Published in final edited form as:
Psychiatry Res. 2014 December 30; 224(3): 254–261. doi:10.1016/j.pscychresns.2014.10.014.

Smoking automaticity and tolerance moderate brain activation during explore-exploit behavior

Merideth A. Addicott[a,b], **John M. Pearson**[c,d], **Brett Froeliger**[f], **Michael L. Platt**[c,d,e], and **F. Joseph McClernon**[a,b]

[a]Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham NC USA 27710

[b]Duke-UNC Brain Imaging and Analysis Center, Duke University School of Medicine, Durham NC USA 27710

[c]Department of Neurobiology, Duke University, Durham NC USA 27710

[d]Center for Cognitive Neuroscience, Duke University, Durham NC USA 27710

[e]Department of Evolutionary Anthropology, Duke University, Durham NC USA 27710

[f]Department of Neurosciences, Medical University of South Carolina, Charleston SC USA 29403

Abstract

The adaptive trade-off between exploration and exploitation is a key component in models of reinforcement learning. Over the past decade, these models have been applied to the study of reward-seeking behavior. Drugs of addiction induce reward-seeking behavior and modify its underlying neurophysiological processes. These neurophysiological changes may underlie a behavioral shift from a flexible, exploratory mode to a focused, exploitative mode, which precedes the development of inflexible, habitual drug use. The goal of this study was to investigate the relationship between explore/exploit behavior and drug addiction by examining the neural correlates of this behavior in cigarette smokers. Participants (n = 22) with a range of smoking behaviors completed a smoking dependence motives questionnaire and played a 6-armed bandit task while undergoing functional magnetic resonance imaging (fMRI). Exploratory behavior produced greater activation in the bilateral superior parietal and bilateral frontal cortices than exploitative behavior. Exploitative behavior produced greater activation in the bilateral superior and middle temporal gyri than exploratory behavior. fMRI data and orthogonalized smoking

Corresponding Author: Merideth Addicott, Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Box 3527, Lakeview Pavilion E Suite 300, Durham NC 27705, merideth.addicott@duke.edu.

Author contributions

MAA and FJM were responsible for the study concept and design. MAA acquired and analyzed imaging data. JMP programmed the bandit task and assisted with the analysis of behavioral data. MAA drafted the manuscript. MAA, JMP, BF, MLP, and FJM provided critical revision of the manuscript for important intellectual content. All authors critically reviewed content and approved the final version for publication.

dependence motive scores were entered into multiple linear regression analyses. After controlling for nicotine tolerance, smoking automaticity positively correlated with activation in the same bilateral parietal regions preferentially activated by exploratory choices. These preliminary results link smoking dependence motives to variation in the neural processes that mediate exploratory decision making.

Keywords

Automaticity; smoking; bandit task; exploit; explore; neuroimaging

1. INTRODUCTION

Drugs of abuse induce reward-seeking behavior, but modify its underlying neurophysiological processes and therefore constitute an aberration of natural reinforcement learning (Schultz, 2011). Over the past decade, theories of reinforcement learning have been applied to the study of reward-seeking behavior (McClure et al., 2003; Ramnani et al., 2004), and these learning theories have direct implications for the study of drug addiction (Redish, 2004; Schultz, 2011). An essential component of reinforcement learning is the explore/exploit trade-off. Preliminary work has shown smoker/nonsmoker differences in explore/exploit behavior (Addicott et al., 2013); however, the underlying neural basis for these differences has not been investigated. This research is necessary in order to understand how substance dependence affects reinforcement learning, which could inform new therapeutic interventions.

Reinforcement learning is a process of interacting with the environment to learn what actions will maximize rewarding outcomes (Sutton and Barto, 1998). Without prior knowledge of the environment, one must explore options to discover the resulting outcomes and subsequently exploit options with a history of some reward. In a constantly changing world, behavior must be adapted to changes in reward outcomes in order to maximize gain and minimize loss. When rewards diminish over time, a decision is needed to either continue exploiting the known option or to risk exploring other options. Exploration maximizes information about alternative options, but runs the risk of not discovering an option with a more desirable outcome. In the case of an animal foraging for food in the natural world, there must be a trade-off between these behaviors for long-term survival because exploitation of a single food source will eventually lead to its depletion. The question of when to stay and exploit versus when to leave and explore is central to the theory of reinforcement learning, and this dilemma is referred to as the explore/exploit trade-off (Cohen et al., 2007; Sutton and Barto, 1998).

The explore/exploit trade-off arises naturally in the context of the multi-armed bandit task (Berry and Fristedt, 1985; Gittins, 1989), named in reference to a “one-armed bandit,” or slot machine. In this task, there are 2 or more arms, or options, for a subject to select from to obtain a reward. The goal of this task is to maximize the number of rewards received over a series of trials. In order to accomplish this, the subject must balance exploring the options to learn about the underlying reward contingencies and exploiting the option with the best

current expectation of reward (Kaelbling et al., 1996). In the non-stationary version of the task, the reward values for each arm increase or decrease across trials independently of other options (Berry and Fristedt, 1985; Daw et al., 2006). As the rewards from a selected arm change across trials, subjects can either continue exploiting the current arm for the anticipated payout, or explore other arms to seek potentially larger payouts. Here, adaptive decision making means minimizing costs (e.g., time, energy expenditure, or missed opportunities) and maximizing gain (e.g., reward). The neural mechanisms underlying explore/exploit behavior have been investigated using a 4-armed bandit task (Daw et al., 2006). Using functional magnetic resonance imaging (fMRI), Daw and colleagues reported greater brain activation in the frontopolar cortex and intraparietal sulcus during exploration compared to exploitation, but no activation greater for exploitation than exploration (Daw et al., 2006).

In a prior study, we administered a 6-armed version of the bandit task to cigarette smokers and nonsmokers. Tobacco addiction results in changes in brain function and motivational behavior; for review, see (Benowitz, 2008). These changes may underlie a behavioral shift from a flexible, exploratory mode to a focused, exploitative mode, which precedes the development of inflexible, habitual drug use (Graybiel, 2008). Therefore, the explore/exploit trade-off may be particularly relevant to the study of drug addiction. We found that smokers made significantly fewer exploratory choices (and thus more exploitative choices) in the first 300 trials of the task (Addicott et al., 2013). Furthermore, we examined the relationship between explore/exploit behavior and smoking dependence motives. Tobacco smoking is primarily motivated by automatic (habitual) tobacco use, a loss of volitional control over this use, cravings, and drug tolerance (Piasecki et al., 2010; Piper et al., 2004). We found a negative correlation between smokers' self-reported tolerance and loss of control and their number of exploratory choices. This suggests that as dependence motives become stronger, smokers become more biased towards exploitative behavior.

