Addiction Biology

SSA SOCIETY FOR THE STUDY OF ADDICTION

ORIGINAL ARTICLE

doi:10.1111/adb.12545

Neural mechanisms of extinguishing drug and pleasant cue associations in human addiction: role of the VMPFC

Anna B. Konova[1], Muhammad A. Parvaz[2], Vladimir Bernstein[2], Anna Zilverstand[2], Scott J. Moeller[4], Mauricio R. Delgado[3], Nelly Alia-Klein[2] & Rita Z. Goldstein[2]

Center for Neural Science, New York University, New York, NY USA[1], Departments of Psychiatry and Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY USA[2], Department of Psychology, Rutgers University, Newark, NJ USA[3] and Department of Psychiatry, Stony Brook University School of Medicine, Stony Brook, NY USA[4]

ABSTRACT

The neurobiological mechanisms that underlie the resistance of drug cue associations to extinction in addiction remain unknown. Fear extinction critically depends on the ventromedial prefrontal cortex (VMPFC). Here, we tested if this same region plays a role in extinction of *non-fear*, drug and pleasant cue associations. Eighteen chronic cocaine users and 15 matched controls completed three functional MRI scans. Participants first learned to associate an abstract cue (the conditioned stimulus, CS) with a drug-related (CS_{D+}) or pleasant (CS_{P+}) image. Extinction immediately followed where each CS was repeatedly presented without the corresponding image. Participants underwent a second identical session 24 hours later to assess retention of extinction learning. Results showed that like fear extinction, non-fear-based extinction relies on the VMPFC. However, extinction-related changes in the VMPFC differed by cue valence and diagnosis. In controls, VMPFC activation to the CS_{D+} (which was unpleasant for participants) gradually increased as in fear extinction, while it decreased to the CS_{P+}, consistent with a more general role of the VMPFC in flexible value updating. Supporting a specific role in extinction retention, we further observed a cross-day association between VMPFC activation and skin conductance, a classic index of conditioned responses. Finally, cocaine users showed VMPFC abnormalities for both CSs, which, in the case of the CS_{D+}, correlated with craving. These data suggest a global deficit in extinction learning in this group that may hinder extinction-based treatment efforts. More broadly, these data show that the VMPFC, when functionally intact, supports extinction learning in diverse contexts in humans.

Keywords cocaine, craving, extinction, functional magnetic resonance imaging, reward, ventromedial prefrontal cortex.

Correspondence to: Rita Z. Goldstein, Department of Psychiatry, 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. E-mail: rita.goldstein@mssm.edu

INTRODUCTION

Extinction is the process by which 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 & Mueller 2008). The predominant view is that extinction does not eliminate the CS–US association; rather, it leads to a lessening in the conditioned response by creating a new (CS-no-US) association that competes for expression, leaving memory vulnerable to recovery of the conditioned response (Quirk & Mueller 2008). Extinction has been extensively studied in humans and non-human animals in the domain of fear learning, e.g. using electric shock (Phelps *et al.* 2004; Milad *et al.* 2005), aversive sounds (Neumann & Waters 2006) and even monetary loss (Schlund *et al.* 2015). However, less is known about extinction in the appetitive domain, and while animal work suggests similarities in the neurobiological mechanisms of extinction of, e.g. drug seeking and fear (Peters, Kalivas & Quirk 2009), the mechanism of non-fear-based extinction in humans remains unknown.

Addiction is characterized by continued drug seeking and use despite reduced pleasure derived from the drug and catastrophic health and social consequences. This behavior is assumed to be at least partly driven by a

learning process whereby cues associated with drug consumption acquire excessive and persistent salience, perpetuating drug seeking. The persistence of drug seeking despite negative consequences and a reduction in the drug's rewarding effects suggests that addicted individuals may have diminished ability to form and/or maintain new associations for cues that previously, though no longer, predict drug rewards (e.g. learning that the drug or drug-associated cues are no longer as valuable). This is also predicted by the neural circuitry supporting extinction learning, which overlaps extensively with that directly impacted by addictive substances and addiction (Goldstein & Volkow 2011), potentially rendering this process especially vulnerable in this population.

Substantial work in the fear domain demonstrates a central role for the ventromedial prefrontal cortex (VMPFC) in the formation, retention and later retrieval of extinction learning (Quirk & Mueller 2008; Milad & Quirk 2012). In humans, VMPFC activity increases during fear extinction (Milad *et al.* 2007) and extinction retrieval (Phelps *et al.* 2004; Kalisch *et al.* 2006), and both neural activity (Phelps *et al.* 2004) and cortical thickness (Milad *et al.* 2005; Hartley, Fischl & Phelps 2011) in this region correlate with psychophysiological indices of extinction success [e.g. lowered skin conductance response (SCR) to the CS]. Beyond fear extinction, the VMPFC along with the striatum form what is known as the 'brain's valuation system', a set of regions that represent (and possibly update) value in a domain-general manner (Bartra, McGuire & Kable 2013). In addiction, the VMPFC (Kober *et al.* 2016) and striatum (Kuhn & Gallinat 2011) are implicated in the experience of craving, a motivational state often triggered by drug-associated cues that can promote drug seeking.

This more general role of the VMPFC in valuation and craving suggests that the VMPFC may also be a candidate region involved in extinction of non-fear-based and secondary reinforcers, including drug-related and appetitive cues. However, while the effect of extinction-based therapy on drug-cue reactivity has just begun to be examined (Vollstadt-Klein *et al.* 2011; Prisciandaro *et al.* 2013), no studies to date have investigated the role of the VMPFC in extinction learning itself in human addiction. Such an investigation has important implications for the potential utility of extinction-based therapies for addiction and for the basic neuroscientific understanding of non-fear-based extinction more generally.

Modeled after classical fear-conditioning studies, here, we examined the neural correlates of extinction learning for drug and pleasant cue associations in non-treatment seeking, chronic cocaine users and sociodemographically matched healthy non-drug users in a 2-day functional magnetic resonance imaging (fMRI) study. The study comprised an acquisition phase, where participants learned to associate an abstract cue with a drug-related (CS_D+) or pleasant (CS_P+) image, and two extinction phases (the latter for assessing the retention of extinction from day 1), where the abstract cues were repeatedly presented without the corresponding images. Throughout, we collected SCR and blood-oxygen-level-dependent (BOLD) response, time-locked to the presentation of the abstract cue, as indices of the conditioned response. We hypothesized that the VMPFC and striatum would exhibit parametric changes across the learning phases as participants form new, affectively neutral, associations with the CS_D+ and CS_P+. We expected cocaine users to show abnormalities in these regions for both cues.

MATERIALS AND METHODS

Participants

The participants were chronic cocaine users and healthy individuals with no history of drug or psychiatric illness, who were native English speakers recruited from the community through advertisements and by word of mouth, and who provided written informed consent to participate in accordance with the local institutional review board. To minimize the influence of factors other than those related to cocaine addiction, the groups were selected to match on multiple sociodemographic characteristics and cigarette smoking status (Table 1). All participants were asked to complete study procedures on two separate days (psychophysiological measures and fMRI methods). The final sample consisted of 18 cocaine users and 15 healthy controls. G*Power 3.1.9.2 (Faul *et al.* 2009) was used to determine whether this sample was sufficiently powered. Given a 2 (group) × 3 (learning phase) mixed design, 80% desired power, α error probability = 0.05 and a within-between subject interaction of a large effect size (Cohen's *d*, henceforth referred to simply as *d*, of 0.8), it was determined that $N = 12$ participants would be needed. For a medium effect size ($d = 0.5$), $N = 28$ participants would be needed. Thus, our sample of $N = 33$ was sufficiently powered for effect sizes of $d \geq 0.5$.

