



Published in final edited form as:

Psychopharmacology (Berl). 2019 January ; 236(1): 407–414. doi:10.1007/s00213-018-5098-y.

The emerging neuroscience of appetitive and drug cue extinction in humans

Anna B. Konova¹ and Rita Z. Goldstein²

¹Department of Psychiatry and University Behavioral Health Care (UBHC), Rutgers University–New Brunswick, Piscataway, NJ

²Departments of Psychiatry and Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY

Abstract

Fear extinction has been extensively studied in both humans and non-human animals, and this work has contributed greatly to our understanding and treatment of anxiety disorders. Yet other psychopathologies like addiction might be associated with impairments selectively in extinction of non-fear based, *appetitive and drug cue associations*, and these processes have been underexplored in clinical translational neuroscience. Important questions regarding similarities and differences in the neurobiological mechanisms underlying aversive and appetitive extinction remain unanswered, particularly those pertaining to cross-species evidence for the role of the ventromedial prefrontal cortex and, to some extent, the striatum. Here, we aim to draw attention to the paucity of studies investigating non-fear based extinction in humans, summarize emerging findings from the available literature, and highlight important directions for future research. We argue that closing these gaps in our understanding could help inform the development of more targeted, and perhaps more durable, forms of extinction-based treatments for addiction and related psychopathologies characterized by abnormally persistent appetitive and drug cue associations.

Keywords

addiction; appetitive; drug cue; reward; extinction; learning; fMRI; ventromedial prefrontal cortex; striatum; amygdala

Extinction is a well-known phenomenon whereby conditioned responses to an otherwise neutral cue (a conditioned stimulus, CS) that has acquired affective properties after being paired with an arousing event (an unconditioned stimulus, US) gradually diminish when the cue is no longer reinforced (Bouton 2004; Quirk and Mueller 2008). The predominant view is that extinction does not eliminate the original CS–US association; rather, it leads to a lessening in the conditioned response by creating a new (CS–no–US) association (an

Correspondence: Correspondence may be addressed to Anna B. Konova, PhD (Department of Psychiatry and University Behavioral Health Care (UBHC), Rutgers University–New Brunswick, 671 Hoes Lane West, Piscataway, NJ 08854, USA. anna.konova@rutgers.edu) or Rita Z. Goldstein, PhD (Departments of Psychiatry and Neuroscience, Icahn School of Medicine at Mount Sinai, The Leon and Norma Hess Center for Science and Medicine, 1470 Madison Ave., New York, NY 10029, USA. rita.goldstein@mssm.edu).

Conflicts of Interest: The authors declare that they have no conflicts of interest.

“extinction memory”) that competes for expression, leaving the organism vulnerable to recovery of the conditioned response. The neurobiology of aversive and fear extinction, e.g., learning that a light no longer predicts electric shock, has been extensively studied in both humans (Fullana et al. 2018; Sehlmeier et al. 2009) and non-human animals (Milad and Quirk 2012; Quirk and Mueller 2008) and this work has informed much of our understanding and treatment of anxiety disorders which are characterized by excessive fear responses and deficits in extinguishing these responses (Graham and Milad 2011). Yet other psychopathologies like addiction and eating disorders, which are similarly theorized to have etiologies stemming from failure in extinction (Bouton 2011), may be impaired selectively in extinction of non-fear based, appetitive and drug cue associations, and these processes have been underexplored in clinical translational neuroscience. Much less is known about extinction in these other domains, and while preclinical work suggests similarities in the neurobiological mechanisms of extinction of, e.g., drug seeking and fear (Peters et al. 2009), cross-domain translation is limited by important differences in experimental design across studies while cross-species translation is limited primarily by the paucity of studies investigating the mechanisms of non-fear based extinction in humans. Addressing these limitations could contribute to our basic understanding of extinction mechanisms as well as the development of improved extinction-based approaches to changing drug use (Taylor et al. 2009) and unhealthy eating (Jansen et al. 2016) behaviors triggered by cue exposure. These neuroscience-informed approaches could, for example, aim to rely less on neural circuitry impacted by addiction pathology and, instead, harness recent developments in the fear extinction literature of more durable forms of extinction training.

Neurobiological mechanisms of aversive extinction

The process of extinction can be broken down into three phases: acquisition (initial reduction in the conditioned response), consolidation (stabilization of extinction memory, occurring over hours), and retrieval of extinction learning (recall or expression of extinction memory, typically assessed 24 hours later). As reviewed in (Milad and Quirk 2012; Quirk and Mueller 2008), data in rodents broadly suggest that acquisition and consolidation of fear extinction require plasticity in the structures involved in the expression of fear (namely, basolateral amygdala), whereas extinction recall requires the infralimbic (IL) cortex, the most ventral aspect of the rodent medial prefrontal cortex [although see (Do-Monte et al. 2015) who draws a distinction between facilitating storage of extinction in target structures vs. retrieval of extinction memory]. More specifically, extinction recall, like extinction acquisition and consolidation, depends on local inhibition within the amygdala, and this inhibition may be driven by IL inputs to the amygdala. By contrast, another medial prefrontal structure in the rodent brain that is situated just dorsal to the IL area—the prelimbic (PL) cortex—instead facilitates fear-related amygdala activity. Thus, the IL cortex appears to selectively support fear extinction memory expression, whereas the PL cortex promotes fear responses.

