Supplement

eMethods.1 Operational Definitions of Amygdala and Hippocampus ROIs.

The amygdala and hippocampus are both identified by visual inspection on subjects' coronally oriented T1-weighted structural images as gray matter tissue in the medial temporal lobe. The anterior boundary of the amygdala is the first coronal slice on which the anterior commissure appears. The amygdala continues to be outlined on coronal slices moving in the posterior direction until the first slice on which the lateral horn of the temporal ventricle appears. A horizontal line (parallel to the anterior commissure) is drawn from the most superior part of the ventricle, and amygdala continues to be identified as the medial temporal gray matter superior to this line. The line at the most superior point of the ventricle and the amygdala continue to be drawn on successive posterior slices until the amygdala is no longer visible. The anterior boundary of the hippocampus is defined as the first (most anterior) coronal slice on which the horizontal line superior to the ventricle is drawn. Hippocampus is defined as the medial temporal lobe gray matter inferior to the line. The hippocampus continues to be drawn on successive posterior slices inferior to the line where the line is identified, and posterior to the amygdala and ventricle until the hippocampus is no longer visible.

eMethods.2 PET Kinetic Method

To address the issues of low signal-to noise ratio of [^{11}C] FLB457 cortical specific binding as well as the poor suitability of the cerebellum as a reference tissue, a kinetic approach utilizing a modified two tissue compartment model was applied. In this method, the rate constant k_4, and V_{ND}, the distribution volume of the nondisplaceable compartment, were fitted to a single value across regions, whereas the delivery constant K_1 and the rate constant k_3 were fitted in each region and condition. The method was applied separately to the 9 cortical regions and the 5 moderate binding subcortical regions examined in the study. The reason for separating the low and moderate binding regions was the possibility that k_3 and k_4 estimates might be correlated, with moderate binding regions having a lower estimated value of k_4 than low binding regions. In theory, k_4 is the dissociation rate of the tracer from the receptor and is independent of k_3, but in reality the estimated quantities tend to be correlated, and this approach allowed for the optimization routine to converge to fitted values appropriate to the two different ranges of k_3. In the simulations, two versions of the method were compared: One in which the same values of k_4 and V_{ND} were estimated for both the baseline and post-amphetamine scan (the 2 scan method), and a second version in which k_4 and V_{ND} were fitted separately for baseline and post-amphetamine (1 scan method). To test which approach is more accurate, simulations were performed in which the model assumptions (uniform k_4 and V_{ND} across regions and/or scans) were intentionally violated and estimated values of V_T, BP_{ND} and their percent change across conditions were compared to the (simulated) true values. For completeness, standard 2TC fits with all 4 rate constants fitted in each region and condition, using cerebellum as a reference tissue, were performed as well.

Simulation Methods. Each modeling assumption was tested with a set of 250 Monte Carlo simulations with 9 simulated cortical regions, a simulated cerebellum, and mean decrease in BP_{ND} following amphetamine of 5%. For each of the 250 runs in a simulation, rate constants [K_1, k_2, k_3, k_4] for the baseline condition were generated as the mean observed across cortical regions in the real data, plus a Gaussian random variable with standard deviation of 10% of the mean parameter value. Rate constants for the post-amphetamine condition were set to the same value as baseline with k_3 decreased across regions by uniform random variables ranging between 0 and 10% of the baseline value (mean decrease of 5%). Pairs of arterial plasma input functions were selected randomly from among the baseline/post-amphetamine pairs of input functions in the real study data, and continuous time activity curves were generated. These were subsampled at the mid-frame times of the collected data to generate discrete time activity curves comparable to the reconstructed PET data, and an additional layer of mean-zero noise with standard deviation equal to 2.5% of the interpolated values was added. These data were then fitted by the 1 scan, 2 scan and 2TC methods, with weighting as in the main study. V_T was computed in every case. BP_{ND} was computed as the estimated k_3/k_4 ratio for the 1 scan and 2 scan methods and both as k_3/k_4 and indirectly

