Nicotine & Tobacco Research, 2020, 164–171
doi:10.1093/ntr/nty136
Original investigation
Received August 17, 2017; Editorial Decision June 25, 2018; Accepted June 28, 2018

Original investigation

Smoking Decisions: Altered Reinforcement Learning Signals Induced by Nicotine State

Travis E. Baker, PhD[1], Yashar Zeighami, PhD[2], Alain Dagher, MD, PhD[2], Clay B. Holroyd, PhD[3]

[1]Center for Molecular and Behavioral Neuroscience, Rutgers University, Newark, NJ; [2]Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, Canada; [3]Department of Psychology, University of Victoria, BC, Canada

Corresponding Author: Travis E. Baker, Center for Molecular and Behavioral Neuroscience, Rutgers University, Newark, 197 University Avenue, Newark, NJ 07102, USA. Telephone: (862)-250-3351; E-mail: travis.e.baker@rutgers.edu

Abstract

Introduction: Alterations in dopamine signaling play a key role in reinforcement learning and nicotine addiction, but the relationship between these two processes has not been well characterized. We investigated this relationship in young adult smokers using a combination of behavioral and computational measures of reinforcement learning.

Methods: We asked moderately dependent smokers to engage in a reinforcement learning task three times: smoking as usual, smoking abstinence, and cigarette consumption. Participants' trial-to-trial training choices were modeled using a reinforcement learning model that calculates separate learning rates associated with positive and negative prediction errors.

Results: We found that learning from positive prediction error signals is reduced during smoking abstinence and enhanced following cigarette consumption. By contrast, learning from negative prediction error signals was enhanced during smoking abstinence and reduced following cigarette consumption. Finally, when tested with novel pairs of stimuli, participants were relatively better at selecting the positive feedback predicting stimuli than avoiding the negative feedback predicting stimuli during the smoking as usual session, a pattern that reversed following cigarette consumption.

Conclusions: These findings provide a specific computational account of altered reinforcement learning induced by smoking state (abstinence and consumption) and may represent a unique target for treatment of nicotine addiction.

Implications: This study illustrates the potential of computational psychiatry for understanding reinforcement learning deficits associated with substance use disorders in general and nicotine addiction in particular. We found that learning from positive prediction error signals is reduced during smoking abstinence and enhanced following cigarette consumption. By contrast, learning from negative prediction error signals was enhanced during smoking abstinence and reduced following cigarette consumption. By highlighting important computational differences between three states of smoking, these findings hold out promise for integrating experimental, computational, and theoretical analyses of decision-making function together with research on addiction-related disorders.

Introduction

Smoking has been linked to devastating health problems and is responsible for more than 480 000 deaths per year in the United States and nearly 6 million deaths worldwide.[1] So, why do individuals repeatedly decide to smoke despite such catastrophic consequences on personal health? Needless to say, the impact of nicotine on the brain's decision-making system is of long-standing interest to addiction researchers. Although the pharmacological effects of nicotine are becoming increasingly clear,[2] implicating an array of transmitter systems including dopamine, noradrenaline, acetylcholine, glutamate, and GABA,[3,4] how nicotine alters neural mechanisms for decision making remains poorly understood.

A likely contributor to addictive behavior is the dopamine system, because transient changes in dopamine that occur during positive and negative reinforcement play a key role in reinforcement learning, which ultimately guides future decisions,[5] and nicotine modulates dopamine levels via its influence on nicotinic acetylcholine receptors.[3] Dopamine activity encodes a reward prediction error (RPE) signal, which guides stimulus and action value learning in standard reinforcement learning models.[6] Accordingly, it has been proposed that phasic bursts of dopamine activity (positive RPE) facilitate learning from positive feedback, whereas transient cessations in dopamine activity (negative RPE) facilitate avoidance learning.[7] Acute nicotine use increases phasic dopamine activity via acetylcholine receptors on dopamine neurons, which is thought to contribute to the nicotine-induced reinforcement of addictive behaviors.[4,8] By contrast, smoking abstinence has been shown to reduce tonic and phasic dopamine activity, contributing to changes in motivation and sensitivity to drug-associated stimuli, thus promoting relapse.[9–11] Although dopaminergic RPE signals have been implicated in both decision-making and addiction processes,[12–14] and individual variability in reinforcement learning may present a core vulnerability to nicotine addiction,[15–19] a complete understanding of how cigarette consumption and smoking abstinence affects positive and negative RPE signals during reinforcement learning remains to be developed.[20]

Adopting a computational psychiatry approach,[21] we examined the impact of smoking abstinence and cigarette consumption on behavioral and computational measures of reinforcement learning. In particular, across 2 days we asked smokers to engage in the Probabilistic Selection Task (PST)—a trial-and-error learning task that has been shown to be sensitive to dopamine dysfunction[22]—three times, during a "smoking as usual" session (day 1), then following a 24-hour period of smoking abstinence, and then immediately after cigarette consumption (day 2). According to an influential theory of reinforcement learning, positive RPE signals facilitate approach learning in this task by reinforcing a striatal "Go" pathway via D1 receptors, whereas negative RPE signals facilitate avoidance learning in this task by reinforcing a striatal "No-go" pathway via D2 receptors.[23] We predicted that if different smoking states (ie, smoking as usual, smoking abstinence, and cigarette consumption) differentially affect RPE signals during reinforcement learning, then performance in this task should differ between these smoking states.