The functional homologue of the rodent IL region in humans is typically identified as the ventromedial prefrontal cortex (VMPFC). The VMPFC is a large, heterogeneous swath of the human prefrontal cortex, extending from the anterior cingulate cortex (ACC) anterior to the genu of the corpus callosum to the medial orbitofrontal cortex, inclusive, and

encompassing Brodmann areas 25, ventral portions of 24 and 32, medial portion of 11, and ventral and medial portions of 10 (Mackey and Petrides 2014). As summarized in (Fullana et al. 2018; Sehlmeier et al. 2009), a central role for the VMPFC in fear extinction has indeed received empirical support in human imaging studies. Using magnetic resonance imaging (MRI), this previous work suggests that VMPFC activity increases during fear extinction (Milad et al. 2007) and extinction retrieval (Kalisch et al. 2006; Phelps et al. 2004), and that both neural activity (Phelps et al. 2004) and cortical thickness (Hartley et al. 2011; Milad et al. 2005) in this region correlate with indices of extinction success (e.g., lowered psychophysiological response to the CS). Most recently, a meta-analysis of neuroimaging studies in healthy human subjects found evidence for selective involvement of the VMPFC in fear extinction recall (Fullana et al. 2018). Further, also in support of animal findings, the broader pattern of activity observed suggested that some circuits implicated in fear extinction overlapped with those implicated in fear conditioning, such as the dorsal ACC [the presumed homologue of the rodent PL region in humans (Milad et al. 2006), although see (Balleine and O'Doherty 2010; Uylings et al. 2003)] and extended salience network regions such as the anterior insula, among others. But as expected, activity in these regions was less robust during extinction than during conditioning, paralleling the changing CS contingencies from conditioning to extinction. Strikingly, however, the authors did not observe significant activation in the amygdala during any extinction phase, which they also failed to observe in a previous meta-analysis of fear acquisition (Fullana et al. 2016). Thus, while work in humans in the aversive domain supports differential roles of the VMPFC and dorsal ACC in fear extinction which parallels findings in rodents for the IL and PL cortex, respectively, findings pertaining to the amygdala are somewhat mixed and may be influenced by methodological limitations associated with imaging this structure.

Neurobiological mechanisms of appetitive and drug cue extinction in rodents: Similarities and differences to aversive extinction

In rodents, the neural circuits that support extinction of conditioned drug reinforcement—the most studied form of non-fear based extinction—have been found to mostly overlap with fear extinction circuits (Millan et al. 2011; Peters et al. 2009) (Figure 1). However, some differences in the regions that form the core of an ‘extinction circuit’ and the nature of their circuit interactions are observed as well, suggesting there is a critical need to study extinction across the aversive and non-aversive domains in humans. This work could not only contribute to our basic neuroscientific understanding of the precise circuit mechanisms of extinction learning and their potential modulation by e.g., stimulus identity or valence, but could also lead to an improved understanding and treatment of psychopathologies characterized by abnormally persistent appetitive and drug cue associations such as drug addiction.

First, the rodent literature is in agreement that, as in fear conditioning, the PL cortex supports expression of conditioned drug seeking responses. Second, this literature suggests that the IL cortex is a candidate locus of extinction learning and recall across aversive and non-aversive extinction including extinction of drug cue associations. On the other hand, while its role in aversive processing is less well understood [although see e.g., (Raczka et al.

2011)], the ventral striatum (where the nucleus accumbens is located) is thought to be critical for both the acquisition and expression of appetitive extinction (Millan et al. 2011). In particular, while the nucleus accumbens core promotes drug seeking responses, the nucleus accumbens shell is thought to *inhibit* these responses (Peters et al. 2008). In this sense, preferential IL excitatory connections to the nucleus accumbens shell (versus nucleus accumbens core) (Haber and Knutson 2010) could underlie expression of drug cue (and appetitive) extinction and inhibition of conditioned approach responses. Also, in contrast to the degree of specificity found in the aversive domain, and perhaps because of the extensive inter-connectedness with both nucleus accumbens core and shell, the IL cortex does not appear to selectively support appetitive extinction recall. Instead, this region has been found to be necessary for both the expression of reward cue driven behavior (Bossert et al. 2011; Bossert et al. 2012), similar to other prefrontal regions such as the neighboring orbitofrontal cortex [e.g., (Moorman and Aston-Jones 2014)], and its extinction (LaLumiere et al. 2010; Peters et al. 2008; Rhodes and Killcross 2007; Rhodes and Killcross 2004). This “dual function” could arise from the simultaneous existence of separate, but intermingled, neural ensembles within the IL region that selectively encode conditioned appetitive and extinction memories (Van den Oever et al. 2013).

However, although tempting to conclude that there are clear points of similarity (e.g., role of the PL cortex, importance of the IL cortex) and difference (e.g., importance of the nucleus accumbens, less selective role of the IL cortex) across rodent studies of aversive and non-aversive extinction, several important confounds should be considered. First, most of the non-fear extinction work has been conducted in the context of preclinical models of addiction. This can obscure similarities to findings in the fear domain by virtue of the presence of pathological brain changes resulting from the effects of drugs of abuse or the development of an addiction phenotype. Indeed, in taking a closer look at the above reviewed work, greater concordance is observed across domains regarding the function of the IL cortex when we limit our analysis to studies that utilize food rewards in otherwise ‘healthy’ rodents (Rhodes and Killcross 2007; Rhodes and Killcross 2004). On the other hand, differences related to stimulus identity or valence can be obscured in these models by the fact that drug predicting cues can elicit both appetitive and aversive responses in pathological states. Other important differences across domains in experimental design also require careful consideration, in particular the requirement for instrumental responding typical of appetitive and drug cue extinction studies but not fear extinction studies that instead tend to utilize classical Pavlovian conditioning. This difference can have profound effects on the broader circuits that might be engaged [for example, parts of the striatum have been implicated in forming stimulus-response-outcome associations and Pavlovian-to-instrumental transfer (Liljeholm and O’Doherty 2012)]. These issues are not unique to rodent work and apply broadly to studies in humans, which we discuss next.

Neurobiological mechanisms of appetitive and drug cue extinction in humans: Emerging findings

Although much progress has been made in the aversive domain in translating rodent work to humans and to psychopathologies characterized by deficits in aversive extinction (Milad et

al. 2014), only a handful of studies in humans have directly examined the neural mechanisms of non-fear based extinction, and even fewer have examined how these mechanisms compare across the aversive and non-aversive domains and across health and psychopathology. To date, only two human imaging studies have directly examined the neural correlates of appetitive extinction in healthy subjects. In the first of these studies, Kruse et al. (2017) combined functional MRI with a monetary incentive delay task in which, during an initial acquisition phase, subjects learned to associate abstract cues with monetary gain ($CS_{\text{APPETITIVE}^+}$) or no gain (CS^-). During extinction, which took place 24 hours later, subjects performed the same task while neither cue was reinforced. Acquisition of the $CS_{\text{APPETITIVE}^+}$ was associated with increased activity (relative to the CS^-) in the ventral striatum and VMPFC, as well as the amygdala and dorsal ACC. Much of the same regions remained engaged during early extinction, eventually diminishing in activity as extinction progressed. Relative exceptions were the ventral striatum and amygdala which continued to respond differentially to the $CS_{\text{APPETITIVE}^+}$. Interestingly, and in line with imaging work on fear extinction, *higher* VMPFC activity during early extinction correlated with greater reductions in subjective arousal post-extinction.