using the cerebellum V_T as an estimate of V_{ND} (overestimated by 20%, by design, assuming specific binding in cerebellum) for the 2TC method, for comparison. The tested conditions were i) variability in k_4 across regions, ii) change in k_4 between scan conditions, iii) change of V_{ND} across scan conditions and combinations of I, ii and iii. To test i) k_4 was varied across regions by a Gaussian random variable with standard deviation equal to 5% of the mean. To test ii) k_4 was varied between scans, with post-amphetamine k_3 readjusted by the same amount such that the post-amphetamine decrease in BP_{ND} was preserved. These were tested at the 5% and 10% standard deviation levels for k_4, as both Gaussian and uniform random variables, with the latter representing the case in which all amphetamine effect was attributable to reduced affinity, i.e. increased k_4. To test iii), V_{ND} was varied between scan conditions, again with a Gaussian random variable with standard deviation equal to 5% or10% of the baseline value. Finally combinations of ii and iii, i.e. simultaneous changes of both k_4 and V_{ND} across conditions, were tested.

Results. Representative results are shown in Table S1. The table shows the ordinary regression slopes and intercepts of the fitted parameters against the simulated true parameters, along with Pearson correlation coefficients, for each of the 2250 fits for a representative set of simulation conditions. The 2 scan and 1 scan methods both outperformed conventional 2TC under all conditions. This includes BP_{ND} and ΔBP_{ND} where the simulated specific binding in cerebellum could be expected to affect the outcome, but also accuracy and precision of baseline and post-amphetamine V_T, where many of the tested conditions would have appeared a priori to favor the greater flexibility of 2TC (data not shown). Under all of the tested conditions, the 2 scan method was more accurate than the 1 scan method. Surprisingly, this included all simulations in which k_4 or V_{ND} was changed across conditions, which might be expected to favor the 1 scan method. We conclude that the simulations support the 2 scan method as being the most sensitive approach for extracting ΔBP_{ND} and ΔV_T using [^{11}C]FLB457 with amphetamine challenge.

Discussion. In this study, we applied a kinetic method that would simultaneously address the low signal in cortex, the low expected change in that signal across the tested conditions, and the lack of a suitable reference tissue. Our method fitted single values of V_{ND} and k_4 across brain regions and scans for each subject. The assumption of single V_{ND} is widely accepted, and in fact implicit in the interpretation of ΔBP_{ND} as reflecting changes in specific binding only. The use of a single k_4 is less conventional, though not unique to this study [1]. There is some evidence suggesting that measured k_4 may vary across regions due to differing microenvironments [2]. Additionally, on statistical grounds, data fitting algorithms could conceivably provide correlated parameter estimates, for example, estimating BP_{ND} in high binding regions by a combination of both large k_3 and small k_4. The simulations here showed that even when the underlying compartment model incorporated all of these conditions that do not conform to the 2 scan model assumptions, the 2 scan method still outperformed all other methods, due to the increased parsimony associated with estimating 38 parameters per subject, rather than 40 for the 1 scan method or 80 for 2TC. One reasonable concern when applying this approach is identifiability of V_{ND} and BP_{ND}, i.e. that the partition of total distribution volume V_T into its component factors V_{ND} and $(1 + BP_{ND})$ from direct estimation of these may not represent the true physiological parameters in the absence of additional information to formulate constraints in the data fitting process. To address this concern, we also measured and reported total V_T and ΔV_T and found these to be most precisely measured by the 2 scan method as well. We expected, a priori, that ΔV_T would be less than ΔBP_{ND} because a portion of V_T is V_{ND} which should not be affected by DA release. In fact, in the study data we did see that $\Delta V_T < \Delta BP_{ND}$ in HC, but the blunting in SCZ and significant group differences were still detectable. As noted in the main article, average estimated V_{ND}, 2.97 ± 0.87 in SCZ and 3.27 ± 0.55, is in approximate accord with the Narendran study [3] in that cerebellum V_T measured here was 4.19 ± 0.95 in SCZ and 4.56 ± 1.08 in HC, corresponding to V_{ND} equaling 70 ± 13% of cerebellum V_T in SCZ and 73 ± 10% in HC. Nevertheless, we cannot claim to have measured physiological V_{ND} by this method without further validation, for example through

blocking studies. Our simulations do however strongly suggest that the approach improves estimation of ΔV_T and ΔBP_{ND} compared to methods that measure within subject regional parameters independently.

eTable.1 Simulation Results for ΔBP_{ND} and ΔV_T

Condition: k_4 varies within scan (Gaussian, 5% SD) across regions but not between scans				
	slope	intercept	R	% Outliers
ΔBP 2TC indirect	1.0077	0.0083	0.0073	14%
ΔBP 2TC	1.1703	0.0270	0.0333	0.3%
ΔBP 1 scan	1.0326	0.0206	0.0405	0
ΔBP 2 scan	1.0157	0.0009	0.5106	0
ΔVT 2TC	1.1550	0.0141	0.0609	0.4%
ΔVT 1 scan	1.0234	0.0033	0.2203	0
ΔVT 2 scan	1.0084	0.0004	0.5078	0
Condition: k_4 varies within scan (Gaussian, 5% SD) and between scans (uniform between 0 and 10%)				
	slope	intercept	R	% Outliers
ΔBP 2TC indirect	1.1776	-0.0031	0.0104	0.3%
ΔBP 2TC	1.0388	0.0205	0.0265	0.0027
ΔBP 1 scan	1.1552	0.0249	0.0478	0
ΔBP 2 scan	1.0238	0.0203	0.4779	0
ΔVT 2TC	1.1816	0.0068	0.0699	0.1%
ΔVT 1 scan	1.0018	-0.0014	0.1920	0
ΔVT 2 scan	1.0224	0.0136	0.4781	
Condition: V_{ND} varies between conditions (Gaussian, SD = 5%)				
	slope	intercept	R	% Outliers
ΔBP 2TC indirect	1.1424	-0.0215	0.0109	15%
ΔBP 2TC	0.9151	0.0082	0.0222	0.1%
ΔBP 1 scan	1.1581	0.0167	0.0812	0
ΔBP 2 scan	1.0656	0.0057	0.1414	0
ΔVT 2TC	0.9993	0.0041	0.2930	0.1%
ΔVT 1 scan	0.9771	-0.0029	0.7182	0
ΔVT 2 scan	0.9626	-0.0017	0.8783	0
Condition: k_4 varies within (Gaussian, SD = 5%), between (uniform 0 to 10%) V_{ND} varies between (Gaussian, SD = 5%)				
	slope	intercept	R	% Outliers
ΔBP 2TC indirect	0.9094	0.0239	0.0079	15%
ΔBP 2TC	0.9851	0.0185	0.0241	0.3%
ΔBP 1 scan	0.9996	0.0188	0.0658	0
ΔBP 2 scan	0.9969	0.0108	0.1245	0
ΔVT 2TC	1.0429	0.0051	0.2859	0.2%
ΔVT 1 scan	1.0269	-0.0016	0.7280	0
ΔVT 2 scan	0.9731	0.0057	0.8606	0

Regression of parameter estimates for each method against true simulation parameters. ΔBP 2TC indirect used cerebellum VT to estimate regional BPND indirectly as a function of distribution volumes. Cerebellum VT was set to 20% greater than VND. The other 2TC method used direct estimation of k3/k4 for BPND. Fits for which |ΔBPND| > 80% or |ΔVT| > 100% were considered outliers and not included in the regression.

eMethods.3 fMRI Methods

SOWT Task. The participants completed 24 trials of the SOWT, with each trial containing eight steps on which response was required. At the start of each trial, eight simple line drawings of three-dimensional objects were presented in a 3 × 3 grid, with the central position of the grid being empty. The stimuli were the same as those used by Curtis and colleagues [4]. Unique stimuli were used on each of the first 12 trials, and stimuli were repeated exactly once during the latter 12 trials. On each step, participants had 7 s to move a mouse cursor to select any object that had not been selected on a previous trial (thus, all responses were correct on the first step). Once a selection was made, a white outline was displayed around the selected object until 9 s had elapsed from the start of the step. At this point, the objects in the display were pseudo-randomly rearranged in the grid, with the blank space appearing in the same location as the most recently selected item (to prevent participants from using a spatial strategy or simply responding in the same location on each trial). If no response was made within 7 s, a white outline was displayed for 2 s around a randomly selected object that would have been a correct response; participants were instructed to remember this object as if they had selected it themselves. If an incorrect selection was made, a red box was displayed over the object until 7 s had elapsed from the start of the step, after which the same procedure was followed as in the case when a participant made no response. The ITI was 9 s.