Furthermore, to examine the computational differences in reinforcement learning across smoking states, we employed a reinforcement learning model known as Q-learning, which provides a means to simulate individual trial-to-trial task choices to determine separate learning rates for positive and negative RPE signals.[24] The model learns the value associated with each stimulus-action pair using an RPE signal and updates the current estimated value on each trial based on the RPE signal. The learning rates for positive and negative RPEs determine the degree to which recent RPEs affect the expected value, which updates action values by the product of the RPE and the learning rate. Thus, if the RPE signals are artificially amplified or attenuated in response to cigarette consumption or smoking abstinence, respectively, then changes in learning rates during training phase of the PST should capture such smoking-induced RPE alterations. In this way, computational modeling is well suited for identifying and characterizing reinforcement learning-based endophenotypes of nicotine addiction, and thus can provide a new level of biological insight into smoking-related dysfunction (eg, relapse, craving, withdrawal, and anhedonia).

Methods

Participants

For this study, a total of 25 cigarette smokers (12 male; mean age = 25 years, SEM = 0.7) completed the experimental sessions as depicted in Figure 1. Participants were undergraduate students at

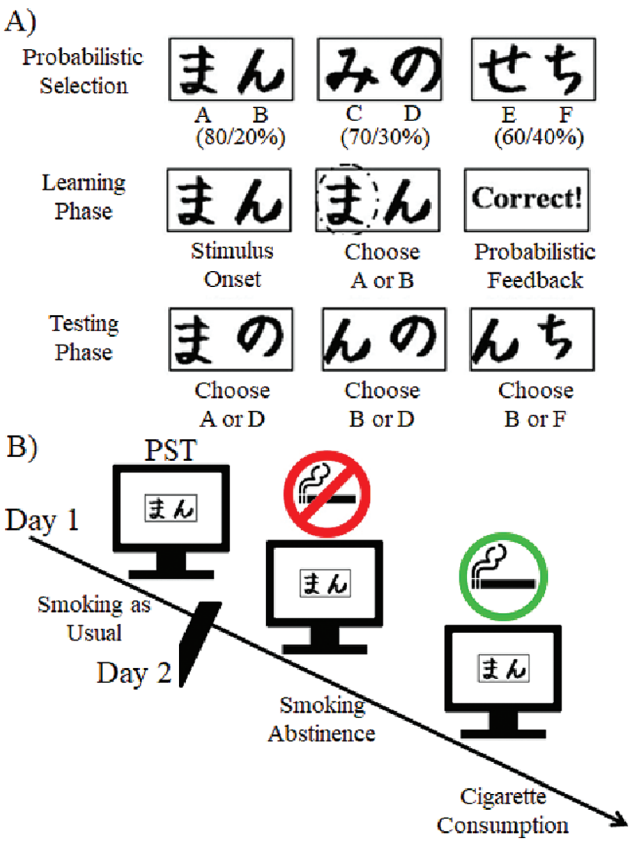

Figure 1. Probabilistic learning task. (A) Top row: stimuli and feedback probabilities (percent positive feedback vs. negative feedback). Middle row: schematic of an example trial during the learning phase. On each trial of the task, they viewed a fixation cross (green circle, 1 s) followed by a pair of visual stimuli that are not easily verbalized by most English speakers (ie, Japanese Hiragana characters) presented in black on a white background in 72 pt font. They then pressed the key corresponding to the stimulus that they believed to be correct. Visual feedback was provided following each choice: the word "Correct!" printed in blue or "Incorrect" printed in red (1 s). If no response was made within 6 seconds, then the words "no response detected" were displayed in red (1 s). Bottom row: schematic of an example trial during the testing phase. (B) Experiment session sequence across 2 days..

the University of Victoria, who participated as part of a smoking electroencephalographic study on the neural mechanisms of reward processing and substance use,[18] the data of which are reanalyzed here. Each subject received course credit for their participation in this two-session study spanning a 2-week interval. Participants were required to be current smokers who were not currently trying or planning to quit smoking. Nicotine dependence scores were assessed using the specific substance involvement score of the tobacco category of the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST),[25] which characterized the current sample of smokers in the moderate nicotine-dependent range (mean = 12.4, SEM = 1.1). To note, mid-range ASSIST scores between 4 and 19 for any substance, including tobacco, are an indication of hazardous or harmful use of that substance.[25] No subject reported concurrent neuropsychiatric symptoms. All participants had normal or corrected-to-normal vision and all gave informed consent. The study was approved by the local research ethics committee and was conducted in accordance with the ethical standards prescribed in the 1964 Declaration of Helsinki.

Procedures

Each subject was asked to participate in a two-part study spanning a 2-week interval (Figure 1). Informed consent, questionnaire data, and a baseline measure of breath carbon monoxide (CO) levels were obtained during the first part of the study. Participants were not instructed to refrain from smoking before this session or asked to smoke on arrival. Current smoking status was confirmed by assessing breath CO levels using a piCO Smokerlyzer CO breathalyzer (Bedfont Scientific Ltd, Kent, England). Participants were then asked to complete a set of computer-based questionnaires used in our previous studies.[14,26] Subjects then completed a paper-and-pen Tobacco Craving Questionnaire (TCQ), a 47-item scale that assesses state levels of craving.[27] All inventories were administered and scored according to their guidelines.

Following the completion of the questionnaires, participants engaged in the PST (hereafter, the "smoking as usual" session). As mentioned earlier,[28] participants are required to learn three concurrent discriminations (stimulus pairs AB, CD, and EF), and are provided with performance feedback (correct-incorrect) with schedules of 80%-20%, 70%-30%, and 60%-40%, respectively (Figure 1A). Participants learned by trial-and-error to choose the more frequently correct stimulus over the alternative in each pair, namely, by selecting stimuli A, C, and E more often than B, D, and F. Critically, they could do so by learning that stimuli A, C, and E were associated with relatively more positive feedback, by learning that stimuli B, D, and F were associated with relatively more negative feedback, or both. Participants advanced to the testing phase of the task if after any block of 60 trials they satisfied performance criteria for the three stimulus pairs (65% A in AB, 60% C in CD, and 50% E in EF) or after six blocks (360 trials) of training if these criteria were not met.