In the second study in healthy subjects, utilizing a similar monetary incentive delay paradigm with both appetitive and aversive outcomes, Ebrahimi et al. (2017) tested if the glutamate agonist d-cycloserine can enhance appetitive extinction recall similar to findings in the aversive domain, and if this enhancement is supported by effects of the drug on appetitive extinction circuits. Over three days, subjects learned to first associate abstract cues with monetary gain ($CS_{\text{APPETITIVE}^+}$), loss (CS_{AVERSIVE^+}), or a neutral outcome (CS^-). Extinction training took place 24 hours later (day 2) following either d-cycloserine or placebo administration. Finally, extinction recall was assessed another 24 hours later (day 3). Collapsing across the drug conditions, acquisition of the $CS_{\text{APPETITIVE}^+}$ was associated with increased activity (relative to the CS^-) in only one region: the hippocampus. Extinction training was instead associated with increased activity in the dorsal ACC, and extinction recall with increased activity in the amygdala, with the latter observed only in subjects receiving placebo. To further interrogate extinction recall of the $CS_{\text{APPETITIVE}^+}$ and its underlying circuit involving the amygdala, the authors conducted a psychophysiological interaction analysis. While in the placebo group, amygdala connectivity was observed with a diffuse set of regions, in the d-cycloserine group stronger amygdala connectivity was observed selectively with the VMPFC, which the authors interpreted as possibly underlying the extinction recall enhancing effects of the drug observed in prior work. Thus, while the VMPFC was not directly observed in this study in the context of appetitive extinction, its involvement may be revealed indirectly through interactions with the amygdala (at least while receiving a glutamatergic agonist).

Extending this work to other affective stimuli (emotional pictures), we recently examined extinction learning in a psychopathology known to adversely impact VMPFC function and structure—cocaine addiction (Konova et al. 2017). For an in-depth discussion of the effects of addictive substances and addiction on the prefrontal cortex, see (Goldstein and Volkow 2011; Konova and Goldstein 2015). A defining feature of addiction is continued drug seeking and use despite reduced pleasure derived from the drug and negative personal and social consequences. This behavior is assumed to be at least partly driven by a failure in

extinction learning such that addicted individuals may have diminished ability to form and/or maintain new associations for cues or situations that were previously, although no longer, predictive of drug rewards. Characterizing the neurobiology of appetitive and drug cue extinction can therefore provide a deeper understanding of what is widely believed to be a central underlying dysfunction in this psychopathology. Here, healthy and cocaine addicted subjects first learned to associate an abstract cue with a drug-related (CS_{DRUG+}), pleasant ($CS_{APPETITIVE+}$), or a neutral ($CS-$) image. Two extinction phases followed, the former immediately after and the latter conducted 24 hours later (day 2) to assess extinction recall from day 1. We expected that, in controls, the CS_{DRUG+} would produce patterns of activity in the VMPFC consistent with aversive processing, while the $CS_{APPETITIVE+}$ (CS paired with a pleasant image) could be used to probe appetitive processing. These hypotheses were based on offline explicit and implicit preference measures that revealed that both controls and cocaine users found the drug-related image used as the “US” as least pleasant, and the image used as the affectively pleasant “US” as most pleasant. Controls additionally found the drug-related image more unpleasant than cocaine users. Defining a functional region of interest in the VMPFC derived from a prior fear extinction study (Phelps et al. 2004), our main finding was a cue type \times learning phase \times diagnostic group interaction in the left VMPFC (Figure 2A). This three-way interaction was explained by differences over the learning phases in response to the CS_{DRUG+} versus $CS_{APPETITIVE+}$ in controls but not cocaine users. As in fear extinction studies in humans, in controls, VMPFC activation tended to be higher during extinction and extinction recall for the CS_{DRUG+} (which was unpleasant), and higher than in cocaine users (significantly during recall). However, it was *lower* during extinction for the $CS_{APPETITIVE+}$ such that parametric reductions in VMPFC activation were observed from acquisition to extinction to extinction recall on day 2. In contrast, in cocaine users, there was no systematic shift in VMPFC response by cue type as extinction progressed. Supporting a specific role in extinction *recall* for drug cue associations as previously found for fear (Phelps et al. 2004), across groups, the magnitude of VMPFC response to the CS_{DRUG+} on day 2 correlated with the success of extinction learning on day 1, as indexed by a reduction in skin conductance response to the CS_{DRUG+} and consistent with activity in this region reflecting expression of the extinction memory (Figure 2B). Furthermore, cocaine users with higher VMPFC responses to the CS_{DRUG+} during extinction training on day 1 (i.e., who looked more like controls) reported greater reductions in cocaine craving, including that triggered by environmental cues, on day 2 relative to day 1 (Figure 2C). Together, these data suggest that the same VMPFC region reported to be engaged during extinction of fear also supports extinction of drug cue and appetitive associations (at least in health). Interestingly, we found a pattern of activation in the striatum in both groups that mirrored that of the VMPFC in controls: striatum responses increased for the CS_{DRUG+} and decreased for the $CS_{APPETITIVE+}$ as extinction progressed. The reverse pattern (decrease for the CS_{DRUG+} and increase for the $CS_{APPETITIVE+}$) was observed in the amygdala in more exploratory whole-brain analyses.

