Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Neurobiology

Negative reinforcement in drug addiction: the darkness within

George F Koob

Drug seeking is associated with the activation of reward neural circuitry, but I argue that drug addiction also involves another major source of reinforcement, specifically negative reinforcement driven by the 'dark side' (i.e., a decrease in the function of normal reward-related neurocircuitry and persistent recruitment of the brain stress systems). This combination forms the antireward system or 'darkness within.' Understanding the neuroplasticity of the neurocircuitry that comprises the negative reinforcement associated with addiction is the key to understanding the vulnerability to the transition to addiction, misery of addiction, and persistence of addiction.

Addresses
Committee on the Neurobiology of Addictive Disorders, The Scripps Research Institute, La Jolla, CA, USA

Corresponding author: Koob, George F (gkoob@scripps.edu)

Current Opinion in Neurobiology 2013, **23**:559–563

This review comes from a themed issue on **Addiction**

Edited by **Barry Everitt** and **Ulrike Heberlein**

For a complete overview see the Issue and the Editorial

Available online 27th April 2013

0959-4388/$ – see front matter,

http://dx.doi.org/10.1016/j.conb.2013.03.011

We conceptualize addiction as a three-stage cycle: *binge/intoxication*, *withdrawal/negative affect*, and *preoccupation/anticipation*. These stages worsen over time and ultimately engage a little or no reward but only a struggle to return to a euthymic state during the terminal condition [1••]. Positive reinforcement is defined as the process by which the presentation of a stimulus increases the probability of a response. Negative reinforcement is defined as the process by which removal of an aversive stimulus (or aversive state in the case of addiction) increases the probability of a response. Different theoretical perspectives, from experimental psychology (positive and negative reinforcement frameworks), social psychology (self-regulation failure framework), and neurobiology (counteradaptive and sensitization framework), can be superimposed on the stages of the addiction cycle [1••]. These stages are thought to feed into each other, become more intense, and ultimately lead to the pathological state known as *addiction*.

In this context, my argument is that drug addiction involves elements of both impulse control disorders and compulsive disorders. In addition to the impulsivity of the early stages of the addiction process, an additional source of motivation is recruited: negative reinforcement. The development of the aversive emotional state that drives the negative reinforcement of addiction is defined here as the 'dark side' of addiction. Subjects with classic atypical impulse control disorders, such as kleptomania, experience an increasing sense of tension or arousal before committing an impulsive act, pleasure, gratification or relief at the time of committing the act, and regret, self-reproach, or guilt following the act [2]. In contrast, subjects with classic compulsive-like disorders, such as obsessive compulsive disorder, experience anxiety and stress before committing a compulsive repetitive behavior and relief from the stress by performing the compulsive behavior [2]. We have argued that drug addiction progresses from a source of positive reinforcement that may indeed involve more elements of impulsivity (defined behaviorally as 'actions which are poorly conceived, *prematurely expressed*, unduly risky, or inappropriate to the situation and that often result in undesirable consequences'; [3]) to sensitization of the brain stress and antireward systems that may involve more elements of compulsivity (defined as actions inappropriate to the situation *that persist*, have no obvious relationship to the overall goal, and which often result in undesirable consequences), both of which set up a powerful negative reinforcement process.

My thesis is that addiction involves long-term, persistent dysregulation of the activity of neural circuits that mediate motivational systems, deriving from two sources: decreased function of the brain reward systems that normally mediate natural rewards and recruitment of brain stress and antireward systems that drive aversive states. Antireward is a concept developed by Koob and Le Moal [4••], based on the hypothesis that brain systems are in place to limit reward (see footnote in [1••]), with an opponent process concept that forms a general feature of biological systems. Our concept of an antireward system is derived from the hypothesis of both within-system and between-system neuroadaptations to excessive activation of the reward system at the neurocircuitry level. Within-system neuroadaptations are defined as the process by which the primary cellular response element to the drug (circuit A) itself adapts to neutralize the drug's effects. Persistence of the opposing effects after the drug disappears produces adaptation. A between-system neuroadaptation is a circuitry change, in which circuit B (i.e., the antireward circuit) is activated, opposing the action of circuit A, potentially at multiple levels (i.e., the reward circuit). More recently,

we hypothesized that a within-system neuroadaptation can directly result from a between-system neuroadaptation, in which circuit B (i.e., the antireward circuit) is activated by excessive activity of circuit A and in turn suppresses the activity of circuit A.

