals act as "rewards" (Kelley and Berridge 2002). Rewards are experienced as pleasurable and as motivating (they are desired). Environmental cues that predict their availability (e.g., the smell of baking bread) are rapidly learned and are imbued with incentive properties: they activate "wanting" and initiate behaviors aimed at obtaining the desired goal. Such goal-directed behaviors tend to increase in frequency over time (reinforcement) and to become highly efficient. Of course rewarding goals for humans can vary enormously in immediacy, complexity, and motivational power, ranging from a well-liked food to seeing a favorite painting in a museum.

The brain has evolved several specialized mechanisms to maximize the ability of an organism to obtain rewards. There are mechanisms to provide internal representations of rewards and to assign them relative values compared with pursuing other possible goals; these mechanisms depend primarily on the orbital prefrontal cortex (Schoenbaum et al.2006). There are mechanisms that permit an organism to learn and to make relatively efficient and automatic, sequences of actions to obtain specific rewards; these depend primarily on the dorsal striatum (Everitt and Robbins 2005). Mechanisms of cognitive control support successful completion of goal-directed behaviors by maintaining the goal representation over time, suppressing distractions, and inhibiting impulsive actions that redirect the organism. Cognitive control is dependent on the prefrontal cortex and its connections to the striatum and thalamus. In humans, the capacity for cognitive control appears to be a relatively stable trait that is an important predictor of life success (Eigsti et al. 2006). Deficits in cognitive control play an important role in

attention deficit hyperactivity disorder (Vaidya et al. 2005) and may increase vulnerability to later substance misuse.

These circuits respond in a coordinated fashion to new information about rewards through the action of the neurotransmitter dopamine (Montague et al. 2004). Dopamine is released from neurons with cell bodies in the ventral tegmental area (VTA) and substantia nigra within the midbrain. These neurons project widely through the forebrain and can influence all of the circuits involved in reward-related learning, as well as in other aspects of cognition and emotion. Dopamine projections from the VTA to the nucleus accumbens bind the pleasurable (hedonic) response to a reward to desire and to goal-directed behavior (Berridge and Robinson 1998; Everitt and Robbins 2005). Dopamine projections from the VTA to the prefrontal cortex play a critical role in the assignment of value and in updating goal representations in response to the state of the organism (Montague et al. 2004). Dopamine projections from the substantia nigra to the dorsal striatum are critical for consolidating new behavioral responses so that reward-related cues come to activate efficient strategies to reach the relevant goal (Everitt and Robbins 2005).

Addictive drugs are Trojan horses. Unlike natural rewards, addictive drugs have no nutritional, reproductive, or other survival value. However, all addictive drugs exert pharmacologic effects that cause release of dopamine. Moreover, the effects of addictive drugs on dopamine release are quantitatively greater than that produced by natural rewards under almost all circumstances.

Normally dopamine serves as a "learning signal" in the brain. Dopamine is released when a reward is new, better than expected, or unpredicted in a particular circumstance (Schultz et al. 1997; Schultz 2006). When the world is exactly as expected, there is nothing new to learn; no new circumstances to connect either to desire or to action—and no increase in dopamine release. Because addictive drugs increase synaptic dopamine by direct pharmacologic action, they short circuit the normal controls over dopamine release that compare the current circumstance with prior experience. Thus, unlike natural rewards, addictive drugs always signal "better than expected." Neural circuits "overlearn" on an excessive and grossly distorted dopamine signal (Montague et al. 2004; Hyman 2005; Hyman et al. 2006). Cues that predict drug availability such as persons, places, or certain bodily sensations gain profound incentive salience and the ability to motivate drug seeking. Because of the excessive dopamine signal in the prefrontal cortex (Volkow and Fowler 2000) drugs become overvalued compared with all other goals. Rational goals such as self-care, working, parenting, and obeying the law are devalued. In addition, normal aspects of cognitive control weaken; even if the addicted person wants to "cut down," prepotent cue-initiated drug-seeking responses are extremely difficult to suppress. If the person is successful in delaying drug seeking (or is, for external reasons unable to seek drugs), intense craving may result (Tiffany 1990). Because the changes in synaptic weight and synaptic structure that underlie memory are among the longest-lived alterations in biology, the ability of drug-related cues to cause relapses may persist for many years, even a lifetime.