Taken together with fear extinction studies, this initial study in addicted individuals and the two aforementioned studies in healthy subjects provide preliminary evidence that the VMPFC, dorsal ACC, striatum, and amygdala may support extinction learning across the aversive and non-aversive domains; however, the specific activity profiles engaged may

depend on the extinction phase, stimulus identity/valence, and other experimental design features. More specifically, in our affective pictures study, we observed decreased VMPFC response to the CS_{APPETITIVE+} during extinction recall (relative to acquisition), a pattern opposite to that found for the CS_{DRUG+} and extinction of fear. Kruse et al. (2017) instead found significantly positive VMPFC responses across acquisition and extinction using monetary rewards. However, these studies cannot be directly compared because Kruse et al. (2017) did not test extinction recall; the VMPFC effect observed in that study was limited to early extinction, which as the authors caution, could index initial acquisition of extinction, recall of the appetitive association from day 1, or both. It is not possible to ascertain how VMPFC response patterns might have changed during extinction recall. Ebrahimi et al. (2017) did not observe changes in VMPFC response despite having a recall phase; rather, in this study, a role for the VMPFC was revealed through circuit-level interactions with the amygdala. It is worth noting that extinction recall here took place after three “reactivation trials” (paired presentations of the CS with the corresponding US), and the effect of this methodological design feature remains uncertain. Further, while Ebrahimi et al. (2017) included a dedicated CS_{AVERSIVE+}, the other two studies did not. However, as the authors did not directly compare findings across CS types (CS_{APPETITIVE+} vs. CS_{AVERSIVE+}), the extent of similarity or difference (both qualitative and quantitative) in the VMPFC response across domains remains to be fully tested. Finally, only one of the three studies included drug addicted individuals and, in this group, VMPFC (but not e.g., striatum) response patterns differed from that of controls. This could reflect specific pathological brain changes but also features of the conditioned stimuli. As mentioned before, drug cues elicit both appetitive and aversive responses in chronic drug users. Thus, further work is needed using appetitive, aversive, and drug-related cues in both healthy and addicted subjects to determine the independent effects of diagnosis and cue type. Parallel work in animal models may also be uniquely able to determine the causal effect of drug use history.

The VMPFC may work in concert with the striatum and amygdala during appetitive (and drug cue) extinction, as indeed supported by findings from all three studies. Activity in the striatum paralleled that of the VMPFC in both our study and in Kruse et al. (2017), which could reflect a role for this region in both acquisition and extinction consistent with rodent work. In both, however, subjects were asked to execute a motor response (button pressing) and in Kruse et al. (2017) this response was instrumental. Thus, striatum activity during the CS_{APPETITIVE+} could signal that a response will be rewarded or that a reward is coming. Some of this concern is partly alleviated by the fact that subjects did not appear to acquire a conditioned behavioral response, which would be reflected in more accurate and/or faster button pressing to the CS_{APPETITIVE+} and/or during acquisition relative to extinction. The impact of having to exert a motor response at all however is not known. Future studies will need to use more potent reinforcers to minimize this important difference in experimental design across domains, as we discuss in more detail in the next section. These studies also reveal an unexpected finding in the amygdala during appetitive extinction and extinction recall. In rodents, the amygdala is found to have mutually excitatory connections with the ventral striatum (Stuber et al. 2011; Wassum and Izquierdo 2015), which could serve as a potential mechanism mediating its role in appetitive extinction. Indeed, human imaging studies indicate that the amygdala represents the full range of valence from unpleasant to

pleasant (Jin et al. 2015), together suggesting that these amygdala findings should be explored further in relation to the striatum and the VMPFC in appetitive extinction.

Considerations for translational work on appetitive and drug cue extinction, outstanding questions, and future directions

Overall, the picture that emerges is that comparison of extinction mechanisms across the aversive and non-aversive domains is limited by important differences in experimental design features across studies, and cross-species translation is severely limited by the paucity of neuroimaging studies in humans. Addressing these limitations in future work is thus of top priority. More generally human appetitive extinction studies should follow recommendations for the design of human fear extinction studies (Lonsdorf et al. 2017). From a practical standpoint, to address critical differences in the requirement for instrumental responding, future studies (at least in humans) may need to use more potent appetitive reinforcers matched in salience to the aversive reinforcers (e.g., using choice indifference procedures). For example, these studies could use appetitive and aversive odor or taste stimuli offered in different concentrations, which have been successfully used in Pavlovian style tasks [e.g., (Bray et al. 2008)]. The use of primary reinforcers (rather than money or more abstract picture stimuli) could also more directly facilitate cross-species translation, which has been a major advantage of fear conditioning paradigms. Alternatively, or in addition, studies in humans could leverage more naturalistic stimuli, particularly those tagging context, delivered for example through immersive virtual reality technology (Kroes et al. 2017), to help facilitate translation of findings from the lab to the clinic (or to other real-world behavior). In addition to using parallel task paradigms across domains, future studies particularly in rodents should examine appetitive extinction processes in ‘healthy’ non-addicted and non-deprived states, to disentangle the effects of pathological state from valence that have often been conflated in previous research.

Other outstanding issues include determining the underlying mechanism through which the VMPFC supports extinction learning and how this knowledge can be leveraged in treatment development and treatment tailoring for addiction. The human VMPFC is thought to subservise diverse functions, which may need to be considered in seeking to understand its role in extinction learning (Delgado et al. 2016; Schneider and Koenigs 2017). For example, an unanswered question pertains to whether the VMPFC has a fundamentally inhibitory role in extinction learning, serving to suppress the original CS-US association in favor of the extinction memory. The prevailing viewpoint stemming from findings in rodents and fear extinction studies in humans is that the VMPFC contributes to learning by inhibiting maladaptive affective responses, regardless of valence, as reflected in increased activity in the VMPFC during extinction and its recall. However, this viewpoint is difficult to reconcile with other findings in rodents and with a range of studies in humans including the imaging findings described in this perspective. The broader role of the VMPFC in subjective valuation and decision making (Bartra et al. 2013) suggests that the VMPFC could instead serve to encode the current value of the CS. For example, in the context of fear extinction, activity in the VMPFC could signal changes in the safety (positive value) of the CS+ (Fullana et al. 2016; Harrison et al. 2017). VMPFC activity may be initially suppressed by