To investigate this relationship further, we recruited smokers who used between 1 and 20 cigarettes per day to undergo fMRI while performing the 6-armed bandit task (6ABT). We expected fronto-parietal activation during exploratory decisions and orbitofrontal activation during the reward outcome, similar to a previous study (Daw et al., 2006). The relations between smokers' explore/exploit brain activation and smoking dependence motives were assessed using multiple linear regression analyses. We hypothesized that there would be a negative correlation between smoking dependence motive scores and the number of exploratory choices (similar to our previous study), and thus, negative correlations between dependence motive scores and exploratory activation during the selection phase of the task. In addition, research has shown diminished responsivity to monetary rewards among dependent smokers (Buhler et al., 2010); therefore, we hypothesized a negative correlation between dependence motive scores and reward-related activation during the outcome phase of the task.

2. MATERIALS AND METHODS

2.1 Participants

Twenty-four participants were recruited via newspaper and internet advertisements from the Durham/Chapel Hill/Raleigh area of North Carolina. Participants were included in the study if they were in general good health, right-handed, and smoked at least 1 cigarette per day for at least two years and were not currently interested in quitting. Participants were excluded if they reported significant health problems (e.g., hypertension), current use of psychoactive medication, current use of smokeless tobacco or nicotine replacement therapy, or if they had a positive breath alcohol concentration or urinary drug screen for illicit drugs during the screening session. This study was designed to investigate the association between smoking characteristics and explore/exploit-related brain activity; therefore, we recruited participants with a range of smoking behaviors and included participants who smoked between 1 and 20 cigarettes per day. Participants provided written informed consent, and this study was approved by the Institutional Review Board at Duke University.

2.2 Procedure

Participants completed an initial screening session in which they completed questionnaires and were trained to perform the bandit task in a mock MRI scanner. The training session consisted of 20 trials of the bandit task. Participants were then scheduled for a single MRI session in the following week. Tobacco use prior to the scan was verified using breath carbon monoxide (CO) levels. To reduce the incidence of withdrawal symptoms, participants were instructed to smoke “as usual” prior to the scan and all participants reported smoking within 30 minutes of the appointed scan time.

2.3 Self-report questionnaires

Nicotine dependence and smoking dependence motives were measured with the Fagerström Test for Nicotine Dependence (FTND) (Heatherton et al., 1991) and the Wisconsin Inventory for Smoking Dependence Motives (WISDM) (Piper et al., 2008). The WISDM is a multidimensional measure of motivational factors that affect tobacco dependence (Piper et al., 2004). Four subscales of the WISDM (automaticity, craving, loss of control, and tolerance) have been identified as primary dependence motives and are strong predictors of relapse and withdrawal (Piper et al., 2008). The subscale “automaticity” is characterized by smoking without awareness or intention (e.g., “I often smoke without thinking about it.”), “loss of control” is characterized by a sense of lost volitional control over cigarette use (e.g., “cigarettes control me”), “craving” is characterized by smoking in response to cravings (e.g., “I frequently crave cigarettes”), and “tolerance” is characterized by the need to smoke increasing amounts over time to experience the desired effects (e.g., “I usually want to smoke right after I wake up”) (Piper et al., 2008; Piper et al., 2004). Other smoking variables (e.g., cigarettes per day and duration of smoking) were obtained from a tobacco use history questionnaire.

2.4 Bandit task

This version of the task was modeled after Daw et al. (2006). In each trial, six slot machines (i.e., targets) were depicted on a computer screen for 1.5 s. Participants selected one target to play by pressing the corresponding response box button. If a selection was made within 1.5 s, the selected target was highlighted for 3 s then the number of points paid off for that target was displayed for 1 s. If a selection was late or absent, a red box appeared in the middle of the screen for 4 s and the trial was excluded from the analysis. Each trial lasted 5.5 s followed by a variable inter-trial interval of 2.1 ± 0.4 s (M ± S.D., range 1.5 to 3.5 s) with an approximate Poisson distribution. The number of points associated with each target changed gradually from trial to trial, independently of other targets; see Figure 1. The payoff on the first trial was 50, and the payoff for the *i*th target on trial *t* was drawn from a Gaussian distribution with a standard deviation of 2.8 around a mean,, and rounded to the nearest integer. Subsequent target values, *r*, were randomly adjusted according to a biased random walk,

$$r_{i,t+1}=\lambda\left(r_{i,t}-\theta\right)+\theta+\eta \quad (1)$$

where central tendency parameter,, is equal to 0.985 and the Gaussian random variable,, has a mean of zero and a standard deviation equal to 2.8. The mean,, is equal to 50. Payoffs were allowed to range between 0 and 100 points. The MRI version of the bandit task was administered in two runs of 100 trials each. Each run had unique payoff schedules and runs were administered in counterbalanced order. The task was programmed in Matlab (MathWorks, Inc. Natick, MA) using the Psychophysics Toolbox (Brainard, 1997).

Target selections were classified as exploratory or exploitative according to a model-based account of each participant's choice behavior (previously described in Addicott et al. 2013).

Participants were instructed to "earn as many points as you can, so … play the machine you think has the highest point value at any given moment." Participants were informed that they would earn up to an additional $10 based on the ratio of points they earned over the total number of points possible.

In the MRI scanner, visual stimuli were projected onto a screen behind the participant's head. The visual stimuli were viewed via mirrors mounted on the headcoil. Participants selected targets using MR-compatible button boxes. Buttons in the left hand selected targets on the left side of the screen, and buttons in the right hand selected targets on the right side of the screen.

2.5 Data analysis

Relations between questionnaire data were investigated with Pearson correlations. WISDM subscale scores were orthogonalized using a principal component analysis (PCA). PCA uses an orthogonal transformation to convert correlated variables into linearly uncorrelated component scores. The resulting WISDM component scores were entered into a multiple linear regression analysis with the fMRI data in order to investigate unique relationships between individual subscales and BOLD activation.

2.6 Image acquisition

Images were acquired with a 3T General Electric Signa EXCITE HD scanner (Milwaukee, WI) equipped with 40 mT/m gradients. Blood oxygenation-level dependent (BOLD) functional images were collected for 32 contiguous slices parallel to the anterior/posterior commissures. A gradient-recalled inward spiral pulse imaging sequence was used to collect functional scans (TR = 1500 ms, TE = 30 ms, FOV = 25.6 cm^2, matrix = 64 × 64, flip angle = 60º, slice thickness = 3.8 mm, resulting in 4 × 4 × 3.8 mm voxels). Each functional run consisted of 518 brain volumes. The first four volumes were collected to achieve steady state equilibrium and were subsequently discarded. The total run length of each run was 12 m and 57 s. Following the functional runs, a high-resolution, three-dimensional, fast-spoiled gradient-recalled echo (3D-FSPGR) anatomical sequence was collected (TR = 7.484 ms, TE = 2.984 ms, FOV = 25.6 cm^2, matrix = 256 × 256, flip angle = 12º, 166 slices, slice thickness = 1 mm).

2.7 Image preprocessing and data analysis

Functional data were preprocessed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm). Images were slice time corrected via sinc interpolation and were realigned to correct for head motion over the course of each functional run. Then, images were coregistered to the individual's anatomical image and standard T1 template, then normalized to the template. During normalization, images were high pass filtered (filter width = 128 s) and voxels were resampled to 2 × 2 × 2 mm. Lastly, images were smoothed with a Gaussian kernel with a full width at half maximum (FWHM) of 8 mm.