All participants were in good health and not currently taking medication. Drug use and psychiatric histories were ascertained by a comprehensive clinical interview administered by trained research staff with extensive experience evaluating drug-addicted populations, consisting of the Structured Clinical Interview for DSM-IV Axis I Disorders (research version; First *et al.* 1996; Ventura *et al.* 1998) and the Addiction Severity Index (McLellan *et al.* 1992). Exclusion criteria for both groups were (1) history of head trauma, neurological disease or loss of consciousness >30 minutes; (2) abnormal

Table 1 Demographic and drug use characteristics of the study sample.

	Test	*Control* (n = 15)	*Cocaine users* (n = 18)
Demographics			
Age (years)	$t_{31} = 0.3$	45.3 ± 1.6	45.8 ± 1.1
Sex (male/female)	$\chi^2 = 0.6$	13/2	17/1
Race (African-American/Caucasian/Hispanic)	$\chi^2 = 0.8$	10/2/3	12/4/2
Education (years)	$t_{31} = 1.7$	13.3 ± 0.4	12.6 ± 0.3
Verbal IQ: Wide Range Achievement Test III—Reading Scale	$t_{31} = 0.7$	94.0 ± 4.2	90.0 ± 3.5
Nonverbal IQ: Wechsler Abbreviated Scale of Intelligence—Matrix Reasoning Scale	$t_{31} = 0.5$	8.4 ± 0.9	8.9 ± 0.7
State Depression: Beck Depression Inventory II[a]	Z = 1.7	3.2 ± 1.3	7.6 ± 2.2
Socioeconomic Status: Hollingshead Index[a]	$t_{28} = 1.3$	32.9 ± 3.4	28.9 ± 2.4
Handedness (laterality quotient)	$t_{31} = 0.5$	0.7 ± 0.2	0.8 ± 0.2
Drug use			
Cigarette smokers (current or past/non-smokers)	$\chi^2 = 3.5$	6/9	13/5
Daily cigarettes (current smokers: n = 5/11)	Z = 1.0	4.8 ± 1.7	8.5 ± 2.3
Alcohol use lifetime (years) (n = 10/11)	$t_{19} = 0.8$	20.0 ± 3.1	22.8 ± 2.2
Cocaine use lifetime (years)	–	–	17.8 ± 1.6
Duration of current abstinence/time since last cocaine use (days)[b]	–	–	13.9 ± 7.4
Days/week of cocaine use during the past 30 days	–	–	2.6 ± 0.6
Cocaine urine status (positive/negative): day 1 | day 2	–	–	6/12 | 4/14
Withdrawal symptoms: 18-item CSSA (0–126): day 1 | day 2[c]	–	–	22.2 ± 3.6 | 11.8 ± 2.1
Cocaine craving: 5-item Questionnaire (0–45): day 1 | day 2[d]	–	–	20.1 ± 3.1 | 15.5 ± 3.5

[a]Data missing for one control and two cocaine users. [b]Data missing for two cocaine users. [c]Day 1 versus day 2 ($t_{17} = 3.9$, $P = 0.001$). [d]Day 1 versus day 2 ($t_{17} = 2.2$, $P = 0.04$). Abbreviations: CSSA, Cocaine Selective Severity Assessment Scale. Values are frequencies or means ± standard error of the mean (SEM).

vital signs; (3) history of major medical conditions; (4) history of major psychiatric disorders (other than substance use disorders in the cocaine group and nicotine use disorder in both groups); (5) positive urine pregnancy test in women; (6) contraindications to MRI; and (7) except for cocaine in the cocaine user group, positive urine screens for psychoactive drugs or their metabolites.

Cocaine users were non-treatment-seeking individuals who reported an average lifetime history of 17 years of cocaine use, 2 days/week of cocaine use in the past 30 days and some cocaine use within the past 4 months (see Table 1 for detailed drug use information). Participants identified cocaine as their primary drug of choice, meeting criteria for cocaine dependence (n = 17) or abuse (n = 1) [in early full (n = 2) or partial (n = 1) remission]. Current co-morbid disorders included alcohol dependence (n = 2) and marijuana abuse (n = 2); one participant also met criteria for a current depressive episode. Thirteen cocaine users and 6 controls were cigarette smokers. Six participants tested positive for cocaine on day 1 (indicating use ≤72 hours) and 4 tested positive for cocaine on day 2. Controls tested negative for all drugs on both study days. Apart from number of days since last use (Z = 1.87, P = 0.06), cocaine urine positive and cocaine urine negative participants (based on day 1 status) did not differ in their clinical profile (i.e. they did not differ in any of the drug use variables listed in Table 1; $P > 0.086$). Nevertheless, because the passage of time from last use (day 1 to day 2) was associated with a reduction in craving and withdrawal (Table 1) and because there is some clinical and preclinical evidence to suggest that recent cocaine exposure impacts learning (Schoenbaum *et al.* 2004; McCracken & Grace 2013; Spronk *et al.* 2016), we tested if cocaine urine status modified any of the observed diagnostic group effects.

Study procedures and fMRI task

All participants completed three fMRI sessions conducted over two consecutive days (99 total sessions; Figure 1). To ensure comparable experiences between sessions and across the two study days, and to control the time in-between, participants were overnighted in the laboratory. On day 1, after a period of acclimation to the scanner environment, participants first learned to associate a colored square, the CS, with different types of 'USs'. This created three cue types: CS_D+ (CS paired with a drug-related image of a person smoking crack), CS_{P+} (CS paired with a pleasant image of a smiling baby) and CS− (CS paired with a neutral image of an unadorned

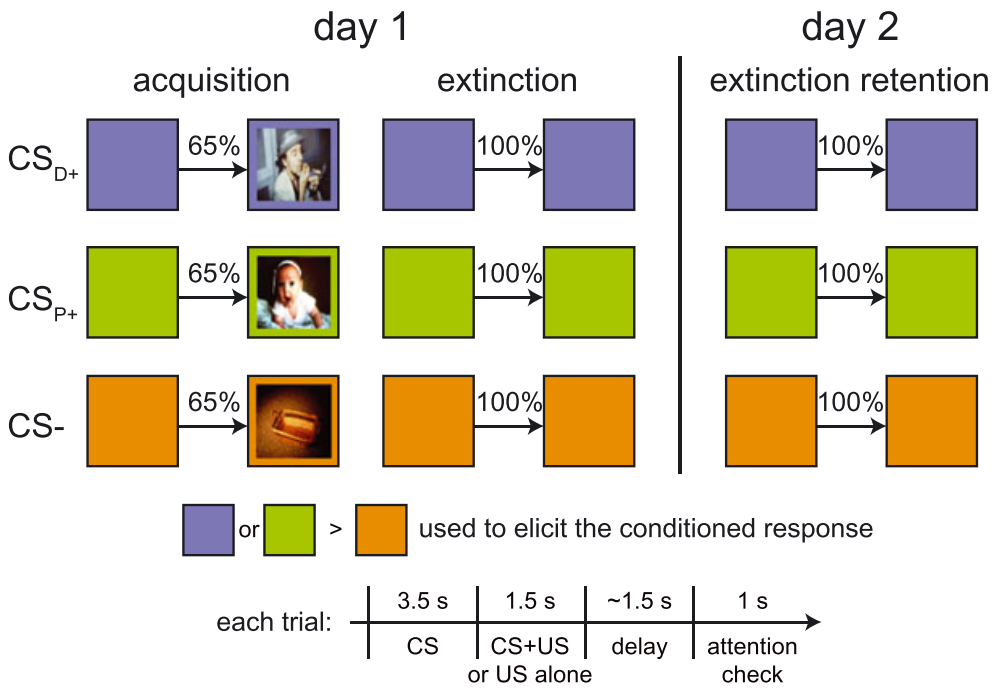

Figure 1 Study overview and conditioning paradigm. While in the MRI scanner and over the course of three scanning sessions, subjects learned to associate a cue (colored square), the conditioned stimulus (CS), with a drug-related (CS$_D$+), affectively pleasant (CS$_P$+) or neutral (CS−) image. Following two reinforced presentations of the CS for each CS type (not shown), extinction training immediately followed acquisition, where the CS was presented repeatedly without the paired image. A second extinction training session took place 24 hours later. A typical paired trial consisted of presentation of the CS for 3.5 seconds, followed by presentation of the unconditioned stimulus (US; the corresponding image) inside the CS for 1.5 seconds, a variable ~1.5-second fixation screen and a 1-second screen requiring a non-contingent button press indicating whether the US appeared or not on that trial. Following another variable ~1.5-second fixation screen, the next trial began. Unpaired trials were identical to the paired trials with the exception that the CS remained on the screen for the entire 5 seconds.