During the testing phase, participants were exposed to all possible combinations of these stimuli (ie, AB, CD, EF, AC, AD, AE, AF, BC, BD, BE, BF, CE, DF) in a random order and were required to select the symbol in each pair that they believed to be correct, but without receiving any feedback about their choices. They were told to use "gut instinct" whenever they did not know which option to choose. Each test pair was presented six times. Participant's accuracy for "approach learning" involves selecting the A stimulus (AC, AD, AE, and AF), and for "avoidance learning" involves avoiding the B stimulus (BC, BD, BE, and BF). Importantly, different stimuli were used in each session to minimize potential learning effects between sessions. At the end of this session, participants were instructed to abstain from smoking for 24 hours before the start of next scheduled session.

On a separate day, participants returned to the laboratory and were asked again to provide their consent, CO levels, and to complete the TCQ. Compliance with the 24-hour abstinence period was defined as breath CO less than 8 ppm, a verified cutoff threshold for confirming smoking abstinence.[29] The data of participants failing to meet this criterion were excluded from further analysis. Following a separate electroencephalographic task,[18] participants engaged in the PST for the second time (hereafter, the "smoking abstinence" session) (Figure 1B). On completion, participants were escorted to a smoking-designated area where they were provided with an opportunity to consume one cigarette. Participants then returned to the laboratory (within 5 minutes postcigarette) and were asked to provide expired CO levels and to complete the third session of the PST (hereafter, the "cigarette consumption" session). Importantly, different stimuli were used across the three sessions to minimize potential learning effects between sessions.

PST Behavioral Analysis

We first analyzed PST learning phase strategic performance (win-stay and lose-shift) using repeated measures analyses of variance (ANOVAs) with smoking state (smoking as usual, smoking abstinence, cigarette consumption) as a within-subject factors. Lose-shift performance was measured on trials in which participants selected a particular stimulus, received negative feedback, and used this feedback to avoid selecting the same stimulus in the next trial in which it appeared. Win-stay performance was assessed similarly (trials in which participants used positive feedback to select the same stimulus in the next trial). Next, we analyzed PST testing phase accuracy using repeated measures ANOVAs with smoking state (smoking as usual, smoking abstinence, cigarette consumption) and stimulus condition (approach and avoidance) as within-subject factors.

Computational Modeling

We further analyzed the learning phase data for each subject using a Q-learning computational model.[24,30] The choice behavior of each subject was separately modeled with individual learning parameters for positive and negative learning rates to explore whether the effects of smoking abstinence and cigarette consumption relate to difference underlying neuronal mechanisms for positive- and negative-feedback learning. The model estimated the expected value for each possible action on each trial based on the subject's choice-outcome histories. Action values $Q(s_t, a_t)$ were initiated at zero and updated at the end of each trial according to the subject's choice and the subsequent outcome via a prediction error signal, which reflected the difference between the received outcome and the expected value:

$$\delta_t = R_t - Q(s_t, a_t) \quad (1)$$

where, δ_t indicates the prediction error to feedback delivery at trial t, R_t is the received outcome at trial t (1 for positive feedback and 0 for negative feedback), and $Q(s_t, a_t)$ is the expected outcome in the state s_t (eg, receiving AB pair in the trial) with the choice a_t (eg, choosing A). Action values were updated separately for positive and negative prediction errors according to the following equation:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha^{+} \delta_t \quad \text{if } \delta_t > 0$$

$$Q(s_t,a_t) = Q(s_t,a_t) + \alpha^{-}\delta_t \quad if\ \delta_t < 0 \tag{2}$$

where, α^{+} and α^{-} are positive and negative learning rates, respectively.

Action selection is based on the "soft-max" equation, in which the probability of each action is calculated on the basis of the expected value of that action in the given state compared with the expected value of the other available action:

$$P(a_t = A, s_t = AB) = \frac{1}{1+\exp\left[\begin{matrix}-\beta(Q(s_t,A)-Q(s_t,B))\\-\varphi(C(s_t,A)-C(s_t,B))\end{matrix}\right]} \tag{3}$$

where, $P(a_t = A, s_t = AB)$ is the probability of choosing A given the AB option, β is the inverse temperature parameter that defines the slope of the soft-max *ranging from* $\beta = 0$ *(select at random with uniform probability) to* $\beta = \infty$ *(always select the action with the higher value)*, $C(s_t,a_t)$ represents whether a_t had been selected the last time subject faced the same choice (eg, in case of choice A in last AB trial, $C(s_t,A) = 1$ *and* $C(s_t,B) = 0$), and φ is a perseveration parameter (0 = no impact of previous choice, ∞ = select the choice on the previous trial).

Each participant's trial-by-trial choices during the learning phase were fit separately for each of the three sessions using a Q-learning model with four free parameters (positive and negative learning rates, the exploration parameter, and the perseveration parameter). To note, learning rates determine the degree that recent RPEs affect the estimated value, the exploration (ß) parameter regulates the probability that the agent will select the action with the higher value (if ß is close to zero, all actions are selected with a nearly uniform probability, ie, random behavior; if ß is large, the action with higher value will usually be selected, ie, a "greedy" policy). The perseveration parameter (φ) encodes the probability of repeating an action on the subsequent presentation of the same stimulus pair irrespective of learned values. Each parameter was analyzed separately using repeated measures ANOVAs with smoking state (smoking as usual, smoking abstinence, cigarette consumption) as a within-subject factor. Assuming a quadratic relationship existed between the three levels of smoking state (high to low and low to high) and PST measures, we took both the quadratic and linear contrasts into consideration. The computational analysis was carried out on the first block of 60 trials, when feedback is maximally informative and plasma nicotine levels following consumption are at their highest[31] (please see Supplementary Online Material [SOM] for analysis across all trials).

Results

Participants

In contrast to session 1, smoking abstinence appeared to increase participants' subjective levels of cigarette craving as measured by the TCQ (Table 1; Figure 2A, right side). In particular, a repeated measures ANOVAs on CO levels with smoking state (smoking as usual, smoking abstinence, cigarette consumption) as a within-subject factor revealed a main effect of smoking state, $F(2, 48) = 11.36$, $p < .001$, $N^2 = 0.32$. Further, a significant quadratic contrast was observed, $F(1, 24) = 23.62$, $p < .001$, $N^2 = 0.50$, with a nonsignificant linear contrast $F(1, 24) = 0.58$, $p = .452$, $N^2 = 0.02$, suggesting that CO level changes across sessions (ie, from high to low and then low to high) were associated with smoking state (Table 1; Figure 2A, left side). Post hoc tests indicated that the CO levels at the start of the smoking abstinence session ($M = 4.2$ ppm, SEM = 0.4) were significantly lower than those taken at the smoking as usual session ($M = 8.2$ ppm, SEM = 1), $t(24) = 3.7$, $p < .001$, and the levels following cigarette consumption ($M = 7.5$ ppm, SEM = 0.7), $t(25) = -5.7$, $p < .001$; the difference between the smoking as usual and cigarette consumption sessions was not statistically significant ($p > .05$).

Table 1. Sample Characteristics

	M	SEM (range)
Sample (*N*)	*n* = 25	
Sex (male/female)	12/13	
Age	21.12	0.6 (18–30)
Smoking behavior		
Nicotine dependence	12.4	1.1 (3–20)
Cigarettes/day	3.8	0.8 (1–17)
Years smoking	5.0	0.7 (1–17)
CO ppm		
Smoking as usual	8.2	1.0 (2–20)
Smoking abstinence	4.2	0.4 (1–7)
Cigarette consumption	7.5	0.7 (2–16)
TCQ score		
Smoking as usual	3.9	0.2 (2–5.4)
Smoking abstinence	4.3	0.2 (2–5.8)
Cigarette consumption	NR	

CO = carbon monoxide; NR = not recorded; SEM = standard errors of the mean; TCQ = Tobacco Craving Questionnaire.
Numbers indicate means and SEM, except for sample, which indicates sample size (*N*), and sex, which indicates participant number (male/female).

Conversely, participants' TCQ scores taken at the smoking as usual session ($M = 3.9$, SEM = 0.19) were significantly smaller than those taken at the start of the smoking abstinent session ($M = 4.3$, SEM = 0.19), $t(24) = -3.4$, $p < .005$ (Figure 2A, right side). Together, these results suggest that smoking abstinence before the experiment increased participants' subjective levels of cigarette craving.

Behavioral Results

Learning Phase Accuracy

We analyzed PST learning phase accuracy (AB, CD, and EF) using a repeated measures ANOVA with smoking state (smoking as usual, smoking abstinence, cigarette consumption) as a within-subject factor. In line with the earlier reported results, this analysis revealed a significant quadratic contrast $F(1, 24) = 4.71$, $p < .05$, $N^2 = 0.15$, with a nonsignificant linear contrast, $F(1, 24) = 0.38$, $p = .54$, $N^2 = 0.01$, suggesting that the change in learning phase accuracy across sessions (ie, from high to low and then low to high) was consistent with an effect of CO levels, and probably not a practice effect (Figure 2B, left side; Table 1). Post hoc tests indicated that the learning phase accuracy was lower during the smoking abstinent session ($M = 0.55$, SEM = 0.03) relative to the cigarette consumption session ($M = 0.64$, SEM = 0.03), $t(24) = -2.1$, $p < .05$, but only slightly different relative to the smoking as usual session ($M = 0.61$, SEM = 0.02), $t(24) = -1.4$, $p = .18$. No differences were observed between the smoking as usual and cigarette consumption states ($p > .05$).

Furthermore, a repeated measures ANOVA on overall win-stay performance with smoking state (smoking as usual, smoking abstinence, cigarette consumption) as a within-subject factor revealed an effect of smoking state, $F(2, 48) = 3.2$, $p < .05$, $N^2 = 0.12$ (Supplementary Figure 1, right panel). Post hoc tests indicated

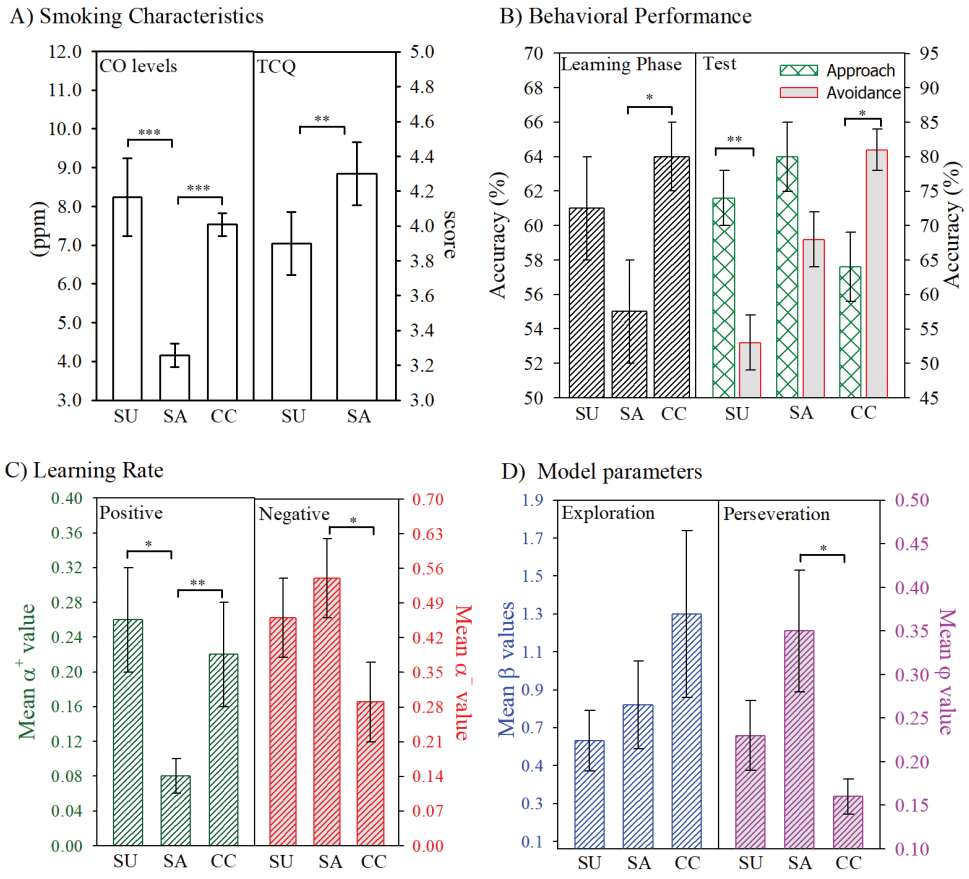

Figure 2. Smoking, behavior, and computational results. Top. (A) Smoking characteristics. Carbon monoxide (CO) levels (left panel) and Tobacco Craving Questionnaire (TCQ) (right panel) taken during smoking as usual (SU), smoking abstinence (SA), and cigarette consumption (CC) session. (B) Learning phase (left side) and testing phase (right side) accuracy in task performance (left side). Bottom. Q-learning results on (C) positive (green bars: left side) and negative (red bars: right side) learning rate values, and (D) exploration (blue bars: left side) and perseveration (purple bars: right side) parameters. Results are shown for the first block of 60 trials. Error bars indicate standard errors of the means. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

that participants chose a win-stay strategy more often during the cigarette consumption session (M = 0.79, SEM = 0.02) than during smoking as usual session (M = 0.70, SEM = 0.02), $t(24) = 3.01$, $p < .006$, with a marginal difference with the smoking abstinence session, (M = 0.72, SEM = 0.03), $t(24) = 1.8$, $p = .08$. In regards to lose-shift performance, no differences were observed across sessions and an analysis on the first 60 trials did not yield any significant results (Supplementary Figure 1, left panel).

PST Testing Phase Results

A repeated measures ANOVA on PST testing phase accuracy with smoking state (smoking as usual, smoking abstinence, cigarette consumption) and stimulus condition (approach and avoidance) as factors revealed a significant interaction, $F(2, 38) = 5.9$, $p < .006$, $N^2 = 0.24$. Post hoc tests indicated that participants performed better on approach trials (M = 0.74, SEM = 0.04) than avoidance trials (M = 0.8, SEM = 0.04), $t(19) = 3.5$, $p < .005$ during the smoking as usual session (Figure 2B, right side). In contrast, participants performed better on avoidance trials (M = 0.81, SEM = 0.04) than approach trials (M = 0.65, SEM = 0.04), $t(19) = –2.2$, $p < .05$, during the cigarette consumption session, with no differences observed between approach and avoidance trials for the smoking abstinence session ($p = .42$). It is also worth noting that compared with the smoking as usual session, avoidance accuracy was higher during the smoking abstinence (M = 0.68, SEM = 0.05), $t(19) = –2.2$, $p < .05$, and cigarette consumption (M = 0.81, SEM = 0.04), $t(24) = –4.6$, $p < .001$, session.

Computational Results

Q-Learning Results

A repeated measures ANOVA on positive learning rate values (Figure 2C, left side) with smoking state (smoking as usual, smoking abstinence, cigarette consumption) as a within-subject factor revealed a significant quadratic contrast $F(1, 24) = 12.71$, $p < .005$, $N^2 = .35$ with a nonsignificant linear contrast $F(1, 24) = .13$, $p = .723$, $N^2 = 0.005$, suggesting that the quadratic change in positive learning rates across sessions (ie, from high to low and then low to high) was consistent with the quadratic change in CO levels across sessions. Post hoc tests indicated that the positive learning rates were reduced during the smoking abstinent session (M = 0.08, SEM = 0.02), relative to the smoking as usual (M = 0.26, SEM = 0.07), $t(24) = 2.3$, $p < .05$, and cigarette consumption (M = 0.22, SEM = 0.06), $t(24) = –2.5$, $p < .01$, session. Supporting this observation, the positive learning rate difference between the smoking abstinence and cigarette consumption session was correlated with the CO level difference between the smoking abstinence and cigarette consumption session, $r = .465$, $p = .01$. No differences were observed between the smoking as usual and cigarette consumption sessions ($p = .72$).

In regards to negative learning rate, a comparable ANOVA (Figure 2C, right side) also revealed a trend for a main effect of smoking state, $F(2, 24) = 3.1$, $p = .07$, $N^2 = 0.10$. However, both the quadratic contrast $F(1, 24) = 3.11$, $p = .09$, $N^2 = 0.12$ and linear contrast $F(1, 24) = 2.3$, $p = .14$, $N^2 = 0.09$, were nonsignificant (Figure 2C, right side). Nevertheless, it is worth noting that the negative learning rate value was enhanced during the smoking abstinence session ($M = 0.54$, SEM = 0.08) relative to the cigarette consumption session ($M = 0.29$, SEM = 0.06), $t(24) = 2.1$, $p < .05$.

Although the main predictions of the computational analysis focused on positive and negative learning rate values, we also evaluated whether individual differences in the exploration and perseveration model parameters could provide insight into smoking-induced modulation of task performance (Figure 2D). In regards to the exploration parameter, a repeated measures ANOVA on ß-parameter values did not reveal any main effects ($p > .05$) (Figure 2D, left side). However, a repeated measures ANOVA on the perseveration parameter revealed a main effect of smoking state, $F(2, 48) = 3.5$, $p < .05$, $N^2 = 0.13$. Post hoc tests indicated that the perseveration values were significantly greater during the smoking abstinent session ($M = 0.35$, SEM = 0.04) than those during the cigarette consumption session ($M = 0.16$, SEM = 0.02), $t(24) = 2.4$, $p < .05$, and were marginally different with respect to the smoking as usual session, ($M = 0.24$, SEM = 0.04), $t(24) = 1.7$, $p = .08$ (Figure 2D, right side). No differences were observed between the smoking as usual and the cigarette consumption session ($p > .05$).

Discussion

We used computational modeling and a decision-making task (PST) to investigate the impact of smoking state (smoking as usual, smoking abstinence, and cigarette consumption) on reinforcement learning signals and choice behavior in smokers. Learning rates for positive RPE signals were significantly higher following cigarette consumption than during smoking abstinence. Consistent with this finding, we found that accuracy and the proportion of win-stay choices during the training phase increased following cigarette consumption. Considerable evidence has shown that phasic bursts in dopamine activity encode positive RPE signals,[32] and that acute nicotine exposure enhances phasic dopamine release through the activation of nicotinic acetylcholine receptors located in the ventral tegmental area and in the striatum.[4,9–11] Evidently, by increasing phasic dopamine release, cigarette consumption increases RPE magnitude. In the Q-learning model, action values are updated by the product of RPE and the learning rate, so our observation that positive learning rates increased following cigarette consumption is consistent with the predictions of the dopaminergic RPE hypothesis.[20,33] By increasing phasic responsiveness of dopamine neurons, nicotine could act to increase the gain on the RPE signal, favoring learning from positive feedback. These findings suggest how future interventions could be tailored for specific computational profiles. For instance, we speculate that excessive RPE signaling related to drug rewards might be alleviated by administration of pharmaceuticals (such as dopamine antagonists) or brain stimulation techniques (such as transcranial magnetic stimulation[19]).

Our results also show that cigarette consumption following an abstinence period restored the positive learning rate to a value indistinguishable from the smoking as usual state, suggesting that smoking restored the positive RPE signals to their normal size. This finding is consistent with the hypothesis that during a period of smoking abstinence, which is associated with symptoms such as anhedonia, craving, and withdrawal, smokers learn through experience that smoking cigarettes reverses the aversive effects of smoking abstinence, thereby facilitating drug seeking behavior to increase dopamine levels.[4,10] Taken together, our results suggest that the positive reinforcement bases of nicotine addiction may be bidirectional. We propose that the augmented positive RPE signals following cigarette consumption may bias the action–selection mechanism to reinforce behaviors that ultimately converge on drug use, while simultaneously providing immediate relief for abstinence deficits such as anhedonia, craving, and withdrawal. These results demonstrate that smoking abstinence and cigarette consumption can alter positive RPE signals in humans and affect reinforcement learning and decision making.

Smoking state also appeared to affect learning from negative RPE signals, although this relationship was not statistically significant (Figure 2C, right side). However, we feel it worth noting that in the early phase of learning (first 60 trials), the learning rate associated with negative feedback was heightened during smoking abstinence relative to cigarette consumption. This finding is consistent with the idea[22] that low levels of striatal dopamine can enhance the ability to learn from negative RPEs by increasing the duration of dopamine dips during negative feedback.[14] If negative RPEs are, in fact, encoded in pauses in dopamine activity, and smoking abstinence significantly affects pause durations and decreases basal dopamine concentration and tonic dopamine release,[9,11] then learning rates for negative RPE should increase during smoking abstinence and decrease during cigarette consumption, as we have found here. This possibility will need to be validated in future studies.

Further, it is interesting to note that the relationship observed between smoking state and learning rates for positive and negative RPE signals dovetail with previous computational findings in people with Parkinson's disease while on and off dopamine medication. In regards to learning rate, studies have revealed that while off medication—a state that is associated with low striatal dopamine levels—people with Parkinson's disease exhibit lower action values when learning from positive feedback and higher action value when learning from negative feedback.[30,34] By contrast, higher positive learning rates were observed in patients with Parkinson's disease on L-DOPA, a medication that increases vesicular dopamine concentration. Consistent with these findings, we have shown that following cigarette consumption and smoking abstinence sessions, smokers exhibit learning patterns that are similar to patients with Parkinson's disease on and off medication, respectively, which supports the idea that the findings reported here may be dopamine related. Furthermore, two recent studies demonstrated that off-medication patients with Parkinson's disease exhibited higher perseveration than on-medication patients, which is also consistent with the results reported here. Nevertheless, whether or not these behavioral modifications reflect alterations by nicotine to the dopamine system,[30,34] the cholinergic system,[35] or other neural systems is unknown. For instance, a recent extension to standard reinforcement learning model suggests that the pause durations of cholinergic interneurons can be dynamically modulated to optimize the trade-off between behavioral flexibility and stability.[35] The compatibility of our results with the results of multiple Parkinson's disease studies provides insight into the impact of changing levels of dopamine on positive and negative reinforcement learning signals and perhaps contributes to an apparent neuroprotective effect of cigarette smoking against Parkinson's disease.[36]

Despite these commonalities across studies, we note that the effects of smoking on PST testing phase performance were at odds with our predictions. Previous work has supported the idea that disruption in positive and negative RPE signaling in the basal ganglia can selectively impair approach and avoidance performance during the testing phase of the PST. For example, people with Parkinson's disease are more accurate on avoid than approach trials of the PST while off medication because of a diminished negative RPE signal, and more accurate on approach than avoid trials while on medication because of an enhanced positive RPE signal, findings that have been confirmed empirically,[37,38] but see Grogan et al.[39] In contrast to such findings, we found that smokers were relatively more accurate at approach performance during the smoking as usual session, relatively more accurate at avoidance performance following the cigarette consumption session, and had no bias during the smoking abstinence session. We suggest a few factors that may have contributed to these discrepancies. First, it has been argued that learning aspects of the task structure in the first task iteration can influence performance of subsequent task iterations,[40] introducing inconsistencies across sessions that contribute to poor test–retest reliability of PST testing-phase performance.[28]

Second, because of the task design used in our initial study,[18] the sessions were not counterbalanced. Future studies should extend this study by counterbalancing across sessions. Nevertheless, it is worth noting that the change in learning phase accuracy and computational results across sessions (ie, from high to low and then low to high) was consistent with an effect of CO levels, which argues against a practice effect. A third explanation may stem from subtle differences between RPEs calculated during the learning phase (with feedback) and testing phase (with no feedback). For instance, the learning rates found by fitting participants' trial-to-trial training choices may reflect rapid adaptation to changing outcomes, but would conceal the learning rates of a learning system more adept at discriminating between subtly different probabilistic reinforcement values at test.[41] Future studies should adopt alternative computational models (eg, dual Q-learning model),[41] as well as alternative approaches to fitting the data (eg, a hierarchical approach)[42] to resolve the ambiguities between the training and testing phases of the PST and across multiple iterations of the task.

Further limitations of this study include its brevity and sample. First, our study consisted of a small sample of male and female smokers in the low to moderate nicotine-dependent range (CO values of 8.2 ppm), who might not be representative of the smoking population in general (eg, greater than 10 ppm), or of severely dependent nicotine users in particular. Although this population provides a reasonable point of departure, future studies should replicate and extend these findings with a wider sample of the general smoking population (eg, a cutoff of 10 ppm[43]), as well as with individuals who abuse other drug types, and with comorbid disorders (eg, depression, impulsivity, and schizophrenia). Future studies should also include measures of nicotine withdrawal to gain a better understanding of RPE signals role in key abstinence symptoms.

Conclusion

In sum, this study illustrates the potential of computational psychiatry for understanding reinforcement learning deficits associated with substance use disorders in general and nicotine addiction in particular. In view of the role of positive and negative RPE signals in decision making, it is possible that the dynamic alteration of RPE signaling during smoking abstinence and cigarette consumption can have seemingly paradoxical roles in addiction. Following drug use, the enhanced positive RPE signals should serve to reinforce and promote drug-related behaviors, whereas disrupted negative RPE signals should interfere with learning about the negative consequences of these behaviors. By contrast, in an abstinent state associated with reduced dopamine levels, a reinforcement mechanism that emphasizes negative relative to positive RPE signals would likely promote increased avoidance learning, and thereby facilitate actions that avoid abstinence symptoms (ie, withdrawal, craving, anhedonia). By highlighting computational differences between two states of addiction, these findings hold out promise for integrating experimental, computational, and theoretical analyses of decision-making function together with research on addiction-related disorders.

Supplementary Material

Supplementary data are available at *Nicotine & Tobacco Research* online.

Funding

This research was supported by Canadian Institutes of Health Research Operating Grant #97750 awarded to C.B.H. The first author was supported by Doctoral Awards from the Integrated Mentor Program in Addictions Research Training (IMPART) and the Canadian Institutes of Health Research #195501.

Declaration of Interests

None declared.

Acknowledgments

We are grateful to the research assistants of the Learning and Cognitive Control Laboratory for help with data collection.

References

1. Rehm J, Taylor B, Room R. Global burden of disease from alcohol, illicit drugs and tobacco. *Drug Alcohol Rev*. 2006;25(6):503–513.
2. Berrendero F, Robledo P, Trigo JM, Martín-García E, Maldonado R. Neurobiological mechanisms involved in nicotine dependence and reward: participation of the endogenous opioid system. *Neurosci Biobehav Rev*. 2010;35(2):220–231.
3. Picciotto MR. Common aspects of the action of nicotine and other drugs of abuse. *Drug Alcohol Depend*. 1998;51(1–2):165–172.
4. Benowitz NL. Pharmacology of nicotine: addiction, smoking-induced disease, and therapeutics. *Annu Rev Pharmacol Toxicol*. 2009;49:57–71.
5. Schultz W. Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioural ecology. *Curr Opin Neurobiol*. 2004;14(2):139–147.
6. Schultz W. Predictive reward signal of dopamine neurons. *J Neurophysiol*. 1998;80(1):1–27.
7. Cohen MX, Frank MJ. Neurocomputational models of basal ganglia function in learning, memory and choice. *Behav Brain Res*. 2009;199(1):141–156.
8. Rice ME, Cragg SJ. Nicotine amplifies reward-related dopamine signals in striatum. *Nat Neurosci*. 2004;7(6):583–584.
9. Grieder TE, George O, Tan H, et al. Phasic D1 and tonic D2 dopamine receptor signaling double dissociate the motivational effects of acute nicotine and chronic nicotine withdrawal. *Proc Natl Acad Sci U S A*. 2012;109(8):3101–3106.