For the *binge/intoxication stage* of the addiction cycle, studies of the acute reinforcing effects of drugs of abuse *per se* have identified key neurobiological substrates. Evidence is strong for a role for dopamine in the acute reinforcing actions of psychostimulants, opioid peptide receptors in the acute reinforcing effects of opioids, γ-aminobutyric acid and opioid peptides in the acute reinforcing actions of alcohol, and dopamine and glutamate in the enhanced incentive salience of repeated drug taking. Important anatomical circuits include the mesocorticolimbic dopamine system that originates in the ventral tegmental area and projects to the nucleus accumbens, opioid peptides in the ventral striatum, extended amygdala, and ventral tegmental area, and glutamate in the prefrontal cortex. However, acute drug use, even binge drug use, is not addiction. We argue that activation of reward and incentive salience circuitry plays a prominent role in this stage of the addiction process and persists in the *preoccupation/anticipation* stage but misses a key element of the addiction process — the *withdrawal/negative affect* stage.

For the present commentary, the *withdrawal/negative affect* stage can be defined as the presence of motivational signs of withdrawal in humans, including chronic irritability, physical pain, emotional pain (i.e., hyperkatifeia; [5]), malaise, dysphoria, alexithymia, and loss of motivation for natural rewards. It is characterized in animals by increases in reward thresholds during withdrawal from all major drugs of abuse. More compelling, in animal models of the transition to addiction, similar changes in brain reward thresholds occur that temporally precede and highly correlate with escalation in drug intake ([6–8]; Figure 1). Such acute withdrawal is associated with decreased activity of the mesocorticolimbic dopamine system, reflected by electrophysiological recordings and *in vivo* microdialysis [9,10•,11]. Human imaging studies of individuals with addiction during withdrawal or protracted abstinence have generated results that are consistent with such animal studies. There are decreases in dopamine D_2 receptors (hypothesized to reflect hypodopaminergic functioning), hyporesponsiveness to dopamine challenge [12], and hypoactivity of the orbitofrontal-infralimbic cortex system [12]. These are hypothesized to be within-system neuroadaptations that may reflect presynaptic release or postsynaptic receptor plasticity.

More importantly for the present thesis, as dependence and withdrawal develop, brain antireward systems, such as corticotropin-releasing factor (CRF), norepinephrine, and dynorphin, are recruited in the extended amygdala. The extended amygdala is composed of the central nucleus of the amygdala, bed nucleus of the stria terminalis, a transition area in the medial (shell) part of the nucleus accumbens, and a massive projection to the lateral hypothalamus. For example, extracellular CRF in the extended amygdala is increased during acute withdrawal from all drugs of abuse. Critically, CRF receptor antagonists injected into the extended amygdala block the anxiety-like effect of withdrawal from all drugs of abuse and blunt excessive drug taking during escalated drug taking with extended access to all drugs of abuse [13,14]. We hypothesize that the brain stress neurotransmitter CRF known to be activated during the development of excessive drug taking comprises a between-system opponent process, and this activation is manifest when the drug in removed (Figure 2), producing anxiety, hyperkatifeia, and irritability symptoms associated with acute and protracted abstinence.

Blockade of the κ opioid system can also block the aversive stimulus effects of drug withdrawal and stress [15••,16,17••,18–22,23•]). Even more compelling is that excessive drug self-administration can be blocked by κ antagonists [24–27] and may be mediated by the shell of the nucleus accumbens [28]. These results suggest a within/between-system neuroadaptation that was originally hypothesized by Carlezon and Nestler [29], in which activation of cyclic adenosine monophosphate response element binding protein by excessive dopamine and opioid peptide receptor activation in the nucleus accumbens (within-system neuroadaptation) triggers the induction of dynorphin (between-system neuroadaptation) to feed back to suppress dopamine release and glutamate release (Figure 2). Thus, our hypothesis is that antireward circuits are recruited as between-system neuroadaptations [30•] during the development of addiction, producing aversive or stress-like states [13,31•,32] via two mechanisms: direct activation of stress-like, fear-like states in the extended amygdala (CRF-norepinephrine) and indirect activation by suppressing dopamine (dynorphin).

A critical problem in drug addiction is chronic relapse, in which addicted individuals return to compulsive drug taking long after acute withdrawal. This corresponds to the *preoccupation/anticipation stage* of the addiction cycle outlined above. Koob and Le Moal also hypothesized that the dysregulations that comprise the 'dark side' of drug addiction persist during protracted abstinence to set the tone for vulnerability to 'craving' by activating drug-induced, cue-induced, and stress-induced reinstatement neurocircuits that are now driven by a reorganized and possibly hypofunctioning prefrontal system [33]. For example, converging lines of evidence suggest that impairment of medial prefrontal cortex cognitive function and overactivation of the central nucleus of the amygdala