Supplementary Information

METHOD

Reinforcement learning task

During the fMRI scan, subjects carried out an instrumental discrimination learning task with probabilistic feedback that required making choices to maximize wins and minimize losses (Figure 1A). In each trial, one of three possible pairs of abstract pictures was randomly presented: rewarding, punishing or neutral. There were 30 trials of each valence (90 trials per visit in total), randomized and interspersed by an inter-trial interval of variable duration (0.5-4.5 s). For each trial, the subject used a button push (first or second finger of the right hand) to indicate a choice of picture. Selection of one of the pictures would lead to a particular outcome (a picture of a £1 coin in rewarding trials, a red cross over a £1 coin in punishing trials, and a purple circle the same size of the coin in neutral trials) with a 70% probability, whereas selection of the other picture led to the outcome with only 30% probability. A “neutral outcome” trial simply entailed presentation of a blue circle in rewarding and punishing trials, and a pink circle in the neutral valence. Subjects learned the task by trial and error. Optimal responding involved learning to choose the high probability and the low probability cues in rewarding and punishing trials, respectively. The relationship of a given abstract picture to the probability of obtaining the expected feedback was counterbalanced across subjects. The position of the high probability stimulus was randomized and counterbalanced across trials within valence. Before starting the task, subjects were informed that any money they won during the task would be paid to them at the end of the experiment.

Rating scales and behavioral analyses

Immediately after the fMRI scan, subjects were interviewed by an experienced psychiatrist who had passed the membership examination of the Royal College of Psychiatrists, to measure the severity of any mild (prodromal) psychotic symptoms (Comprehensive Assessment of At Risk Mental States, CAARMS, subscales 1.1 Unusual thought content, 1.2 Non-bizarre ideas and 1.3 Perceptual abnormalities) or manic symptoms (Young mania rating scale) (1–3). Slightly modified version of the scales were used in order to tailor them for use in an acute experimental setting rather than the original clinical use in determining mental state over several days or weeks at a time.

fMRI data acquisition and preprocessing

A 3T Siemens TIM Trio system was used to collect imaging data. Gradient-echo echo planar T2*-weighted images measuring BOLD contrast were acquired from 35 non-contiguous near axial planes, with a TR=1.62 s and a TE=30 ms, flip angle=65^o, in-plane resolution=3 x 3 mm, matrix size 64 x 64, bandwidth 2442 Hz/Px. A total of 530 volumes per subject and session were acquired (35 slices each of 2 mm thickness). The first 5 volumes were discarded to allow for T1 equilibration effects.

Imaging data was analyzed using FSL software (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl) (4). First, we ran a Multivariate Exploratory Linear Optimized

Decomposition into Independent Components (MELODIC) (5) analysis to detect task independent sources of noise, caused by subject's movement or inhomogeneities in the magnetic field. The selected artefactual components were included as nuisance regressors in the subject level analysis.

Individual subjects' data were analyzed using FEAT (FMRI Expert Analysis Tool). EPI images were realigned, motion-corrected, slice-timing corrected and spatially smoothed with a Gaussian kernel (3 mm, full-width half-maximum). The time series in each session was high-pass filtered (128 s cut off), and the images were registered first to a whole brain echo-planar image, then to an anatomical scan obtained from the corresponding subject, and finally normalized to a standard template (MNI).

Region of interest definition

We expected the main drug effects to be mediated by the output targets of dopamine neurons, rather than in their cell bodies. Based on extensive previous evidence that reward prediction errors are encoded in the ventral striatum combined with evidence that amphetamines most strongly affect dopamine release in the ventral (limbic) striatum (6), we restricted our RPE analyses to a region of interest that included the nucleus accumbens and ventral aspects of caudate nucleus and putamen, containing 398 voxels (3184 mm^3). We restricted our analysis of incentive value computation to the ventromedial prefrontal cortex, as extensive previous evidence implicates this region, which is sometimes referred to as the ventromedial prefrontal cortex/orbitofrontal cortex, as being involved in representing values of actions or stimuli (7). Recently, it has been demonstrated that ventromedial prefrontal cortex is particularly sensitive to dopaminergic modulation in humans (8). Our specific ventromedial prefrontal cortex region of interest was based on a recent study demonstrating action value computation in this region (9). This study reported the greatest effect in three different clusters located around coordinates 6, 34, -8 (MNI space). We created a region of interest using a 8 mm radius sphere centered on these y and z coordinates and in the midline (0, 34, -8), containing 257 voxels (2056 mm^3) (Figure S2).

Computational model.

We estimated reward prediction error and incentive value parameters for each trial by following a basic Q learning algorithm. The decisions of all subjects throughout the task were extracted and analyzed in Matlab. The algorithm assigned an action value (Q) to the action of selecting a picture (A or B) in each trial (t). Since the subjects were unaware of the structure of the task, Q_A and Q_B were initially set to 0.

$$Q_A(t_1) = Q_B(t_1) = 0$$

The actual outcomes (£1 gain and £1 loss) were coded as +1 and -1 in rewarding and punishing trials, respectively, whereas the purple circle was coded as +1 in neutral trials. The "non-outcome" trial (blue circle in rewarding and punishing trials, and pink circle in neutral trials) was coded as 0. Therefore, the prediction error for each trial was:

$$\delta(t) = R(t) – Q(t)$$

where R(t) was the outcome (or reinforcement) for that particular trial. The vector of reward prediction error values included positive numbers when the outcome was better than expected (for example avoiding a loss in the potentially punishing trials), and negative values when it was worse than predicted (for example a neutral outcome trial when a reward was expected, or a loss when expecting a neutral outcome).

After obtaining the outcome and estimating RPE, the Q value of the chosen cue (say A) was updated as follows:

$Q_A(t+1) = Q_A(t) + \alpha * \delta\ (t)$

Conventionally, in Q-learning, one models the value of an action; in our experiment the value of the action of choosing a cue is equivalent to the value of the cue (and to the expected value), hence in our report we use the term incentive value, which is equivalent to the action value and the cue value. We obtained two vectors of incentive values for each subject: one for reward and another for punishment trials, including only positive and only negative numbers, respectively.

α is a constant known as the learning rate, and represents the extent to which RPE is used to update the value of an option. A high learning rate is often, but not always, beneficial. For example, a subject is selecting a rewarding action and at some point the reward is not delivered: a subject with a very high learning rate might immediately start selecting the alternative option, whilst a subject with a lower learning rate might repeat the selection of previously rewarded object. Given the Q values for each of the actions in a particular trial, the associated probability of selecting one of them was estimated by implementing the softmax rule. For example, the probability of choosing A in a trial t was:

$$P_A(t) = \frac{e^{Q_A(t)/\beta}}{e^{Q_A(t)/\beta} + e^{Q_B(t)/\beta}}$$

β is another constant that can be thought of as representing the balance between the exploration among different options and the exploitation of a particular one. Both constants (α and β) were adjusted to maximize the likelihood of the actual choices of the subjects during the task. A Q learning algorithm was implemented in Matlab with all possible combinations of 100 α values (from 0.01 to 1) and 100 β values (from 0.05 to 5), for all subjects and pharmacological conditions. For the fMRI analyses, the pair of α and β values that offered a negative log likelihood closest to zero was selected as the best to explain the subjects performance for use in calculating incentive value and reward prediction error regressors in the fMRI analysis. We used the same method to calculate the pair of values that best explained the subjects' performance in reward and punishment trials, and in all pharmacological conditions.

RESULTS

Whole brain fMRI results.

Placebo. The RPE signal was represented bilaterally in the striatum and the occipital cortex (P<0.01, corrected). No cluster survived correction for multiple comparisons in the analysis of the incentive value signal. We present uncorrected results in this data supplement in case they are useful for future hypothesis generation or meta-analyses. Two clusters were located in ventromedial prefrontal cortex and anterior frontal cortex at a more lenient threshold (P<0.005 uncorrected, cluster minimum size (k) >15 voxels) (Figure S3; Table S1).

Placebo vs Methamphetamine. There were no significant results at a whole brain level, P<0.05 corrected. At a more relaxed threshold (P<0.005 uncorrected, k>50), several areas displayed a disrupted RPE signal in methamphetamine in the bilateral occipital lobe, left ventral striatum and right orbitofrontal cortex/anterior superior temporal gyrus (Figure S4; Table S1). At a similar statistical threshold (P<0.005 uncorrected, k>25), methamphetamine disrupted incentive value signal in a cluster located in the anterior prefrontal cortex (Figure S4; Table S1).

Methamphetamine + amisulpride vs Methamphetamine. At a whole brain level, there were no significant voxels where amisulpride reverted the effects of methamphetamine on RPE and incentive value, even at a relaxed threshold (P<0.005, uncorrected). There were no voxels where the learning signals were significantly improved in the methamphetamine visit, compared with the methamphetamine + amisulpride condition.

Table S1. Summary of fMRI results at a whole brain level for the placebo visit and drug effect

		Region	Voxels	X,Y,Z (mm)	Peak P
Placebo	RPE	R Putamen/Insula	1079	32,-10,4	<0.001
		L Putamen	250	-28,-16,0	0.002
		R Occipital pole	250	22,-102,-4	<0.001
		L amygdala	82	-26,-6,-28	0.007
		Postcentral gyrus	51	-2,-18,52	0.005
	Incentive value	L VMPFC	27	-4,34,-12	0.002
		R anterior FC	18	34,58,4	0.003
Placebo > M	RPE	Medial OC	380	-2,-82,28	<0.001
		L operculum	140	-36,-20,26	0.001
		L lateral OC	127	-50,-78,20	0.001
		L medial OC	98	-20,-70,14	0.001
		R OFC/STG	90	34,8,-22	0.001
		L OC	74	-34,-78,-4	0.002
		L N Accumbens	67	-6,2,-6	0.001
		R dorsal OC	52	28,-90,32	0.001
	Incentive value	R anterior FC	29	26,58,0	0.002

FC, frontal cortex; L, left; M, methamphetamine; OC, occipital cortex; OFC, orbitofrontal cortex; R, right; STG, superior temporal gyrus; VMPFC, ventromedial prefrontal cortex.
See text for statistical thresholds.

REFERENCES

1. Yung AR, Yuen HP, McGorry PD, Phillips LJ, Kelly D, Dell'Olio M, Francey SM, Cosgrave EM, Killackey E, Stanford C, Godfrey K, Buckby J: Mapping the onset of psychosis: the Comprehensive Assessment of At-Risk Mental States. Aust N Z J Psychiatry 2005; 39:964–71

2. Young RC, Biggs JT, Ziegler VE, Meyer DA: A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry 1978; 133:429–35

3. Morrison a. P, French P, Stewart SLK, Birchwood M, Fowler D, Gumley a. I, Jones PB, Bentall RP, Lewis SW, Murray GK, Patterson P, Brunet K, Conroy J, Parker S, Reilly T, Byrne R, Davies LM, Dunn G: Early detection and intervention evaluation for people at risk of psychosis: multisite randomised controlled trial. BMJ 2012; 344:e2233–e2233

4. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM: Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 2004; 23 Suppl 1:S208–19

5. Beckmann CF, DeLuca M, Devlin JT, Smith SM: Investigations into resting-state connectivity using independent component analysis. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 2005; 360:1001–13

6. Martinez D, Slifstein M, Broft A, Mawlawi O, Hwang D, Huang Y, Cooper T, Kegeles L, Zarahn E, Abi-dargham A, Haber SN, Laruelle M: Imaging Human Mesolimbic Dopamine Transmission With Positron Emission Tomography . Part II : Amphetamine-Induced Dopamine Release in the Functional Subdivisions of the Striatum. Blood 2003; 285–300

7. Levy DJ, Glimcher PW: The root of all value: a neural common currency for choice. Current opinion in neurobiology 2012; In Press. DOI: 10.1016/j.conb.2012.06.001

8. Jocham G, Klein TA, Ullsperger M: Dopamine-Mediated Reinforcement Learning Signals in the Striatum and Ventromedial Prefrontal Cortex Underlie Value-Based Choices. Journal of Neuroscience 2011; 31:1606–1613

9. Gläscher J, Hampton AN, O'Doherty JP: Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. Cerebral cortex 2009; 19:483–95

Figure S1. Performance and individual learning parameters in all drug conditions. Methamphetamine did not affect the subjects' performance (A). However, learning rate (α) was diminished in reward trials after methamphetamine infusion (B) [t(16)=2.78, P=0.013], and recovered by amisulpride [t(16)=3.32, P=0.004. Exploration/exploitation parameter (β) remained unchanged in all conditions. Whiskers represent standard deviations.