For the first-level analysis, trials were modeled with 2 time points: the time of the decision to select a target (set arbitrarily half-way between the target presentation and the individual's reaction time per trial) and the presentation of the reward outcome (similar to Daw et al. 2006). Decisions were coded as exploratory or exploitative according to the behavioral model described above. Reward outcomes were divided into 4 bins: 50, 51-60, 61-70 71 points and were modeled with a step-wise increase across bins in the first-level analysis. These selection and outcome conditions were convolved with a canonical hemodynamic response function and used as explanatory variables in a general linear model (GLM). The motion parameters produced by the image realignment were included as additional regressors of no interest to account for motion.

Individual contrast images of explore/exploit trials and reward outcomes were included into two second-level random effects analyses using one-sample t-tests. To examine the relationship between WISDM subscales and brain activation, the main contrast images (explore/exploit and reward outcomes) were entered into separate multiple regression analyses with the 4 transformed component scores of the WISDM. Main effects were considered significant if they passed a statistical threshold of $p < 0.05$ cluster-corrected. Cluster size for the comparisons was determined using AlphaSim and running 1000 Monte Carlo simulations (Ward, 2000) ($p < 0.005$, uncorrected; 190 contiguous significant voxels). Figures were made using MRIcro (Rorden and Brett, 2000).

In order to report correlation coefficients, each subject's percent BOLD signal change within the significant group-level activation clusters was extracted using MarsBaR (Brett et al. 2002) and entered into Pearson correlation analyses with the WISDM component scores.

Two participants were removed from the data set for not following task instructions during the MR scan. This resulted in a final sample of 22 participants (14 men).

3. RESULTS

3.1 Participants

Participants were 36 ± 11 (M ± SD) years of age and smoked 12 ± 6 cigarettes per day (range 1-20). FTND scores were 4 ± 3 and CO levels on the scan day were 12 ± 8 ppm. The racial distribution of the participants was 50% African American, 36.4% Caucasian, 4.5% Hispanic, 4.5% Asian, and 4.5% Other.

3.2 Self-report measures

There was a high degree of intercorrelation among WISDM subscale scores (r's 0.67, all p's 0.001). The PCA resulted in 4 components. The first component loaded positively with the mean of all four subscales, the second component loaded positively with automaticity and negatively with tolerance, the third component loaded positively with craving, and the fourth component loaded positively with loss of control, see Table 1.

3.3 Behavioral Results

Overall, participants completed 97.5% of trials. Of these trials, 31 ± 8% (M ± SD) were exploratory. Participants earned 90 ± 2% of the total possible points resulting in an average bonus of $9. There was a negative correlation between the percentage of exploratory trials and the ratio of points ($r = -.48$, $p = 0.026$). In other words, participants tended to exploit targets with the greatest point values. The percentage of exploratory choices and the ratio of points earned were not significantly correlated with any of the participants' demographic data, smoking characteristics, or WISDM subscale scores (p's > 0.05).

3.4 fMRI Results

A main effect analysis of the selection phase showed a significant difference between exploratory and exploitative trial types. In general, there was more broadly distributed activation during exploratory trials, see Figure 2. In particular, exploratory trials were associated with significant bilateral activation in superior parietal lobules, including the intraparietal sulcus, precuneus, and supramarginal gyri, with activation extending into the lateral occipital cortices. Exploration was also associated with activation in the right and left superior, middle, and precentral frontal gyri extending into the frontal poles, the right paracingulate gyrus, and subcortical areas including the right caudate nucleus and thalamus, and the left putamen, see Table 2. In contrast, exploitative trials were associated with significant bilateral activation in superior and middle temporal gyri, planum temporale, and the left angular gyrus, see Figure 2 and Table 2.

A main effect analysis of the outcome phase showed that increasing reward values across trials were associated with activation in the bilateral caudate nucleus, bilateral anterior cingulate gyri, left posterior cingulate gyrus, and regions in the parietal and occipital cortices. Decreasing reward values were associated with activation in the bilateral superior and middle frontal gyri, and bilateral precuneus and superior occipital cortex, see Figure 3 and Table 2.

Significant positive regression coefficients between the second WISDM component score (automaticity) and explore>exploit brain activation were found in the left and right parietal cortices including bilateral postcentral and supramarginal gyri; see Figure 4a and Table 3. Individual's percent BOLD signal change correlated with the second WISDM component score ($r = 0.36$ and $r = 0.35$ for the right and left activation clusters, respectively). No significant correlations were found among the other three components for the selection phase.

Significant positive regression coefficients between the third WISDM component score (craving) and reward outcomes were found in bilateral posterior cingulate gyri, and in left occipital, temporal, and frontal cortical regions, see Figure 4b and Table 3. No significant correlations were found among the other three components for the outcome phase.

4. DISCUSSION

The present study investigated explore/exploit-related brain activation in smokers with a range of smoking characteristics. Explore>exploit activation was found in several brain regions including the bilateral superior parietal lobes, intraparietal sulci (IPS), superior and middle frontal gyri, frontal poles, and cerebellum. Exploit>explore activation was found in the bilateral superior and middle temporal gyri. Multiple regression analyses investigated the relation between brain activation and smokers' self-reported smoking dependence motives. The results revealed a relationship between the automaticity and tolerance motives and explore>exploit activation in the bilateral postcentral (PCG) and supramarginal gyri (SMG) of the parietal cortices. Importantly, the automaticity and tolerance scores had been orthogonalized by a principal components analysis; therefore, these results indicate that, after controlling for tolerance, increases in automatic smoking behavior predict greater brain activation during exploratory decision making. Brain activation during the reward outcome phase of the task was also investigated: as reward values increased, there was greater activation in the caudate nucleus, and the anterior and posterior cingulate gyri. Regression analyses with the dependence motives indicated a relationship between craving scores and reward-related activation in the posterior cingulate and other regions. These preliminary results suggest a link between smoking motives and the variation in the neural processes that mediate exploratory decision making.

We had originally hypothesized that smoking dependence motives would negatively correlate with the number of exploratory choices, as shown previously (Addicott et al., 2013). However, we did not find such a correlation in the present data. This may be due to the shortened length of the task; the previous task had 1500 trials, compared to 200 trials in the current study. Additionally, in the earlier study we reported an average of 21%

exploratory choices among smokers (10 cigarettes per day) in the first 300 trials (and an average of 45% exploratory choices among nonsmokers). The current subjects had an average of 31% exploratory choices. This intermediate extent of exploratory behavior could be expected given this sample's range of cigarette use (1 – 20 cigarettes per day); however, there was no correlation between explore/exploit behavior and number of cigarettes smoked per day. The shorter length of the current task could have also affected the ratio of explore/exploit choices and their relationship with dependence motives.