the fear generating CS+ relative to the CS– during conditioning, a difference that gradually diminishes over the course of extinction learning as the CS+ becomes less threatening/safer and thus more positively valenced. Preliminary support for this perspective comes from work on cognitive regulation, an emotion regulation strategy thought to depend on overlapping circuitry (Schiller and Delgado 2010), which suggests that rather than tracking the engagement of regulatory processes, the VMPFC represents updated value *subsequent* to regulation (Winecoff et al. 2013). In this case, the experience of positive emotion was found to activate the VMPFC, whereas the regulation of positive emotion decreased VMPFC activation (a finding that was not anticipated based on prior work on regulation of negative emotion). More recently, these diverging accounts of VMPFC function were tested in another study using reversal learning (Zhang et al. 2015), a paradigm in which CS-US contingencies switch such that an initially rewarded cue becomes non-rewarded while the reverse is true for a different, initially neutral cue. The inhibition hypothesis predicts that VMPFC activity should increase to inhibit an inappropriate ‘reward’ response during reversal, despite the reduced value of the no-longer-rewarded stimulus, which the valuation hypothesis predicts should instead be associated with decreased VMPFC activity. Interestingly, the authors observed VMPFC activity profiles consistent with both hypotheses, although in partly distinct VMPFC subregions. Thus, it is still possible that distinct groups of neurons or subregions within the VMPFC support inhibition versus value representation, and possibly aversive versus appetitive extinction—a hypothesis that could be tested in parallel rodent work.

In addition to the need for further studies interrogating the neural correlates of appetitive extinction with brain imaging in humans, at least three other future directions should be considered. First, as with fear extinction work in anxiety disorders, more studies are needed to examine the nature of abnormalities in appetitive and drug cue extinction learning across addiction, gambling, and eating disorders. One hypothesis that will be important to test is if differences in the neurobiology of these extinction processes are universally present or if they are specific to the phenomenological characteristics of each disorders (e.g., Would gamblers show selective impairments in extinction of gambling cues?). This could prove critical in identifying novel ways to facilitate the effectiveness of extinction manipulations in specific disorders (e.g., addiction) or specific domains leaving others untouched (e.g., based on our findings in cocaine users, we may wish to harness the relatively intact striatum or provide effective scaffolding to the impaired function in the VMPFC in developing extinction-based treatments for addiction). Second, given that drug cue extinction as a treatment for addiction often fails in the clinic (Conklin and Tiffany 2002; Torregrossa and Taylor 2013), these studies will also need to simultaneously leverage developments in the literature on fear extinction in humans and appetitive extinction in rodents that have identified viable means to enhance the standard extinction approach. This can be achieved by supplementing standard extinction training with pharmacological agents that target appetitive extinction circuitry or by directly stimulating these circuits (e.g., with TMS) and examining how/whether these manipulations change the neural implementation of extinction learning. One intuitive hypothesis to test is whether changing appetitive extinction-specific activity changes appetitive extinction recall, and if this effect is specific or generalizable to other domains. In addition, future studies in humans should examine alternative, more

durable formulations of extinction learning and the neural circuits that support their effectiveness across the aversive and appetitive domains such as post-retrieval extinction, more ‘gradual’ extinction, and novel neuroimaging-based approaches such as unconscious extinction facilitated by neurofeedback [e.g., (Taschereau-Dumouchel et al. 2018)]. These latter complementary investigations can help identify the boundary conditions of traditional approaches and motivate translation to the clinic and to a broader range of healthy and psychiatric populations.

Conclusions

Further work is clearly needed to understand extinction—at a behavioral and neural level—in *humans in the context of appetitive, drug, and other disease-related cues*, an understanding that we argue could inform treatment development in psychopathologies thought to be associated with deficits in non fear-based extinction learning such as addiction and eating disorders.

Acknowledgments

Acknowledgments: The authors were supported by grants from the National Institute on Drug Abuse (F32DA039648 to A.B.K. and R01DA041528 and R21DA020626 to R.Z.G.).

References

- Balleine BW, O'Doherty JP (2010) Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* 35: 48–69. [PubMed: 19776734]
- Bartra O, McGuire JT, Kable JW (2013) The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* 76: 412–27. [PubMed: 23507394]
- Bossert JM, Stern AL, Theberge FR, Cifani C, Koya E, Hope BT, Shaham Y (2011) Ventral medial prefrontal cortex neuronal ensembles mediate context-induced relapse to heroin. *Nat Neurosci* 14: 420–2. [PubMed: 21336273]
- Bossert JM, Stern AL, Theberge FR, Marchant NJ, Wang HL, Morales M, Shaham Y (2012) Role of projections from ventral medial prefrontal cortex to nucleus accumbens shell in context-induced reinstatement of heroin seeking. *J Neurosci* 32: 4982–91. [PubMed: 22492053]
- Bouton ME (2004) Context and behavioral processes in extinction. *Learn Mem* 11: 485–94. [PubMed: 15466298]
- Bouton ME (2011) Learning and the persistence of appetite: extinction and the motivation to eat and overeat. *Physiol Behav* 103: 51–8. [PubMed: 21134389]
- Bray S, Rangel A, Shimojo S, Balleine B, O'Doherty JP (2008) The neural mechanisms underlying the influence of pavlovian cues on human decision making. *J Neurosci* 28: 5861–6. [PubMed: 18509047]
- Conklin CA, Tiffany ST (2002) Applying extinction research and theory to cue-exposure addiction treatments. *Addiction* 97: 155–67. [PubMed: 11860387]
- Delgado MR, Beer JS, Fellows LK, Huettel SA, Platt ML, Quirk GJ, Schiller D (2016) Viewpoints: Dialogues on the functional role of the ventromedial prefrontal cortex. *Nat Neurosci* 19: 1545–1552. [PubMed: 27898086]
- Do-Monte FH, Manzano-Nieves G, Quinones-Laracuenta K, Ramos-Medina L, Quirk GJ (2015) Revisiting the role of infralimbic cortex in fear extinction with optogenetics. *J Neurosci* 35: 3607–15. [PubMed: 25716859]