Data acquisition. Imaging was carried out on a Philips 1.5 Tesla Intera scanner at the Columbia Radiology MRI Center at the Neurological Institute of New York. Participants lay supine on the scanner bed while viewing stimuli projected onto a screen located at the foot of the scanner bed through a mirror mounted on the head coil. The cursor was controlled with a hand-held fiber optic trackball with buttons on either side. T1-weighted images were obtained with an SPGR sequence with a 256 mm FOV, 200 slices, and 1 mm isotropic voxels. Whole-brain functional EPIs were obtained using an 8-channel SENSE coil with a 2 s TR, 28 ms TE, 77° flip angle, 192 mm field of view, 40 slices, 3 mm isotropic voxels, and SENSE factor of 1.5. The relatively small slice thickness precluded acquisition of data on the ventral most portion of the cerebellum in a large majority of participants; however, this region was not of particular interest in the present study. Participants completed nine runs of 160 volumes each that included either three task trials or two task trials and one control trial between the two task trials. Thirty seconds of rest occurred after each trial, and 32 seconds before the first trial in each run.

Data preprocessing. Preprocessing was performed with a combination of SPM8 and custom Matlab scripts. Reconstructed PAR/REC format files were obtained from the scanner and converted to 32-bit floating point precision Analyze format files. A rough in-brain mask was computed and in-brain signal values were used to identify artifactual volumes: any volume departing from a sliding window by more than eight mean absolute deviations in terms of either mean global signal or Mahalanobis distance was identified and treated as a nuisance regressor during first-level statistical analyses.

Data then underwent slice-timing correction using SPM8 and motion realignment using INRIAlign (Freire, Roche, & Mangin, 2002). Motion realignment parameters were inspected to detect excessive motion; one run from each of two participants (one patient and one control) were excluded from analyses due to translation greater than 2.5 mm or rotation greater than 2.5° from their median position during the run. EPI images were then coregistered to the individual subjects' T1 image used for drawing of ROIs for PET analyses. PET ROIs were then resliced into a 3 mm isotropic voxel space to match the EPI data, and these resliced ROIs were used on an individual-subject basis to extract BOLD activation to the task. For calculation of group-level activation maps, images underwent spatial normalization to the ICBM template using the segmentation algorithm in SPM8 and spatial smoothing with an 8 mm FWHM kernel.

First-level statistical modeling. Prior to first-level modeling, the value in each voxel in each volume was divided by the mean value in that voxel over the entire time-series, expressed as a percent, so that the magnitude of the first-level hemodynamic response function (HRF) estimates were equivalently scaled across runs. Data for each participant were modeled in the GLM framework implemented in SPM8. A three parameter HRF model (with temporal and dispersion derivatives) was used to estimate the Blood Oxygen Level Dependent (BOLD) response to each modeled event. An explicit mask was calculated by using the conjunction of the smoothed (6 mm FWHM) gray matter segmentation and the skull-stripped mean EPI for each subject, to restrict the analysis to regions of gray matter and regions not suffering from excessive signal dropout due to susceptibility artifacts, respectively.

A separate set of regressors was used to model each of the eight steps of the SOWT and the control task as a nine second boxcar (resulting in 27 regressors in total, due to the three parameter HRF

model). Although participants had only seven seconds in which to make a response, we opted to use the full nine seconds of each step in order to capture the period of time in which participants were presumably attempting to maintain previously selected items in WM. Error trials on any of the task steps and on the control task made up two separate sets of regressors (six total), again modeled as nine second boxcars. A two second boxcar was used to model the presentation of textual instructions prior to each trial, and motor responses (button presses) and error feedback (red square appearing over the incorrectly-selected item) were modeled as instantaneous events in order to prevent motor and visual activity from being confounded with other modeled events. Finally, nuisance regressors included all six motion parameters estimates (three translation parameters and three rotation parameters), the squared motion parameters, the first derivative of the motion parameters, the squared derivative of the motion parameters, and dummy regressors for artifactual scans identified as outlined above. Activation at each step of the SOWT and during the control task was quantified as the area under the curve (AUC) in a temporal window ranging from 2 s to 9 s with respect to the three basis functions defining the canonical HRF; this window corresponds to the rise and fall of the HRF following the initial dip and prior to the undershoot.