Secondly, we hypothesized there would be negative correlations between dependence motives and brain activation. During the selection phase, if subjects with higher dependence motive scores made fewer exploratory decisions, then there would most likely be a corresponding decrease in exploratory-related activation. However, as discussed above, dependence motives did not predict the extent of exploratory behavior, yet we found a positive relationship between two dependence motives and exploratory activation. The second principal component represented positively weighted automaticity scores and negatively weighted tolerance scores; in other words, the orthogonalized scores reflect automaticity after controlling for tolerance. This component positively correlated with activation in the bilateral PCG and SMG, these regions were also active during the main explore>exploit contrast. The PCG and IPS are commonly activated during strategic choice selection and task switching, and these regions are believed to be involved in preparatory control (Ruge et al., 2009; Slagter et al., 2006). Our results suggest that as smoking becomes more motivated by automatic behavior rather than by nicotine tolerance, there is more neural activation in parietal regions during exploratory decision making, and this activation may be suggestive of increased cognitive effort needed to explore. Increased effort may be related to a diminished cognitive capacity for sustained exploratory behavior, which may bias an individual towards exploitative behavior over time.

Lastly, we hypothesized a negative correlation between dependence motive scores and reward-related activation because earlier studies have shown that dependent smokers have decreased neural responses to monetary rewards (Buhler et al., 2010; Martin-Soelch et al., 2003). Contrary to our hypothesis, we found a positive correlation between reward-related activation and the craving dependence motive in the posterior cingulate gyrus, as well as other regions throughout the brain. Craving is a complex phenomenon involving memory, attention, and motor preparation (Franken, 2003), which may account for the large number of brain regions related to craving. In particular, the posterior cingulate integrates reward processing, attention and memory, and activity in the posterior cingulate has been shown to predict exploration during bandit performance (Pearson et al., 2009). This relationship suggests that as craving becomes an increasingly strong motive for smoking, there is an associated increase in the monitoring of outcomes for decision making.

As expected, we found a greater extent of activation during explore trials than exploit trials, most markedly in the bilateral parietal and frontal cortices. Our results are consistent with the outcomes of similar studies. Daw and colleagues (2006) reported greater activation in the bilateral IPS and frontal poles during exploratory compared to exploitative trials. Additionally, Boorman and colleagues (2009) looked at activation related to switching between targets in a task where targets had independent reward probabilities that changed

slowly across trials. Similar to the bandit task, participants had to decide when to switch targets based on recent reward history. Boorman et al. reported activation in the IPS during switch trials, and also reported that frontal pole activation correlated with the relative unchosen action probability, which informs switches to alternative courses of action (Boorman et al., 2009). This evidence suggests a fronto-parietal circuit underlies exploratory decision making. In the parietal cortex, the IPS and SMG have been implicated in the selection of action rules during task-switching (Philipp et al., 2012; Slagter et al., 2006) and in spatial working memory and spatial attention (Silk et al., 2010). The frontopolar cortex is believed to support attention to both external stimuli and internal representations (Burgess et al., 2007) and it is also thought to subserve the switching in and out of behavioral options while maintaining other options that are pending (Boorman et al., 2009; Koechlin and Hyafil, 2007). Perhaps the IPS monitors the location and value of alternative options while the frontopolar cortex executes task-switching during exploratory decision making.

Unexpectedly, the exploit>explore contrast revealed a small extent of activation in the bilateral temporal lobes, including the middle and superior temporal gyri, planum temporale, and the left angular gyrus. These regions are mostly known for auditory, language, and semantic processing (Price, 2010). However, the angular gyrus has been implicated in number comparisons (Gobel et al., 2001). It is possible that this temporal lobe activation is related to the numerical monitoring of point values during exploitation, but more research is needed to understand the cognitive and neural mechanisms that underlie this behavior. Exploration, on the other hand, most likely engages a variety of cognitive processes, including spatial and numerical working memory, attention, and executive control. The greater extent of activation during explore trials supports the idea that exploration is more cognitively demanding, or recruits more cortical energy, than exploitation.

During the reward outcome phase, Daw and colleagues found activation in the medial orbitofrontal cortex associated with increasing point values (Daw et al., 2006). In our study, increasing point values were associated with activation in the caudate nucleus, nucleus accumbens, anterior cingulate and paracingulate gyrus. This pattern of activation appears reasonable, since the dorsal striatum has been frequently associated with winning monetary rewards (Delgado, 2007), and the anterior cingulate is known to play a role in encoding the relationship between actions and the reinforcement value of their outcomes (Rushworth et al., 2004). More interestingly, decreasing point values were associated with activation in the lateral premotor cortex and precuneus; both of these regions were also preferentially activated by exploratory trials. The premotor cortex is involved in early movement selection (Schluter et al., 1998) and the precuneus has been implicated in ambiguous decision making (Krain et al., 2006). This association suggests that as point values decrease, there is increased activity in motor and cognitive preparation areas that may anticipate target switching.

A limitation of the current study is the absence of a nonsmoking control group. With nonsmokers for comparison, we could investigate between-group differences and assess whether the duration of smoking affects neural activation during explore/exploit decision making. In addition, we did not test or control for differences in executive function or working memory that could have affected bandit performance. Future studies should also

determine whether smoking states (i.e., withdrawal) as well as traits relate to bandit performance.

To our knowledge, this is the first study to explore the association between tobacco use and explore/exploit brain activation, and it is one of very few studies to apply principles of reinforcement learning to the investigation of substance dependence. Neurobiological research suggests that reinforcement learning, and explore/exploit behavior in particular, are mediated by dopamine (Schultz, 2011), acetylcholine, and norepinephrine (Aston-Jones and Cohen, 2005; Usher et al., 1999; Yu and Dayan, 2005). All three of these neurotransmitters are directly or indirectly affected by nicotine (Di Chiara, 2000; Picciotto and Corrigall, 2002; Watkins et al., 2000), which underscores the utility of reinforcement learning models for the study of tobacco addiction and the need for more research in this area. The potential clinical applications of this work include incorporating explore/exploit behavioral therapies into nicotine addiction interventions, or possibly taking advantage of existing tendencies toward automatic behaviors to help curb smoking behaviors.

Acknowledgments

This research was supported by National Institute on Drug Abuse (NIDA) Grant P30 DA023026, K01 DA033347 (MAA), and R01 DA023516 (FJM). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIDA. This study was also supported by Duke University-University of North Carolina Chapel Hill Brain Imaging and Analysis Center (BIAC).

REFERENCES

Addicott MA, Pearson JM, Wilson J, Platt ML, McClernon FJ. 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. 2013; 21:66–73. doi: Doi 10.1037/A0030843. [PubMed: 23245198]

Aston-Jones G, Cohen JD. Adaptive gain and the role of the locus coeruleus-norepinephrine system in optimal performance. Journal of Comparative Neurology. 2005; 493:99–110. doi: Doi 10.1002/Cne.20723. [PubMed: 16254995]

Benowitz NL. Neurobiology of nicotine addiction: Implications for smoking cessation treatment. The American Journal of Medicine. 2008; 121:s3–s10. [PubMed: 18342164]

Berry, DA.; Fristedt, B. Bandit problems: sequential allocation of experiments. Chapman and Hall; London, New York: 1985.

Boorman ED, Behrens TEJ, Woolrich MW, Rushworth MFS. How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. Neuron. 2009; 62:733–743. doi: DOI 10.1016/j.neuron.2009.05.014. [PubMed: 19524531]

Brainard DH. The psychophysics toolbox. Spatial Vision. 1997; 10:433–436. [PubMed: 9176952]

Brett, M.; Anton, JL.; Valabregue, R.; Poline, JB. Region of interest analysis using an SPM toolbox [abstract]; Presented at the 8th International Conference on Functional Mapping of the Human Brain; Sendai, Japan. June 2-6, 2002; Available on CD-ROM in NeuroImage

Buhler M, Vollstadt-Klein S, Kobiella A, Budde H, Reed LJ, Braus DF, Smolka MN. Nicotine dependence is characterized by disordered reward processing in a network driving motivation. Biological Psychiatry. 2010; 67:745–752. doi: DOI 10.1016/j.biopsych.2009.10.029. [PubMed: 20044075]

Burgess PW, Dumontheil I, Gilbert SJ. The gateway hypothesis of rostral prefrontal cortex (area 10) function. Trends in Cognitive Sciences. 2007; 11:290–298. doi: DOI 10.1016/j.tics.2007.05.004. [PubMed: 17548231]

Cohen JD, McClure SM, Yu AJ. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. Philosophical Transactions of the Royal Society B. 2007; 362:933–942.

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. Cortical substrates for exploratory decisions in humans. Nature. 2006; 441:876–879. doi: 10.1038/nature04766. [PubMed: 16778890]

Delgado MR. Reward-related responses in the human striatum. Reward and Decision Making in Corticobasal Ganglia Networks. 2007; 1104:70–88. doi: DOI 10.1196/annals.1390.002.

Di Chiara G. Role of dopamine in the behavioural actions of nicotine related to addiction. European Journal of Pharmacology. 2000; 393:295–314. [PubMed: 10771025]

Franken IHA. Drug craving and addiction: integrating psychological and neuropsychopharmacological approaches. Progress in Neuro-Psychopharmacology & Biological Psychiatry. 2003; 27:563–579. doi: Doi 10.1016/S0278-5846(03)00081-2. [PubMed: 12787841]

Gittins, JC. Multi-armed bandit allocation indices. Wiley, Chichester; New York: 1989.

Gobel S, Walsh V, Rushworth MFS. The mental number line and the human angular gyrus. Neuroimage. 2001; 14:1278–1289. doi: DOI 10.1006/nimg.2001.0927. [PubMed: 11707084]

Graybiel AM. Habits, rituals, and the evaluative brain. Annual Review of Neuroscience. 2008; 31:359–387. doi: DOI 10.1146/annurev.neuro.29.051605.112851.

Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. British Journal of Addiction. 1991; 86:1119–1127. [PubMed: 1932883]

Kaelbling LP, Littman ML, Moore AW. Reinforcement learning: A survey. Journal of Artificial Intelligence Research. 1996; 4:237–285.

Koechlin E, Hyafil A. Anterior prefrontal function and the limits of human decision-making. Science. 2007; 318:594–598. doi: 318/5850/594 [pii]10.1126/science.1142995. [PubMed: 17962551]

Koob GF, Volkow ND. Neurocircuitry of Addiction. Neuropsychopharmacology. 2010; 35:217–238. doi: Doi 10.1038/Npp.2009.110. [PubMed: 19710631]

Krain AL, Wilson AM, Arbuckle R, Castellanos FX, Milham MP. Distinct neural mechanisms of risk and ambiguity: A meta-analysis of decision-making. Neuroimage. 2006; 32:477–484. doi: DOI 10.1016/j.neuroimage.2006.02.047. [PubMed: 16632383]

Martin-Soelch C, Missimer J, Leenders KL, Schultz W. Neural activity related to the processing of increasing monetary reward in smokers and nonsmokers. European Journal of Neuroscience. 2003; 18:680–688. [PubMed: 12911764]

McClure SM, Berns GS, Montague PR. Temporal prediction errors in a passive learning task activate human striatum. Neuron. 2003; 38:339–346. [PubMed: 12718866]

Pearson JM, Hayden BY, Raghavachari S, Platt ML. Neurons in posterior cingulate cortex signal exploratory decisions in a dynamic multioption choice task. Current Biology. 2009; 19:1532–1537. doi: 10.1016/j.cub.2009.07.048. [PubMed: 19733074]

Philipp AM, Weidner R, Koch I, Fink GR. Differential roles of inferior frontal and inferior parietal cortex in task switching: Evidence from stimulus-categorization switching and response-modality switching. Human Brain Mapping. 2012 doi: 10.1002/hbm.22036.

Piasecki TM, Piper ME, Baker TB. Tobacco Dependence: Insights from Investigations of Self-Reported Motives. Current Directions in Psychological Science. 2010; 19:395–401. doi: Doi 10.1177/0963721410389460. [PubMed: 21552361]

Picciotto MR, Corrigall WA. Neuronal systems underlying behaviors related to nicotine addiction: Neural circuits and molecular genetics. Journal of Neuroscience. 2002; 22:3338–3341. [PubMed: 11978809]

Piper ME, Bolt DM, Kim SY, Japuntich SJ, Smith SS, Niederdeppe J, Baker TB. Refining the tobacco dependence phenotype using the Wisconsin Inventory of Smoking Dependence Motives. Journal of Abnormal Psychology. 2008; 117:747–761. doi: Doi 10.1037/A0013298. [PubMed: 19025223]

Piper ME, Piasecki TM, Federman EB, Bolt DM, Smith SS, Fiore MC, Baker TB. A multiple motives approach to tobacco dependence: the Wisconsin Inventory of Smoking Dependence Motives (WISDM-68). Journal of Consulting and Clinical Psychology. 2004; 72:139–154. [PubMed: 15065950]

Price CJ. The anatomy of language: a review of 100 fMRI studies published in 2009. Year in Cognitive Neuroscience 2010. 2010; 1191:62–88. doi: DOI 10.1111/j.1749-6632.2010.05444.x.

Ramnani N, Elliott R, Athwal BS, Passinghm RE. Prediction error for free monetary reward in the human prefrontal cortex. Neuroimage. 2004; 23:777–786. doi: DOI 10.1016/j.neoruimage.2004.07.028. [PubMed: 15528079]

Redish AD. Addiction as a computational process gone awry. Science. 2004; 306:1944–1947. doi: DOI 10.1126/science.1102384. [PubMed: 15591205]

Rorden C, Brett M. Stereotaxic display of brain lesions. Behavioural Neurology. 2000; 12:191–200. [PubMed: 11568431]

Ruge H, Braver T, Meiran N. Attention, intention, and strategy in preparatory control. Neuropsychologia. 2009; 47:1670–1685. doi: DOI 10.1016/j.neuropsychologia.2009.02.004. [PubMed: 19397862]

Rushworth MFS, Walton ME, Kennerley SW, Bannerman DM. Action sets and decisions in the medial frontal cortex. Trends in Cognitive Sciences. 2004; 8:410–417. doi: DOI 10.1016/j.tics.2004.07.009. [PubMed: 15350242]