- Ebrahimi C, Koch SP, Friedel E, Crespo I, Fydrich T, Strohle A, Heinz A, Schlagenhauf F (2017) Combining D-cycloserine with appetitive extinction learning modulates amygdala activity during recall. *Neurobiol Learn Mem* 142: 209–217. [PubMed: 28512009]
- Fullana MA, Albajes-Eizagirre A, Soriano-Mas C, Vervliet B, Cardoner N, Benet O, Radua J, Harrison BJ (2018) Fear extinction in the human brain: A meta-analysis of fMRI studies in healthy participants. *Neurosci Biobehav Rev* 88: 16–25. [PubMed: 29530516]
- Fullana MA, Harrison BJ, Soriano-Mas C, Vervliet B, Cardoner N, Avila-Parcet A, Radua J (2016) Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. *Mol Psychiatry* 21: 500–8. [PubMed: 26122585]
- Goldstein RZ, Volkow ND (2011) Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nature reviews Neuroscience* 12: 652–69. [PubMed: 22011681]
- Graham BM, Milad MR (2011) The study of fear extinction: implications for anxiety disorders. *Am J Psychiatry* 168: 1255–65. [PubMed: 21865528]
- Haber SN, Knutson B (2010) The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology* 35: 4–26. [PubMed: 19812543]
- Harrison BJ, Fullana MA, Via E, Soriano-Mas C, Vervliet B, Martinez-Zalacain I, Pujol J, Davey CG, Kircher T, Straube B, Cardoner N (2017) Human ventromedial prefrontal cortex and the positive affective processing of safety signals. *Neuroimage* 152: 12–18. [PubMed: 28254509]
- Hartley CA, Fischl B, Phelps EA (2011) Brain structure correlates of individual differences in the acquisition and inhibition of conditioned fear. *Cerebral cortex* 21: 1954–62. [PubMed: 21263037]
- Jansen A, Schyns G, Bongers P, van den Akker K (2016) From lab to clinic: Extinction of cued cravings to reduce overeating. *Physiol Behav* 162: 174–80. [PubMed: 26994737]
- Jin J, Zelano C, Gottfried JA, Mohanty A (2015) Human Amygdala Represents the Complete Spectrum of Subjective Valence. *J Neurosci* 35: 15145–56. [PubMed: 26558785]
- Kalisch R, Korenfeld E, Stephan KE, Weiskopf N, Seymour B, Dolan RJ (2006) Context-dependent human extinction memory is mediated by a ventromedial prefrontal and hippocampal network. *J Neurosci* 26: 9503–11. [PubMed: 16971534]
- Konova AB, Goldstein RZ (2015) Addiction and Addiction Treatment. *The Wiley Handbook on the Cognitive Neuroscience of Addiction*: 109.
- Konova AB, Parvaz MA, Bernstein V, Zilverstand A, Moeller SJ, Delgado MR, Alia-Klein N, Goldstein RZ (2017) Neural mechanisms of extinguishing drug and pleasant cue associations in human addiction: role of the VMPFC. *Addict Biol*.
- Kroes MCW, Dunsmoor JE, Mackey WE, McClay M, Phelps EA (2017) Context conditioning in humans using commercially available immersive Virtual Reality. *Sci Rep* 7: 8640. [PubMed: 28819155]
- Kruse O, Tapia Leon I, Stark R, Klucken T (2017) Neural correlates of appetitive extinction in humans. *Soc Cogn Affect Neurosci* 12: 106–115. [PubMed: 27803289]
- LaLumiere RT, Niehoff KE, Kalivas PW (2010) The infralimbic cortex regulates the consolidation of extinction after cocaine self-administration. *Learn Mem* 17: 168–75. [PubMed: 20332188]
- Liljeholm M, O'Doherty JP (2012) Contributions of the striatum to learning, motivation, and performance: an associative account. *Trends Cogn Sci* 16: 467–75. [PubMed: 22890090]
- Lonsdorf TB, Menz MM, Andreatta M, Fullana MA, Golkar A, Haaker J, Heitland I, Hermann A, Kuhn M, Kruse O, Meir Drexler S, Meulders A, Nees F, Pittig A, Richter J, Romer S, Shiban Y, Schmitz A, Straube B, Vervliet B, Wendt J, Baas JMP, Merz CJ (2017) Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neurosci Biobehav Rev* 77: 247–285. [PubMed: 28263758]
- Mackey S, Petrides M (2014) Architecture and morphology of the human ventromedial prefrontal cortex. *Eur J Neurosci* 40: 2777–96. [PubMed: 25123211]
- Milad MR, Quinn BT, Pitman RK, Orr SP, Fischl B, Rauch SL (2005) Thickness of ventromedial prefrontal cortex in humans is correlated with extinction memory. *Proceedings of the National Academy of Sciences of the United States of America* 102: 10706–11. [PubMed: 16024728]
- Milad MR, Quirk GJ (2012) Fear extinction as a model for translational neuroscience: ten years of progress. *Annual review of psychology* 63: 129–51.