wooden basket). The acquisition phase consisted of 13 paired trials in which each CS presentation co-terminated with presentation of the corresponding US and 7 unpaired trials in which each CS was presented without the US, for a total of 20 trials per cue type (60 total). Participants were instructed to 'try and figure out which color predicted which image'. Day 1 extinction immediately followed and began with 2 paired trials followed by 20 unpaired trials for each cue type (66 total). The participants completed another, identical extinction session ~24 hours later but without the paired reminder trials. Trial order was pseudo-randomized in all sessions, and CS color assignment was counterbalanced across participants. SCR and fMRI data were acquired throughout. Participants earned $25/session (max $75).

Participants also completed two-alternative forced choice tasks and subjective ratings for the task stimuli (see Supporting Information). No SCR or fMRI data were collected during these tasks. These data confirmed that, as expected based on our prior work in independent samples of cocaine-addicted and control subjects (Moeller *et al.* 2009; Moeller *et al.* 2010; Moeller *et al.* 2013), participants found the drug-related image as least pleasant (and chose to view it least often) and the affectively pleasant image as most pleasant (and chose to view it most often). Controls additionally rated the drug-related image as more unpleasant (and chose to view it less often) than cocaine users (Figure S1a and b).

SCR acquisition and analysis

Skin conductance was acquired with shielded Ag-AgCl electrodes (AD Instruments, Inc.) attached to the second and big toes of the left foot. The electrode cables were grounded through an RF filter panel. Data were continuously recorded at 200 samples/s. Offline data analysis was performed in MATLAB. The continuous data were low-pass filtered (1 Hz) and then divided into epochs. As in previous studies, the SCR amplitude on each trial was computed as the peak amplitude in the 0.5 to 4.5-second time window following CS onset minus the average amplitude in the 0.5 seconds prior to CS onset. Thus, SCR to the CS$_D$+, CS$_P$+ and CS− reflected changes in skin conductance level beyond changes in this measure produced by the preceding trial or task phase.

After square-root transformation, the data were analyzed in a 2 (cue type: drug, pleasant) × 3 (learning phase: acquisition, day 1 extinction, day 2 extinction) × 2 (group: cocaine users, controls) mixed analysis of variance (ANOVA) on the differential SCR values

(CS_D+ versus CS− and CS_P+ versus CS−). Considering that we expected across-session as well as *within*-session learning, the ANOVA was restricted to all unpaired trials during acquisition and the last half of all extinction trials on days 1 and 2. Due to artifacts in the SCR signal, the subsample with complete SCR data for all three learning phases consisted of $n = 11$ cocaine users and $n = 6$ controls. Hence, our SCR analyses comparing the diagnostic groups across the learning phases were only powered to detect large effects (specifically, $d \geq 0.68$).

Image acquisition and analysis

Functional images were acquired with a 4T Varian/Siemens MRI scanner by using a coronal T_2*-weighted single shot gradient-echo echo planar imaging sequence (TE/TR = 20/1600 ms, 3.125 × 3.125-mm^2 in-plane resolution, 4-mm slice thickness, 1-mm gap, 33 coronal slices, 20-cm field of view, 64 × 64 matrix size, 90° flip angle, 200 kHz bandwidth with ramp sampling). Image processing and analyses were performed in SPM8 (Wellcome Trust Centre for Neuroimaging, London UK). The data were first realigned, co-registered and spatially normalized to a standard echo planar imaging template in the Montreal Neurological Institute (MNI) frame, resulting in a final voxel size of 3 × 3 × 3 mm. Criteria for acceptable motion were ≤2 mm translation or ≤2° rotation in any direction. The data were spatially smoothed with an 8-mm full-width-at-half-maximum Gaussian kernel.

Three random-effects general linear models (GLMs) were specified for each participant, corresponding to acquisition, day 1 extinction and day 2 extinction, each with a session-specific intercept, the 6 motion parameters as regressors of no interest and all task conditions convolved with a canonical HRF and high-pass filter (cut-off frequency: 1/1500 seconds). The GLM for acquisition included trial onsets for the CS_D+, CS_P+ and CS–, separately for paired and unpaired trials (six conditions in total), modeled as epochs with duration equal to the length of CS presentation (3.5 or 5 seconds; i.e. terminating at presentation of the US). The GLMs for day 1 and day 2 extinction each included trial onsets for each CS (all unpaired trials, epoch duration = 5 seconds). The GLM for day 1 extinction additionally included a fourth condition for the 'reminder trials'. Beta maps were computed for each participant for each CS and learning phase.

Given that previous studies have identified a circuit centered on the VMPFC in extinction learning, and the VMPFC and striatum in the representation and updating of values including those for drug cues, we focused on these two regions, although exploratory whole-brain and targeted control region analyses were also performed (see Supporting Information). As with SCR, we conducted 2 (cue type) × 3 (learning phase) × 2 (group) mixed ANOVAs on the differential BOLD responses (CS_D+ versus CS− and CS_P+ versus CS−), our neural measure of the conditioned response, extracted as average beta estimates from unbiased regions of interest (ROIs). All $N = 33$ participants were included in the fMRI analyses.

The striatum (entire caudate and putamen) ROI was anatomically defined in PICKATLAS (ANSIR Laboratory; http://fmri.wfubmc.edu/software/PickAtlas). The VMPFC was defined as a 12-mm radius sphere centered on the coordinates reported in Phelps *et al.* (2004) after transformation to MNI space (Talairach: $x = \pm 2$, $y = 38$, $z = -3$; MNI: $x = \pm 3$, $y = 42$, $z = -12$). The term VMPFC is used to describe a large, heterogeneous region of the medial prefrontal cortex that spans parts of the anterior cingulate cortex (anterior to the genu of the corpus callosum) and the medial orbitofrontal cortex, encompassing Brodmann areas (BAs) 25, ventral portions of 24 and 32, medial portion of 11 and ventral and medial portions of 10 (Mackey & Petrides 2014). The particular aspect of the VMPFC included in our ROI is the medial portion of BA 11 and the ventral portion of BA 10 (see inset in Figure 2); this aspect has been linked to emotion regulation including the use of extinction strategies (Diekhof *et al.* 2011). In the anterior–posterior direction, our VMPFC ROI falls centrally, touching on posterior aspects traditionally linked to the representation of negative affect and anterior aspects representing positive affect (Grabenhorst & Rolls 2011; Myers-Schulz & Koenigs 2012). Notably, our ROI almost fully overlaps with the VMPFC locus identified in Bartra *et al.* (2013) to represent value in diverse contexts. An initial analysis that included laterality as a factor revealed differential responses in the left versus right VMPFC ROIs. Therefore, the ANOVAs reported were performed on averaged left and right side activation values for the striatum but not the VMPFC.

RESULTS

Psychophysiological and self-reported measures

The 2 (cue type) × 3 (learning phase) × 2 (group) mixed ANOVA revealed a reduction in SCR (to the CS_D+ and CS_P+ relative to the CS−) over the learning phases ($F_{2,30} = 2.71$, $P = 0.08$, $d = 0.85$), an effect that reached significance for the linear contrast (acquisition > day 1 extinction > day 2 extinction: $F = 5.93$, $P = 0.028$, $d = 1.26$; Figure S2), but no significant diagnostic group, cue type or interaction effects ($F < 0.77$, $P > 0.47$, $d < 0.45$). In addition, there were no

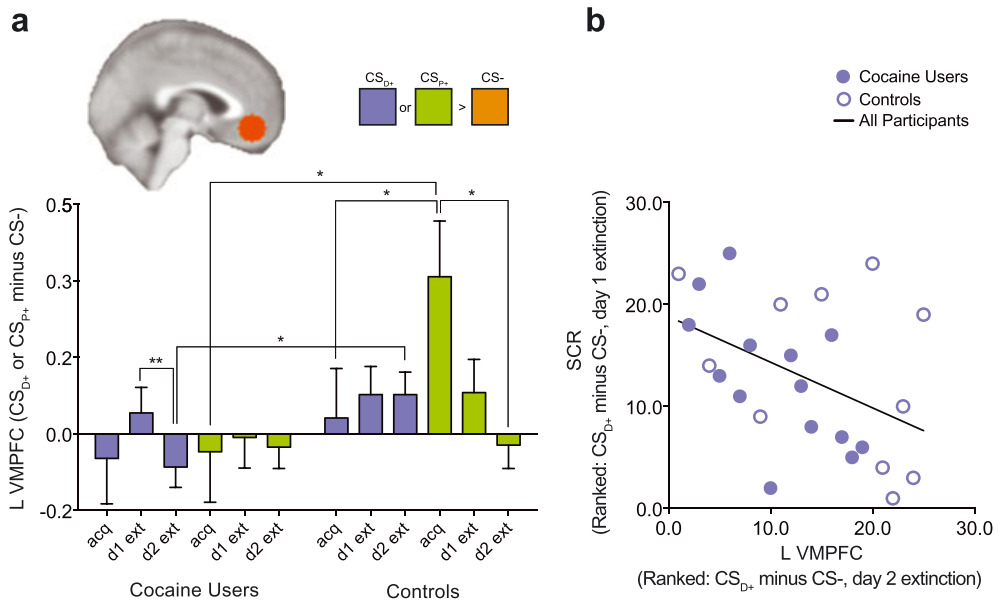

Figure 2 Modulation of left ventromedial prefrontal cortex (VMPFC) activation by learning phase is valence-specific and correlates with psychophysiological response to the conditioned stimulus (CS_D+) paired with a drug cue. (a) Plot shows left VMPFC activation for each learning phase (acquisition, day 1 extinction, day 2 extinction) as a function of cue-type [CS_D+ (cue paired with a drug-related image) and CS_P+ (cue paired with a pleasant image), both relative to the CS− (cue paired with a neutral image)] and diagnostic group (cocaine users and controls). There was a 3-way interaction such that left VMPFC activation decreased in response to the CS_P+ but increased in response to the CS_D+ with extinction training in controls but not cocaine-addicted participants. (b) Left VMPFC activation on day 2 extinction correlated with the success of day 1 extinction as indexed by reductions in skin conductance response (SCR) to the CS_D+ relative to the CS− across subjects, pointing to a role of this region in the recall of extinction learning. The overall pattern of results in the right VMPFC, as well as the correlation between right VMPFC activation and SCR, was similar albeit weaker (see Results). $+P \leq 0.10$, $*P \leq 0.05$, $**P \leq 0.01$. See also Figures S3–S5 and Tables S1 and S2

differences between the groups in attention or subjective ratings/choice for the CSs (see Figure S1c and d).

Neural correlates of extinction learning for drug and pleasant cue associations

We hypothesized that, paralleling the SCR data, the striatum and VMPFC would show similar progressive increased or decreased activation over the learning phases as tested with 2 × 3 × 2 mixed ANOVAs on the differential BOLD responses (CS_D+ and CS_P+, both relative to the CS−) in each ROI, followed by linear contrasts specifically testing for this progression.

The main finding we observed was a diagnostic group main effect ($F_{1,31} = 5.44$, $P = 0.026$, $d = 0.84$) and a cue type × learning phase interaction in the left VMPFC ($F_{1.37,42.44} = 4.05$, $P = 0.039$, $d = 0.72$), which were both qualified by a significant cue type × learning phase × group interaction ($F_{2,62} = 4.21$, $P = 0.019$, $d = 0.74$; all other effects, $P > 0.31$, $d < 0.37$). A similar albeit statistically weaker pattern was observed in the right VMPFC (group main effect: $F_{1,31} = 4.47$, $P = 0.043$, $d = 0.76$; cue type × learning phase interaction: $F_{1.27,39.47} = 3.07$, $P = 0.054$, $d = 0.63$; cue type × learning phase × group interaction: $F_{2,62} = 1.94$, $P = 0.15$, $d = 0.50$). All other effects in the right VMPFC were non-significant ($P > 0.32$, $d < 0.38$). The 3-way interaction in the left VMPFC was explained by differences over the learning phases in response to the CS_D+ versus CS_P+ in controls but not cocaine users (Figure 2a). As in fear extinction studies, in controls, VMPFC activation was higher during extinction for the CS_D+ (which was rated as unpleasant). However, it was *lower* during extinction for the CS_P+ (which was rated as pleasant; cue × learning phase interaction in controls: $F_{1.32,18.48} = 5.00$, $P = 0.029$, cue × learning phase linear effect: $P = 0.021$). In contrast, in cocaine users, there was no such shift as extinction progressed (cue × learning phase interaction in the cocaine group: $F_{1.44,24.45} = 0.72$, $P = 0.49$). See the Supporting Information for preliminary data showing that these VMPFC findings do not appear to be specific to the abstract image cues used in the present study but rather extend to alternate USs (i.e. gain of real money).

As in Phelps *et al.* 2004, we also tested the cross-day association between SCR and VMPFC activity. The success of extinction learning on day 1, as indexed by a reduction in SCR (average SCR over the last half of day 1 extinction) to the CS_D+ versus CS−, correlated with the magnitude of left VMPFC activation to the CS_D+ versus CS− during day 2 extinction ($n = 25$, $R_S = -0.45$, $P = 0.025$; Figure 2b) and neither group alone drove this

effect. In cocaine users (n = 14), this relationship was $R_S = -0.66$, while in controls (n = 11), it was $R_S = -0.44$. The relationship between SCR and right VMPFC activation was similar ($R_S = -0.39$, $P = 0.053$; cocaine users: $R_S = -0.57$, controls: $R_S = -0.41$), altogether showing that the VMPFC might have a specific role in the *retrieval* of extinction learning for drug cues as previously found for fear.

Finally, supporting a role for the striatum in extinction learning, but not supporting our hypothesis of group differences, we observed a significant cue × learning phase interaction in this region ($F_{2,62} = 7.97$, $P = 0.001$, $d = 1.01$) and no significant diagnostic group or additional interaction effects (all $F < 0.61$, $P > 0.54$, $d < 0.28$). Similar to findings in the VMPFC, the 2-way interaction was explained by higher striatum activation to the CS_D+ but lower activation to the CS_P+ during extinction relative to acquisition (cue × learning phase linear effect: $P = 0.003$; Figure 3). There was no significant correlation with SCR for either day 1 or day 2 extinction ($R_S < 0.22$, $P > 0.31$).

See Tables S1 and S2 and Figures S3 and S4 for results of whole-brain analyses and the Supporting Information for control region analyses. In addition to providing independent support for our ROI findings, the whole-brain analyses showed additional involvement of the amygdala and parahippocampal gyrus, among other regions such as the inferior frontal gyrus and sensory cortices, during extinction learning, and as expected, no significant differential task modulation or task condition by diagnostic group interactions in our negative control (auditory cortex) region. Finally, because a subset (n = 6 on day 1 and n = 4 on day 2) of participants tested positive for cocaine, indicating recent (≤72 hours) exposure to the drug, we also tested whether cocaine urine status had any bearing on our main results. The linear learning phase main effect on SCR, the cue type × learning phase × group interaction in the left VMPFC and the cue type × learning phase interaction in the striatum all remained significant when we excluded the n = 4 participants who were cocaine positive on both study days ($P < 0.038$). SCR and activation in the two ROIs also did not differ by cocaine urine status (positive/negative) at acquisition and day 1 extinction (n = 6 versus n = 12, respectively; $P > 0.066$) or day 2 extinction (n = 4 versus n = 14, respectively; $P > 0.084$). These control analyses suggest that the effects of recent cocaine use are not likely to have confounded those of diagnosis.

Relationship to craving

Given its role in cue-induced craving, we tested whether VMPFC activation during extinction correlated with participants' current and past 24-hour desire for cocaine, including that triggered by drug-related cues (total craving score; Table 1). Those cocaine users with higher VMPFC activation to the CS_D+ versus CS− during day 1 extinction (i.e. who looked more like controls) reported a greater reduction in craving on day 2 relative to day 1 ($R = -0.49$, $P = 0.04$; Figure 4), as driven by the relationship to day 2 craving ($R = -0.53$, $P = 0.02$; day 1 craving: $R = -0.26$, $P = 0.29$).

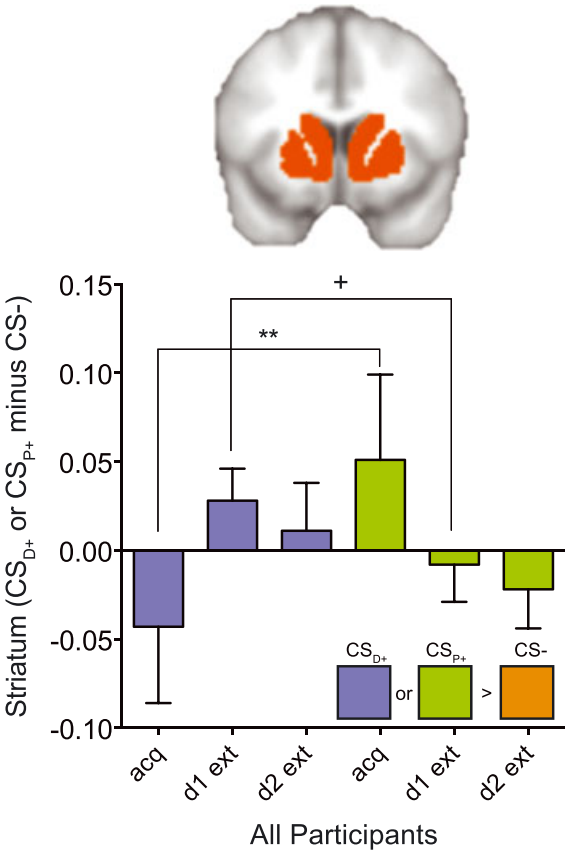

Figure 3 Modulation of striatum activation by learning phase is valence-specific across participants. (A) Plot shows average left and right striatum activation for each learning phase (acquisition, day 1 extinction, day 2 extinction) as a function of cue-type [CS_D+ (cue paired with a drug-related image) and CS_P+ (cue paired with a pleasant image), both relative to the CS− (cue paired with a neutral image)] across controls and cocaine users. There was a 2-way interaction such that, across participants, striatum activation decreased in response to the CS_P+ but increased in response to the CS_D+ with extinction training. +$P \leq 0.10$, *$P \leq 0.05$, **$P \leq 0.01$. See also Figures S3 and S4 and Tables S1 and S2

DISCUSSION

Extinction of fear critically depends on the VMPFC (Phelps *et al.* 2004; Milad *et al.* 2005; Milad *et al.* 2007; Hartley *et al.* 2011). Here, we show that this same region plays a role also in extinction of drug cue associations.

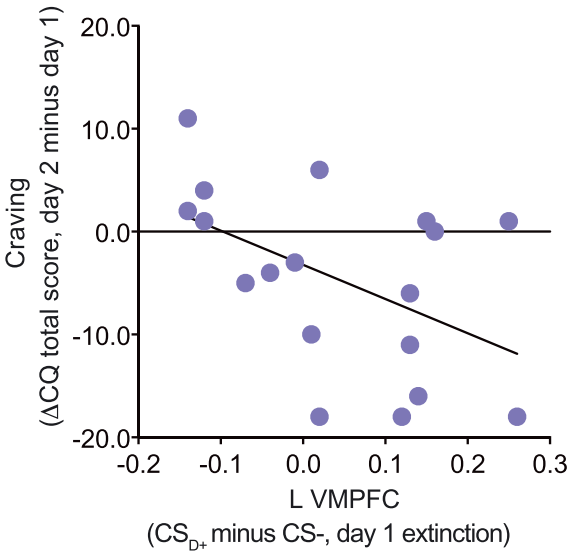

Figure 4 Left ventromedial prefrontal cortex (VMPFC) activation to the cue paired with a drug-related image (CS_D+) relative to the cue paired with a neutral image (CS−) on day 1 extinction correlated with a reduction in craving from day 1 to day 2, showing that participants who were more successful at modulating activation in this region on day 1 extinction in response to the drug-relevant cue experienced *less severe* drug cravings a day later

During extinction learning, VMPFC response to the CS_D+ (cue associated with the drug-related image) increased, while psychophysiological arousal decreased. These two measures were correlated such that participants who showed the most success in extinguishing arousal responses to the CS_D+ on day 1 also showed the greatest VMPFC increases to the CS_D+ on day 2, presumably when extinction learning from day 1 is recalled. While a similar (increased) VMPFC activation is observed in fear extinction studies (Phelps *et al.* 2004; Milad *et al.* 2007; Schiller *et al.* 2013), VMPFC response to the CS_P+ (cue associated with the affectively pleasant image) and an alternate appetitive CS (cue associated with monetary gain; see Supporting Information) instead *decreased* during extinction learning, consistent with a value updating process. Finally, while controls showed these distinct VMPFC response profiles, cocaine-addicted individuals, who manifest deficits in the VMPFC, did not, suggesting that the VMPFC, when intact, supports extinction learning in diverse contexts including of drug cue and pleasant associations.

That activation in the VMPFC might reflect a shift from a more valenced state (unpleasant as for the CS_D+ or pleasant as for the CS_P+) to a less valenced or neutral state follows from a large body of work showing that the VMPFC represents the value of a wide range of (appetitive and aversive) stimuli to guide behavior (Bartra *et al.* 2013). While this more general role of the VMPFC in extinction has been previously hypothesized (Schiller *et al.* 2008; Schiller & Delgado 2010), direct empirical support has been limited as neuroimaging studies have almost exclusively focused on fear extinction. Animal work shows that the infralimbic cortex (the rodent homologue of the VMPFC) is involved in extinction of both appetitive and aversive CSs (Peters *et al.* 2009). In these studies, inactivation of the infralimbic cortex impairs extinction as well as extinction recall for CSs that during acquisition predicted shocks (e.g. Sierra-Mercado, Padilla-Coreano & Quirk 2011) but seems to *facilitate* extinction for CSs that during acquisition predicted appetitive reinforcers (e.g. sugar; Mendoza, Sanio & Chaudhri 2015). While additional studies are clearly needed to determine if the same VMPFC-mediated mechanism underlies extinction for all types of CSs, we speculate that the VMPFC stores the current value of the CS, likely in concert with other regions which themselves represent specific features of the CS (e.g. the amygdala in the case of aversive CSs and the striatum in the case of appetitive CSs).

In a previous fear extinction study (Phelps *et al.* 2004), reduction in SCR during extinction correlated with VMPFC activation a day later, suggesting that the VMPFC is specifically involved in the retrieval of extinction learning. We saw this same relationship here: lower SCR during day 1 extinction correlated with higher VMPFC activation during day 2 extinction. This relationship is further supported by evidence that the VMPFC, unlike a more dorsal medial region of the prefrontal cortex typically implicated in the expression of conditioned associations (the dorsal anterior cingulate), granger causally drives changes in skin conductance (Zhang *et al.* 2014). Despite differences in the activity profiles of the VMPFC for the CS_D+ versus the CS_P+, however, for both CSs, SCR decreased during extinction relative to acquisition. This cue-*insensitive* SCR pattern is not surprising given that SCR indexes arousal and is consistent with that observed in extinction learning studies of food- and shock-paired CSs (Andreatta & Pauli 2015). Importantly, despite eliciting similar SCRs, valence ratings clearly differentiated the CSs as appetitive or aversive in this prior study. Thus, as indices of the conditioned response, VMPFC BOLD and SCR might represent partly distinct aspects of the CS (valence versus arousal).

Perhaps most strikingly, while controls showed parametric VMPFC changes over the learning phases, that were modulated by CS type, chronic cocaine users did not. This was the case for both CSs, pointing to a generalized VMPFC abnormality in this group. Speaking to the clinical relevance of this finding, cocaine users who more successfully modulated their VMPFC response to the CS_D + during day 1 extinction (i.e. who looked more like controls) reported greater reductions in craving 24 hours later. While these data predict cocaine-addicted individuals with greater VMPFC impairments may be

particularly vulnerable in real-world situations involving drug cues, the SCR data did not reveal differences from controls. We used SCR as a passive measure of the conditioned response following a large body of literature. However, SCR is a noisy measurement in the MRI environment (indeed, there was substantial data loss that may have reduced our power to detect group differences in SCR if these differences existed). Alternatively, however, these neural differences might exist outside overt SCR differences. In studies of anxiety disorders, for example, reproducible VMPFC impairments during fear extinction (Milad *et al.* 2008; Milad *et al.* 2009) and its later retrieval (Milad *et al.* 2013) are observed in the absence of abnormalities in SCR. That is, VMPFC activation might be more sensitive in detecting differences from health, although this possibility remains to be tested. To directly test this possibility, future studies could incorporate trial-by-trial expectancy ratings as explicit measures of learning and additional physiological indices of arousal (e.g. pupil dilation) and valence [e.g. startle potentiation/attenuation (Andreatta & Pauli 2015)].

Contrary to expectations, the groups differed only in the VMPFC. Activation in the striatum mirrored that of the VMPFC, increasing in response to the CS_D+ and decreasing in response to the CS_P+ as extinction progressed, but there were no differences between cocaine users and controls. This coordinated pattern of activation in the VMPFC and striatum, another central node within the valuation system, further supports a value-sensitive account of extinction learning. But while we and others consistently find structural and functional impairments in stimulant users in the VMPFC (Alia-Klein *et al.* 2011; Konova *et al.* 2012; Parvaz *et al.* 2012; Ersche *et al.* 2013), that in some cases persist long after drug use ceases (Tanabe *et al.* 2009), a similar consensus finding regarding the striatum has been difficult to ascertain from the human neuroimaging literature, with a bulk of studies showing intact or even enhanced striatum function [see Balodis *et al.* (2012) and Konova & Goldstein (2015) for a detailed discussion on this topic]. Thus, the specific contribution of the striatum to extinction learning in drug addiction requires further study.

Given the theoretical impetus for extinction-based therapy in addiction (Taylor *et al.* 2009), it is important to consider how our VMPFC findings might inform treatment development, considering at least two potentially meaningful aims: enhancing VMPFC function or 'bypassing' it. For the former, the indirect dopamine agonist methylphenidate is shown to facilitate extinction and its retention, possibly via local effects in the infralimbic cortex (Abraham, Cunningham & Lattal 2012; Luo *et al.* 2015). This pharmacological approach may be particularly well-suited for cocaine-addicted individuals as prior studies in this population show that methylphenidate bolsters VMPFC function on both emotionally salient (Goldstein *et al.* 2010; Volkow *et al.* 2010) and emotionally neutral (Li *et al.* 2010; Moeller *et al.* 2014) tasks. Methylphenidate is also shown to modulate resting functional connectivity with the VMPFC (Konova *et al.* 2013). For the latter aim (decreasing reliance on the VMPFC), post-retrieval extinction, which interferes with memory reconsolidation, may offer a more efficacious method for targeting drug-related associations (Auber *et al.* 2013; Hutton-Bedbrook & McNally 2013). Post-retrieval extinction is shown to effectively reduce reinstatement and/or renewal of drug, sugar- and fear-related associations in rodents and humans (Monfils *et al.* 2009; Clem & Huganir 2010; Schiller *et al.* 2010; Flavell, Barber & Lee 2011; Rao-Ruiz *et al.* 2011; Xue *et al.* 2012; Sartor & Aston-Jones 2014). Some of this procedure's success is attributed to its relative independence of the VMPFC (Schiller *et al.* 2013) and may therefore represent a viable alternative for disorders with compromises in this region such as addiction.

In summary, a hallmark feature of drug addiction is an inability to discontinue drug seeking and use despite reduced pleasure derived from the drug and a range of negative consequences including the foregoing of other potentially rewarding outcomes. Here, we show that this inability may stem from a VMPFC-mediated impairment in forming and maintaining new associations for stimuli that were previously, though no longer, predictive of both drug and non-drug related outcomes. As this impairment may hinder the success of extinction-based therapies for addiction, future work could aim to concomitantly bolster VMPFC function and/or develop treatments that minimize reliance on this region.

ACKNOWLEDGEMENTS

This research was supported by grants from the National Institute on Drug Abuse (1F32DA039648 to A.B.K.; 5F32DA033088 to M.A.P.; 1K01DA037452 to S.J.M.; and 5R21DA020626, 2R21DA034954-01 and 5R01DA023579 to R.Z.G.) and the Netherlands Organization for Scientific Research (Rubicon 446-14-015 to A.Z.). The authors would like to thank Thomas Maloney, Daniel Carrero and Xufeng Han for help with data collection and analysis and Patricia A. Woicik for help with participant recruitment. The authors would also like to thank Elizabeth A. Phelps and Candace M. Raio for helpful discussions on the study design and early versions of the manuscript.

DISCLOSURE/CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

MAP, MRD, NA-K and RZG were responsible for the study concept and design. ABK and MAP contributed to data acquisition. ABK, MAP and VB analyzed the data. ABK drafted the manuscript. MAP, AZ, SJM, MRD and RZG provided critical revisions of the manuscript for important intellectual content. All authors critically reviewed content and approved the final version for journal submission.

REFERENCES

Abraham AD, Cunningham CL, Lattal KM (2012) Methylphenidate enhances extinction of contextual fear. Learn Mem 19:67–72.

Alia-Klein N, Parvaz MA, Woicik PA, Konova AB, Maloney T, Shumay E, Wang R, Telang F, Biegon A, Wang GJ, Fowler JS, Tomasi D, Volkow ND, Goldstein RZ (2011) Gene × disease interaction on orbitofrontal gray matter in cocaine addiction. Arch Gen Psychiatry 68:283–294.

Andreatta M, Pauli P (2015) Appetitive vs. aversive conditioning in humans. Front Behav Neurosci 9:128.

Auber A, Tedesco V, Jones CE, Monfils MH, Chiamulera C (2013) Post-retrieval extinction as reconsolidation interference: methodological issues or boundary conditions? Psychopharmacology (Berl) 226:631–647.

Balodis IM, Kober H, Worhunsky PD, Stevens MC, Pearlson GD, Potenza MN (2012) Attending to striatal ups and downs in addictions. Biol Psychiatry 72:e25–e26.

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 76:412–427.

Bouton ME (2004) Context and behavioral processes in extinction. Learn Mem 11:485–494.

Clem RL, Huganir RL (2010) Calcium-permeable AMPA receptor dynamics mediate fear memory erasure. Science 330:1108–1112.

Diekhof EK, Geier K, Falkai P, Gruber O (2011) Fear is only as deep as the mind allows: a coordinate-based meta-analysis of neuroimaging studies on the regulation of negative affect. Neuroimage 58:275–285.

Ersche KD, Williams GB, Robbins TW, Bullmore ET (2013) Meta-analysis of structural brain abnormalities associated with stimulant drug dependence and neuroimaging of addiction vulnerability and resilience. Curr Opin Neurobiol 23:615–624.

