Supplementary Online Content

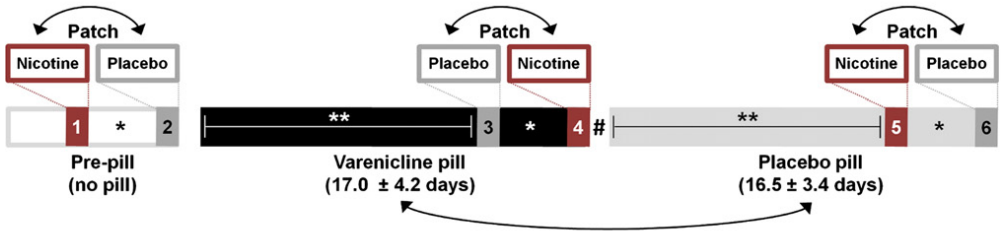
Lesage E, Aronson SE, Sutherland MT, Ross TJ, Salmeron BJ, Stein EA. Neural signatures of cognitive flexibility and reward sensitivity following nicotinic receptor stimulation in dependent smokers: a randomized trial. *JAMA Psychiatry.* Published online April 12, 2017. doi:10.1001/jamapsychiatry.2017.0400

This supplementary material has been provided by the authors to give readers additional information about their work.

1 eAppendix 1. Methods

1.1 Study design

This study was a randomized, double-blind, placebo-controlled, crossover design involving two drugs: varenicline pills and nicotine patches. Subjects completed a total of 6 fMRI sessions: 2 baseline sessions (pre-pill), and 2 sessions each after a varenicline or placebo pill regimen. In each of the 3 varenicline conditions, the two sessions were randomized to either nicotine or placebo patch. Randomization was maintained by the study physician, while researchers, technicians, and participants remained blinded. This is in accordance with the protocol described in previous publications[1–4].



eFigure 1. Study design. The current study reports data from the four completely counterbalanced sessions (sessions 3-6). Varenicline and placebo pill sessions were separated by more than two weeks (**). Nicotine and placebo patch scans were separated by an average of 2.9 ± 1.7 days (*). No washout interval separated medication periods (#). Double-headed arrows explain indicate randomized and counterbalanced order of the sessions. Figure reproduced with permission from Sutherland et al.[3] See also eAppendix 1: Study design.

1.2 Drug regimen

Varenicline and placebo pills were administered in accordance with standard guidelines (http://www.pfizer.com/products) over a ~17-day schedule. The varenicline regimen began at an 0.5 mg once-daily dosage for days 1-3, stepped up to 0.5 mg twice daily at days 4-7, and remained at 1 mg twice daily for days 8-17. Active and placebo medication appeared identical and were distributed in blister packs. Scanning sessions occurred at the end of each regimen (varenicline 17.0 ± 4.2 days; placebo pill 16.5 ± 3.4 days). No washout interval separated medication periods. For those participants whose placebo regimen followed the varenicline regimen, carryover effects were assumed negligible given the ~24-hour elimination half-life of varenicline[5] and the fact that placebo varenicline scanning sessions and active varenicline scanning sessions were separated by more than 2 weeks.

For each of the pill conditions, nicotine or placebo patches were applied to the upper back at the start of fMRI visits (separated by 2.9 ± 1.7 days). Nonsmokers received a 7mg nicotine patch dose, while smokers received a dose that matched daily nicotine intake (21, 28, 35, or 42 mg patches for 10-15, 16-20, 21-25, and >25 cigarettes/day, respectively). The patch was worn for the duration of the 9-hour visit, which consisted of 2 MRI scans.

The PRL task was completed in the second of the 2 MRI scans and began approximately 6-7 hours after initial patch application. It was assumed that data collected within the 2-9 hour post-patch window was associated with steady plasma nicotine levels, in accordance with pharmacokinetic data.[6]

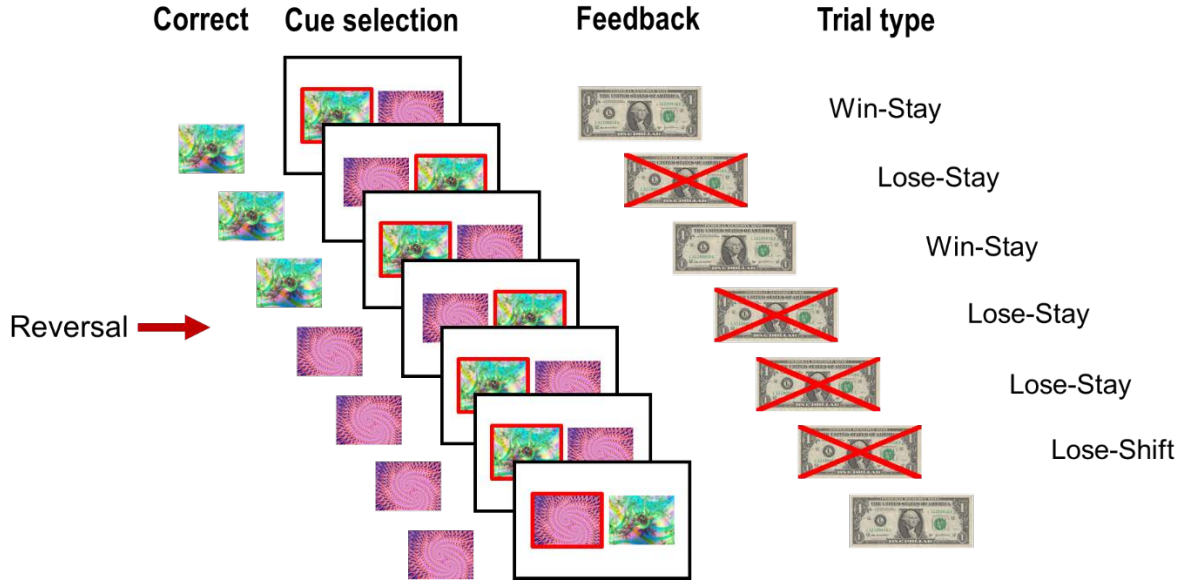
Before each scanning session, subjects were asked to abstain from alcohol for 24 hours, and to moderate caffeine intake for 12 hours. Smokers were required to abstain from cigarette use for 12 hours before and also during scanning days, but were not otherwise restricted on smoking behavior during the 6-8 week course of the entire study. At the start of each session, all participants underwent testing for recent drug and alcohol use, and expired carbon monoxide (CO) levels. CO levels of less than or equal to 15 parts per million (ppm) indicated abstinence. Medication dose was confirmed by the participant on the morning of the scanning session. Medication side effects and adherence were monitored by regular telephone assessments and at in-person visits.

1.3 Probabilistic reversal learning task

Task structure: The probabilistic reversal learning task was based on that used by Cools et al.[7] Participants were presented with two easily-distinguishable, abstract color patterns, and were required to select the correct one (eFigure 2). One pattern probabilistically led to a monetary reward (75% chance of a reward, 25% chance of a punishment) while the other had the reverse contingencies. The goal of the task was for participants to earn as much money as possible by making the “correct” choice through trial-and-error. The contingencies reversed when the participant had selected the correct pattern 5 times consecutively, or when 20 trials elapsed since the last reversal (i.e., a failed reversal). Participants thus had to infer contingency switches based on the probabilistic outcomes.

Monetary reward: There was a monetary incentive for PRL task performance. Participants were paid 50% of the amount “earned” in the task (see eTable 2: Overall Score), typically in the $50-$80 range (mean=$66.37, SD=$20.67).

Trials and timing: The task consisted of three runs, each consisting of 120 trials. The total task took ~35 minutes. Cues were presented for 1500ms, feedback for 1000ms. Trial onsets were temporally jittered (1500-5500ms) to allow independent estimation of the BOLD response for each trial.



eFigure 2. Probabilistic reversal learning task and trial types. Each line represents a trial. The participant selects (red frame) one of the two cues presented on the screen and immediately receives feedback indicating win (reward) or loss (punishment). Contingencies (left-most image indicates reward-predicting cue) reverse at regular intervals during the task. Trials are categorized based on the outcome (win or lose) and on whether the participant chooses the same or a different cue the next trial (stay or shift). See also eAppendix 1: Probabilistic reversal learning task.

Training and instructions: Participants received explicit instruction and thorough training on the PRL task. During a separate orientation session, participants were trained in multiple phases to maximize their understanding of the task structure. Participants were first instructed that one of two images would lead to a monetary reward, that these contingencies would reverse, and that the participant should learn which image is rewarding through trial and error. Participants then practiced this deterministic version of the reversal learning task. Second, after this basic structure of the task was well-understood and participants learned how to maximize monetary gain by selecting the correct image and shifting their selections when appropriate, the probabilistic element was introduced. Participants were explicitly told that the computer would sometimes give the wrong feedback to make it harder for the participant to identify the correct image and earn money. Participants then performed a training block without reversals but with

probabilistic feedback, to expose them to the probabilistic feedback aspect. Third, all elements of the PRL were recapped by talking the participant through a diagram that depicted a series of 15 hypothetical trials and responses. This series of trials illustrated the probabilistic nature of the task and the reversals further. Participants were instructed that "the hardest part of this task is deciding if you should switch your response or stay with the same picture. In other words, the hardest part of the task is determining if the rule has switched or if the computer is lying to you. Just because the computer told you that you won a dollar does not mean that the image you chose is the correct one. And just because you lost a dollar does not necessarily mean that you chose the wrong image. Remember that you should only switch your response when you are sure that the rule has changed." Finally, participants performed a 120-trial practice block of the PRL task in the mock scanner.

1.4 Computational modeling

Our modeling approach was based on Hampton et al,[8] who found that a Hidden Markov Model (HMM), which incorporates model-bases knowledge of the task structure (namely the anti-correlated nature of the values of the two cues) fits choice data in a PRL task better than a simple Rescorla-Wagner (RW) model. The RW model relies solely on experience and therefore learns only about the cue that is encountered (model-free learning), whereas the HMM incorporates the task structure and therefore constitutes model-based learning. Since then, other researchers have also found that computational models that incorporate the underlying task structure explain choice behavior in PRL better than simple RW models.[9,10] We fit three models: one Rescorla-Wagner (RW) model and two HMMs.

1.4.1 Rescorla-Wagner model

In a RW model, participants update the value of cues (A and B) through experience on each trial (t). The value associated with the selected cue (V_A) is updated proportionally to the difference between the predicted outcome (o) and the actual outcome (prediction error δ), weighted by the learning rate η. Higher learning rates and larger prediction errors lead to larger updates in the values associated with each cue. In this model, the participant updates only the value of the selected cue; in a trial where A is selected, the participant learns nothing about B.

$$V_A\,(t+1) = V_A(t) + \eta\delta(t) \quad (1)$$

$$\delta(t) = o(t) - V_A(t) \quad (2)$$

On the basis of the difference between these values V_A and V_B, the probability of choosing the cue A is described by Luce's rule[11]:

$$P(A) = \frac{1}{1+ e^{-\beta((V_a - V_b)-\alpha)}} \quad (3)$$

Here, the probability of selecting cue A depends on the value difference between the two cues, the bias α towards one value, and the inverse temperature β. In the RW model, positive values for α indicate a bias toward selecting B, as a larger difference between V_A and V_B is required to increase the chance of selecting cue A. Conversely, a negative α implies a bias toward selecting cue A. The inverse temperature β can be interpreted as the sensitivity to the available evidence. A small β pulls the probability of selecting a particular cue toward 0.5 - relying less on the evidence (the V_A minus V_B term), while a large β amplifies the impact of the difference in value between the two cues. For example, if $V_A > V_B$ and α is 0, then a high β would make it extremely likely that cue A would be chosen, while a low β pulls this probability toward indifference. In fitting this model to the data, we estimated the learning rate η, α and β. The probability of selecting cue B is the complement of the probability of selecting cue A.

1.4.2 Hidden Markov Models

The other two models were HMMs whereby the model of the task is considered. In a PRL task, there are two states (correct: having selected the correct cue; or incorrect: having selected the incorrect cue), each of which is associated with a certain outcome Y (a reward or a punishment; e.g. a 75% probability of a reward in state $X_{correct}$).

At any time point t, participants must infer a hidden state (X_t = correct or X_t = incorrect) from imperfect (probabilistic) outcomes (Y; Y_t = reward or Y_t = punishment). The probability of being correct is initially set at 0.5 and updated each trial. Each trial participants make a choice S (S_t = stay or S_t = shift). Reward contingencies reverse with transition probability δ (δ therefore has a different meaning than in the RW model). That is, if $X_{t=1}$ = correct, and the participant shifts their response, then $X_{t=2}$ is correct with probability δ and $X_{t=2}$ is incorrect with probability 1-δ. If $X_{t=1}$ = correct, and the participant does not shift, then $X_{t=2}$ is correct with probability 1-δ and $X_{t=2}$ is incorrect with probability δ. A high δ means reversals are likely, while a low δ means reversals are unlikely.

$$Prior(X_t = correct) = \sum_{X_{t-1}} P(X_t = correct \,|X_{t-1}, S_t) * Posterior(X_{t-1}) \quad (4)$$

The probability of being correct (probability that X_t = correct) given the evidence Y_t is calculated per Bayes' rule

$$Posterior(X_t = correct) = \frac{P(Y_t|X_t=correct)*Prior(X_t=correct)}{\sum_{X_t\ states} P(Y_t|X_t)*Prior(X_t)} \quad (5)$$

The probability of being incorrect state (posterior probability that X_t = incorrect; $P_{incorrect}$) informs the probability of a shift, which can be calculated per Luce's choice rule:

$$P(shift) = \frac{1}{1+ e^{-\beta(P_{incorrect}-\alpha)}} \quad (6)$$

Two nested HMMs were fitted. The first HMM (HMM1) had three free parameters: δ (the probability of a reversal), α (the bias toward shifting (<0.5) or staying (>0.5)), and β, the reliance on the available evidence.

The third model, HMM2, incorporated a fourth free parameter that captures the extent to which the time since the last switch increases the probability of a switch. That is, as the number of trials increases, a reversal becomes more likely, regardless of the valence of the outcomes. Per HMM1, one should become increasingly confident that one is in the correct state, and increasingly reluctant to shift as the number of trials since the last reversal grows and outcomes are positive. However, the task structure is such that, in fact, the longer it has been since the last reversal, the more likely a new reversal becomes. HMM2 includes a free parameter that weights to what extent the time since the last reversal increases the probability of a reversal.

For HMM2, this is implemented into the calculation of the posterior probability of being in the correct state or the incorrect state (with Posterior(X_t=incorrect) = 1- Posterior(X_t=correct)):

$$Posterior(X_t = correct) = \frac{P(Y_t|X_t=correct)*Prior(X_t=correct)}{\sum_{X_t\ states} P(Y_t|X_t)*Prior(X_t)} * e^{-lag_t*delay-weight} \quad (7)$$

Where the fixed parameter *lag* has value 0 for the first four trials following a reversal (where a reversal is impossible, even if the participant selects the correct cue on each trial), and then increases linearly. Non-zero values for the free parameter delay-weight mean that the longer one gets away from the previous reversal, the more likely one is in the incorrect state. Therefore, delay-weight signals that a participant factors in the structure of the task, whereby switches occur with some regularity.

1.4.3 Model comparison

Following Hampton et al,[8] we calculated a log likelihood for each of the models as follows:

$$\log L = \frac{\sum B_{switch} * \log P_{switch}}{N_{switch}} + \frac{\sum B_{stay} * \log P_{stay}}{N_{stay}} \quad (8)$$

Because there are far more stay trials than switch trials, the likelihood is estimated for stays and switches separately and divided by the number of stays and switches. To account for the number of free parameters in comparing model fits (three in RW and HMM1, four in HMM2), we calculated the Bayesian Information Criterion (BIC), which accounts for this increase in free parameters by "punishing" models with more free parameters.

$$BIC = -2 \log L + M * \frac{Log\ N}{N} \quad (9)$$

Here M is the number of free parameters and N is the number of observations. Lower BIC values indicate a better fitting model. Models were fitted to the choice data and free parameters were estimated with a simulated annealing method in R (https://www.r-project.org/). Model fits were compared by submitting the BIC to paired t-tests.

1.5 MRI acquisition

MRI images were acquired on a Siemens 3T Magnetom Allegra scanner (Erlangen, Germany). To capture the blood oxygenation level-dependent (BOLD) response during the task, we acquired T2*-weighted, single-shot gradient-echo, echo-planar imaging (EPI) scans (1104 volumes over three 368-volume runs; 39 axial slices, slice thickness 4mm, oblique-axial slices (30° to AC-PC), TR=2000ms; TE=27ms; flip angle =80°; field of view=220×220mm; image matrix=64×64). In addition, a high-resolution T1-weighted structural scan was acquired (MPRAGE, TR=2500ms; TE=4.38ms; FA=8°; voxel size=1 mm3, oblique-axial).

1.6 MRI analysis

1.6.1 Preprocessing and first-level modeling

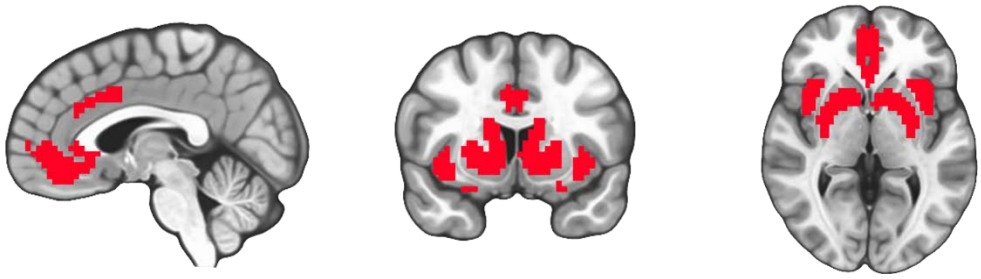
Preprocessing: Functional images were slice-time corrected, registered to the anatomy, motion-corrected, spatially registered to the Talairach template and smoothed to 8mm at FWHM.

First-level general linear model: Event types of interest were modeled according to their outcome and the behavioral response on the next trial, as in previous work[12] (win-stay, lose-stay, and lose-shift) . Win-shift trials were not modeled separately due to their very small number (mean occurrence = 10.0, SD=11.2 out of 360 trials). Regressors were convolved with a haemodynamic response function (HRF) and its temporal derivative. Six regressors captured head motion. 5th order polynomial regressors were included to capture baseline trends.

First-level contrasts of interest: Conditions were estimated against the implicit baseline and two contrasts of interest were calculated to capture reward-sensitivity and flexibility. The reward-sensitivity contrast (win-stay minus lose-stay) was calculated by subtracting lose-stay trials from win-stay trials; cognitive flexibility (shift minus stay) was calculated by subtracting lose-shift trials from lose-stay trials. This approach was taken to ensure that the reward-sensitivity contrast was not confounded by differences in the behavioral response and that the flexibility contrast was not confounded by differences in valence. Thus, in the reward-sensitivity contrast the behavioral response was kept constant (stay only), while in the flexibility contrast the outcome was kept constant (lose only).

1.6.2 Regions of interest and correction for multiple comparisons

We used a region of interest approach based on brain areas that are implicated in addiction and nicotine withdrawal, and in PRL[13]. In addition to classic MCL regions (orbitofrontal cortex, ventral and dorsal striatum), we also included the anterior cingulate cortex and the anterior insular cortices, as these are densely packed with $\alpha 4\beta 2$ receptors, as well as the bilateral amygdala, due to its pivotal role in withdrawal [1,2,7,14–16]. Anatomical masks of the nucleus accumbens, caudate nucleus, putamen, amygdala, orbitofrontal cortex, anterior cingulate cortex, and anterior insula were derived from Desai maximum probability maps.[17] All 7 masks were summed, and a 3dClustSim algorithm (AFNI) was used to estimate the appropriate correction level to arrive at family-wise error (FWE) correction level of $\alpha<0.05$ across the composite mask. Within the small volume (1,978 voxels), we corrected at a voxel-wise level of $p<0.05$, with a cluster size of 53 voxels.



eFigure 3: composite mask of the regions of interest. Imaging results were FWE corrected within these regions. Talairach coordinates X=2, Y=10, Z=0. Left is displayed on the right.

For the whole-brain task maps, we constructed a 90% probability masks of the functional data (44,413 voxels). Using a 3dClustSim algorithm, we arrived at a voxel-wise threshold of $p<0.001$ that was combined with a minimum cluster size of 19 voxels (FWE $\alpha<0.05$).

2 eAppendix 2. Results

2.1 Demographics

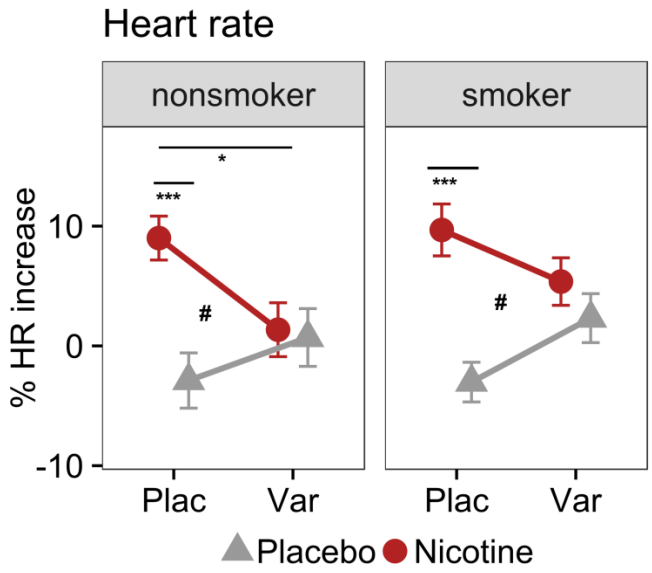
Groups were matched on gender and ethnicity. Smokers were moderately nicotine dependent (Fagerström scores: 5±2) smoked 18±8 cigarettes/day, and reported daily cigarette use for 18±11 years. Smokers were significantly older than controls (eTable 1), and age was therefore included as a covariate in all between-group analyses. Data from one male nonsmoker was excluded from all analyses due to poor behavioral performance (see below for description and quantification) and excessive head motion.

eTable 1: Demographics.

	smokers (N=24)	nonsmokers (N=20)	group differences
Gender (F/M)	12 / 12	10 / 10	t(42) = 0, p=1
Age (mean +/- SD)	35.8 +/- 9.9	30.4 +/- 7.2	t(41.327) = 2.12, p=0.040
IQ (mean +/- SD)	105.95 +/-12.37	113.37 +/- 11.57	t(37.0) = -1.93, p = 0.061
race (AA/C/A/more than 1)	6 / 14 / 3 / 1	8 / 8 / 3 /1	F(3,40) = 0.49, p=0.6886
Fagerstrom index	5.00 +/- 1.9	n/a	n/a
Years daily smoking	18.0 +/- 10.6	n/a	n/a
Cigarettes per day	17.7 +/- 7.9	n/a	n/a

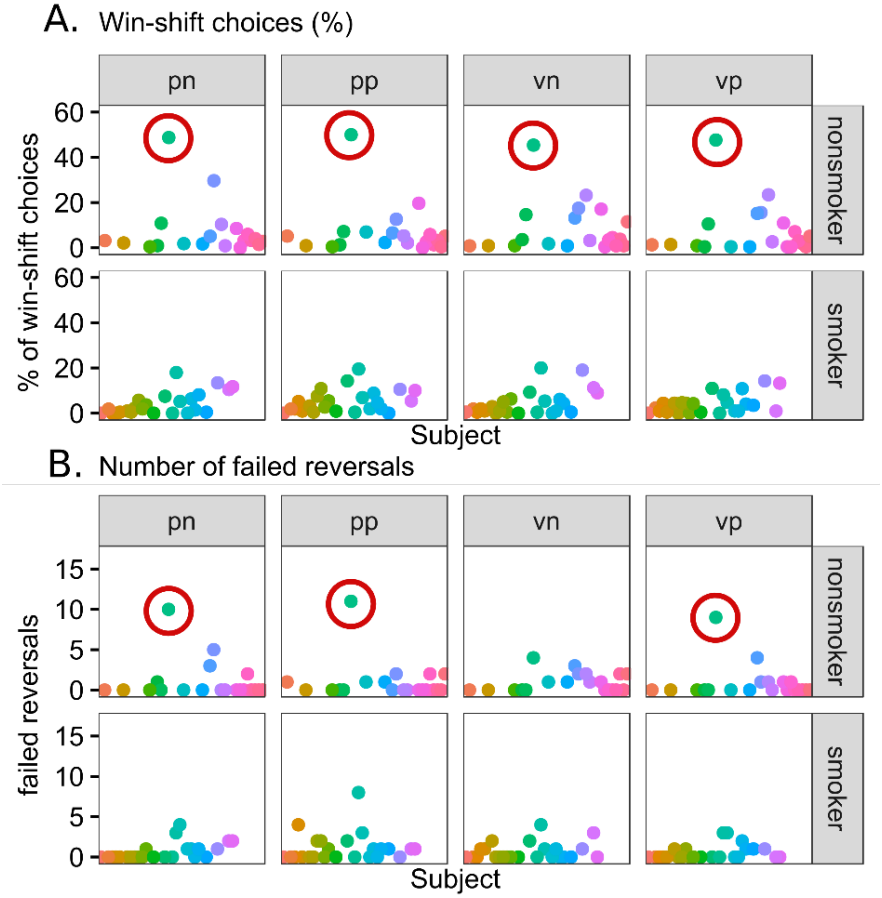
Note. AA: African American, C: Caucasian, A: Asian. * p<0.05.

2.2 QA: effectiveness of nicotinic receptor stimulation



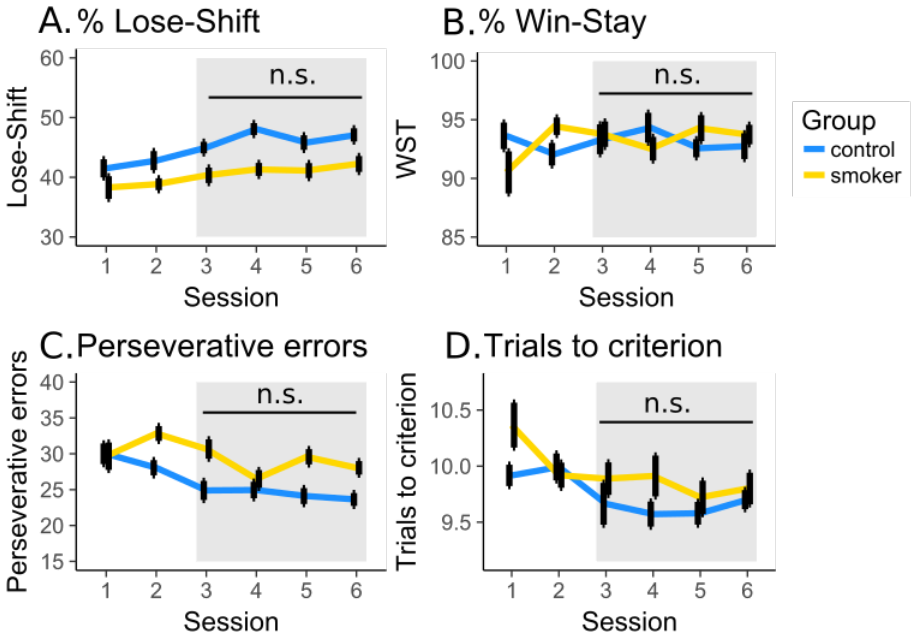
eFigure 4. QA: Effectiveness of nicotinic receptor stimulation. Nicotine and varenicline similarly modulated smokers' and nonsmokers' heart rates. Examination of the heart rate modulation after drug administration indicates that the drugs and doses employed in the two groups were physiologically significant, and similarly effective in both groups. Nonsmokers (NICOTINE-by-VARENICLINE interaction: χ_1^2=6.45, p=0.013) and smokers (NICOTINE-by-VARENICLINE interaction: χ_1^2=5.99, p=0.019) displayed similar cardiovascular responses to pharmacological manipulations (NICOTINE-by-VARENICLE-by-GROUP interaction χ_1^2=0.076, p=0.78)[3]. Nicotine-induced HR increases were observed under placebo pill conditions (NICOTINE effect in smokers: χ_1^2=21.67, p<0.001; NICOTINE effect in nonsmokers: χ_1^2=15.23, p<0.001), and were diminished or absent under varenicline pill conditions (no NICOTINE in either group; p's > 0.26). # p<0.05 interaction effect; * p<0.05 pairwise comparison; *** p<0.001 pairwise comparison.

2.3 QA: task engagement



eFigure 5. QA: Assessment for task engagement. We examined two metrics (win-shifts and failed reversals) to identify individuals or sessions where participants may not have been appropriately engaged in the task or may not have been 'paying attention'. A) Win-shift behavior. If a participant was not engaged with the task, we expected them to make erratic or random choices, even in "easy" task conditions. This would be reflected in a high proportion of win-shift choices (with chance being 50%, and our definition of a cut-off). B) Failed reversals. The task was constructed such that if a participant did not learn the reversal (select the correct stimulus 5 times consecutively) within 20 trials, the rule changed. If a participant was not engaged in the task, one would expect many failed reversals. The average number of reversals was 36.3 ±4.1), and we used 25% of this (9 failed reversals) as a cut-off. One participant (one male control, data point circled) performed considerably worse (responding at chance level after a win, over 9 failed reversals in 3 of the sessions), was deemed not to have engaged with the task and was therefore removed from behavioral and imaging analyses. PN: placebo-nicotine, PP: placebo-placebo, VN: varenicline-nicotine, VP: varenicline-placebo.

2.4 QA: absence of learning effects



eFigure 6. QA: Absence of learning effects. No order effects were observed across the sessions included in the behavioral and imaging analysis. It is possible that complex tasks such as PRL can show learning effects across sessions. However, most of the learning is likely to occur during the training session (see eMethods) and first two sessions. Indeed, any behavioral improvements appeared to level off over the first sessions, and behavior did not demonstrate significant performance differences within sessions 3 to 6. In addition, the order of the four pharmacological manipulations (nicotine versus placebo, varenicline versus placebo) was counterbalanced (see eFigure 1), indicating that session order was unlikely to confound pharmacological effects. Grey area indicates sessions that were included in this report.

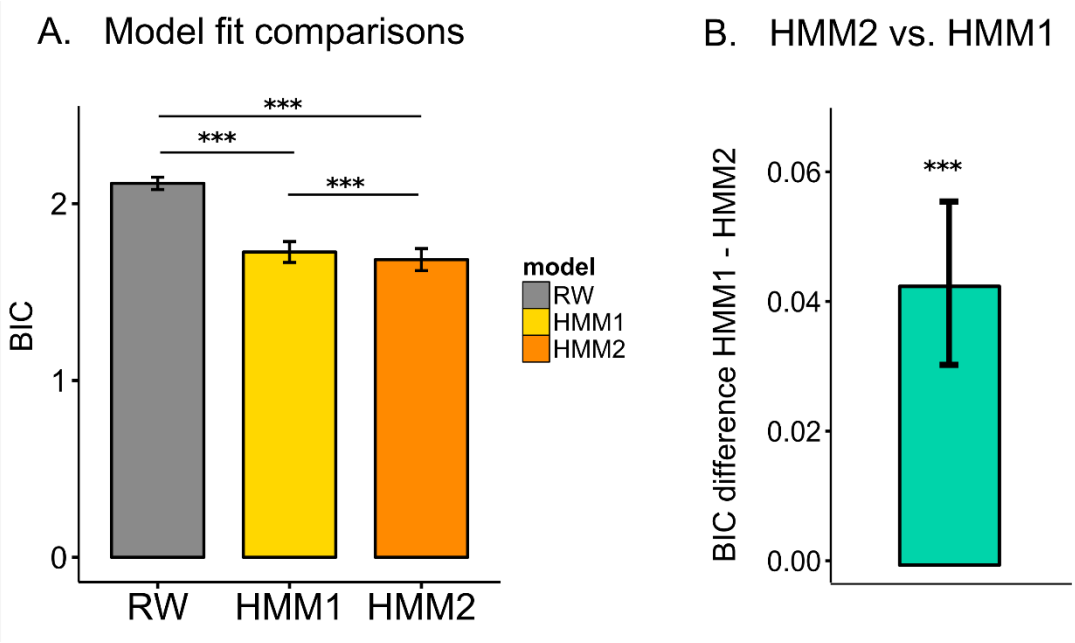
2.5 Behavioral results

eTable 2. Behavioral results.

		Overall Score		Lose-Shift (%)		Win-Stay (%)		Trials to Crit.		Persev. err.		RT (ms)	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Smokers	Plac-Plac	124.38	6.12	44.83	0.92	94.43	0.57	10.15	0.21	25.13	1.20	656.64	9.79
	Plac-Nic	122.33	6.12	41.67	1.31	95.75	0.44	9.86	0.11	30.33	1.11	629.36	9.25
	Var-Plac	120.46	5.11	41.10	1.31	95.83	0.66	9.68	0.12	29.63	0.99	654.10	9.43
	Var-Nic	126.17	6.06	41.01	0.96	95.46	0.45	9.63	0.12	29.71	0.90	648.76	8.86
	Average	123.34	5.85	42.15	1.13	95.37	0.53	9.83	0.14	28.70	1.05	647.22	9.33
Nonsmokers	Plac-Plac	142.42	7.64	47.04	0.70	95.15	0.93	9.66	0.17	24.21	1.04	609.32	8.94
	Plac-Nic	145.21	6.42	46.82	1.14	94.89	0.93	9.54	0.12	25.32	1.30	605.04	9.71
	Var-Plac	143.89	5.78	48.12	1.06	94.56	0.73	9.59	0.12	24.63	1.23	626.39	11.47
	Var-Nic	146.89	5.50	46.72	0.87	93.41	0.67	9.73	0.08	23.42	0.57	613.72	12.70
	Average	144.60	6.34	47.18	0.94	94.50	0.82	9.63	0.12	24.40	1.04	613.62	10.71
	Grand mean	133.97	6.09	44.66	1.03	94.94	0.67	9.73	0.13	26.55	1.04	630.42	10.02

Note. Overall score: number of wins – number of losses. Lose-Shift: percentage of shift responses after a loss. Win-Stay: percentage of stay responses after a win. Trials to Crit.: average number of trials until the correct stimulus is selected 5 times consecutively. Persev. err.: number of perseverative errors; selections of a previously rewarded cue at least three trials (including two losses) after a reversal. RT: reaction time.

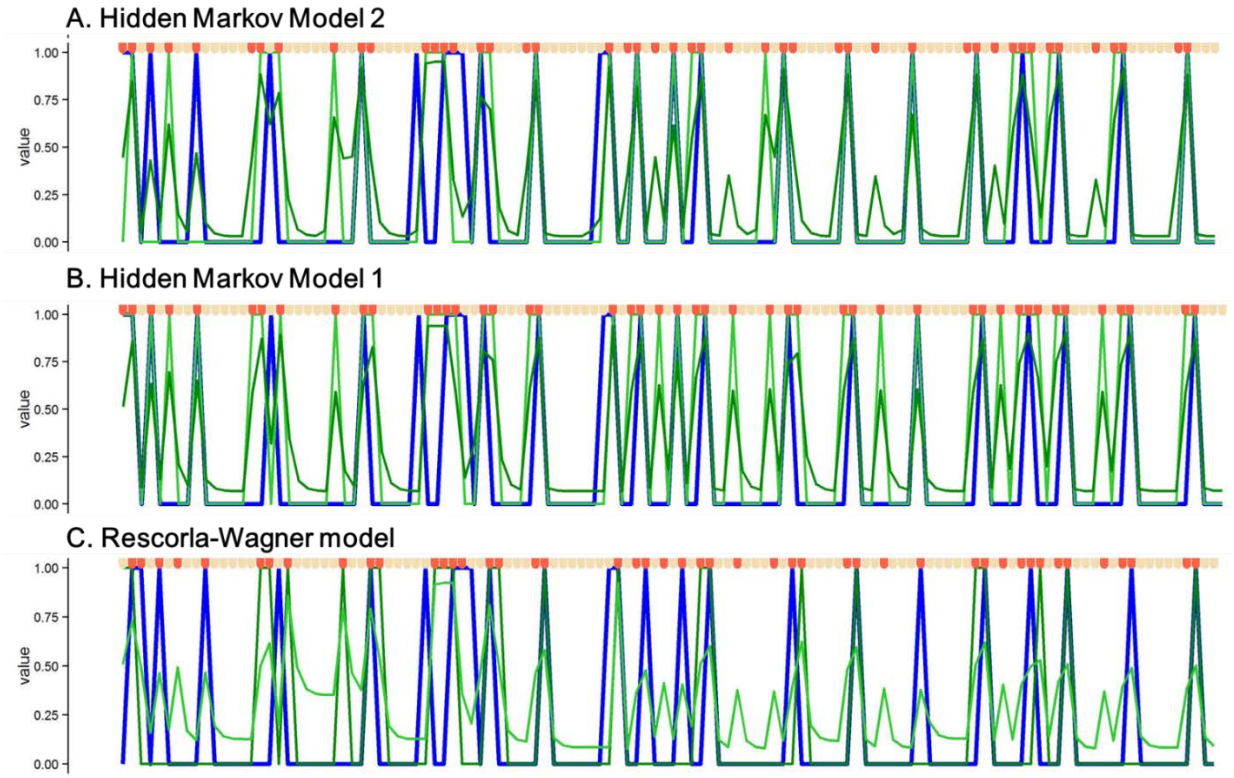
2.6 Computational modeling results



eFigure 7. Computational modeling results: Model Comparisons. A. Bayesian information criterions for the three models. B. Average difference in BIC for HMM1 and HMM2. Error bars denote standard error of the mean. *** p<0.001 in paired t-tests. See also eAppendix 2: Computational modeling results.

Both HMM1 (Paired t-test of BIC RW vs. HMM1: t_{168}=25.69, p<0.001) and HMM2 (Paired t-test of BIC RW vs. HMM2: t_{168}=25.55, p<0.001) performed better than the Rescorla-Wagner model (eFigure 7A). HMM2, which included a fourth free parameter which indexed how much the chance of shifting was affected by the time since the last reversal, resulted in slightly better model fits than HMM1 (eFigure 7B; Paired t-test of BIC HMM1 minus BIC HMM2: t_{168}=6.74, p<0.001). Therefore, we based our inference on the parameter estimates using HMM2. eFigure 8 shows model fit data for a representative subject. Parameter estimates for each model can be found in eTables 3-5.

Out of 172 sessions fitted to the HMM2, the model generated erratic parameter estimates for only three sessions (three different conditions: one on nicotine only, one on nicotine and varenicline, and one on neither; placebo-placebo). These data, from three different smoker participants, were excluded from further computational analyses assessing GROUP, NICOTINE, and VARENICLINE effects.



eFigure 8. Computational modeling results: Example model fit. Fits for the three models considered are shown using one run (120 trials) of a representative participant's data. Ribbon above indicates the outcome on each trial (beige is reward, red is punishment), the bold blue line is the participant's behavior (1 is a shift, 0 is stay). The light green is the probability of a switch as estimated by the model. The dark green is a binary version of this switch probability (1 if the $P_{switch} > 0.5$, 0 if $P_{switch} < 0.5$) and represents the "guess" the model makes. As is quantified in eFigure 7, the HMM1 and HMM2 fit the behavioral data better than RW. HMM2, in turn, provides a better fit to the choice data than HMM1, notably after longer periods of positive reinforcement, the probability of a switch increases. HMM1, which does not factor in delay weight, overweights the importance of negative outcomes to optimize for. Under HMM2 it is recognized that negative outcomes right after a reversal should incentivize a shift less than negative outcomes long after a reversal). Negative outcomes are therefore processed more moderately, resulting in a better fit to the participant's choice data. See also eAppendix 2: Computational modeling results.

eTable 3. Parameters estimates and goodness of fit measure for HMM2.

HMM2		Alpha		log(Beta)		Delta		Delay Weight		NLL		BIC	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Smokers	PP	0.46	0.02	0.77	0.024	0.19	0.01	0.15	0.02	0.85	0.021	1.78	0.04
	PN	0.53	0.01	0.91	0.032	0.22	0.02	0.20	0.02	0.77	0.029	1.61	0.06
	VP	0.52	0.01	0.88	0.037	0.18	0.01	0.13	0.02	0.78	0.034	1.63	0.07
	VN	0.50	0.01	0.82	0.024	0.17	0.01	0.17	0.01	0.81	0.025	1.68	0.05
Nonsmokers	PP	0.49	0.01	0.84	0.020	0.20	0.02	0.18	0.02	0.83	0.022	1.72	0.04
	PN	0.49	0.01	0.88	0.020	0.18	0.009	0.19	0.01	0.79	0.020	1.64	0.04
	VP	0.51	0.01	0.88	0.018	0.21	0.02	0.21	0.02	0.79	0.017	1.64	0.03
	VN	0.50	0.01	0.82	0.022	0.20	0.007	0.21	0.03	0.86	0.025	1.78	0.05

eTable 4: Parameters estimates and goodness of fit measure for HMM1.

HMM1		Alpha		log(Beta)		Delta		NLL		BIC	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Smokers	PP	0.45	0.02	0.77	0.024	0.22	0.02	0.87	0.021	1.80	0.04
	PN	0.52	0.01	0.91	0.039	0.24	0.02	0.80	0.029	1.65	0.06
	VP	0.51	0.01	0.91	0.039	0.20	0.01	0.80	0.033	1.65	0.07
	VN	0.49	0.009	0.85	0.025	0.20	0.01	0.83	0.025	1.72	0.05
Nonsmokers	PP	0.48	0.009	0.83	0.019	0.23	0.02	0.86	0.021	1.76	0.04
	PN	0.48	0.008	0.88	0.018	0.22	0.01	0.82	0.018	1.68	0.04
	VP	0.50	0.009	0.85	0.014	0.26	0.02	0.84	0.014	1.72	0.03
	VN	0.49	0.006	0.80	0.021	0.25	0.01	0.90	0.024	1.84	0.05

eTable 5: Parameters estimates and goodness of fit measure for the Rescorla-Wagner model.

RW		Alpha		log(Beta)		Eta		NLL		BIC	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Smokers	PP	-0.003	0.014	0.414	0.023	0.778	0.029	1.022	0.014	2.093	0.029
	PN	-0.010	0.011	0.527	0.050	0.806	0.018	0.014	0.018	2.078	0.036
	VP	0.014	0.009	0.518	0.042	0.766	0.022	1.019	0.018	2.087	0.036
	VN	-0.015	0.010	0.487	0.024	0.779	0.025	1.025	0.014	2.099	0.027
Nonsmokers	PP	-0.008	0.010	0.418	0.022	0.811	0.024	1.043	0.015	2.136	0.031
	PN	0.002	0.009	0.462	0.021	0.822	0.020	1.022	0.012	2.093	0.024
	VP	0.002	0.012	0.417	0.022	0.842	0.023	1.036	0.012	2.121	0.025
	VN	-0.023	0.018	0.383	0.022	0.827	0.014	1.062	0.016	2.174	0.032

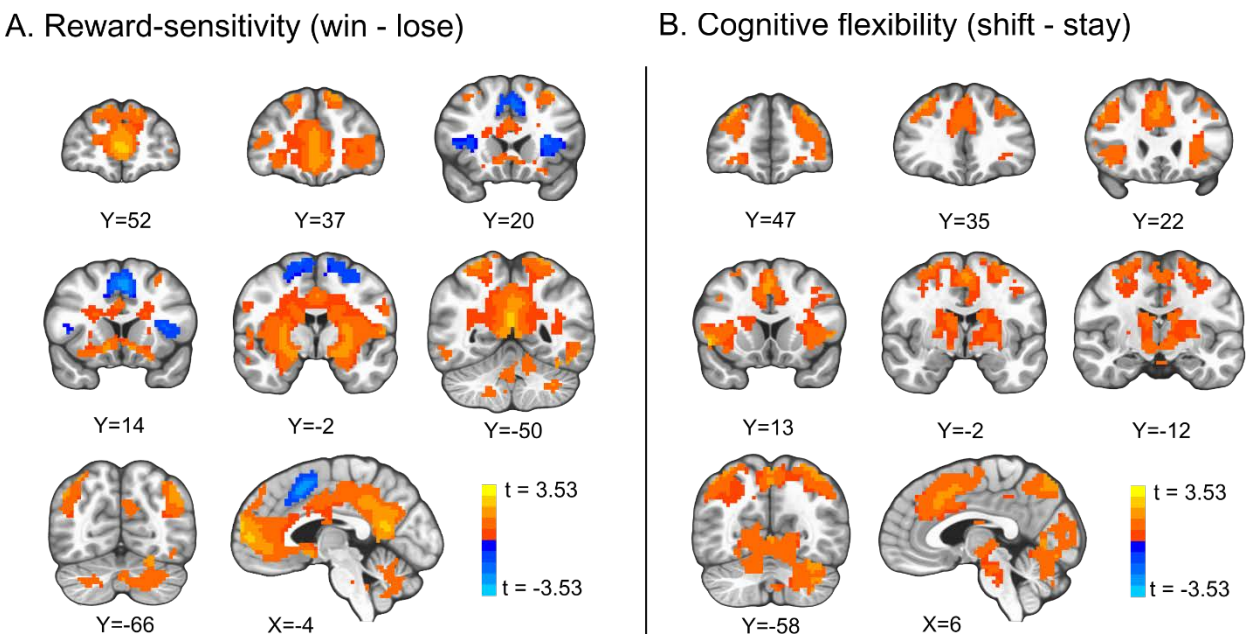
2.7 fMRI results

eTable 6. Reward sensitivity contrast: table of results.

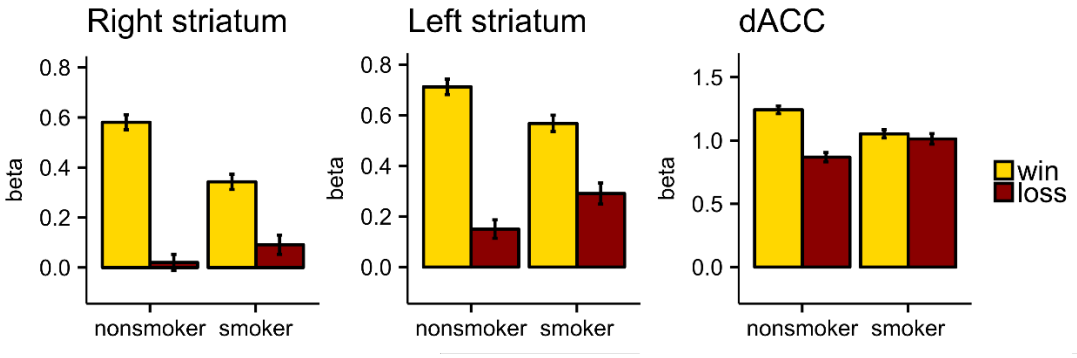
Gross anatomical region	cluster size (voxels)	Talairach coordinates (LPI) x	y	z
REWARD SENSITIVITY (whole brain corrected FWE alpha < 0.05)				
Cingulate / vmPFC / occipitotemporal gyrus / striatum	12959	-1.5	-52.5	14.5
dorsal ACC	610	-1.5	10.5	47.5
right fusiform gyrus	163	22.5	-85.5	-12.5
left fusiform gyrus	161	-34.5	-70.5	-15.5
left anterior insula	136	-31.5	16.5	11.5
right posterior middle temporal gyrus	68	52.5	-46.5	-15.5
right anterior insula	67	31.5	19.5	11.5
right cerebellar Crus II	45	31.5	-67.5	-36.5
right middle frontal gyrus	40	46.5	43.5	11.5
left inferior parietal lobule	21	-55.5	-31.5	44.5
GROUP DIFFERENCE - REWARD SENSITIVITY (small volume corrected FWE alpha < 0.05)				
Right putamen	114	34.5	-7.5	2.5
Left putamen	82	-25.5	-13.5	8.5
Dacc	75	7.5	16.5	29.5

eTable 7. Cognitive flexibility contrast: table of results.