- Milad MR, Rauch SL, Pitman RK, Quirk GJ (2006) Fear extinction in rats: implications for human brain imaging and anxiety disorders. *Biol Psychol* 73: 61–71. [PubMed: 16476517]
- Milad MR, Rosenbaum BL, Simon NM (2014) Neuroscience of fear extinction: implications for assessment and treatment of fear-based and anxiety related disorders. *Behav Res Ther* 62: 17–23. [PubMed: 25204715]
- Milad MR, Wright CI, Orr SP, Pitman RK, Quirk GJ, Rauch SL (2007) Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biological psychiatry* 62: 446–54. [PubMed: 17217927]
- Millan EZ, Marchant NJ, McNally GP (2011) Extinction of drug seeking. *Behav Brain Res* 217: 454–62. [PubMed: 21073904]
- Moorman DE, Aston-Jones G (2014) Orbitofrontal cortical neurons encode expectation-driven initiation of reward-seeking. *J Neurosci* 34: 10234–46. [PubMed: 25080585]
- Peters J, Kalivas PW, Quirk GJ (2009) Extinction circuits for fear and addiction overlap in prefrontal cortex. *Learn Mem* 16: 279–88. [PubMed: 19380710]
- Peters J, LaLumiere RT, Kalivas PW (2008) Infralimbic prefrontal cortex is responsible for inhibiting cocaine seeking in extinguished rats. *J Neurosci* 28: 6046–53. [PubMed: 18524910]
- Phelps EA, Delgado MR, Nearing KI, LeDoux JE (2004) Extinction learning in humans: role of the amygdala and vmPFC. *Neuron* 43: 897–905. [PubMed: 15363399]
- Quirk GJ, Mueller D (2008) Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology* 33: 56–72. [PubMed: 17882236]
- Raczka KA, Mechias ML, Gartmann N, Reif A, Deckert J, Pessiglione M, Kalisch R (2011) Empirical support for an involvement of the mesostriatal dopamine system in human fear extinction. *Transl Psychiatry* 1: e12. [PubMed: 22832428]
- Rhodes SE, Killcross AS (2007) Lesions of rat infralimbic cortex enhance renewal of extinguished appetitive Pavlovian responding. *Eur J Neurosci* 25: 2498–503. [PubMed: 17445245]
- Rhodes SE, Killcross S (2004) Lesions of rat infralimbic cortex enhance recovery and reinstatement of an appetitive Pavlovian response. *Learn Mem* 11: 611–6. [PubMed: 15466316]
- Schiller D, Delgado MR (2010) Overlapping neural systems mediating extinction, reversal and regulation of fear. *Trends Cogn Sci* 14: 268–76. [PubMed: 20493762]
- Schiller D, Kanen JW, LeDoux JE, Monfils MH, Phelps EA (2013) Extinction during reconsolidation of threat memory diminishes prefrontal cortex involvement. *Proc Natl Acad Sci U S A* 110: 20040–5. [PubMed: 24277809]
- Schneider B, Koenigs M (2017) Human lesion studies of ventromedial prefrontal cortex. *Neuropsychologia* 107: 84–93. [PubMed: 28966138]
- Sehlmeyer C, Schoning S, Zwitterlood P, Pfliederer B, Kircher T, Arolt V, Konrad C (2009) Human fear conditioning and extinction in neuroimaging: a systematic review. *PLoS One* 4: e5865. [PubMed: 19517024]
- Stuber GD, Sparta DR, Stamatakis AM, van Leeuwen WA, Hardjoprajitno JE, Cho S, Tye KM, Kempadoo KA, Zhang F, Deisseroth K, Bonci A (2011) Excitatory transmission from the amygdala to nucleus accumbens facilitates reward seeking. *Nature* 475: 377–80. [PubMed: 21716290]
- Taschereau-Dumouchel V, Cortese A, Chiba T, Knotts JD, Kawato M, Lau H (2018) Towards an unconscious neural reinforcement intervention for common fears. *Proc Natl Acad Sci U S A* 115: 3470–3475. [PubMed: 29511106]
- Taylor JR, Olausson P, Quinn JJ, Torregrossa MM (2009) Targeting extinction and reconsolidation mechanisms to combat the impact of drug cues on addiction. *Neuropharmacology* 56 Suppl 1: 186–95. [PubMed: 18708077]
- Torregrossa MM, Taylor JR (2013) Learning to forget: manipulating extinction and reconsolidation processes to treat addiction. *Psychopharmacology (Berl)* 226: 659–72. [PubMed: 22638814]
- Uytings HB, Groenewegen HJ, Kolb B (2003) Do rats have a prefrontal cortex? *Behav Brain Res* 146: 3–17. [PubMed: 14643455]
- Van den Oever MC, Rotaru DC, Heinsbroek JA, Gouwenberg Y, Deisseroth K, Stuber GD, Mansvelder HD, Smit AB (2013) Ventromedial prefrontal cortex pyramidal cells have a temporal dynamic role

in recall and extinction of cocaine-associated memory. *J Neurosci* 33: 18225–33. [PubMed: 24227731]

Wassum KM, Izquierdo A (2015) The basolateral amygdala in reward learning and addiction. *Neurosci Biobehav Rev* 57: 271–83. [PubMed: 26341938]

Winecoff A, Clithero JA, Carter RM, Bergman SR, Wang L, Huettel SA (2013) Ventromedial prefrontal cortex encodes emotional value. *J Neurosci* 33: 11032–9. [PubMed: 23825408]

Zhang Z, Mendelsohn A, Manson KF, Schiller D, Levy I (2015) Dissociating Value Representation and Inhibition of Inappropriate Affective Response during Reversal Learning in the Ventromedial Prefrontal Cortex. *eNeuro* 2.

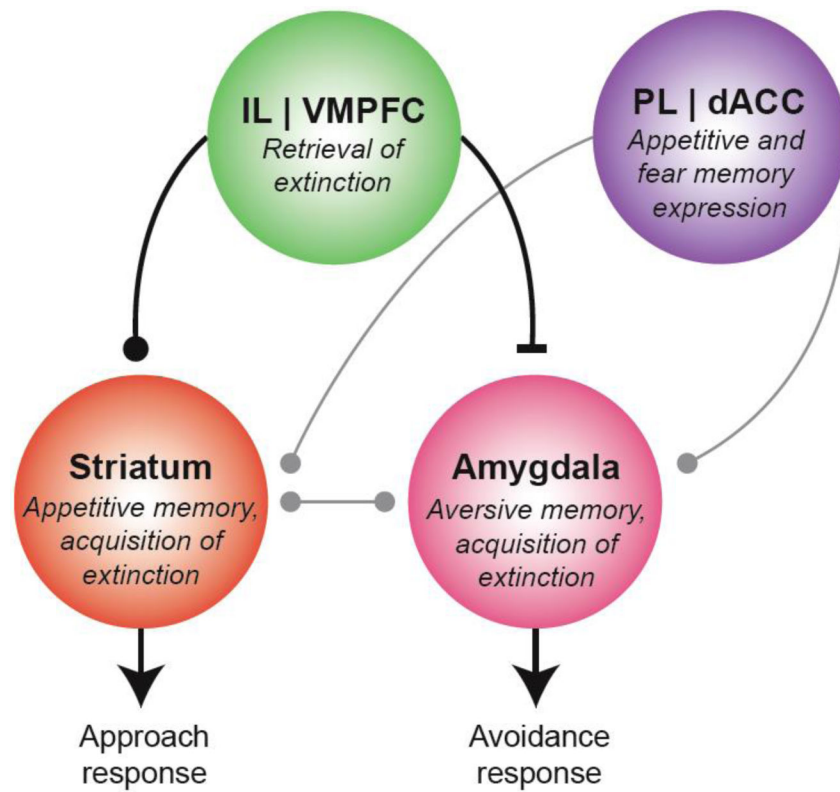


Figure 1.