Second-level statistical modeling. Contrast images of overall task-related activation were calculated for each participant by taking the mean activation at each voxel across steps one through eight in the SOWT and subtracting the activation in the corresponding voxel in the control task. These contrast images were then tested for significance using robust regression[5] and thresholded at $P < 0.05$ after false discovery rate (FDR) correction[6].

Relationship between fMRI BOLD with PET ΔBP_{ND} in DLPFC. Voxels showing significant task-related activation in either group were returned to individual-subject spaces by using the inverse of the normalization transformation produced during segmentation. BOLD percent signal change values were extracted for each subject from voxels that were significant at the group level and also fell within the DLPFC ROI drawn for a given participant. These percent signal changes values were then used in a regression model with ΔBP_{ND}. Model selection was determined based on Akaike information criterion[7].

eMethods.4 Adjusted Hit Rate Formula for the N-Back Task

The hit rate (HR) was calculated as the number of correct responses divided by the number of targets (maximum is 1, minimum is 0). The error rate (ER) was calculated as the number of errors divided by the number of non-targets (maximum is 1, minimum is 0). The adjusted HR (AHR) was calculated as HR – ER. AHR ranges from 1 (if the subject provided all the correct responses and no incorrect response) to –1 (if the subject provided no correct response and all the incorrect responses). Operating at chance level corresponds to an AHR of 0..

eResults.1 Drug Naïve vs Drug Free Patients

The drug-naïve (DN) patients had higher BP_{ND} ($t = 2.96$, $p = 0.008$) and lower ΔBP_{ND} ($t = -2.12$, $p = 0.048$) and ΔV_T ($t = -2.28$, $p = 0.035$) in the DLPFC than the drug-free (DF) patients. However, these groups differed in age (DN: 25.9 ± 7.2 years; DF: 36.1 ± 10.0 years; $t = -2.24$, $p = 0.038$). Given the well documented effect of age on cortical D2R parameters[8,9] analyses were repeated with age as a covariate. With age included, there are no significant group differences for BP_{ND}, ΔBP_{ND} or ΔV_T. DLPFC BP_{ND} and ΔBP_{ND} do not correlate significantly with duration of illness or drug-free interval. DLPFC V_T correlates negatively with duration of illness in the total group of patients ($r = -0.538$, $p = 0.014$), which is largely driven by the drug-free patients ($r = -0.547$, $p = 0.043$).

eTable.2 V_T Results

eTable 2. Distribution volumes (V_T)

	HC (n = 21)			SCZ (n = 20)		
	Baseline	Post-Amph	ΔV_T	Baseline	Post-Amph	ΔV_T
CEREBELLUM	4.6 ± 1.1	4.4 ± 1.1	-2.9 ± 7.5%	4.2 ± 1.0	4.3 ± 1.1	2.1 ± 7.8%
DLPFC	9.2 ± 2.6	8.7 ± 2.4	-4.8 ± 6.9%	8.1 ± 2.8	8.2 ± 2.8	0.9 ± 7.1%
OFC	9.4 ± 2.6	8.9 ± 2.4	-3.9 ± 8.4%	8.4 ± 3.0	8.5 ± 3.1	1.1 ± 6.8%
MFC	9.5 ± 2.5	9.1 ± 2.4	-3.2 ± 8.0%	8.6 ± 2.8	8.6 ± 2.8	0.2 ± 7.1%
A. CING	10.5 ± 2.9	10 ± 2.7	-3.8 ± 9.6%	9.4 ± 2.9	9.6 ± 3.0	2.2 ± 8.6%
OCC CTX	9.1 ± 2.8	8.7 ± 2.5	-4.3 ± 7.4%	8.2 ± 3.0	8.3 ± 3.2	1.1 ± 6.5%
PAR CTX	9.8 ± 2.8	9.4 ± 2.6	-4.2 ± 6.3%	8.8 ± 3.3	8.8 ± 3.3	0.5 ± 6.7%
TEMP CT	14.6 ± 4.1	14 ± 3.9	-4.1 ± 7.6%	13.1 ± 5.0	13.1 ± 5.1	0.1 ± 6.9%
SUB GEN	11.2 ± 3.3	10.8 ± 2.6	-1.6 ± 11.6%	10.3 ± 3.7	10.5 ± 3.4	3.1 ± 11.9%
INSULA	14.4 ± 3.9	13.9 ± 3.7	-3.2 ± 7.8%	13.4 ± 4.5	13.7 ± 4.9	1.8 ± 7.7%
AMYGDALA	24.5 ± 8.9	23.7 ± 7.9	-1.7 ± 13.8%	21.4 ± 7.2	21.8 ± 8.0	1.8 ± 9.2%
HIPPOCAMPUS	13 ± 4.6	12.3 ± 3.9	-3.9 ± 10.1%	11.2 ± 2.8	11.2 ± 3.3	0.1 ± 8.0%
SN/VTA	20.8 ± 6.5	19.9 ± 6.1	-3.7 ± 10.5%	18.5 ± 5.7	18.9 ± 5.8	2.3 ± 8.9%
THALAMUS	26.8 ± 9.3	25.7 ± 8.4	-2.7 ± 10.1%	23.4 ± 8.7	23.9 ± 9.2	2.4 ± 8.6%
UNCUS	17.9 ± 6.0	16.9 ± 5.7	-4.7 ± 13.1%	15.9 ± 5.6	15.8 ± 6.1	-0.9 ± 9.5%

Region Distribution Volumes at baseline, following amphetamine, and their percent change

Supplement References

1. Ashworth S, Rabiner EA, Gunn RN, et al. Evaluation of 11C-GSK189254 as a novel radioligand for the H3 receptor in humans using PET. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine.* Jul 2010;51(7):1021-1029.
2. Votaw JR, Kessler RM, de Paulis T. Failure of the three compartment model to describe the pharmacokinetics in brain of a high affinity substituted benzamide. *Synapse.* 1993;15(3):177-190.
3. Narendran R, Mason NS, Chen CM, et al. Evaluation of dopamine D(2)/(3) specific binding in the cerebellum for the positron emission tomography radiotracer [(1)(1)C]FLB 457: implications for measuring cortical dopamine release. *Synapse.* Oct 2011;65(10):991-997.
4. Curtis CE, Zald DH, Pardo JV. Organization of working memory within the human prefrontal cortex: a PET study of self-ordered object working memory. *Neuropsychologia.* 2000;38(11):1503-1510.
5. Wager TD, Keller MC, Lacey SC, Jonides J. Increased sensitivity in neuroimaging analyses using robust regression. *NeuroImage.* May 15 2005;26(1):99-113.
6. Benjamini Y, Hochberg Y. CONTROLLING THE FALSE DISCOVERY RATE - A PRACTICAL AND POWERFUL APPROACH TO MULTIPLE TESTING. *Journal of the Royal Statistical Society Series B-Methodological.* 1995;57(1):289-300.
7. Akaike H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 1974;AC19:716-723.
8. Inoue M, Suhara T, Sudo Y, et al. Age-related reduction of extrastriatal dopamine D2 receptor measured by PET. *Life Sci.* Jul 20 2001;69(9):1079-1084.
9. Kaasinen V, Vilkman H, Hietala J, et al. Age-related dopamine D2/D3 receptor loss in extrastriatal regions of the human brain. *Neurobiol Aging.* Sep-Oct 2000;21(5):683-688.

e Figure.1 Scatterplots of fitted vs true results from a representative simulation where all 3 assumptions were violated: k_4 varied across regions and increased between scan conditions and V_{ND} changed randomly between scan conditions. BP_{ND} decreased following amphetamine randomly between 0 and 10%. In this set of simulations the BP_{ND} change was always contributed to by increased k_4 across conditions but the 2 scan method still gave the best estimates of the changes in BP_{ND} and V_T across conditions even though the approach fits a single k_4 across scans.