Faul F, Erdfelder E, Buchner A, Lang AG (2009) Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 41:1149–1160.

First MB, Spitzer RL, Gibbon M, Williams J (1996) Williams J. Structured Clinical Interview for DSM-IV Axis I Disorders—Patient Edition (SCID-I/P, Version 2.0). Biometrics Research Department, New York State Psychiatric Institute: New York.

Flavell CR, Barber DJ, Lee JL (2011) Behavioural memory reconsolidation of food and fear memories. Nat Commun 2:504.

Goldstein RZ, Volkow ND (2011) Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. Nat Rev Neurosci 12:652–669.

Goldstein RZ, Woicik PA, Maloney T, Tomasi D, Alia-Klein N, Shan J, Honorio J, Samaras D, Wang R, Telang F, Wang GJ, Volkow ND (2010) Oral methylphenidate normalizes cingulate activity in cocaine addiction during a salient cognitive task. Proc Natl Acad Sci U S A 107:16667–16672.

Grabenhorst F, Rolls ET (2011) Value, pleasure and choice in the ventral prefrontal cortex. Trends Cogn Sci 15:56–67.

Hartley CA, Fischl B, Phelps EA (2011) Brain structure correlates of individual differences in the acquisition and inhibition of conditioned fear. Cereb Cortex 21:1954–1962.

Hutton-Bedbrook K, McNally GP (2013) The promises and pitfalls of retrieval-extinction procedures in preventing relapse to drug seeking. Front Psych 4:14.

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 26:9503–9511.

Kober H, Lacadie CM, Wexler BE, Malison RT, Sinha R, Potenza MN (2016) Brain activity during cocaine craving and gambling urges: an fMRI study. Neuropsychopharmacology 41:628–637.

Konova AB, Goldstein RZ (2015) Addiction and addiction treatment. In: The Wiley Handbook on the Cognitive Neuroscience of Addiction, p. 109.Wiley - Blackwell: West Sussex, UK.

Konova AB, Moeller SJ, Tomasi D, Parvaz MA, Alia-Klein N, Volkow ND, Goldstein RZ (2012) Structural and behavioral correlates of abnormal encoding of money value in the sensorimotor striatum in cocaine addiction. Eur J Neurosci 36:2979–2988.

Konova AB, Moeller SJ, Tomasi D, Volkow ND, Goldstein RZ (2013) Effects of methylphenidate on resting-state functional connectivity of the mesocorticolimbic dopamine pathways in cocaine addiction. JAMA Psychiat 70:857–868.

Kuhn S, Gallinat J (2011) Common biology of craving across legal and illegal drugs—a quantitative meta-analysis of cue-reactivity brain response. Eur J Neurosci 33:1318–1326.

Li CS, Morgan PT, Matuskey D, Abdelghany O, Luo X, Chang JL, Rounsaville BJ, Ding YS, Malison RT (2010) Biological markers of the effects of intravenous methylphenidate on improving inhibitory control in cocaine-dependent patients. Proc Natl Acad Sci U S A 107:14455–14459.

Luo YX, Xue YX, Liu JF, Shi HS, Jian M, Han Y, Zhu WL, Bao YP, Wu P, Ding ZB, Shen HW, Shi J, Shaham Y, Lu L (2015) A novel UCS memory retrieval-extinction procedure to inhibit relapse to drug seeking. Nat Commun 6:7675.

Mackey S, Petrides M (2014) Architecture and morphology of the human ventromedial prefrontal cortex. Eur J Neurosci 40:2777–2796.

McCracken CB, Grace AA (2013) Persistent cocaine-induced reversal learning deficits are associated with altered limbic cortico-striatal local field potential synchronization. J Neurosci 33:17469–17482.

McLellan AT, Kushner H, Metzger D, Peters R, Smith I, Grissom G, Pettinati H, Argeriou M (1992) The fifth edition of the Addiction Severity Index. J Subst Abuse Treat 9:199–213.

Mendoza J, Sanio C, Chaudhri N (2015) Inactivating the infralimbic but not prelimbic medial prefrontal cortex facilitates the extinction of appetitive Pavlovian conditioning in Long-Evans rats. Neurobiol Learn Mem 118:198–208.

Milad MR, Furtak SC, Greenberg JL, Keshaviah A, Im JJ, Falkenstein MJ, Jenike M, Rauch SL, Wilhelm S (2013) Deficits in conditioned fear extinction in obsessive-compulsive disorder and neurobiological changes in the fear circuit. JAMA Psychiat 70:608–618.

Milad MR, Orr SP, Lasko NB, Chang Y, Rauch SL, Pitman RK (2008) Presence and acquired origin of reduced recall for fear extinction in PTSD: results of a twin study. J Psychiatr Res 42:515–520.

Milad MR, Pitman RK, Ellis CB, Gold AL, Shin LM, Lasko NB, Zeidan MA, Handwerger K, Orr SP, Rauch SL (2009) Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. Biol Psychiatry 66:1075–1082.

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. Proc Natl Acad Sci U S A 102:10 706–10 711.

Milad MR, Quirk GJ (2012) Fear extinction as a model for translational neuroscience: ten years of progress. Annu Rev Psychol 63:129–151.

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. Biol Psychiatry 62:446–454.

Moeller SJ, Honorio J, Tomasi D, Parvaz MA, Woicik PA, Volkow ND, Goldstein RZ (2014) Methylphenidate enhances executive function and optimizes prefrontal function in both health and cocaine addiction. Cereb Cortex 24:643–653.

Moeller SJ, Maloney T, Parvaz MA, Alia-Klein N, Woicik PA, Telang F, Wang GJ, Volkow ND, Goldstein RZ (2010) Impaired insight in cocaine addiction: laboratory evidence and effects on cocaine-seeking behaviour. Brain 133:1484–1493.

Moeller SJ, Maloney T, Parvaz MA, Dunning JP, Alia-Klein N, Woicik PA, Hajcak G, Telang F, Wang GJ, Volkow ND, Goldstein RZ (2009) Enhanced choice for viewing cocaine pictures in cocaine addiction. Biol Psychiatry 66:169–176.

Moeller SJ, Parvaz MA, Shumay E, Beebe-Wang N, Konova AB, Alia-Klein N, Volkow ND, Goldstein RZ (2013) Gene × abstinence effects on drug cue reactivity in addiction: multimodal evidence. J Neurosci 33:10027–10036.

Monfils MH, Cowansage KK, Klann E, LeDoux JE (2009) Extinction-reconsolidation boundaries: key to persistent attenuation of fear memories. Science 324:951–955.

Myers-Schulz B, Koenigs M (2012) Functional anatomy of ventromedial prefrontal cortex: implications for mood and anxiety disorders. Mol Psychiatry 17:132–141.

Neumann DL, Waters AM (2006) The use of an unpleasant sound as an unconditional stimulus in a human aversive Pavlovian conditioning procedure. Biol Psychol 73:175–185.

Parvaz MA, Konova AB, Tomasi D, Volkow ND, Goldstein RZ (2012) Structural integrity of the prefrontal cortex modulates electrocortical sensitivity to reward. J Cogn Neurosci 24:1560–1570.

Peters J, Kalivas PW, Quirk GJ (2009) Extinction circuits for fear and addiction overlap in prefrontal cortex. Learn Mem 16:279–288.

Phelps EA, Delgado MR, Nearing KI, LeDoux JE (2004) Extinction learning in humans: role of the amygdala and vmPFC. Neuron 43:897–905.

Prisciandaro JJ, Myrick H, Henderson S, McRae-Clark AL, Santa Ana EJ, Saladin ME, Brady KT (2013) Impact of DCS-facilitated cue exposure therapy on brain activation to cocaine cues in cocaine dependence. Drug Alcohol Depend 132:195–201.