Aversive and appetitive/drug cue extinction circuits. Regulation of conditioned fear avoidance and appetitive approach and drug seeking behaviors through extinction learning and recall both rely on the infralimbic (IL) cortex, the rodent homologue of the human ventromedial prefrontal cortex (VMPFC). Aversive extinction is instantiated through IL inhibition of the amygdala, which in turn is associated with the expression of conditioned fear. Appetitive extinction might instead be instantiated through IL excitatory connections to the striatum (specifically, the nucleus accumbens), which is associated with the expression of conditioned appetitive approach and drug seeking behavior. The prelimbic (PL) cortex, the presumed rodent homologue of the human dorsal anterior cingulate cortex (ACC), facilitates both aversive and appetitive conditioned responses.

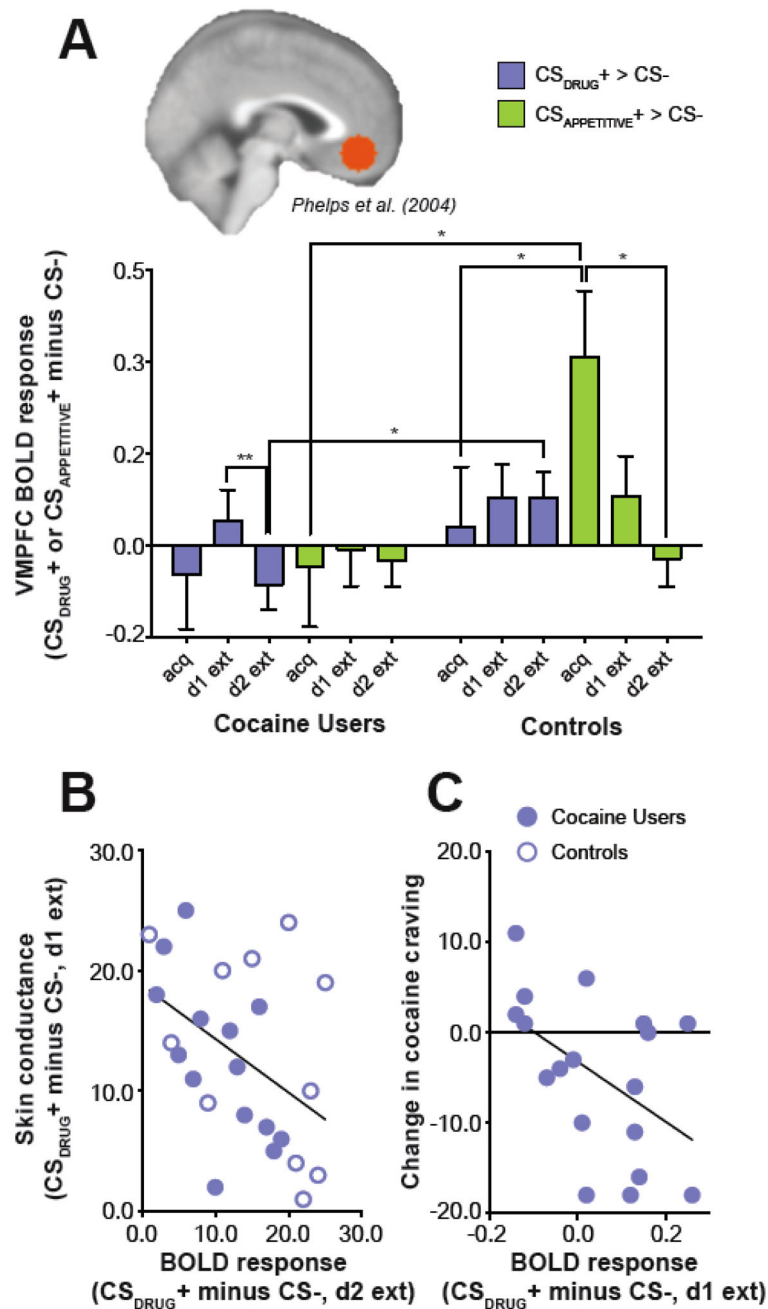


Figure 2.

Role of the ventromedial prefrontal cortex (VMPFC) in appetitive and drug cue extinction in humans and evidence for an impairment in drug addiction. **(A)** VMPFC blood-oxygen-level-dependent (BOLD) responses decrease for the CS_{APPETITIVE}⁺ (cue paired with a pleasant image) but tend to increase for the CS_{DRUG}⁺ (cue paired with a drug-related image) with the progression of extinction training (from acquisition to day 1 extinction to day 2 extinction recall) in healthy subjects but not cocaine addicted subjects. **(B)** Across both groups, VMPFC activation during day 2 extinction correlates with the success of day 1 extinction as indexed by reductions in skin conductance response (SCR) for the CS_{DRUG}⁺, pointing to a

role of this region in the recall of extinction learning as previously shown in fear extinction studies. (C) In cocaine users, VMPFC activation during day 1 extinction (extinction training) correlates with a reduction in craving from day 1 to day 2, such that subjects who were more successful at modulating activation in this region for the drug-relevant cue experienced less severe drug cravings a day later, further suggesting drug cue extinction might be relevant for regulation of craving.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript