

**Association between Reward Sensitivity and Smoking Status in
Major Depressive Disorder**

Shengchuang Feng

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State
University in partial fulfillment of the requirements for the degree of

Master of Science
In
Psychology

Brooks King-Casas, Chair
Pearl H. Chiu
Jian Li

May 10, 2017
Blacksburg, Virginia

Keywords: reward sensitivity, dopamine, prediction error, nicotine, depression, fMRI

Association between Reward Sensitivity and Smoking Status in Major Depressive Disorder

Shengchuang Feng

Abstract

Nicotine use has been linked to increased sensitivity to nondrug rewards as well as improvement in mood among individuals with depression, and these effects have been hypothesized to be mediated through alternations in striatal dopamine activity. Similarly, nicotine use is hypothesized to influence the mechanisms by which healthy and depressed individuals learn about rewards in their environment. However, the specific behavioral and neural mechanisms by which chronic nicotine influences the learning process is poorly understood. Here, we use a probabilistic learning task, functional magnetic resonance imaging and neurocomputational analyses, to show that chronic smoking is associated with higher reward sensitivity, along with lower learning rate and striatal prediction error signal. Further, we show that these effects do not differ between individuals with and without major depressive disorder (MDD). In addition, a negative correlation between reward sensitivity and striatal prediction error signal was found among smokers, consistent with the suggestion that enhanced tonic dopamine associated with increased reward sensitivity leads to an attenuation of phasic dopamine activity necessary for updating of reward value during learning.

Keywords: reward sensitivity, dopamine, prediction error, nicotine, depression, fMRI

Association between Reward Sensitivity and Smoking Status in Major Depressive Disorder

Shengchuang Feng

General Audience Abstract

Nicotine use has been associated with increased sensitivity to nondrug rewards and improvement in mood among individuals with depression, and these effects may be attributed to neural activity in the striatum, a brain region related to reward perception and learning. Chronic nicotine may also influence reward learning in healthy and depressed individuals but this influence is poorly understood. Therefore, we use a computerized learning task to measure participants' behavioral performance, functional magnetic resonance imaging to collect their brain activity data and neurocomputational analyses to examine their learning process. We show that, compared to nonsmokers, chronic smokers perceive rewards as more rewarding (higher reward sensitivity), they learn the value of a reward slower (lower learning rate) and the neural activity in their striatum is weaker in the learning process. Moreover, we show that these effects do not differ between individuals with and without depression.

TABLE OF CONTENTS

Abstract..... ii

General Audience Abstract..... iii

Table of Contents..... iv

List of Tables..... vi

List of Figures..... vii

Introduction..... 1

Methods..... 2

 Participants..... 2

 Assessments..... 3

 Experimental paradigm..... 3

 Model-free analysis..... 3

 Model-based analysis..... 4

Results..... 7

 Behavior..... 7

 Model-free performance..... 7

 Group differences of learning rate/reward sensitivity..... 7

 Trade-off between learning rate and reward sensitivity..... 7

 Parameter recovery..... 8

 Imaging..... 8

 Group differences of striatal prediction error signal..... 8

 Correlations between learning parameters and striatal prediction error signal..... 9

Discussion..... 9

References..... 14

Tables..... 18

Figures..... 20

Supplementary Information..... 24

Supplementary methods.....	24
Participants.....	24
Experiment paradigm.....	25
Behavioral analysis.....	26
Model-free analysis.....	26
Model-based analysis.....	26
Imaging acquisition and fMRI analysis.....	30
Supplementary results.....	31
Supplementary references.....	33
Supplementary tables.....	34
Supplementary figures.....	36
Supplementary tables and figures showing results after groups were matched on sex.....	38

LIST OF TABLES

1. Demographic information and scale measures18

2. Model fit indices of each model19

S1. Model fit indices of *Multiplicative_Both model* and *Multiplicative_Both_GL model*.....34

S2. Brain regions that show prediction error signal.....35

S3. Demographic information and scale measures.....38

LIST OF FIGURES

1. Learning task, behavioral performance and parameter estimates.....20

2. Simulation and parameter recovery.....21

3. Neural substrates of prediction error (PE) and associations among learning parameters and striatal PE signal.....23

S1. Model-free performance.....36

S2. Parameter estimates in gain and loss framing.....37

S3. Model-free performance.....39

S4. Behavioral performance and parameter estimates.....40

S5. Group differences in striatal prediction error (PE) signal and associations among learning parameters and PE signal.....41

Association between Reward Sensitivity and Smoking Status in Major Depressive Disorder

Introduction

Cigarette smoking is a worldwide public health issue, with tobacco use directly linked to the causes of more than 5 million deaths each year (World Health Organization, 2016). Nicotine, the primary driver of smoking behavior, has been found to induce long-lasting increases in sensitivity to nondrug rewards (Attwood, Penton-Voak, & Munafò, 2009; Barr, Pizzagalli, Culhane, Goff, & Evins, 2008; Kenny & Markou, 2006) and alleviate depressive symptoms in general (Haro & Drucker-Colin, 2004; McClernon, Hiott, Westman, Rose, & Levin, 2006; Salín-Pascual, Rosas, Jimenez-Genchi, & Rivera-Meza, 1996), and anhedonia in particular (Janes et al., 2015; Liverant et al., 2014). Evidence suggests that the reward enhancing and antidepressive effects of nicotine are physiologically mediated by dopaminergic activity (Barr et al., 2008; Benowitz, 2008; Janes et al., 2015; Liverant et al., 2014). Acute nicotine administration has been shown to increase dopamine levels in the striatum of rats (Imperato, Mulas, & Di Chiara, 1986; Pontieri, Tanda, Orzi, & Di Chiara, 1996) and humans (Brody, et al, 2009; Brody, et al, 2004; Takahashi et al., 2008), and is attributable to phasic activity of midbrain dopamine neurons projecting to ventral striatum (Rice & Cragg, 2004; Zhang & Sulzer, 2004; Zhang et al., 2009). Phasic dopamine signaling has also been shown to encode the prediction error (the difference between obtained and predicted rewards) that underlie reward learning (Shultz, 2007; Shultz, Dayan & Montague, 1997), suggesting a potential effect of nicotine use on reward guided choice. However, the effect of chronic nicotine use on reward learning in humans, as well as its role in alleviating symptoms of depression, has received much less attention.

The two prior studies of smoking effects on reward learning suggest a potential influence on both the sensitivity to rewards during choice, and the updating of reward values in response to prediction errors. Addicott, Pearson, Wilson, Platt and McClernon (2013) found both diminished exploratory behavior, consistent with greater sensitivity to reward as well as enhanced learning rate among smokers relative to nonsmokers. In contrast, Chiu, Lohrenz and Montague (2008) found striatal prediction error signals predicted choice behavior to a greater extent among nonsmokers relative to smokers, suggesting a disruption in dopamine signaling among smokers. To address these inconsistencies, and examine the effect of chronic nicotine use on both behavioral and neural mechanisms of reward learning in healthy and depressed individuals, we employed a probabilistic learning task, functional magnetic resonance imaging (fMRI), and a reinforcement learning (RL) model incorporating independent effects of learning rate and reward sensitivity.

Methods

Participants

One hundred and two participants were included in the analyses [**Table 1**; see **Supplementary Information (SI)** for details of inclusion and exclusion criteria]. MDD smokers and MDD nonsmokers were matched on measures of depression symptoms. All four groups were matched on age and intelligence. MDD smokers, MDD nonsmokers and control nonsmokers were matched on sex ratio, but the sex ratio in control nonsmokers was different from the other three groups. To account for this and other potential imbalances, sex, age and intelligence were added as covariates in both behavioral and imaging analyses. Results of this sample were confirmed after all groups were matched on sex, age and intelligence (see **SI**). Participants had normal or corrected-to-normal vision and were able to follow verbal or written instructions.

Smokers were allowed to smoke normally before the experiment. Participants were informed of the study requirements and provided written consent before participation. The study was approved by the Institutional Review Board at the Virginia Tech.

Assessments

The following self-report measures were administered to all included participants prior to the scanning session: demographic information, Snaith-Hamilton Pleasure Scale (SHAPS ; Snaith et al., 1995), Beck Depression Inventory (BDI-II; Beck, Steer, & Brown, 1996), Mood and Anxiety Symptom Questionnaire (MASQ; Watson & Clark, 1991), a homemade smoking information questionnaire, and Fagerstrom Test for Nicotine Dependence (FTND; Heatherton, Kozlowski, Frecker, & Fagerstrom, 1991). Exhaled carbon monoxide was also obtained from all participants.

Experimental paradigm

During the scanning session, participants performed a probabilistic instrumental learning task with monetary outcomes, adapted from Pessiglione, Seymour, Flandin, Dolan & Frith (2006). The task was composed of alternating blocks of gain framing (participants only gain money) and loss framing (participants only lose money). In every block of either framing, participants were instructed to learn the reward contingencies of two abstract patterns by making choices between them. Both patterns were associated with a certain probability of better outcomes and worse outcomes, but one pattern was always associated with a higher probability of better outcomes than the other one (see **SI** and **Figure 1a** for details).

Behavioral analysis

Model-free analysis

In the probabilistic learning task, average accuracy, response time and earnings per trial across gains and losses were calculated for each participant. A “correct” trial

was defined as a trial in which a participant chose the better pattern, no matter what outcome that pattern yielded in that particular trial. Group differences in these performance measures and MDD/smoking scores were tested using 2×2 analyses of variance (ANOVAs) with smoking status and MDD status as two factors.

Trial-by-trial learning curves were also plotted for visual inspection (see **SI** and **Figure 1b**).

Model-based analysis

The basic form of an RL model (Sutton & Barto, 1998) describes an updating rule (Rescorla & Wagner, 1972) of expected value (Q) of a certain option.

$$Q_{(t+1)} = Q_{(t)} + \alpha * (V_{(t)} - Q_{(t)}) \quad (\text{Equation 1}).$$

In this equation, $Q_{(t+1)}$ and $Q_{(t)}$ denote expected value of an option at time (trial) $t+1$ and t , respectively; $V_{(t)}$ denotes the actual value of the option at time t ; $V_{(t)} - Q_{(t)}$ is termed prediction error, denoting the difference between the actual value and the expected value at time t . α is the learning rate, which describes how rapidly the agent uses the prediction error to update its expected value of an option. According to this rule, the agent would continuously update its expected value of a certain option at each time point. The agent's probability of choosing a certain option is calculated with a standard softmax function (see **SI**; Luce, 1959).

In our model-based analysis, following Huys, Pizzagalli, Bogdan and Dayan (2013), a parameter (ρ) was added as a multiplier of the outcome value V (Equation 2). This parameter is a measure of reward sensitivity. A larger ρ indicates greater internal worth of an external reward. This model was denoted as *Multiplicative_Both*, meaning ρ was multiplied on both the better and the worse outcomes.

$$Q_{(t+1)} = Q_{(t)} + \alpha * (\rho * V_{(t)} - Q_{(t)}) \quad (\text{Equation 2}).$$

It is possible that ρ is multiplicative or additive to both the better and the worse outcomes or only the better outcomes. Therefore, three alternative reward sensitivity models were also constructed: In *Multiplicative_Better model*, ρ was multiplied only on the better outcomes. In *Additive_Both model*, ρ was additive to both the better and worse outcomes. In *Additive_Better model*, ρ was only additive to the better outcomes (see **SI**).

The basic RL model and four reward sensitivity models were estimated to fit the participants' choice data. Hierarchical Bayesian analysis (performed with the Stan software package; Stan Development Team, 2016) was used to estimate the free parameters (learning rate and reward sensitivity) of each model for individual participants. After model estimation, model fit indices, including Akaike information criterion, Bayesian information criterion and widely applicable information criterion, were calculated for all five models.

To decide whether to use two separate sets of parameters or a single set of parameters for gain and loss framing, another model similar to the best-fitting model (*Multiplicative_Both model*) of the above five models was constructed. This model had reward sensitivity multiplied on both the better outcomes and the worse outcomes, but it had separate learning rate and reward sensitivity for gain and loss framing (denoted *Multiplicative_Both_GL model*). The model fit indices were also compared between *Multiplicative_Both model* and *Multiplicative_Both_GL model*.

For the winning model (*Multiplicative_Both model*), the estimated free parameters were extracted. Group differences of individual estimates of learning rate and reward sensitivity were tested using 2×2 ANOVAs with smoking status and MDD status as two factors. Parameters from gain and loss (estimated with *Multiplicative_Both_GL model*) were tested with the same ANOVAs to validate the

combination of two types of framing. The correlations between learning rate/reward sensitivity and smoking/depression measures were also conducted.

To examine the relationship between learning rate, reward sensitivity and accuracy of participants' choices, a simulation analysis was performed, in which values of accuracy were calculated for different combinations of learning rate (ranging from 0 to 1 in a step size of 0.005) and reward sensitivity (ranging from 0 to 5 in a step size of 0.02) through the best-fitting model. To check whether the winning model had clearly identifiable parameters, a parameter recovery test was also conducted by estimating the learning parameters of 100 simulated participants. A detailed description of the simulation and the parameters recovery can be found in **SI**.

Image acquisition and fMRI analysis

The functional and anatomical imaging was conducted on 3 T Siemens Trio MR scanners (Siemens, Munich, Germany) at Baylor College of Medicine. SPM8 software package (<http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>) was used for imaging data preprocessing and analysis (see **SI** for details of image acquisition, preprocessing and analysis).

Trial-by-trial prediction errors ($V - Q$) calculated from the best-fitting model were entered into the first-level general linear model (GLM) as a parametric modulator of the outcome event. The contrast and beta maps of prediction error were used in a second-level whole-brain analysis and a region of interest (ROI) analysis. In the whole-brain analysis, the contrast maps of prediction errors were entered into a one sample t-test to assess the prediction error-related activation. In the ROI analysis, the ventral striatum prediction error signal for each participant was extracted from the beta maps with the method of leave-one-subject-out (Esterman, Tamber-Rosenau, Chiu & Yantis, 2010; see **SI**).

Ventral striatum prediction error signal was compared among the four groups using a 2×2 ANOVA. Correlation analyses between this striatal prediction error signal and behavioral measures were also conducted within smokers. .

Results

Behavior

Model-free performance

Accuracy, response time and earnings were not significantly different among groups (F 's ≤ 2.29 , P 's ≥ 0.13), indicating neither MDD nor smoking status was related to performance in the task (**Figure S1**).

Group differences of learning rate/reward sensitivity

The model comparison result indicated that *Multiplicative_Both model* was the winning model, which had the smallest model fit indices (**Table 2; Table S1**). ANOVAs of individual estimates of learning rate and reward sensitivity from this model showed that the main effect of smoking was significant: smokers had lower learning rate [$F(1, 95) = 9.81$, $P < 0.01$] but higher reward sensitivity [$F(1, 95) = 13.02$, $P < 0.001$] than nonsmokers (**Figure 1c**). The main effect of MDD was not significant for learning rate or reward sensitivity (F 's ≤ 0.01 , P 's ≥ 0.92), and there was no interaction between MDD and smoking (F 's ≤ 0.13 , P 's ≥ 0.71). Learning rate and reward sensitivity had the similar pattern in gain and loss framing (**Figure S2**).

Trade-off between learning rate and reward sensitivity

In addition to the robust effects of smoking, learning rate and reward sensitivity were themselves related. Across all four groups, a strong negative correlation was found between learning rate and reward sensitivity ($r = -0.51$, $P < 0.001$; **Figure 2a** right panel). Collectively, the ANOVA results, the negative correlation and

comparable performance of the four groups suggested a trade-off relationship between learning rate and reward sensitivity, which in turn determined the performance.

The simulation was specifically aimed to test the learning rate-reward sensitivity trade-off. It turned out that there were some optimal combinations of learning rate and reward sensitivity leading to a high level of accuracy (**Figure 2a** left panel). The combination of a low reward sensitivity and a high learning rate or vice versa is more likely to yield good performance. The heatmap in the right panel of **Figure 2a** shows the empirical relationship between the learning parameters and accuracy and was highly similar to the simulation heatmap in the left panel. The scatterplot of the empirical learning rate-reward sensitivity correlation also shows that smokers had lower learning rate but higher reward sensitivity, while nonsmokers had lower reward sensitivity but higher learning rate. In addition, the centroids of the two groups of dots seem to be located in regions with similar colors (similar accuracy), which is consistent with the ANOVA results showing no group differences in accuracy (**Figure S1a**).

Parameter recovery

Despite the strong correlation and the trade-off between learning rate and reward sensitivity, the parameter recovery results indicated that both parameters were fully recoverable, with a high correlation between simulated and recovered parameters (r 's ≥ 0.89 , P 's < 0.001 ; **Figure 2b**).

Imaging

Group differences of striatal prediction error signal

The whole-brain analysis of all participants revealed robust activation of the bilateral ventral striatum ($P < 0.05$, family-wise error corrected for multiple comparisons at voxel-level over the whole brain; **Figure 3a**; **Table S2**). The ROI

analysis showed that prediction error signal within the ventral striatum had a similar pattern as learning rate, with smokers lower than nonsmokers [$F(1, 95) = 5.87, P < 0.05$; **Figure 3b**]. There was no main effect of MDD or interaction between MDD and smoking (F 's $\leq 0.36, P \geq 0.55$).

Correlations between learning parameters and striatal prediction error signal

Within smokers, learning rate and reward sensitivity were negatively correlated with each other ($r = -0.34, P < 0.05$). Striatal prediction error was positively correlated with learning rate but negatively correlated with reward sensitivity ($|r$'s $\geq 0.33, P$'s < 0.05 ; **Figure 3c**).

Discussion

As our behavioral results show, chronic smokers had higher reward sensitivity than nonsmokers, which corroborates previous nicotine and reward sensitivity studies (Barr et al., 2008; Janes et al., 2015; Kenny & Markou, 2006; Liverant et al., 2014). Nicotine is believed to increase the salience of non-drug incentives (Barr et al., 2008), thus it magnifies the rewarding effect of stimuli. Learning rate had an opposite pattern, with chronic smokers had lower learning rate than nonsmokers, which drives a strong negative correlation between learning rate and reward sensitivity across groups. Interestingly, the performance of the four groups was comparable. Both the simulation and the empirical heatmaps suggest a trade-off relationship between the reward sensitivity to stimuli and how fast the values of stimuli are learned. This trade-off may be made to achieve a relatively optimal level of performance, which was about 70% according to the mean accuracy of the four groups. The decreased learning rate in smokers may also reflect smoking-induced learning deficits. This altered learning can be found in the model-free learning curves as well, which show a

visual trend that smokers reached the relatively high performance later than nonsmokers.

In the imaging results, striatal prediction error signal showed a similar pattern of group differences to learning rate, but an opposite pattern to reward sensitivity. In addition, a negative correlation between reward sensitivity and striatal signal was found in smokers. It is proposed that repeated exposure to drugs of abuse inhibits the removal of extracellular dopamine, leading to a higher level of extracellular dopamine concentration (also termed tonic activity; Grace, 2000), and lower level of postsynaptic dopamine receptor availability in ventral striatum (Nutt, Lingford-Hughes, Erritzoe & Stokes, 2015). High tonic dopamine can impose a sustained suppression on phasic responses (Leknes & Tracey, 2008), through dopamine-releasing-inhibiting autoreceptors (Grace, 2000; Seeman & Madras, 1998). Meanwhile, lower striatal dopamine receptor availability may also diminish the phasic signaling. Our results are supportive of this notion, in that smokers showed decreased striatal prediction error signal, suggesting diminished phasic activity. The authors of most previous studies focused on the effect of acute nicotine administration and they found an enhancement of striatal phasic activity while a suppression of tonic activity (Rice & Cragg, 2004; Zhang & Sulzer, 2004; Zhang et al., 2009). As a complement to previous findings, our results provide evidence that chronic smoking diminishes phasic activity in the ventral striatum, presumably through heightened tonic activity and/or lowered postsynaptic dopamine receptor availability.

Moreover, studies show that increases in tonic dopamine, which is mimicked by oral administration of low dose methylphenidate, can enhance the salience of environmental stimuli (Volkow, Fowler, Wang & Swanson, 2004a). Here we propose that the higher tonic dopamine activity of smokers contributes to their higher reward

sensitivity, reflecting increased reward salience. The phasic activity, however, is weakened in smokers, resulting in diminished responsiveness to prediction error and decreased learning rate. The former is about reward perception, while the latter is about reward learning. Although smokers have heightened salience of reward, they may have impaired reward learning, which is consistent with a recent meta-analysis (Luijten et al., 2017), or they may have an altered trade-off pattern between learning rate and reward sensitivity, which is evidenced by our simulation and empirical heatmaps.

Although the smoking effect is strong for both learning rate and reward sensitivity, we found no MDD effect or interaction between smoking and MDD. Most previous reward sensitivity studies employed a signal detection paradigm, and they found participants with anhedonic symptoms showed an impaired response bias toward the stimulus with more frequent reward, indicating reduced reward sensitivity in depressed individuals (Huys et al., 2013; Pizzagalli et al., 2009; Pizzagalli, Jahn & O'Shea, 2005). The missing MDD effect in our study may be due to different stimuli presentation durations between the signal detection task (100 ms) and our two-armed bandit task (present on the screen until participants made their choices). Results from information-processing bias studies suggest that depression-associated group differences can be moderated by exposure durations (Browning, Holmes & Harmer, 2010; McCabe & Gotlib, 1995; Mogg & Bradley, 2005). For example, MDD patients had significantly lower discrimination accuracy for sad and happy facial expressions than controls when the stimuli presented for 100 ms, but the group differences were not significant when the stimuli presented for 2,000 ms (Surguladze et al., 2004). This moderation can be attributed to the general slowing of cognitive processes in depressed patients (Cooley & Nowicki, 1989). An alternative reason may be that

MDD patients have a tendency to attend negative stimuli (Browning et al., 2010) and the combination of gain and loss framing in our data analysis weakened a possible MDD effect in the loss framing. However, when we separated gain and loss framing, no main effect of MDD was observed in loss (**Figure S2**). Therefore, this explanation can be excluded.

Another null result is also worth noting. Although reward sensitivity is a measure of reward perception, no correlations were found between reward sensitivity and anhedonia measures within MDD smokers, MDD nonsmokers or within all MDD participants. The lack of association between reward task-measured reward sensitivity and self-report reward responsiveness/anhedonia also occurred in previous studies (Janes et al., 2015; Liverant et al., 2014). This suggests that the reward sensitivity estimated from the reward tasks is not isomorphic to the anhedonia measured with self-report scales, but rather an index of reward salience.

The results of this study have two important implications. Firstly, it provides supportive evidence of dopamine's role in encoding both reward learning and reward salience/motivation (Volkow et al., 2004b). Specifically, phasic dopamine activity encodes reward learning and tonic dopamine activity encodes reward salience. The relationship between tonic and phasic activity maps the trade-off between reward learning and reward perception. Secondly, it provides a mechanistic explanation of the reward salience-enhancing effect of chronic nicotine, namely, through increased tonic dopamine. This echoes the medications for attention deficit hyperactivity disorder and Parkinson's disease. Oral administration of low dose methylphenidate can enhance the salience of a monetarily rewarded mathematical task by increasing tonic dopamine in healthy participants (Volkow et al., 2004a). Dopamine medication in Parkinson's disease patients is associated with lower striatal phasic prediction error

signal and learning rate but higher inverse temperature (consistent with higher reward sensitivity) in both gain (Schmidt, Braun, Wager & Shohamy, 2014) and loss conditions (Voon et al., 2010). The neural relationship between tonic and phasic activity as well as the behavioral trade-off between reward sensitivity and learning rate should be targeted in the treatment of nicotine addiction or using nicotine as a medication of other mental disorders. In spite of the implications, some limitations should also be noted. Since the present study is correlational rather than causal, we can't exclude the possibility that the imbalance between reward learning and reward perception predisposes to smoking initiation. Another limitation is that the health consequences of the imbalanced reward learning and perception were not examined in this study. Future studies should test whether and to what extent the imbalance is related to other psychological components, which will shed light on the treatment of addiction and other disorders.

References

- Addicott, M. A., Pearson, J. M., Wilson, J., Platt, M. L., & McClernon, F. J. (2013). Smoking and the bandit: A preliminary study of smoker and nonsmoker differences in exploratory behavior measured with a multiarmed bandit task. *Experimental and Clinical Psychopharmacology*, *21*(1), 66.
- Attwood, A. S., Penton-Voak, I. S., & Munafò, M. R. (2009). Effects of acute nicotine administration on ratings of attractiveness of facial cues. *Nicotine & Tobacco Research*, *11*(1), 44-48.
- Barr, R. S., Pizzagalli, D. A., Culhane, M. A., Goff, D. C., & Evins, A. E. (2008). A single dose of nicotine enhances reward responsiveness in nonsmokers: Implications for development of dependence. *Biological Psychiatry*, *63*(11), 1061-1065.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck depression inventory-II*. San Antonio, TX: The Psychological Corporation.
- Benowitz, N. L. (2008). Neurobiology of nicotine addiction: Implications for smoking cessation treatment. *The American Journal of Medicine*, *121*(4), S3-S10.
- Brody, A. L., Mandelkern, M. A., Olmstead, R. E., Allen-Martinez, Z., Scheibal, D., Abrams, A. L., ... & London, E. D. (2009). Ventral striatal dopamine release in response to smoking a regular vs a denicotinized cigarette. *Neuropsychopharmacology*, *34*(2), 282-289.
- Brody, A. L., Olmstead, R. E., London, E. D., Farahi, J., Meyer, J. H., Grossman, P., ... & Mandelkern, M. A. (2004). Smoking-induced ventral striatum dopamine release. *American Journal of Psychiatry*, *161*(7), 1211-1218.
- Browning, M., Holmes, E. A., & Harmer, C. J. (2010). The modification of attentional bias to emotional information: A review of the techniques, mechanisms, and relevance to emotional disorders. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(1), 8-20.
- Chen, C., Takahashi, T., Nakagawa, S., Inoue, T., & Kusumi, I. (2015). Reinforcement learning in depression: A review of computational research. *Neuroscience & Biobehavioral Reviews*, *55*, 247-267.
- Chiu, P. H., Lohrenz, T. M., & Montague, P. R. (2008). Smokers' brains compute, but ignore, a fictive error signal in a sequential investment task. *Nature Neuroscience*, *11*(4), 514-520.
- Cooley, E. L., & Nowicki, S. (1989). Discrimination of facial expressions of emotion by depressed subjects. *Genetic, Social, and General Psychology Monographs*, *115*, 449-465.
- Esterman, M., Tamber-Rosenau, B. J., Chiu, Y. C., & Yantis, S. (2010). Avoiding non-independence in fMRI data analysis: Leave one subject out. *Neuroimage*, *50*(2), 572-576.
- Grace, A.A. (2000). The tonic/phasic model of dopamine system regulation and its implications for understanding alcohol and psychostimulant craving. *Addiction*, *95*:S119-S128.

- Haro, R., & Drucker-Colin, R. (2004). Effects of long-term administration of nicotine and fluoxetine on sleep in depressed patients. *Archives of Medical Research*, 35(6), 499-506.
- Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerstrom, K. O. (1991). The Fagerström test for nicotine dependence: a revision of the Fagerstrom Tolerance Questionnaire. *British Journal of Addiction*, 86(9), 1119-1127.
- Huys, Q. J., Pizzagalli, D. A., Bogdan, R., & Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: A behavioural meta-analysis. *Biology of Mood & Anxiety Disorders*, 3(12), 1-16.
- Imperato, A., Mulas, A., & Di Chiara, G. (1986). Nicotine preferentially stimulates dopamine release in the limbic system of freely moving rats. *European Journal of Pharmacology*, 132(2-3), 337-338.
- Janes, A. C., Pedrelli, P., Whitton, A. E., Pechtel, P., Douglas, S., Martinson, M. A., ... & Evins, A. E. (2015). Reward responsiveness varies by smoking status in women with a history of major depressive disorder. *Neuropsychopharmacology*, 40(8), 1940-1946.
- Kenny, P. J., & Markou, A. (2006). Nicotine self-administration acutely activates brain reward systems and induces a long-lasting increase in reward sensitivity. *Neuropsychopharmacology*, 31(6), 1203-1211.
- Leknes, S., & Tracey, I. (2008). A common neurobiology for pain and pleasure. *Nature Reviews Neuroscience*, 9(4), 314-320.
- Liverant, G. I., Sloan, D. M., Pizzagalli, D. A., Harte, C. B., Kamholz, B. W., Rosebrock, L. E., ... & Kaplan, G. B. (2014). Associations among smoking, anhedonia, and reward learning in depression. *Behavior Therapy*, 45(5), 651-663.
- Luce, R. D. (1959). *Individual Choice Behavior*. New York, NY: John Wiley & Sons, Inc.
- Luijten, M., Schellekens, A. F., Kühn, S., Machielse, M. W., & Sescousse, G. (2017). Disruption of Reward Processing in Addiction: An Image-Based Meta-analysis of Functional Magnetic Resonance Imaging Studies. *JAMA Psychiatry*, 74(4), 387-398.
- McCabe, S. B., & Gotlib, I. H. (1995). Selective attention and clinical depression: Performance on a deployment-of-attention task. *Journal of Abnormal Psychology*, 104(1), 241-245.
- McClernon, F. J., Hiott, F. B., Westman, E. C., Rose, J. E., & Levin, E. D. (2006). Transdermal nicotine attenuates depression symptoms in nonsmokers: A double-blind, placebo-controlled trial. *Psychopharmacology*, 189(1), 125-133.
- Mogg, K., & Bradley, B. P. (2005). Attentional bias in generalized anxiety disorder versus depressive disorder. *Cognitive Therapy and Research*, 29(1), 29-45.
- Nutt, D. J., Lingford-Hughes, A., Erritzoe, D., & Stokes, P. R. (2015). The dopamine theory of addiction: 40 years of highs and lows. *Nature Reviews Neuroscience*, 16(5), 305-312.
- Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, 442(7106), 1042-1045.

- Pizzagalli, D. A., Holmes, A. J., Dillon, D. G., Goetz, E. L., Birk, J. L., Bogdan, R., ... & Fava, M. (2009). Reduced caudate and nucleus accumbens response to rewards in unmedicated individuals with major depressive disorder. *American Journal of Psychiatry*, *166*(6), 702-710.
- Pizzagalli, D. A., Jahn, A. L., & O'Shea, J. P. (2005). Toward an objective characterization of an anhedonic phenotype: A signal-detection approach. *Biological Psychiatry*, *57*(4), 319-327.
- Pontieri, F. E., Tanda, G., Orzi, F., & Di Chiara, G. (1996). Effects of nicotine on the nucleus accumbens and similarity to those of addictive drugs. *Nature*, *382*(6588), 255.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A.H. Black & W.F. Prokasy (eds.), *Classical Conditioning II: Current Research and Theory* (pp. 64-99). New York: Appleton-Century-Crofts.
- Rice, M. E., & Cragg, S. J. (2004). Nicotine amplifies reward-related dopamine signals in striatum. *Nature Neuroscience*, *7*(6), 583-584.
- Salín-Pascual, R. J., Rosas, M., Jimenez-Genchi, A., & Rivera-Meza, B. L. (1996). Antidepressant effect of transdermal nicotine patches in nonsmoking patients with major depression. *Journal of Clinical Psychiatry*, *57*(9), 387-389.
- Seeman, P., & Madras, B. K. (1998). Anti-hyperactivity medication: Methylphenidate and amphetamine. *Molecular Psychiatry*, *3*(5), 386-396.
- Schmidt, L., Braun, E. K., Wager, T. D., & Shohamy, D. (2014). Mind matters: Placebo enhances reward learning in Parkinson's disease. *Nature Neuroscience*, *17*(12), 1793-1797.
- Schultz, W. (2007). Behavioral dopamine signals. *Trends in Neurosciences*, *30*(5), 203-210.
- Schultz, W. (2010). Dopamine signals for reward value and risk: Basic and recent data. *Behavioral and Brain Functions*, *6*(24), 1-9.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593-1599.
- Snaith, R. P., Hamilton, M., Morley, S., Humayan, A., Hargreaves, D., & Trigwell, P. (1995). A scale for the assessment of hedonic tone the Snaith-Hamilton Pleasure Scale. *The British Journal of Psychiatry*, *167*(1), 99-103.
- Surguladze, S. A., Young, A. W., Senior, C., Brébion, G., Travis, M. J., & Phillips, M. L. (2004). Recognition accuracy and response bias to happy and sad facial expressions in patients with major depression. *Neuropsychology*, *18*(2), 212-218.
- Sutton, R. S. & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.
- Takahashi, H., Fujimura, Y., Hayashi, M., Takano, H., Kato, M., Okubo, Y., ... & Suhara, T. (2008). Enhanced dopamine release by nicotine in cigarette smokers: A double-blind, randomized, placebo-controlled pilot study. *International Journal of Neuropsychopharmacology*, *11*(3), 413-417.

- Volkow, N. D., Fowler, J. S., Wang, G. J., & Swanson, J. M. (2004a). Dopamine in drug abuse and addiction: results from imaging studies and treatment implications. *Molecular Psychiatry*, 9, 557-569.
- Volkow, N. D., Wang, G. J., Fowler, J. S., Telang, F., Maynard, L., Logan, J., ... & Zhu, W. (2004b). Evidence that methylphenidate enhances the saliency of a mathematical task by increasing dopamine in the human brain. *American Journal of Psychiatry*, 161(7), 1173-1180.
- Voon, V., Pessiglione, M., Brezing, C., Gallea, C., Fernandez, H. H., Dolan, R. J., & Hallett, M. (2010). Mechanisms underlying dopamine-mediated reward bias in compulsive behaviors. *Neuron*, 65(1), 135-142.
- Watson, D., & Clark, L. A. (1991). The mood and anxiety symptom questionnaire. *Unpublished manuscript, University of Iowa, Department of Psychology, Iowa City.*
- Wechsler, D. (2001). *Wechsler Test of Adult Reading (WTAR)*. San Antonio, TX: The Psychological Corporation.
- World Health Organization. (2016). Tobacco fact sheet. Available from <http://www.who.int/mediacentre/factsheets/fs339/en/>.
- Zhang, H., & Sulzer, D. (2004). Frequency-dependent modulation of dopamine release by nicotine. *Nature Neuroscience*, 7(6), 581-582.
- Zhang, T., Zhang, L., Liang, Y., Siapas, A. G., Zhou, F. M., & Dani, J. A. (2009). Dopamine signaling differences in the nucleus accumbens and dorsal striatum exploited by nicotine. *Journal of Neuroscience*, 29(13), 4035-4043.

Tables

Table 1. Demographic information and scale measures

Group	MDD smokers (n=25)	Control smokers (n=20)	MDD nonsmokers (n=25)	Control nonsmokers (n=32)
Sex	13 females	3 females	17 females	21 females
Age (years)	36.64±1.91	35.80±2.25	35.24±2.38	33.75±2.08
WTAR	103.28±2.31	102.40±3.22	107.44±2.08	105.16±2.25
Exhaled CO (ppmv)	7.80±1.31 ^a	10.14±2.26 ^b	2.12±0.15	1.97±0.19
Mean cigarettes per day	14.28±1.88	12.20±1.37	0±0	0±0
Years smoking began	19.76±2.02	19.40±2.22	0±0	0±0
Years regular smoking	18.28±1.99	16.6±2.14	0±0	0±0
FTND	4.92±0.60 ^c	4.86±0.74 ^b	0±0	0±0
BDI	35.56±2.04	2.75±0.89	34.68±1.16	2.88±0.89
SHAPS	6.84±0.83	0.90±0.53 ^d	6.40±0.47	0.59±0.44
MASQ anhedonia	90.04±1.94	47.30±4.31 ^d	88.72±1.26	44.28±1.79

Abbreviations: WTAR, Wechsler Test of Adult Reading (Wechsler, 2001); CO, carbon monoxide; ppmv, parts per million by volume; FTND, Fagerstrom Test for Nicotine Dependence; BDI, Beck Depression Index; SHAPS, Snaith-Hamilton Pleasure Scale; MASQ, Mood and Anxiety Symptom Questionnaire. ^a This measure is available for 22 participants in MDD smokers. ^b This measure is available for 7 participants in control smokers. ^c This measure is available for 24 participants in MDD smokers. ^d This measure is available for 10 participants in control smokers. Data are represented as mean± standard error (SE).

Table 2. Model fit indices of each model

Model	AIC^a	BIC^a	WAIC
Basic RL	122.20	127.51	12189.0
Additive_Better	116.63	121.95	11535.5
Additive_Both	113.98	119.29	11522.0
Multiplicative_Better	116.68	121.99	11548.1
Multiplicative_Both	<u>112.95</u>	<u>118.27</u>	<u>11347.9</u>

Abbreviations: AIC, Akaike information criterion; BIC, Bayesian information criterion; WAIC, widely applicable information criterion. ^a These are mean AIC and BIC for each participant. AIC and BIC were used as cross-validation of WAIC. We asked if the individual parameters from the Bayesian analysis were estimated from MLE method for each individual subject, was the best-fitting model still winning. We calculated one AIC and BIC value for each subject and took the mean as a predictive accuracy measure of the choice data. Model fit indices for the best-fitting model are underlined.

Figures

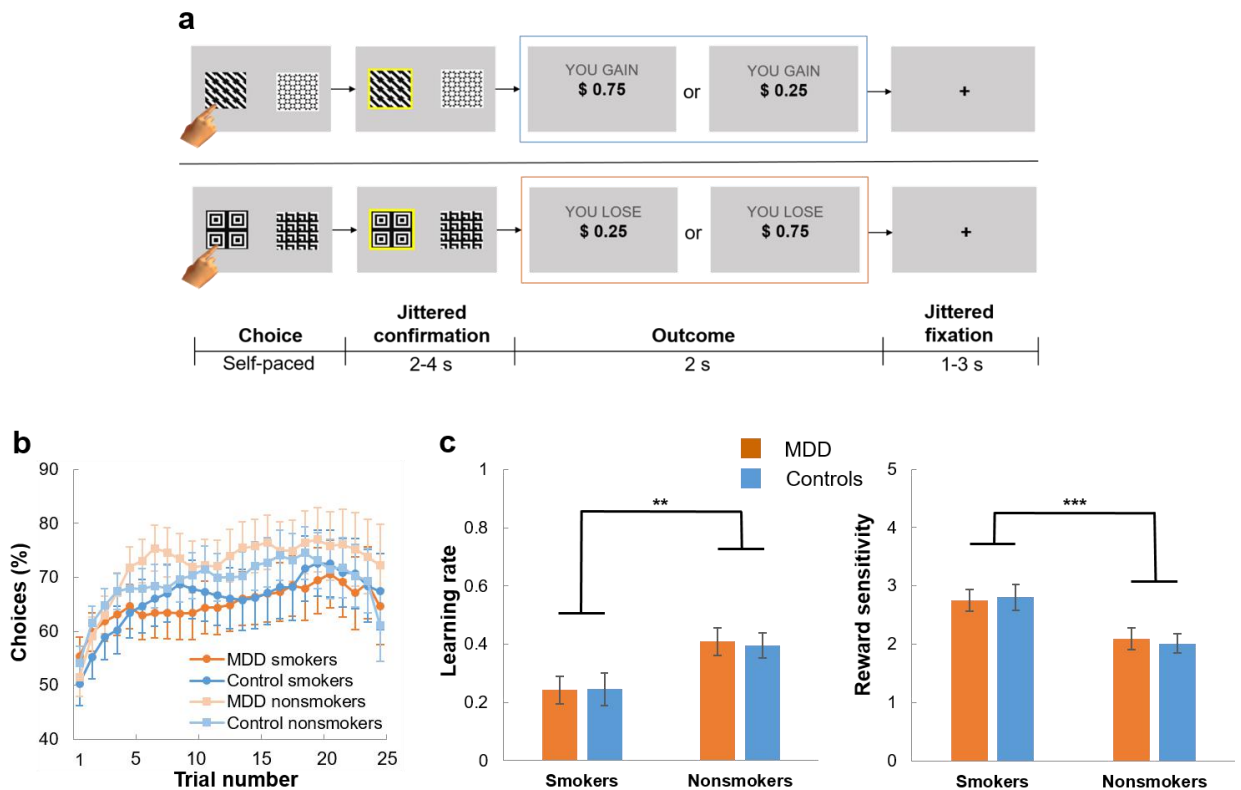


Figure 1. Learning task, behavioral performance and parameter estimates. (a) Probabilistic reward and punishment learning task. Participants choose between two patterns, view the monetary outcome of the choice, and learn over time which is the better option. **(b)** The learning curves depict the running average [window size = 5, averaged over all blocks, mean \pm standard error (SE)] of the trial-by-trial proportion of choices in each group that selected the better option. The four lines represent four groups and show learning. **(c)** Model estimation of learning rate and reward sensitivity. Both parameters show significant main effect of smoking status, with smokers have lower learning rate but higher reward sensitivity than nonsmokers. The main effect of depression status and interaction between depression and smoking are nonsignificant (P 's ≥ 0.92). ** $P < 0.01$. *** $P < 0.001$. Error bars indicate 1 SE.

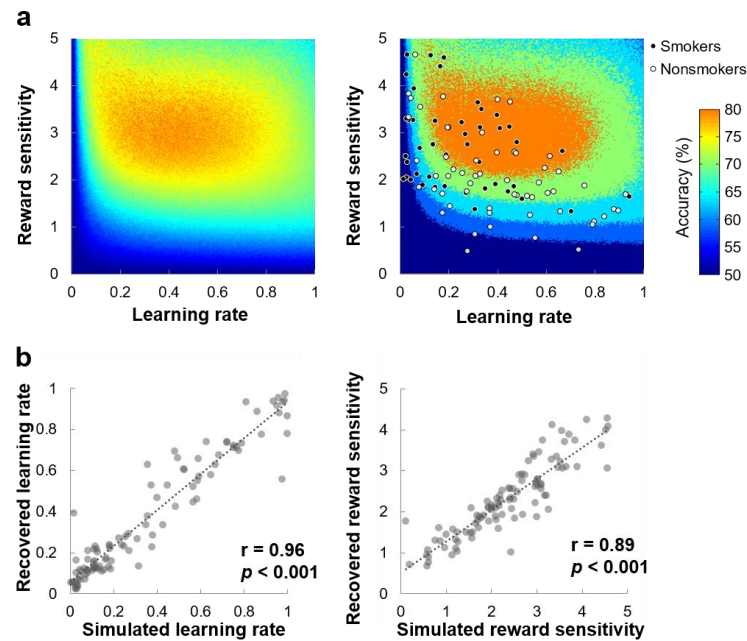


Figure 2. Simulation and parameter recovery. (a) simulated (left) and empirical (right) relationship between learning rate, reward sensitivity and performance. The simulated heatmap is a 201×251 matrix with each element as an accuracy value (percentage of trials a simulated participant chooses the better patterns) calculated from a combination of one learning rate value and one reward sensitivity value. For the heatmap on the right, the simulation heatmap is divided into five areas according to the simulated accuracy $[[0, 60\%), [60\%, 65\%), [65\%, 70\%), [70\%, 75\%),$ and $[75\%, 100\%]]$, and the color of each area is determined by the mean accuracy of all dots (empirical participants) falling within each area. The empirical scatterplot also shows a negative correlation between learning rate and reward sensitivity ($r = -0.51, P < 0.001$). (b) Parameter recovery for learning parameters, showing the ability of model estimation approach to recover the “true” parameter values. We randomly drew 100 learning rate values and 100 reward sensitivity values from two distributions based on empirical group-level estimates (see SI). These “true” or simulated

parameters were randomly paired to form 100 simulated participants and to generate choice data. These 100 learning rate values and 100 reward sensitivity values can be fully recovered with the best-fitting model, as shown by the strong correlations.

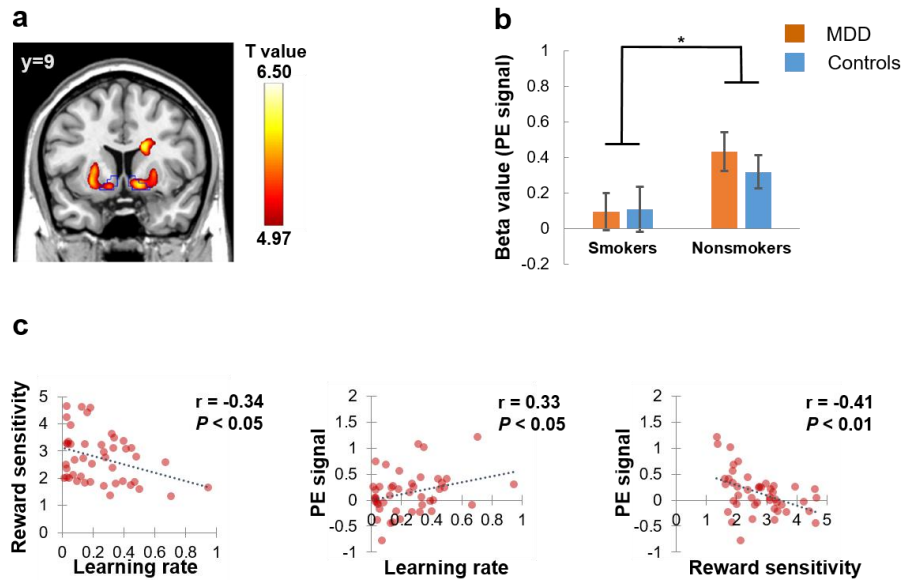


Figure 3. Neural substrates of prediction error (PE) and associations among learning parameters and striatal PE signal. (a) PE signal for all subjects ($P < 0.05$, family-wise error corrected for multiple comparisons at voxel-level over the whole brain, cluster > 50 voxels). The ventral striatum ROI is outlined in blue. (b) PE signal in ventral striatum of four groups shows the main effect of smoking status, with smokers have lower PE signal than nonsmokers. The main effect of depression status and interaction between depression and smoking are nonsignificant (P 's ≥ 0.55). Error bars indicate 1 SE. * $P < 0.05$. (c) Correlations within smokers.

Supplementary Information

Supplementary methods

Participants

One hundred and forty-eight participants were recruited, including 29 major depressive disorder (MDD) smokers, 27 nondepressed control smokers, 54 MDD nonsmokers and 38 control nonsmokers. MDD diagnoses were determined through the full Structured Clinical Interview for DSM-IV Axis I Disorders (SCID; First, Spitzer, Gibbon & William, 2002); in addition, participants with depression were required to be free from additional Axis-I psychopathology, as determined by the SCID. Participants were categorized as smokers if they indicated were daily smokers for at least the past 3 years. No past smokers were included in this sample. Other exclusion criteria were: younger than 18 or older than 55; history of seizure disorder, stroke, or head injury resulting in more than 10 minutes of unconsciousness or with neurological sequelae; hormone disorder; history of electroconvulsive therapy or chemotherapy for cancer; current psychotropic medication use; current pregnancy or menopause; substance abuse within the past 6 months or lifetime dependence (except for nicotine dependence in smokers); functional magnetic resonance imaging (fMRI) contraindications; acute suicidal ideation; concurrent diagnosis of bipolar disorder, schizophrenia, schizoaffective disorder, delusional disorder and organic psychosis.

Nineteen participants were excluded due to missing data or failure of the synchronization between behavioral and imaging data. Five participants were excluded due to excessive head motion within the scanners (cumulative translation > 4 mm and rotation $> 4^\circ$). One participant was excluded due to not switching between different options and one was excluded due to low accuracy (out of the range of mean accuracy - 2.58 standard deviation). To ensure the groups of participants with

depression were matched on depression severity and anhedonia, 20 additional participants were excluded. This yielded a final sample of 102 participants. All four groups were matched [one-way analyses of variance (ANOVAs): $F(3,98)$'s ≤ 0.78 , P 's ≥ 0.51 ; t 's ≤ 1.37 , $P \geq 0.17$] on age and intelligence. MDD smokers, MDD nonsmokers and control nonsmokers were matched on sex ratio [$\chi^2(1)$'s ≤ 0.75 , P 's ≥ 0.39], but the sex ratio in control nonsmokers was different from the other three groups [$\chi^2(1)$'s ≥ 5.12 , P 's < 0.05].

Experiment paradigm

The procedure of one trial in the gain framing was as follows: at the start of each trial, two abstract square patterns were shown on the screen. They were present until the participant made a choice via a scanner-compatible button box. After the participant chose one option, that pattern was highlighted to confirm the choice. After a jittered period, the outcome was shown. Unknown to the participant, one of the patterns was 'better', in that it had a higher probability (75% versus 25%) of leading to the higher outcome. The higher outcome ranged from \$ 0.70 to \$ 0.80, and the lower outcome ranged from \$ 0.20 to \$ 0.30. The outcome was present for 2 s, and then a jittered fixation was shown. After this fixation, another trial was started. This new trial also had the same pair of patterns and the participant continued to choose between them until the contingencies were learned, which was determined by an adaptive algorithm, or until the participant completed 36 trials in the same block. Different patterns were shown in each block and the location of the patterns was randomized for each trial. In the loss framing, the procedure was the same except for the outcome screen, on which participants were shown the amount of money they lost. For example, for a better pattern, there was a probability of 75% to lose a lower

amount (-\$0.20 to -\$0.30) and a probability of 25 % to lose a higher amount (-\$0.70 to -\$0.80).

As implemented, this task took approximately 20 minutes and each participant underwent a thorough practice run to ensure task comprehension prior to entering the scanner. Participants were endowed with an initial amount of money for participating in the study and were paid an amount proportional to their earnings in the game to maximize engagement in the task.

Behavioral analysis

Model-free analysis

The learning curves depict the running average of the trial-by-trial proportion of choices in each group that selected the better option. The window size of the running average from trial 3 to trial 23 is 5, with two trials before and two trials after the current trial. For the first trial and the last trial, the window size is 1. For trial 2 and trial 24, the window size is 3, with one trial before and one trial after them.

Model-based analysis

Given the expected value based on the updating rule in a basic RL model, there is also a function for calculating the agent's probability of choosing a certain option (Luce, 1959):

$$P_{A(t)} = \frac{\exp(\beta * Q_{A(t)})}{\exp(\beta * Q_{A(t)}) + \exp(\beta * Q_{B(t)})} \quad (\text{Equation S1}).$$

In this equation, $P_{A(t)}$ denotes the probability of choosing option A out of two options (A and B) at time t; β is the exploitation-exploration parameter (termed as inverse temperature) and can be used as a measure of the noisiness or randomness of the agent's choice.

The three alternative reward sensitivity models were also constructed as follows: In *Multiplicative_Better model*, ρ was multiplied only on the better outcomes.

$$Q_{(t+1)} = \begin{cases} Q_{(t)} + \alpha * (\rho * V_{(t)} - Q_{(t)}) & \text{when } V_{(t)} > 0.5 \text{ in gain or } V_{(t)} > -0.5 \text{ in loss} \\ Q_{(t)} + \alpha * (V_{(t)} - Q_{(t)}) & \text{when } V_{(t)} < 0.5 \text{ in gain or } V_{(t)} < -0.5 \text{ in loss} \end{cases}$$

(Equation S2).

In *Additive_Both model*, instead of being a multiplier, ρ was additive to both the better and worse outcomes.

$$Q_{(t+1)} = Q_{(t)} + \alpha * (\rho + V_{(t)} - Q_{(t)}) \quad \text{(Equation S3).}$$

In *Additive_Better model*, ρ was only additive to the better outcomes.

$$Q_{(t+1)} = \begin{cases} Q_{(t)} + \alpha * (\rho + V_{(t)} - Q_{(t)}) & \text{when } V_{(t)} > 0.5 \text{ in gain or } V_{(t)} > -0.5 \text{ in loss} \\ Q_{(t)} + \alpha * (V_{(t)} - Q_{(t)}) & \text{when } V_{(t)} < 0.5 \text{ in gain or } V_{(t)} < -0.5 \text{ in loss} \end{cases}$$

(Equation S4).

The basic RL model and four reward sensitivity models were estimated to fit the participants' choice data. The free parameters (learning rate and reward sensitivity) of each model were estimated using hierarchical Bayesian analysis. Bayesian methods estimate the entire posterior distributions of the free parameters, instead of a single best value. The distribution captures the uncertainty of the parameter estimate. In hierarchical Bayesian analysis, a hierarchical model has parameters for each individual, and a distribution of individual parameters at the group level. This group-level distribution has its own parameters that describe the tendency of the group. Different levels can inform each other, allowing more precise estimation of parameters and the ability to capture individual differences in model parameters (Kruschke & Vanpaemel, 2015).

The Stan software package (Stan Development Team, 2016) was used to perform the hierarchical Bayesian analysis. Stan applies Hamiltonian Monte Carlo in its model estimation, which is a Markov chain Monte Carlo (MCMC) sampling algorithm and allows efficient sampling even for complex models with multilevel structures and

highly correlated parameters (Ahn et al, 2014). In our model estimation, individual parameters were assumed to be drawn from group-level distributions. The prior of group-level learning rate was set as a normal distribution transformed through an inverse logit function, which was bounded between 0 and 1. For the basic RL model, the prior of group-level inverse temperature was set as a normal distribution, which was bounded between 0 and 50. For the four reward sensitivity models, the prior of group-level reward sensitivity was set as a normal distribution, which was bounded between 0 and 50.

In reward sensitivity models, β and ρ are difficult to separate (Huys, Pizzagalli, Bogdan, & Dayan, 2013), as they influence agent's probability of choosing an option in a similar way (e.g., both higher β and ρ result in more deterministic/exploitative choices.). In order to get a valid estimation of ρ , β is typically fixed as a constant in models containing ρ . Therefore, in the model estimation, the inverse temperature for all participants was fixed to 3.15, which was the group-level inverse temperature estimated from a separate basic RL model (only contained group-level learning rate and inverse temperature) with all four groups of participants in it.

For the estimation of the five candidate models, four groups were estimated together with MDD status and smoking status as two dummy variables. Therefore, the parameters' differences between MDD and controls or the differences between smokers and nonsmokers at the group level could be estimated unbiasedly. Four MCMC chains were run for each model, with 4000 samples per chain (2500 after discarding warm up samples). All chains were inspected for convergence and showed good mixing, with all values of the potential scale reduction factor (Gelman & Rubin, 1992) less than 1.10.

In the simulation, learning rate had 201 values, ranging from 0 to 1 in a step size of 0.005; reward sensitivity had 251 values, ranging from 0 to 5 in a step size of 0.02. A 201×251 matrix was plotted as a heatmap, with each element as an accuracy value computed from its corresponding learning rate and reward sensitivity through the reward structure of our learning task and the best-fitting model (**Figure 2a** left panel). Another plot containing a scatterplot and a heatmap was drawn to show the empirical relationship between learning rate, reward sensitivity and accuracy. The scatterplot shows the correlation between learning rate and reward sensitivity for all participants. For the heatmap, the aforementioned simulation heatmap was divided into five areas according to the simulated accuracy $[[0, 60\%), [60\%, 65\%), [65\%, 70\%), [70\%, 75\%),$ and $[75\%, 100\%]$, but the colors of which were determined by the mean accuracy of all dots (participants) falling within each area (**Figure 2a** right panel).

In the parameter recovery test, for learning rate, we constructed a normal distribution with an inverse logit transformation using a mean and a variance, which were average values of the group-level mean and variance from the best-fitting model over the four groups. In the same way, for reward sensitivity, we constructed a normal distribution using the mean and variance, which were average values of these two group-level parameters of the four groups. We randomly drew 100 values from the transformed normal distribution as learning rate and 100 values from the normal distribution as reward sensitivity. These 100 learning rate values and 100 reward sensitivity values were paired randomly to form 100 simulated participants. Each pair of learning rate and reward sensitivity were used to generate simulated choices on the probabilistic reward task using the best-fitting model. Parameters were then recovered from the simulated choice data using the same hierarchical Bayesian estimation

method and winning model. Two correlation scatterplots were drawn to show whether learning rate and reward sensitivity could be recovered (**Figure 2b**).

Imaging acquisition and fMRI analysis

Functional images were obtained using a T2*-weighted echo-planar imaging (EPI) sequence [repetition time (TR) = 2000 ms, echo time (TE) = 30 ms, flip angle = 90°, field of view (FOV) = 220 × 220 mm²]. Each volume contained 34 interleaved axial slices (matrix = 64 × 64, in-plane spatial resolution = 3.44 × 3.44 mm², thickness = 4 mm), which were angled 30° with respect to the anterior-posterior commissural line. High-resolution anatomical images were obtained using a T1-weighted 3D magnetization-prepared rapid gradient-echo (MP-RAGE) sequence (TR = 1200 ms, TE = 2.66 ms, flip angle = 8°, FOV = 245 × 245 mm). The anatomical volume contained 192 axial slices (matrix = 245 × 245, in-plane spatial resolution = 1 × 1 mm², thickness = 1 mm).

SPM8 software package was used for imaging data preprocessing. Slice timing artifacts of functional images were corrected and then the imaging data were realigned to correct for head movement between scans, and each participants' anatomical scan was coregistered to the mean functional image produced in the realignment stage. The anatomical scans were then segmented and spatial normalization parameters matrices were generated based on the Montreal Neurological Institute (MNI) template. Functional images were transformed into the MNI space using these matrices. Normalized functional data were then spatially smoothed using an isotropic Gaussian filter with a full-width at half-maximum parameter set to 8 mm.

The first-level statistical analysis was performed by including preprocessed functional data of each participant into a general linear model (GLM) using an event-related analysis procedure. Events in the GLM included choices, confirmation

of choices and outcomes. Prediction errors calculated from the best-fitting model were entered into the GLM as a parametric modulator of the outcome event. Session number and 6 head motion parameters were also included in the GLM as regressors of no interest. The stimulus parameters specific to each event were convolved with a canonical hemodynamic response function, and standard linear regression and parametric modulation analyses were performed to obtain regression coefficients (beta-weights) and t-statistic maps.

In the region of interest (ROI) analysis, the anatomical mask of bilateral ventral striatum from the Oxford-GSK-Imanova structural and connectivity striatal atlases (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases>) was used. The ROI value for one participant was extracted from his/her prediction error-related beta map with a spherical ROI (radius = 6 mm) centered at the peak coordinates within the ventral striatum mask. The peak coordinates were from the group level analysis without this particular participant.

Supplementary results

Depression and anhedonia scores showed a main effect of MDD status in the ANOVAs [BDI: $F(1, 95) = 584.99, P < 0.001$; SHAPS : $F(1, 85) = 74.58, P < 0.001$; MASQ anhedonia: $F(1, 85) = 386.73, P < 0.001$], with no main effect of smoking status or interaction ($F's \leq 1.24, P's \geq 0.27$). Similarly, smoking measures showed a main effect of smoking status [exhaled CO: $F(1, 79) = 54.58, P < 0.001$; cigarettes per day: $F(1,95) = 139.89, P < 0.001$; FTND: $F(1, 81) = 143.42, P < 0.001$] without a main effect of MDD status or interaction ($F's \leq 2.39, P's \geq 0.13$). Post-hoc multiple comparison of means also showed good match of anhedonia between the two MDD groups [SHAPS : $t(85) = 0.62, P = 0.54$; MASQ anhedonia: $t(85) = 0.74, P = 0.46$] and of smoking measures between the two smoking groups (exhaled CO: $t(79) = 0.86,$

$P = 0.10$; cigarettes per day: $t(95) = 0.86$, $P = 0.39$; FTND: $t(81) = -0.19$, $P = 0.85$)

(Table 1).

In the winning model, both the group-level learning rate and reward sensitivity showed significant differences between smokers and nonsmokers. Learning rate was higher for nonsmokers than smokers (smokers – nonsmokers: 95% confidence interval (CI) = [-2.26 -0.76]), while reward sensitivity was higher for smokers than nonsmokers (smokers – nonsmokers: 95% CI = [0.08 1.14]). The group-level parameters were not significantly different between MDD and control participants (MDD-controls: 95% CI = [-1.28 0.19] for learning rate; 95% CI = [-0.46 0.59] for reward sensitivity).

Supplementary references

- Ahn, W. Y., Vasilev, G., Lee, S. H., Busemeyer, J. R., Kruschke, J. K., Bechara, A., & Vassileva, J. (2014). Decision-making in stimulant and opiate addicts in protracted abstinence: Evidence from computational modeling with pure users. *Frontiers in Psychology*, 5, 849.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2007) *Structured Clinical Interview for DSM-IV-TR Axis I Disorders - Patient Edition (With Psychotic Screen) (SCID-I/P (W/ PSYCHOTIC SCREEN), 1/2007 revision)*. Biometrics Research Department, New York State Psychiatric Institute.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4),457-472.
- Huys, Q. J., Pizzagalli, D. A., Bogdan, R., & Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: A behavioural meta-analysis. *Biology of Mood & Anxiety Disorders*, 3(12), 1-16.
- Kruschke, J. K. & Vanpaemel W. (2015) Bayesian estimation in hierarchical models. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.). *The Oxford Handbook of Computational and Mathematical Psychology* (pp.279-299). Oxford University Press.
- Luce, R. D. (1959). *Individual Choice Behavior*. New York, NY: John Wiley & Sons, Inc.

Supplementary tables**Table S1.** Model fit indices of *Multiplicative_Both model* and*Multiplicative_Both_GL model*

Model	AIC^a	BIC^a	WAIC
Multiplicative_Both	<u>115.94</u>	<u>121.23</u>	<u>9941.7</u>
Multiplicative_Both_GL	119.76	130.35	10021.5

Abbreviations: AIC, Akaike information criterion; BIC, Bayesian information criterion; WAIC, widely applicable information criterion. ^a These are mean AIC and BIC for each participant. Model fit indices for the best-fitting model are underlined. Participants with low switching between different options (less than 3 times) or with low accuracy (out of the range of mean accuracy - 2.58 standard deviation) in either gain or loss framing were excluded, resulting in 18 MDD smokers, 17 control smokers, 24 MDD nonsmokers and 28 control nonsmokers.

Table S2. Brain regions that show prediction error signal, $P < 0.05$, family-wise error corrected for multiple comparisons at voxel-level over the whole brain, cluster > 50 voxels

Regions	Voxels	BA	MNI coordinates			T-value
			X	Y	Z	
R Caudate	121	NA	18	3	24	6.89
L Middle cingulate gyrus R Middle cingulate gyrus R Precuneus L Post cingulate gyrus L Paracentral lobule R Post cingulate gyrus	290	31/7/23	-3	-33	45	6.70
L Putamen L Amygdala L Caudate	211	NA	-24	6	-9	6.68
R Cerebellum	210	NA	9	-78	-30	6.42
L Angular gyrus L Inferior parietal gyrus L Middle occipital gyrus	228	40/39/19/7	-48	-63	42	6.40
R Cerebellum	175	NA	36	-66	-36	6.31
R Putamen R Caudate	149	NA	15	9	-9	6.18
L Precuneus L Superior parietal gyrus	87	7	-3	-69	33	6.11
L Inferior orbital frontal gyrus L Middle orbital frontal gyrus	109	11/47	-18	30	0	5.90
L Inferior temporal gyrus L Middle temporal gyrus	65	19	-48	-60	-6	5.64
R Medial orbital frontal gyrus L Medial orbital frontal gyrus	68	11/32/10	9	45	-9	5.60

Abbreviations: Regions = brain regions in an activation cluster (Peak voxel is located in the 1st region.); Voxels = number of voxels in a cluster; BA = Brodmann's area; MNI = Montreal Neurological Institute; T-value = T-value of the peak voxel in a cluster.

Supplementary figures

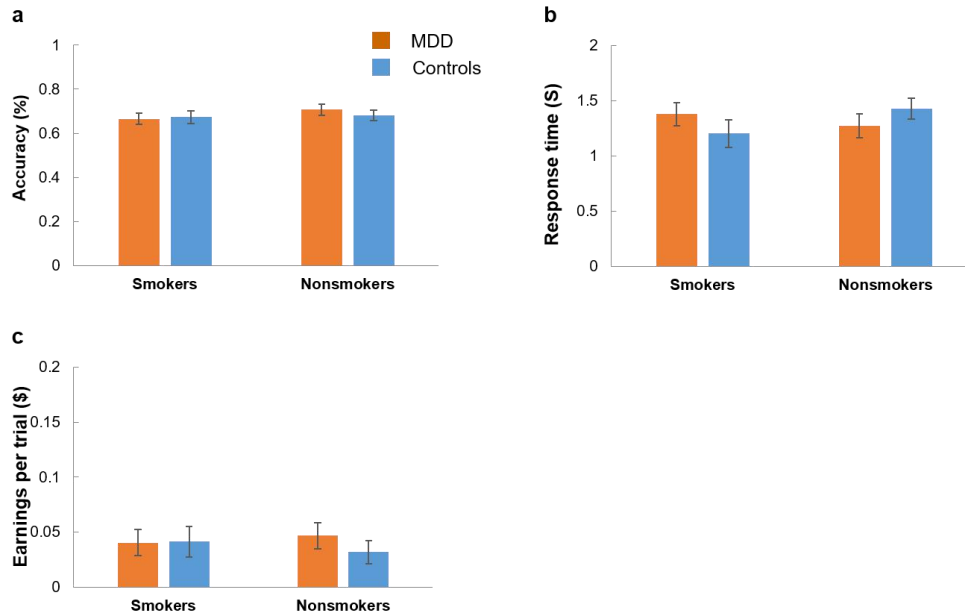


Figure S1. Model-free performance. (a) Group differences of accuracy were not significant [MDD: $F(1, 95) = 0.00, P = 1.00$; smoking: $F(1, 95) = 1.19, P = 0.28$; interaction: $F(1, 95) = 0.46, P = 0.50$]. (b) Group differences of response time were not significant [MDD: $F(1, 95) = 0.01, P = 0.92$; smoking: $F(1, 95) = 0.27, P = 0.60$; interaction: $F(1, 95) = 2.29, P = 0.13$]. (c) Group differences of earnings per trial were not significant [MDD: $F(1, 95) = 0.32, P = 0.57$; smoking: $F(1, 95) = 0.01, P = 0.91$; interaction: $F(1, 95) = 0.43, P = 0.51$]. Error bars indicate 1 standard error (SE).

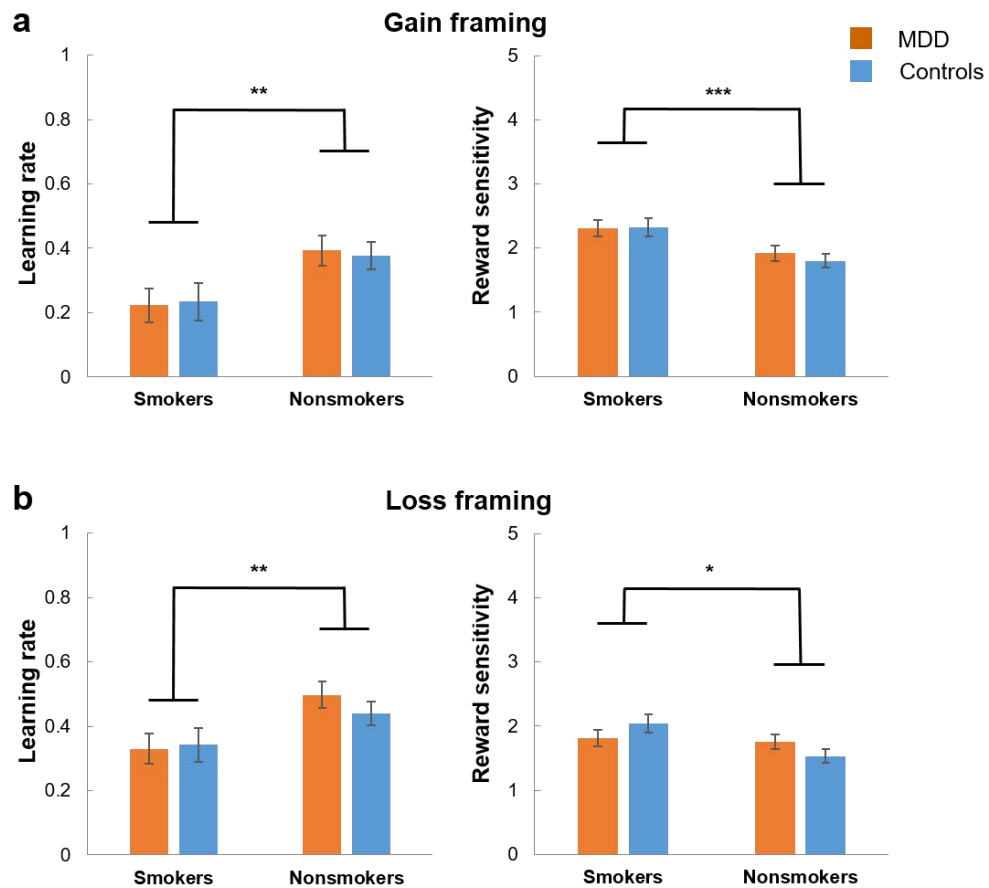


Figure S2. Parameter estimates in gain and loss framing. (a) The main effect of smoking was significant for learning rate and reward sensitivity in gain framing [For learning rate, MDD: $F(1, 80) = 0.001, P = 0.97$; smoking: $F(1, 80) = 9.19, P = 0.003$; interaction: $F(1, 80) = 0.08, P = 0.78$. For reward sensitivity, MDD: $F(1, 80) = 0.17, P = 0.68$; smoking: $F(1, 80) = 12.22, P < 0.001$; interaction: $F(1, 80) = 0.31, P = 0.58$.]. (b) The main effect of smoking was significant for learning rate and reward sensitivity in loss framing [For learning rate, MDD: $F(1, 80) = 0.25, P = 0.62$; smoking: $F(1, 80) = 8.15, P = 0.005$; interaction: $F(1, 80) = 0.64, P = 0.42$. For reward sensitivity, MDD: $F(1, 80) = 0.001, P = 0.98$; smoking: $F(1, 80) = 4.89, P = 0.03$; interaction: $F(1, 80) = 3.67, P = 0.06$.]. * $P < 0.05$. ** $P < 0.01$. *** $P < 0.001$. Error bars indicate 1SE.

Supplementary tables and figures showing results after groups were matched on sex

The following tables and figures show results after groups were matched on sex [$\chi^2(1)$'s ≤ 0.96 , P 's ≥ 0.33], age and intelligence [$F(3,83)$'s ≤ 0.42 , P 's ≥ 0.74 ; t 's ≤ 1.19 , $P \geq 0.24$]. Learning parameters were estimated using this smaller sample and the *Multiplicative_Both model*.

Table S3. Demographic information and scale measures

Group	MDD smokers (n=25)	Control smokers (n=8)	MDD nonsmokers (n=23)	Control nonsmokers (n=31)
Sex	13 females	3 females	15 females	20 females
Age (years)	36.64±1.91	32.25±3.54	35.52±2.56	34.10±2.12
WTAR	103.28±2.31	105.38±4.85	106.30±2.09	105.71±2.25
Exhaled CO (ppmv)	7.80±1.31 ^a	10.50±3.23 ^b	2.13±0.16	2.00±0.19
Mean cigarettes per day	14.28±1.88	14.00±2.01	0±0	0±0
Years smoking began	19.76±2.02	17.13±3.37	0±0	0±0
Years regular smoking	18.28±1.99	13.63±3.63	0±0	0±0
FTND	4.92±0.60 ^c	5.25±1.03 ^b	0±0	0±0
BDI	35.56±2.04	2.50±1.04	34.52±1.25	2.94±0.69
SHAPS	6.84±0.83	0.67±0.42 ^d	6.35±0.49	0.61±0.45
MASQ anhedonia	90.04±1.94	47.33±6.38 ^d	88.87±1.27	43.97±1.82

Abbreviations: WTAR, Wechsler Test of Adult Reading; CO, carbon monoxide; ppmv, parts per million by volume; FTND, Fagerstrom Test for Nicotine Dependence; BDI, Beck

Depression Index; SHAPS, Snaith-Hamilton Pleasure Scale; MASQ, Mood and Anxiety

Symptom Questionnaire. ^a This measure is available for 22 participants in MDD smokers.

^b This measure is available for 4 participants in control smokers. ^c This measure is available for 24 participants in MDD smokers. ^d This measure is available for 6 participants in control smokers. Data are represented as mean± standard error (SE).

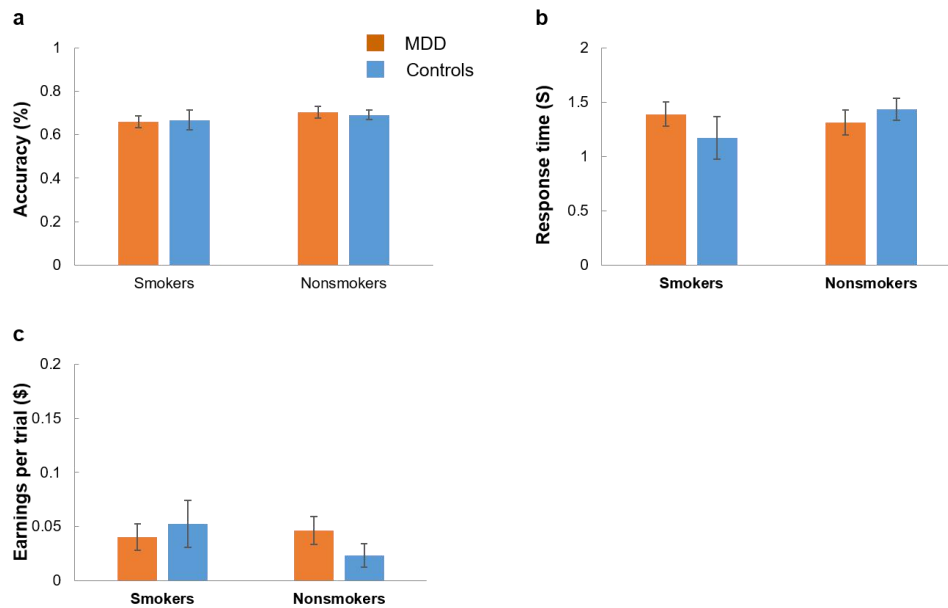


Figure S3. Model-free performance. (a) Group differences of accuracy were not significant [MDD: $F(1, 80) = 0.001$, $P = 0.97$; smoking: $F(1, 80) = 1.08$, $P = 0.30$; interaction: $F(1, 80) = 0.09$, $P = 0.77$]. (b) Group differences of response time were not significant [MDD: $F(1, 80) = 0.14$, $P = 0.70$; smoking: $F(1, 80) = 0.49$, $P = 0.49$; interaction: $F(1, 80) = 1.57$, $P = 0.21$]. (c) Group differences of earnings per trial were not significant [MDD: $F(1, 80) = 0.003$, $P = 0.95$; smoking: $F(1, 80) = 0.23$, $P = 0.63$; interaction: $F(1, 80) = 0.80$, $P = 0.37$]. Error bars indicate 1 SE.

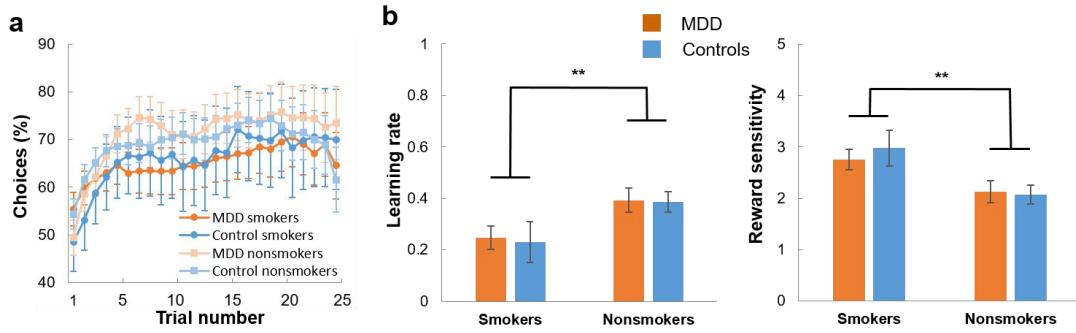


Figure S4. Behavioral performance and parameter estimates. (a) The learning curves depict the running average (window size = 5, averaged over all blocks, mean \pm SE) of the trial-by-trial proportion of participants that chose the better option. The four lines represent four groups and show learning. (b) Model estimation of learning rate and reward sensitivity. Both parameters show significant main effect of smoking status, with smokers have lower learning rate [$F(1, 80) = 7.14, P < 0.01$] but higher reward sensitivity [$F(1, 80) = 9.54, P < 0.01$] than nonsmokers. The main effect of depression status and interaction between depression and smoking are nonsignificant (P 's ≥ 0.54). ** $P < 0.01$. Error bars indicate 1 SE.

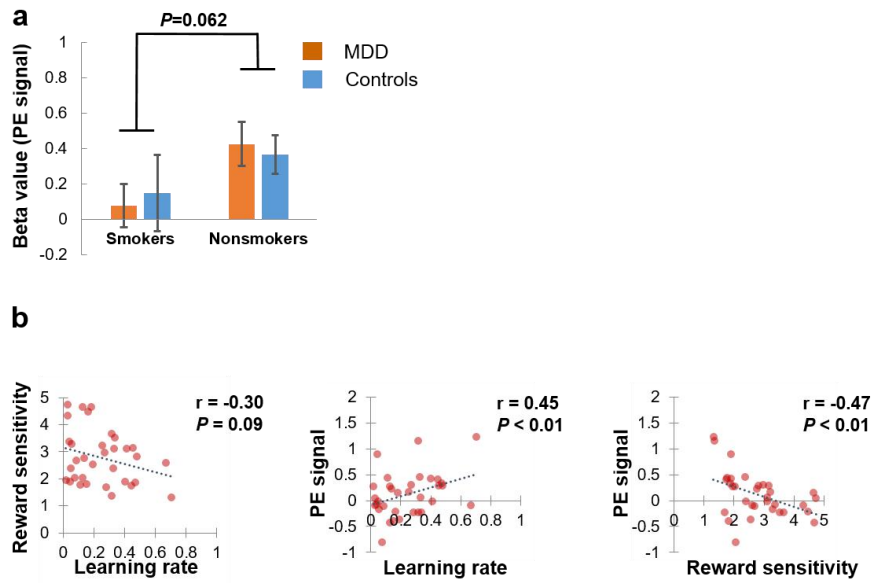


Figure S5. Group differences in striatal prediction error (PE) signal and associations among learning parameters and PE signal. (a) Prediction error (PE) signal in ventral striatum showed the main effect of smoking status, with smokers have lower PE signal than nonsmokers [$F(1, 80) = 3.58$, $P = 0.062$]. The main effect of depression status and interaction between depression and smoking are nonsignificant (F 's ≤ 0.19 , P 's ≥ 0.66). **(b)** Correlations within smokers.