10. Le Foll B, Goldberg SR. Effects of nicotine in experimental animals and humans: An update on addictive properties. *Handb Exp Pharmacol*. 2009(192):335–367.
11. Zhang L, Dong Y, Doyon WM, Dani JA. Withdrawal from chronic nicotine exposure alters dopamine signaling dynamics in the nucleus accumbens. *Biol Psychiatry*. 2012;71(3):184–191.
12. Redish AD. Addiction as a computational process gone awry. *Science*. 2004;306(5703):1944–1947.
13. García-García I, Zeighami Y, Dagher A. Reward prediction errors in drug addiction and Parkinson's disease: from neurophysiology to neuroimaging. *Curr Neurol Neurosci Rep*. 2017;17(6):46.
14. Baker TE, Stockwell T, Barnes G, Holroyd CB. Individual differences in substance dependence: at the intersection of brain, behaviour and cognition. *Addict Biol*. 2011;16(3):458–466.
15. Potts GF, Bloom EL, Evans DE, Drobes DJ. Neural reward and punishment sensitivity in cigarette smokers. *Drug Alcohol Depend*. 2014;144:245–253.
16. Luijten M, van Meel CS, Franken IH. Diminished error processing in smokers during smoking cue exposure. *Pharmacol Biochem Behav*. 2011;97(3):514–520.
17. Schlienz NJ, Hawk LW Jr, Rosch KS. The effects of acute abstinence from smoking and performance-based rewards on performance monitoring. *Psychopharmacology (Berl)*. 2013;229(4):701–711.
18. Baker TE, Wood JMA, Holroyd CB. Atypical valuation of monetary and cigarette rewards in substance dependent smokers. *Clin Neurophysiol*. 2016;127(2):1358–1365.
19. Baker TE, Lesperance P, Tucholka A, et al. Reversing the atypical valuation of drug and nondrug rewards in smokers using multimodal neuroimaging. *Biol Psychiatry*. 2017;82(11):819–827.
20. Keiflin R, Janak PH. Dopamine prediction errors in reward learning and addiction: from theory to neural circuitry. *Neuron*. 2015;88(2):247–263.
21. Maia TV, Frank MJ. From reinforcement learning models to psychiatric and neurological disorders. *Nat Neurosci*. 2011;14(2):154–162.
22. Frank MJ, Seeberger LC, O'reilly RC. By carrot or by stick: cognitive reinforcement learning in Parkinsonism. *Science*. 2004;306(5703):1940–1943.
23. Cohen MX. Neurocomputational mechanisms of reinforcement-guided learning in humans: a review. *Cogn Affect Behav Neurosci*. 2008;8(2):113–125.
24. Sutton RS, Barto AG. Reinforcement learning: an introduction. *IEEE Trans Neural Netw*. 1998;9(5):1054.
25. Humeniuk R, Ali R, Babor TF, et al. Validation of the alcohol, smoking and substance involvement screening test (ASSIST). *Addiction*. 2008;103(6):1039–1047.
26. Baker TE, Stockwell T, Barnes G, Haesevoets R, Holroyd CB. Reward sensitivity of ACC as an intermediate phenotype between DRD4-521T and substance misuse. *J Cogn Neurosci*. 2016;28(3):460–471.
27. Heishman SJ, Singleton EG, Moolchan ET. Tobacco craving questionnaire: reliability and validity of a new multifactorial instrument. *Nicotine Tob Res*. 2003;5(5):645–654.
28. Baker TE, Stockwell T, Holroyd CB. Constraints on decision making: implications from genetics, personality, and addiction. *Cogn Affect Behav Neurosci*. 2013;13(3):417–436.
29. Javors MA, Hatch JP, Lamb RJ. Cut-off levels for breath carbon monoxide as a marker for cigarette smoking. *Addiction*. 2005;100(2):159–167.
30. Piray P, Zeighami Y, Bahrami F, Eissa AM, Hewedi DH, Moustafa AA. Impulse control disorders in Parkinson's disease are associated with dysfunction in stimulus valuation but not action valuation. *J Neurosci*. 2014;34(23):7814–7824.
31. Isaac PF, Rand MJ. Cigarette smoking and plasma levels of nicotine. *Nature*. 1972;236(5345):308–310.
32. Schultz W. Potential vulnerabilities of neuronal reward, risk, and decision mechanisms to addictive drugs. *Neuron*. 2011;69(4):603–617.
33. Cockburn J, Holroyd CB. Focus on the positive: computational simulations implicate asymmetrical reward prediction error signals in childhood attention-deficit/hyperactivity disorder. *Brain Res*. 2010;1365:18–34.
34. Rutledge RB, Lazzaro SC, Lau B, Myers CE, Gluck MA, Glimcher PW. Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. *J Neurosci*. 2009;29(48):15104–15114.
35. Franklin NT, Frank MJ. A cholinergic feedback circuit to regulate striatal population uncertainty and optimize reinforcement learning. *Elife*. 2015;4:1–29.
36. Ross GW, Petrovitch H. Current evidence for neuroprotective effects of nicotine and caffeine against Parkinson's disease. *Drugs Aging*. 2001;18(11):797–806.
37. Frank MJ. Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. *Neural Netw*. 2006;19(8):1120–1136.
38. Frank MJ, Samanta J, Moustafa AA, Sherman SJ. Hold your horses: impulsivity, deep brain stimulation, and medication in Parkinsonism. *Science*. 2007;318(5854):1309–1312.
39. Grogan JP, Tsivos D, Smith L, et al. Effects of dopamine on reinforcement learning and consolidation in Parkinson's disease. *Elife*. 2017;6:1–23.
40. Doll BB, Hutchison KE, Frank MJ. Dopaminergic genes predict individual differences in susceptibility to confirmation bias. *J Neurosci*. 2011;31(16):6188–6198.
41. Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc Natl Acad Sci U S A*. 2007;104(41):16311–16316.
42. Ahn WY, Krawitz A, Kim W, Busmeyer JR, Brown JW. A model-based fMRI analysis with hierarchical Bayesian parameter estimation. *J Neurosci Psychol Econ*. 2011;4(2):95–110.
43. West R, Hajek P, Stead L, Stapleton J. Outcome criteria in smoking cessation trials: proposal for a common standard. *Addiction*. 2005;100(3):299–303.