Schluter ND, Rushworth MFS, Passingham RE, Mills KR. Temporary interference in human lateral premotor cortex suggests dominance for the selection of movements - A study using transcranial magnetic stimulation. Brain. 1998; 121:785–799. doi: DOI 10.1093/brain/121.5.785. [PubMed: 9619185]

Schultz W. Potential vulnerabilities of neuronal reward, risk, and decision mechanisms to addictive drugs. Neuron. 2011; 69:603–617. doi: DOI 10.1016/j.neuron.2011.02.014. [PubMed: 21338874]

Silk TJ, Bellgrove MA, Wrafter P, Mattingley JB, Cunnington R. Spatial working memory and spatial attention rely on common neural processes in the intraparietal sulcus. Neuroimage. 2010; 53:718–724. doi: DOI 10.1016/j.neuroimage.2010.06.068. [PubMed: 20615473]

Slagter HA, Weissman DH, Giesbrecht B, Kenemans JL, Mangun GR, Kok A, Woldorff MG. Brain regions activated by endogenous preparatory set shifting as revealed by fMR1. Cognitive Affective & Behavioral Neuroscience. 2006; 6:175–189. doi: Doi 10.3758/Cabn.6.3.175.

Sutton, RS.; Barto, AG. Reinforcement learning: an introduction. MIT Press; Cambridge, Mass.: 1998.

Usher M, Cohen JD, Servan-Schreiber D, Rajkowski J, Aston-Jones G. The role of locus coeruleus in the regulation of cognitive performance. Science. 1999; 283:549–554. [PubMed: 9915705]

Volkow ND, Wang GJ, Fowler JS, Tomasi D. Addiction Circuitry in the Human Brain. Annual Review of Pharmacology and Toxicology. 2012; 52:321–336. 52. doi: DOI 10.1146/annurev-pharmtox-010611-134625.

Ward, B. Simultaneous inference for fMRI data. Biophysics Research Institute, Medical College of Wisconsin; Milwaukee: 2000.

Watkins SS, Koob GF, Markou A. Neural mechanisms underlying nicotine addiction: acute positive reinforcement and withdrawal. Nicotine & Tobacco Research. 2000; 2:19–37. [PubMed: 11072438]

Yu AJ, Dayan P. Uncertainty, neuromodulation, and attention. Neuron. 2005; 46:681–692. doi: DOI 10.1016/j.neuron.2005.04.026. [PubMed: 15944135]

- We studied the relationship between explore/exploit behavior and tobacco addiction.
- Cigarette smokers completed a 6-armed bandit task while undergoing fMRI.
- Exploratory activation was greater in the parietal and frontal cortices.
- Exploitative activation was greater in the temporal cortices.
- Smoking dependence motives associated with activation in the parietal cortices.

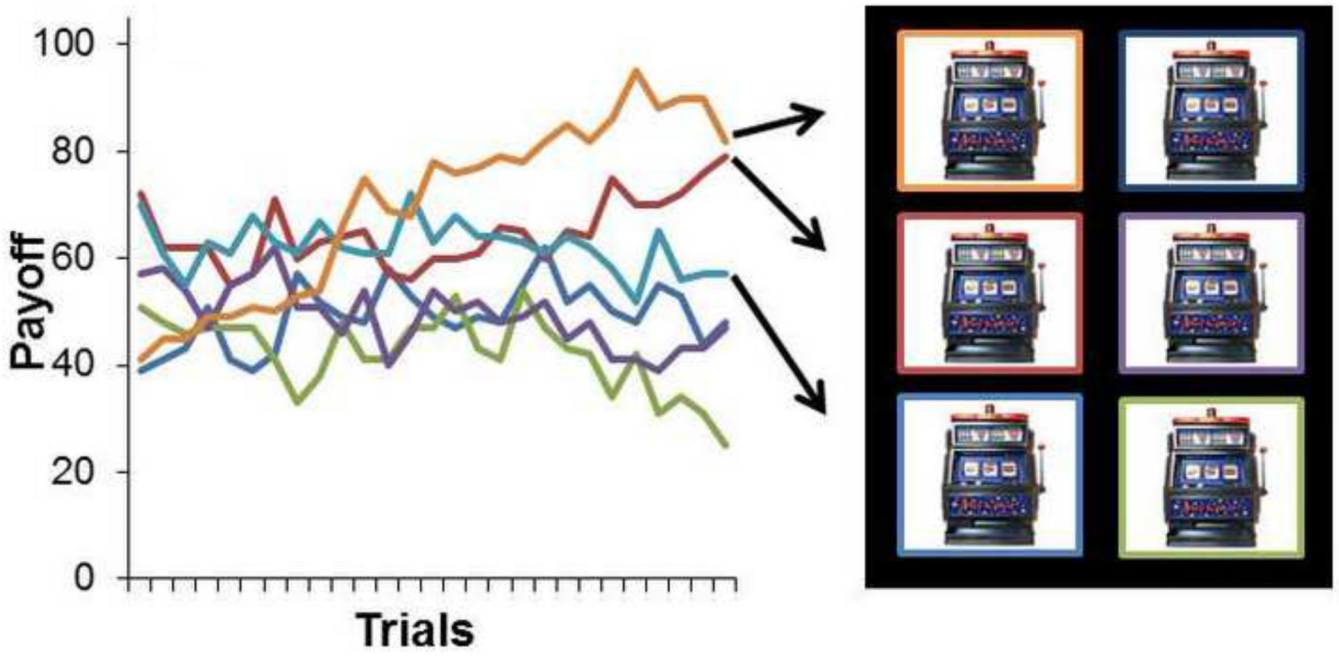

Figure 1.

Example of payoff values per trial for each arm (i.e., slot machine) of the bandit task. Each colored line represents a different arm. Participants selected one arm to play each trial. Payoff values were determined by a biased random walk (Equation 1).

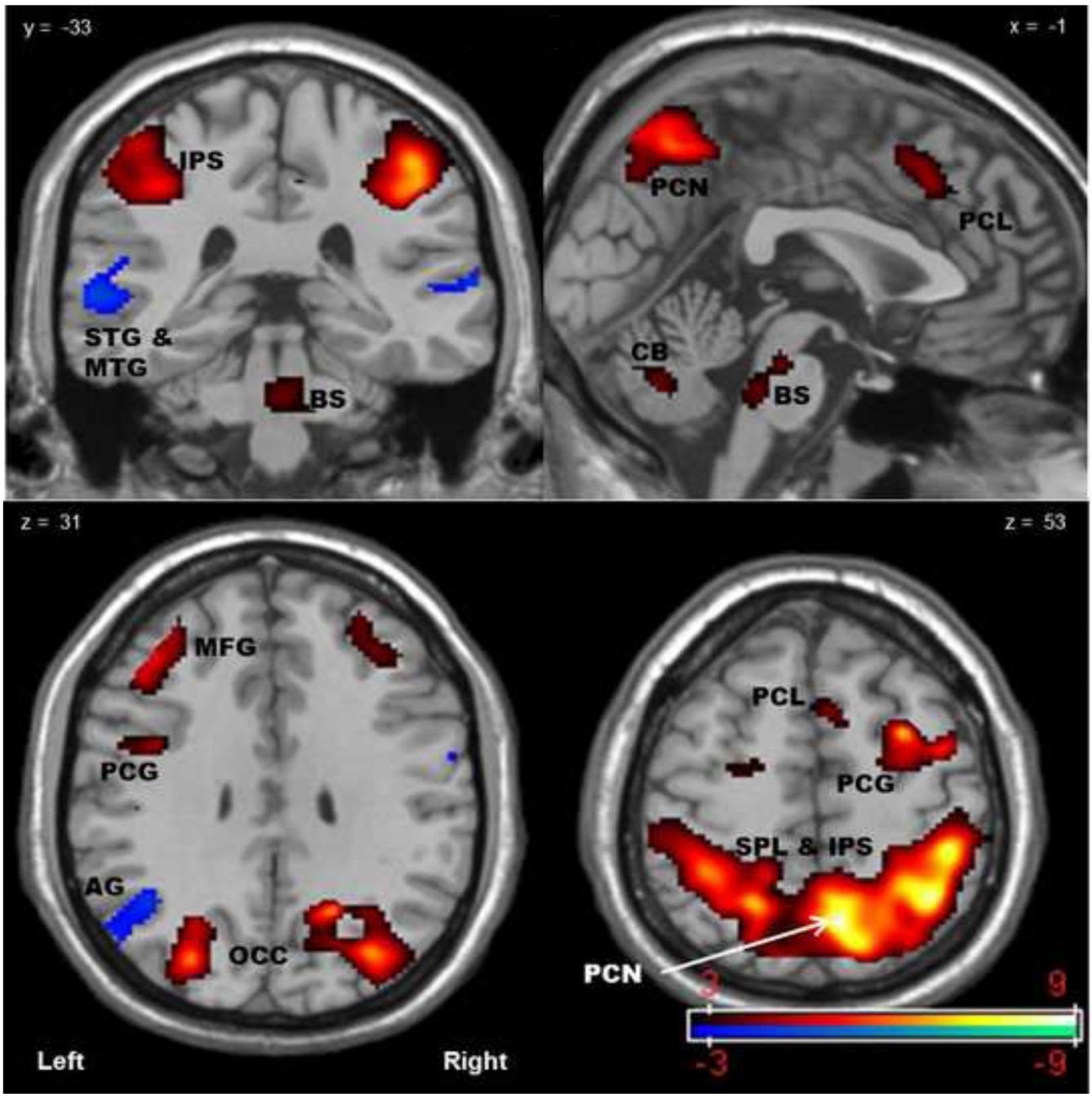

Figure 2.

Statistical parametric maps for the selection phase of the bandit task showing regions of activation for Explore>Exploit (red) and Exploit>Explore (blue). AG: angular gyrus; BS: brainstem; CB: cerebellum; MFG: middle frontal gyrus; IPS: intraparietal sulcus, OCC: occipital cortex; PCG: precentral gyrus; PCL: paracingulate gyrus; PCN: precuneus; SPL: superior parietal lobule; STG & MTG: superior & middle temporal gyrus. Color bar represents T-scores, $p < .005$, cluster size 190 voxels.

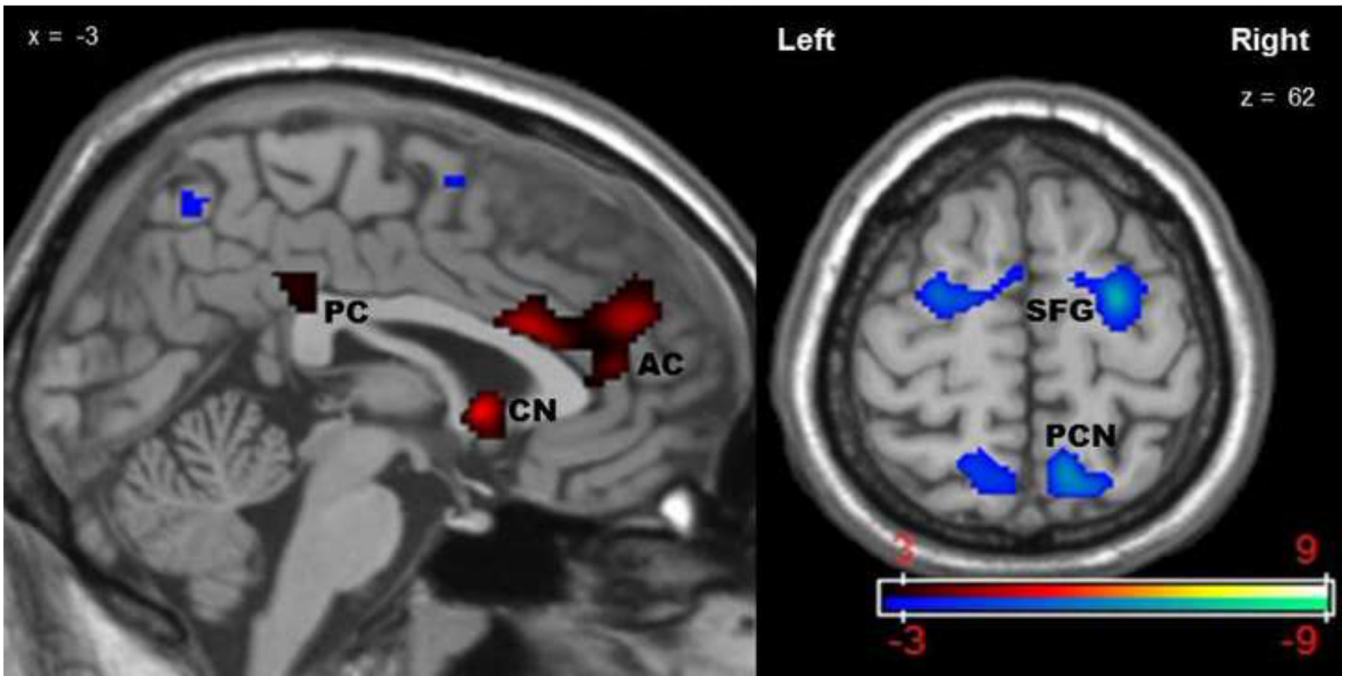

Figure 3.

Statistical parametric maps for the outcome phase of the bandit task showing regions where reward values were associated with increased (red) and decreased (blue) activation. AC: anterior cingulate; CN: caudate nucleus; PC: posterior cingulate; PCN: precuneus; SFG: superior frontal gyrus. Color bar represents T-scores, $p < 0.005$, cluster size 190 voxels.

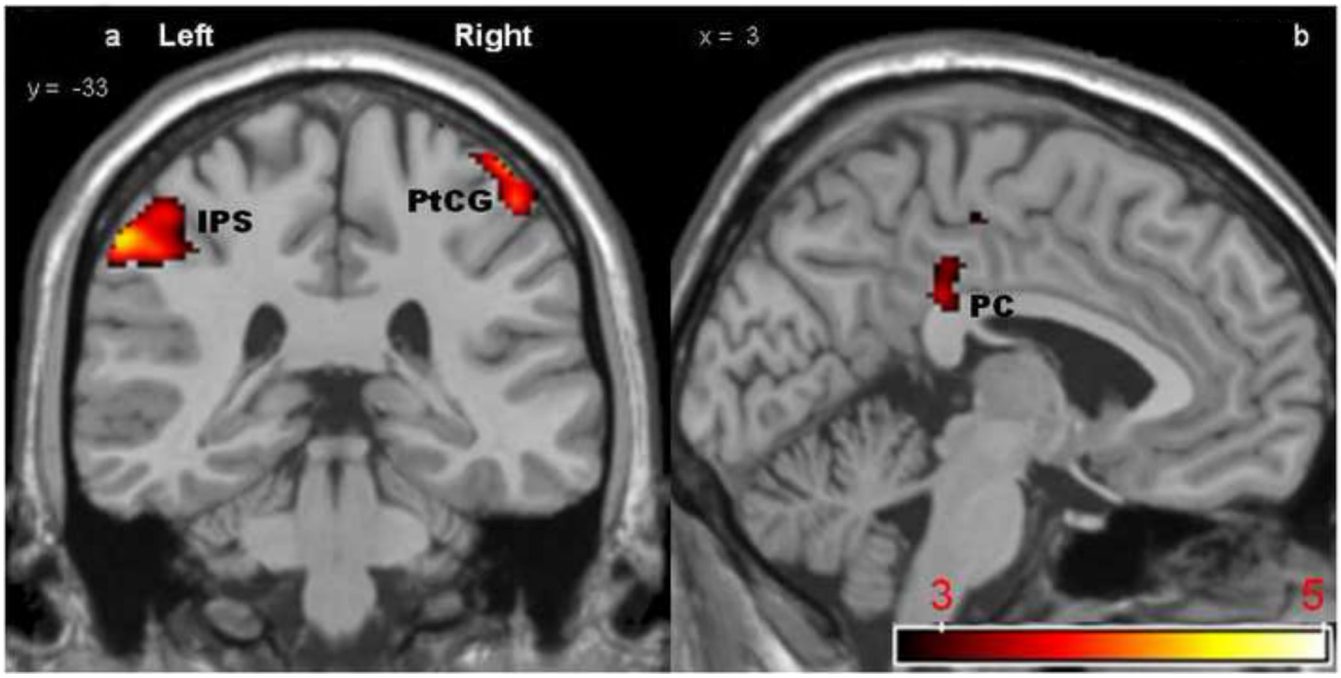

Figure 4.

Statistical parametric maps for the (a) selection and (b) outcome phases of the bandit task showing regions where WISDM component scores correlated with brain activation. Positive automaticity scores and negative tolerance scores correlated with exploratory activation in the bilateral parietal regions (a), and positive craving scores correlated with increasing reward values in the posterior cingulate (b). IPS: intraparietal sulcus; PC: posterior cingulate; PtCG: postcentral gyrus. Color bar represents T-scores, $p < 0.005$, cluster size 190 voxels.

Table 1

Principle component analysis of the 4 primary WISDM subscales.

	Components			
	1	**2**	**3**	**4**
Automaticity	**0.49**	**0.76**	−0.13	−0.40
Loss of control	**0.52**	0.05	−0.26	**0.81**
Tolerance	**0.53**	**−0.63**	−0.38	−0.43
Craving	**0.46**	−0.14	**0.88**	0.00

Table 2

Statistical parametric mapping of the selection and outcome phases of the 6ABT task. Threshold $p < 0.005$ uncorrected, $K_E = 190$. R: right, L: left.

Brain Region		Location of peak voxel (MNI coordinates)	Volume of cluster (μL)	T-score	cluster level, uncorrected
Selection phase: explore > exploit					
Parietal cortex	R & L superior parietal lobule, intraparietal sulcus, precuneus, supramarginal gyrus, lateral occipital cortex	36 −54 56	101448	10.57	<0.001
Frontal cortex	R superior, middle, and precentral frontal gyrus	30 0 62	11104	8.55	<0.001
	L superior, middle, and precentral frontal gyrus	−24 −4 66	4800	6.41	0.003
	R middle frontal gyrus, frontal pole	28 38 38	4800	5.94	0.003
	L middle frontal gyrus, frontal pole	−42 34 28	4760	5.66	0.003
	L precentral, middle frontal gyrus	−40 2 36	4632	4.90	0.003
	R paracingulate gyrus	4 22 42	2856	4.35	0.016
Cerebellum	R cerebellum, brain stem	22 −56 −34	9776	5.30	<0.001
	L cerebellum	−28 −50 −32	4688	5.79	0.003
Subcortical	L pallidum, thalamus, putamen	−16 −4 6	2296	5.55	0.028
	L frontal operculum, insula, putamen	−32 16 12	1784	4.17	0.049
	R caudate, thalamus	18 −6 18	1680	4.15	0.055
Selection phase: exploit > explore					
Temporal cortex	R superior, middle temporal gyrus, planum temporale	66 −30 6	2512	4.96	0.023
	L superior, middle temporal gyrus, planum temporale	−60 −44 0	3504	5.53	0.009
	L angular gyrus, occipital cortex	−42 −54 26	2520	4.19	0.022
Outcome phase: positive correlation with reward value					
Subcortical	L & R caudate nucleus & nucleus accumbens	−6 8 −2	2152	6.07	0.039
Parietal cortex	L supramarginal gyrus	−56 −52 40	3480	4.39	0.012
	L posterior cingulate gyrus	−10 −36 42	1536	3.79	0.075
Frontal cortex	L & R anterior cingulate gyrus & paracingulate gyrus	0 18 22	6336	4.69	0.001
Occipital	R inferior occipital cortex	52 −70 −8	1656	4.37	0.066
Outcome phase: negative correlation with reward value					
Frontal cortex	R superior & middle frontal gyrus	26 −6 62	3696	7.42	0.010
	L superior & middle frontal gyrus	−26 −4 60	2680	5.36	0.024
Parietal cortex	R & L precuneus & superior occipital cortex	10 −60 58	9592	5.01	<0.001

Table 3

Regression analyses of the selection and outcome phases of the 6ABT task. Threshold $p < 0.005$ (uncorrected) $K_E = 190$. R: right, L: left.

Brain Region		Location of peak voxel (MNI coordinates)	Volume of cluster (μL)	T-score	cluster level, uncorrected
Selection phase: positive correlation between explore > exploit & principle component 2					
Parietal cortex	R postcentral gyrus, supramarginal gyrus	52 −28 52	3176	4.99	0.013
	L postcentral gyrus, supramarginal gyrus & intraparietal sulcus	−58 −34 46	2608	3.54	0.022
Outcome phase: positive correlation between reward value & principle component 3					
Parietal cortex	R & L posterior cingulate gyrus	12 −38 30	2704	4.03	0.020
Occipital cortex	L inferior occipital cortex	−38 −70 −4	2216	3.96	0.032
Temporal cortex	L inferior & middle temporal gyrus	−50 −38 −14	1816	4.61	0.049
Frontal cortex	L precentral & superior frontal gyrus	−28 −18 74	1912	3.17	0.044