There remains much to learn about the pathophysiology of addiction. Currently, much research is attempting to demonstrate that drug-induced changes in synaptic connectivity and drug-induced changes in the expression of neuronal genes and proteins are causally involved in addiction-related behaviors (Chao and Nestler 2004; Hyman et al. 2006). This model of pathogenesis, and the research on reward-related learning on which it rests, suggest highly plausible mechanisms by which addicted individuals may "lose control" over drug seeking and drug taking (Montague et al. 2004; Hyman 2005; Kalivas and Volkow 2005; Hyman et al. 2006). Mechanisms that evolved to motivate survival behaviors, the pursuit of natural rewards, are usurped by the potent and abnormal dopamine signal produced by addictive drugs. The result is a brain in which drug cues powerfully activate drug seeking, and in which attempts to suppress drugseeking result in intense craving. This model does not, however, reduce addicted individuals to zombies who are permanently controlled by external cues. As overvalued as drugs become, as potent as the effects of drug cues on behavior, other goals are not extirpated. Perhaps in a drug-free context, perhaps with a good measure of initial coercion, perhaps with family, friends, and caregivers acting as external "prostheses" to strengthen and partially replace damaged frontal mechanisms of cognitive control, and often despite multiple relapses, addicts can cease drug use and regain a good measure of control over their drug taking. Our current models help explain why recovery is difficult and why relapses occur even long after detoxification and rehabilitation. The long experience of humanity with addiction does not counsel fatalism, but implacable efforts to overcome the behavioral effects of neural circuits hijacked by drugs. Finally, views based on cognitive neuroscience and studies of addiction pathogenesis suggest that some apparently voluntary behaviors may not be as freely planned and executed as they first appear. Such cognitive views have not yet penetrated folk psychology, and it is premature for these views to have any place in the courtroom (Morse 2004a; Greene and Cohen 2004). Nonetheless these cognitive views deserve a place in current ethical discussions of personal responsibility. For many reasons, it may be wise for societies to err on the side of holding addicted individuals responsible for their behavior and to act as if they are capable of exerting more control than perhaps they can; however, if the ideas expressed in this review are right, it should be with a view to rehabilitation of the addicted person and protection of society rather than moral opprobrium.

REFERENCES

American Psychiatric Association. 1994. *Diagnostic and statistical manual of mental disorders*, 4th Edition. Washington, DC: American Psychiatric Association.

Berridge, K. C., and T. E. Robinson. 1998. What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Research Reviews* 28: 309–369.

Cacioppo, J. T., G. G. Berntson, R. Adolphs, and C. S. Carter, eds. 2002. *Foundations in social neuroscience*. Cambridge, MA: The MIT Press.

Chao, J., and E. J. Nestler. 2004. Molecular neurobiology of drug addiction. *Annual Review of Medicine* 55: 113–132.

Churchland, P. S. 2006. Moral decision-making and the brain. In *Neuroethics: Defining the issues in theory, practice, and policy*, ed. J. Illes: 3–16. Oxford, UK: Oxford University Press.

Eigsti, I. M., V. Zayas, W. Mischel, Y. Schoda, O. Ayduk, M. B. Dadlani, M. C. Davidson, J. Lawrence Aber, and B. J. Casey. 2006. Predicting cognitive control from preschool to late adolescence and young adulthood. *Psychological Science* 17(6): 478–484.

Everitt, B. J., and T. W. Robbins. 2005. Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nature Neuroscience* 8: 1481–1489.

Gazzaniga, M. S., Ed. 2004. *The cognitive neurosciences III, 4th Edition*. Cambridge, MA: The MIT Press.

Goldman, D., G. Oroszi, and F. Ducci. 2005. The genetics of addictions: Uncovering the genes. *Nature Reviews Genetics* 6: 521–532.

Greene, J., and J. Cohen. 2004. For the law, neuroscience changes nothing and everything. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 359: 1775–1785.

Hser, Y. I., V. Hoffman, C. E. Grella, and M. D. Anglin. 2001. A 33-year follow-up of narcotics addicts. *Archives of General Psychiatry* 58(5): 503–508.

Hyman, S. E. 2005. Addiction: A disease of learning and memory. *American Journal of Psychiatry* 162(8): 1414–1422.

Hyman, S. E., R. C. Malenka, and E. J. Nestler. 2006. Neural mechanisms of addiction: The role of reward-related learning and memory. *Annual Review of Neuroscience* 21(29): 565–598.

Kalivas, P. W., and N. D. Volkow. 2005. The neural basis of addiction: A pathology of motivation and choice. *American Journal of Psychiatry* 162(8): 1403–1413.

Kelley, A. E., and K. C. Berridge. 2002. The neuroscience of natural rewards: relevance to addictive drugs. *The Journal of Neuroscience* 22: 3306–3311.

Koob, G. F, and M. Le Moal. 2005. *Neurobiology of Addiction*. New York: Academic Press.

Leshner, A. I. 1997. Addiction is a brain disease, and it matters. *Science* 278(5335): 45–47.

McLellan, A. T., D. C. Lewis, C. P. O'Brien, and H. D. Kleber. 2000. Drug dependence, a chronic medical illness: Implications for treatment, insurance, and outcomes evaluation. *Journal of the American Medical Association* 284(13): 1689–1689.

Miller, B. T., and M. D'Esposito. 2005. Searching for "the top" in top-down control. *Neuron*. 48: 535–538.

Miller, E. K., and J. D. Cohen. An integrative theory of prefrontal cortex function. 2001. *Annual Review of Neuroscience*. 24: 167–202.

Montague, P. R., S. E., Hyman, and J. D. Cohen. 2004. Computational roles for dopamine in behavioural control. *Nature* 431: 760–767.

Morse, S. J. 2004a. New neuroscience, old problems: Legal implications of brain science. *Cerebrum* 6(4): 81–90.

Morse, S. J. 2004b. Medicine and morals, craving and compulsion. *Substance Use & Misuse* 39(3): 437–460.

O'Brien, C. P., A. R. Childress, R. Ehrman, and S. J. Robbins. 1998. Conditioning factors in drug abuse: Can they explain compulsion? *Journal of Psychopharmacology* 12: 15–22.

Robbins, T. W., and B. J. Everitt. 2002. Limbic-striatal memory systems and drug addiction. *Neurobiology of Learning and Memory* 78: 625–636.

Robinson, T. E., and K. C. Berridge. 2003. Addiction. *Annual Review of Psychology* 54: 25–53.

Roskies, A. 2002. Neuroethics for the new millennium. *Neuron* 35(1): 21–23.

Satel, S. L. 1999. What should we expect from drug abusers? *Psychiatric Services* 50(7): 861.

Schoenbaum, G., M. R. Roesch, and T. A. Stalnaker. 2006. Orbitofrontal cortex, decision-making and drug addiction. *Trends in Neurosciences* 29(2): 116–124.

Schultz, W. 2006. Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology* 57: 87–115.

Schultz, W., P. Dayan, and P. R. Montague. 1997. A neural substrate of prediction and reward. *Science* 275: 1593–1599.

Tiffany, S. T. 1990. A cognitive model of drug urges and drug-use behavior: Role of automatic and nonautomatic processes. *Psychological Review* 97: 147–168.

Vaidya, C. J., S. A., Bunge, N. M., Dudukovic, C. A., Zalecki, G. R., Elliott, and J. D. Gabrieli. 2005. Altered neural substrates of cognitive control in childhood ADHD: Evidence from functional magnetic resonance imaging. *American Journal of Psychiatry*. 162(9): 1605–1613.

Volkow, N. D., and J. S. Fowler. 2000. Addiction, a disease of compulsion and drive: Involvement of the orbitofrontal cortex. *Cerebral Cortex* 10: 318–325.

Wenger, D. M. 2002. *The illusion of conscious will*. Cambridge, MA: The MIT Press.