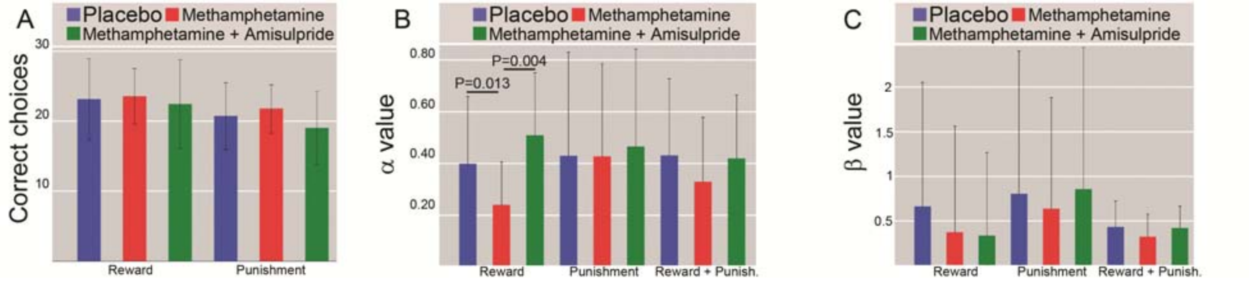

Figure S2. The ventromedial prefrontal cortex region of interest. See text for details.

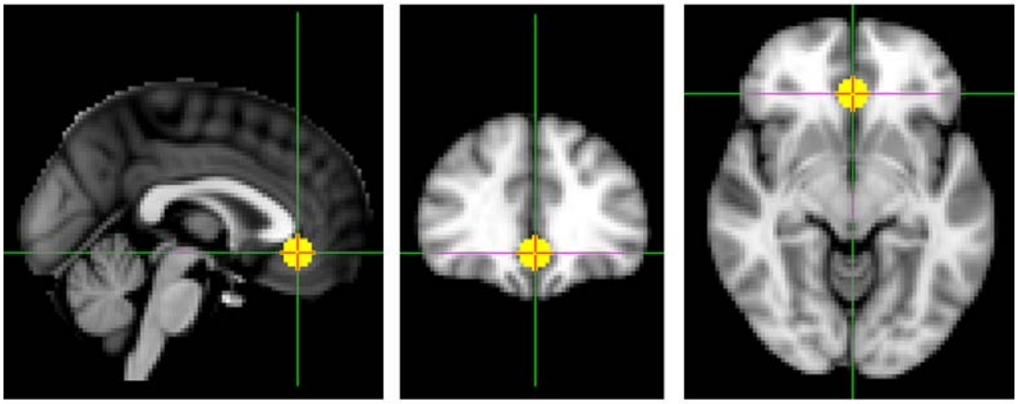

Figure S3. Placebo results at a whole brain level. RPE results are thresholded at P<0.01, corrected. Incentive value signal, at P<0.005, uncorrected. Left hemisphere is on the right side of the image.

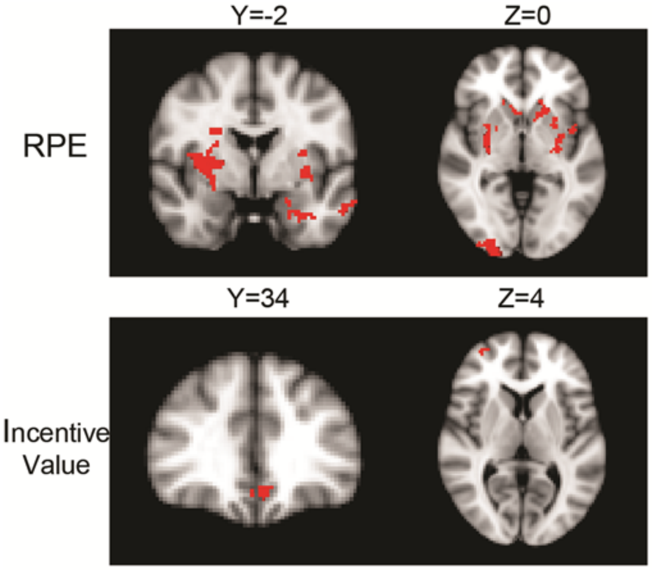

Figure S4. Placebo versus methamphetamine results at a whole brain level. All images at thresholded at P<0.005 uncorrected. Results for RPE are shown for clusters larger than 50 voxels, and larger than 25 voxels for incentive value.