Quirk GJ, Mueller D (2008) Neural mechanisms of extinction learning and retrieval. Neuropsychopharmacology 33: 56–72.

Rao-Ruiz P, Rotaru DC, van der Loo RJ, Mansvelder HD, Stiedl O, Smit AB, Spijker S (2011) Retrieval-specific endocytosis of GluA2-AMPARs underlies adaptive reconsolidation of contextual fear. Nat Neurosci 14:1302–1308.

Sartor GC, Aston-Jones G (2014) Post-retrieval extinction attenuates cocaine memories. Neuropsychopharmacology 39:1059–1065.

Schiller D, Delgado MR (2010) Overlapping neural systems mediating extinction, reversal and regulation of fear. Trends Cogn Sci 14:268–276.

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 110:20040–20045.

Schiller D, Levy I, Niv Y, LeDoux JE, Phelps EA (2008) From fear to safety and back: reversal of fear in the human brain. J Neurosci 28:11517–11525.

Schiller D, Monfils MH, Raio CM, Johnson DC, Ledoux JE, Phelps EA (2010) Preventing the return of fear in humans using reconsolidation update mechanisms. Nature 463: 49–53.

Schlund MW, Brewer AT, Richman DM, Magee SK, Dymond S (2015) Not so bad: avoidance and aversive discounting modulate threat appraisal in anterior cingulate and medial prefrontal cortex. Front Behav Neurosci 9:142.

Schoenbaum G, Saddoris MP, Ramus SJ, Shaham Y, Setlow B (2004) Cocaine-experienced rats exhibit learning deficits in a task sensitive to orbitofrontal cortex lesions. Eur J Neurosci 19:1997–2002.

Sierra-Mercado D, Padilla-Coreano N, Quirk GJ (2011) Dissociable roles of prelimbic and infralimbic cortices, ventral hippocampus, and basolateral amygdala in the expression and extinction of conditioned fear. Neuropsychopharmacology 36:529–538.

Spronk DB, Van der Schaaf ME, Cools R, De Bruijn ER, Franke B, van Wel JH, Ramaekers JG, Verkes RJ (2016) Acute effects of cocaine and cannabis on reversal learning as a function of COMT and DRD2 genotype. Psychopharmacology (Berl) 233:199–211.

Tanabe J, Tregellas JR, Dalwani M, Thompson L, Owens E, Crowley T, Banich M (2009) Medial orbitofrontal cortex gray matter is reduced in abstinent substance-dependent individuals. Biol Psychiatry 65:160–164.

Taylor JR, Olausson P, Quinn JJ, Torregrossa MM (2009) Targeting extinction and reconsolidation mechanisms to combat the impact of drug cues on addiction. Neuropharmacology 56:186–195.

Ventura J, Liberman RP, Green MF, Shaner A, Mintz J (1998) Training and quality assurance with the Structured Clinical Interview for DSM-IV (SCID-I/P). Psychiatry Res 79: 163–173.

Volkow ND, Wang GJ, Tomasi D, Telang F, Fowler JS, Pradhan K, Jayne M, Logan J, Goldstein RZ, Alia-Klein N, Wong C (2010) Methylphenidate attenuates limbic brain inhibition after cocaine-cues exposure in cocaine abusers. PLoS One 5: e11509.

Vollstadt-Klein S, Loeber S, Kirsch M, Bach P, Richter A, Buhler M, von der Goltz C, Hermann D, Mann K, Kiefer F (2011) Effects of cue-exposure treatment on neural cue reactivity in alcohol dependence: a randomized trial. Biol Psychiatry 69:1060–1066.

Xue YX, Luo YX, Wu P, Shi HS, Xue LF, Chen C, Zhu WL, Ding ZB, Bao YP, Shi J, Epstein DH, Shaham Y, Lu L (2012) A memory retrieval-extinction procedure to prevent drug craving and relapse. Science 336:241–245.

Zhang S, Hu S, Chao HH, Ide JS, Luo X, Farr OM, Li CS (2014) Ventromedial prefrontal cortex and the regulation of physiological arousal. Soc Cogn Affect Neurosci 9:900–908.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

Figure S1 Subjective ratings and two-alternative forced choice for the unconditioned stimuli (US, drug-related, pleasant, and neutral images; top) and the conditioned stimuli (CS, colored squares associated with the USs; bottom). Ratings and choice were collected for each US and each CS before and after day 1 acquisition, after day 1 extinction, and before and after day 2 extinction. The top panel shows average (a) valence ratings and (b) percent choice for the US_D+ (in blue; drug-related image) and US_P+ (in green; pleasant image) relative to the US− (neutral image) for each diagnostic group. The bottom panel shows (c) valence ratings and (d) percent choice for the CS_D+ (in blue; colored square associated with the drug-related image) and CS_P+ (in green; colored square associated with the pleasant image) relative to the CS− (colored square associated with the neutral image) for each time point across all participants. Data available for $n = 17$ cocaine users and $n = 13$ controls.

Figure S2 Skin conductance response (SCR) on unpaired trials where the conditioned stimulus (CS), a colored square, was presented without the paired unconditioned stimulus (a drug-related or pleasant image). SCR values shown are averaged over the 7 unpaired trials during acquisition (ACQ), the first 10 trials (early day 1-EXT) and second 10 trials (late day 1-EXT) during day 1 extinction, and the first 10 trials (early day 2-EXT) and second 10 trials (late day 2-EXT) during day 2 extinction, across all participants. Data available for $n = 11$ cocaine users and $n = 6$ controls.

Figure S3 Neural correlates of extinction learning for drug and affectively pleasant cues (wholebrain main effects, $P < 0.05$ cluster-level corrected shown at T range of 1 to 5). Images are shown in radiological convention (left=right). See Table S1. Abbreviations: ACC, anterior cingulate cortex; CER, cerebellum; HIPP, hippocampus; IFG, inferior frontal gyrus; INS, insula; IPL, inferior parietal lobe; SFG, superior frontal gyrus; STG, superior temporal gyrus.

Figure S4 Neural correlates of extinction learning for drug and affectively pleasant cues (wholebrain interaction effects, $P < 0.05$ cluster-level corrected shown at T range of 1 to 5). Images are shown in radiological convention (left=right). See Table S2. Abbreviations: ACC, anterior cingulate cortex; AMY, amygdala; CALC, calcarine; CUN, cuneus; IFG, inferior frontal gyrus; IPL, inferior parietal lobe; MCC, middle cingulate cortex; SMA, supplementary motor area; THAL, thalamus.

Figure S5 Role of the VMPFC in extinction learning for monetary rewards and association with skin conductance response (SCR). (a) Plot shows left VMPFC activation for each learning phase (acquisition, day 1 extinction, day 2 extinction) for the CS_M+ (cue paired with \$4) relative to the CS_M- (cue paired with \$0) as a function of diagnostic group (cocaine users, controls). There was a trend for a learning phase × diagnosis interaction in the ROI analyses suggesting left VMPFC activation decreased in response to the CS_M+ in controls but increased non-significantly in the cocaine user group. Whole-brain analyses revealed a significant and overlapping VMPFC region for the same interaction (shown in navy on the brain slice; MNI coordinates: x=-6, y=38, z=4, peak Z=3.9, 105 voxels, $P < 0.05$ cluster-level corrected). (b) The relationship between left VMPFC activation and the success of extinction on day 1 extinction as indexed by a reduction in skin conductance response (SCR) to the CS_M+ relative to the CS_M- on day 2 extinction across participants. VMPFC values plotted in panels a and b are from the unbiased ROI mask.

Table S1 Main Effects of Cue, Learning Phase, and Diagnosis on Conditioned Response to Drug (CS_D+) or Pleasant (CS_P+) versus Neutral (CS-) Cues (see also Figures 2 & 3)

Table S2 Interaction Effects of Cue, Learning Phase, and Diagnosis on Conditioned Response to Drug (CS_D+) or Pleasant (CS_P+) versus Neutral (CS-) Cues (see also Figures 2 & 3)