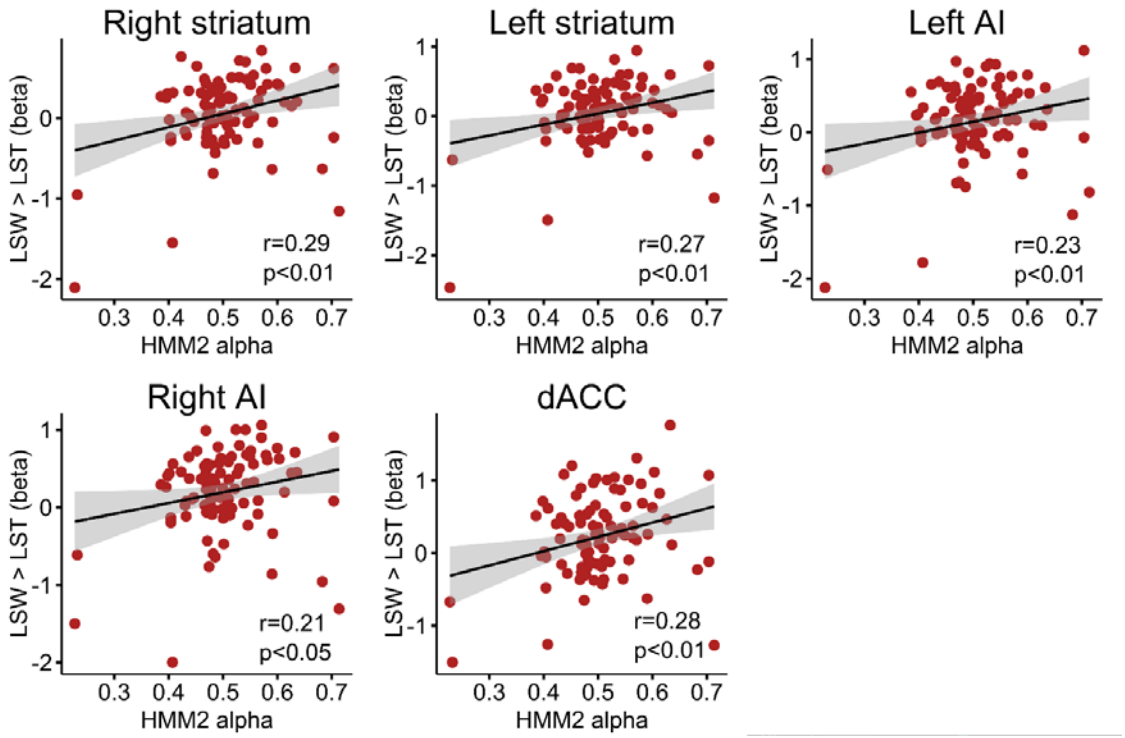
Gross anatomical region	cluster size (voxels)	Talairach coordinates (LPI) x	y	z
FLEXIBILITY (whole brain corrected FWE alpha < 0.05)				
dACC, left AI, bilateral striatum, bilateral lingual gyrus	8128	-31.5	58.5	17.5
Right middle frontal gyrus	582	31.5	61.5	11.5
Right anterior insula / IFG	377	49.5	16.5	-0.5
Right posterior middle temporal gyrus	99	64.5	-31.5	-3.5
Right cerebellar Crus I	27	40.5	-40.5	-33.5
SMOKERS: NICOTINE EFFECT ON FLEXIBILITY (small volume corrected FWE alpha < 0.05)				
Right dorsal and ventral striatum	296	16.5	19.5	2.5
Left dorsal and ventral striatum	251	-7.5	7.5	-6.5
Left anterior insula / IFG	102	-37.5	22.5	-6.5
vmPFC / subgenual ACC	96	4.5	22.5	-3.5
Right anterior insula / IFG	89	37.5	1.5	5.5
dACC	56	-4.5	31.5	23.5



eFigure 9. Imaging results: Task Maps for Reward-Sensitivity and Cognitive Flexibility Contrasts Across Groups and Sessions. A. Brain areas that respond differently to rewarding versus punishing outcomes (warm colors: reward > punishment; cool colors: punishment > reward). B. Brain areas with higher activity before a behavioral shift than before staying (red: shift > stay). Whole-brain corrected at FWE <0.05 (p<0.001 voxelwise, cluster size 19). Left is on the right.



eFigure 10. Imaging results: Reward vs. punishment group differences. Extracted regression weights from brain areas with a significant group difference for wins (yellow) and losses (red). Extracted regression weights and error bars (SEM) presented to aid interpretation only, no statistical inference should be drawn.



eFigure 11. Imaging results: Cognitive Flexibility is associated with Bias to Shift in Smokers in Areas That Show Decreased Shift-Related activity. Neural signatures of cognitive flexibility correlate with the bias toward staying as defined in the computational model (alpha). Regression weights extracted from 5 MCL regions that showed nicotinic effects in smokers (Figure 5 in main paper). Regression weights from the 6th vmPFC cluster were not related to alpha.

3 eREFERENCES

1. Fedota JR, Sutherland MT, Salmeron BJ, Ross TJ, Hong LE, Stein EA. Reward Anticipation Is Differentially Modulated by Varenicline and Nicotine in Smokers. *Neuropsychopharmacology*. 2015;40(8):2038-2046.
2. Sutherland MT, Carroll AJ, Salmeron BJ, Ross TJ, Hong LE, Stein EA. Down-Regulation of Amygdala and Insula Functional Circuits by Varenicline and Nicotine in Abstinent Cigarette Smokers. *Biol Psychiatry*. 2013;74(7):538-546.
3. Sutherland MT, Carroll AJ, Salmeron BJ, Ross TJ, Hong LE, Stein EA. Individual differences in amygdala reactivity following nicotinic receptor stimulation in abstinent smokers. *NeuroImage*. 2013;66:585-593.
4. Sutherland MT, Carroll AJ, Salmeron BJ, Ross TJ, Stein EA. Insula's functional connectivity with ventromedial prefrontal cortex mediates the impact of trait alexithymia on state tobacco craving. *Psychopharmacology (Berl)*. 2013;228(1):143-155.
5. Faessel HM, Smith BJ, Gibbs MA, Gobey JS, Clark DJ, Burstein AH. Single-Dose Pharmacokinetics of Varenicline, a Selective Nicotinic Receptor Partial Agonist, in Healthy Smokers and Nonsmokers. *J Clin Pharmacol*. 2006;46(9):991-998.
6. Palmer KJ, Buckley MM, Faulds D. Transdermal Nicotine. A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic efficacy as an aid to smoking cessation. Drugs. 1992;44(3):498-529.
7. Cools R, Clark L, Owen AM, Robbins TW. Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *J Neurosci*. 2002;22(11):4563–4567.
8. Hampton AN, Bossaerts P, O'Doherty JP. The Role of the Ventromedial Prefrontal Cortex in Abstract State-Based Inference during Decision Making in Humans. *J Neurosci*. 2006;26(32):8360-8367.
9. Patzelt EH, Kurth-Nelson Z, Lim KO, MacDonald III AW. Excessive state switching underlies reversal learning deficits in cocaine users. *Drug Alcohol Depend*. 2014;134:211-217.
10. Sohn H, Kim S. Simple Reinforcement Learning Models Are Not Always Appropriate. *J Neurosci*. 2006;26(45):11511-11512.
11. Luce RD. On the possible psychophysical laws. *Psychol Rev*. 1959;66(2):81-95.
12. O'Doherty J, Critchley H, Deichmann R, Dolan RJ. Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *J Neurosci*. 2003;23(21):7931–7939.
13. Izquierdo A, Jentsch JD. Reversal learning as a measure of impulsive and compulsive behavior in addictions. *Psychopharmacology (Berl)*. 2011;219(2):607-620.

14. Rose EJ, Ross TJ, Salmeron BJ, et al. Acute Nicotine Differentially Impacts Anticipatory Valence- and Magnitude-Related Striatal Activity. *Biol Psychiatry*. 2013;73(3):280-288.
15. Sweitzer MM, Geier CF, Joel DL, et al. Dissociated Effects of Anticipating Smoking versus Monetary Reward in the Caudate as a Function of Smoking Abstinence. *Biol Psychiatry*. 2014;76(9):681-688.
16. O'Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C. Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat Neurosci*. 2001;4(1):95-102.
17. Destrieux C, Fischl B, Dale A, Halgren E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*. 2010;53(1):1-15.