Supporting Information

Glässcher et al. 10.1073/pnas.1206608109

SI Materials and Methods

Participants. All 344 subjects in this study were recruited into the Iowa Patient Registry over the past three decades after a neurological incident of the following frequencies: stroke (n = 253), temporal lobectomy (n = 42), focal surgical resection (benign tumor, n = 28; vascular abnormality, n = 6), encephalitis (n = 7), and other focal pathology (n = 8). All patients underwent comprehensive neuropsychological evaluation by following the methods of the Benton Neuropsychology Laboratory of the Department of Neurology at the University of Iowa Hospitals and Clinics (1), which included the tests that were selected as target indices for this study (cognitive-dependent measures). The Iowa Patient Registry served as the source of neuropsychological and neuroanatomical data. Sample sizes for each neuropsychological test are listed in bold type in Table S1.

Target Neuropsychological Tests. Tests from our battery that are classically associated with executive functions, cognitive control, and decision-making are described below (see also ref. 2). Variables from these tests were used as dependent cognitive measures in the current study.

The Trail-Making Test (TMT) consists of two connect-the-dots tasks scored by time to completion. Part A involves a random arrangement of dots containing numbers 1–25, which the participant connects in ascending order. Part B involves a random arrangement of dots containing numbers 1–13 and letters A–L, which the participant must connect in ascending but alternating order, introducing dual-task and switching demands. We used the score B – A (in time units) as a measure of executive response switching.

In the Wisconsin Card Sorting Test (WCST), participants sort cards by matching each card to one of four targets. Cards may be matched according to number, color, or shape; however, matching criteria are not explicitly explained to the participant. The participant must infer the correct dimension on which to sort by evaluating the examiner’s feedback (correct/incorrect) on each attempted match. After a certain number of correct responses, the sorting criterion is changed (e.g., from color to shape) unannounced to the participant, and the participant must learn the new contingency by trial and error. Perseverative errors (PE) result when a participant fails to adjust to the new sorting rule. We used the number of PEs as an index of the executive function of set shifting.

In the Stroop Test, participants are shown stimuli printed in different ink colors and tasks are scored by time to completion. In the noninterference conditions, the tasks are to (i) read the color word (e.g., RED, GREEN, BLUE) printed in black; and (ii) identify the color of four XXXX’s printed in different ink colors. In the color-word (CW) interference condition, the participant has to name the ink color of a color word, where the two are discrepant (e.g., the word “RED” printed in green ink). We used time to complete the Stroop CW condition as an index of response inhibition.

The Controlled Oral Word Association Test (COWA) requires participants to generate as many unique words as possible that begin with a given letter within 1 min. We used the number of generated words as a measure of verbal fluency and, more generally, as an index of divergent thinking and response creativity.

The Iowa Gambling Task (IGT) is a complex decision-making task in which participants select cards from four different decks over a total of 100 trials and win (and occasionally lose) variable amounts of money depending on the choice. Two of the decks (the disadvantageous decks) provide higher trial-by-trial wins (e.g., $100) but occasionally very large losses with a long-term negative outcome, whereas the other two decks (the advantageous decks) provide lower trial-by-trial wins (e.g., $50) together with smaller losses resulting in a positive long-term outcome. We used the Net Score (number of choices from the advantageous decks minus the number of choices from the disadvantageous decks) as a measure of reward learning, taken as an estimate of the participant’s expected reward value for the advantageous decks.

Preprocessing of Neuropsychological Data. Here we provide a detailed description of the preprocessing steps that isolated cognitive control and value-based decision-making from our target scores.

Covariates for nonexecutive cognitive functions. Executive function tasks require multiple cognitive processes (e.g., memory, language, and perception) for successful performance. Performance scores indexing verbal skills, visual-spatial reasoning, and both verbal and visual memory were derived from additional neuropsychological tests and used as covariates in our analyses by using multiple linear regression.

The Wechsler Adult Intelligence Scale (3, 4) (see ref. 5) for details on how the scores of two version were combined provides four index scores for major cognitive domains: Verbal Comprehension (VC), Perceptual Reasoning (PR), Working Memory (WM), and Processing Speed (PS). Verbal comprehension is not a pure measure of verbal abilities (5). Therefore, we removed the variance in VC explained by PR, WM, and PS via multiple linear regression and used the residuals as a measure of pure verbal abilities. Analogously, we removed the variance in PR explained by VC, WM, and PS and used the residuals as a measure of pure spatial abilities.

A verbal memory index was calculated from the Rey Auditory-Verbal Learning Test (AVLT) (6) by comparing the number of words recalled on the final of five learning trials with the number of words recalled from that list after a time delay [(AVLT.recall/AVLT.run5) × 100]. This index expresses verbal memory as the percent recall performance relative to the last learning block. Similarly, a visual retention score was computed from the Complex Figure Test (CFT) (2) [(CFT.recall / CFT.copy) × 100] and used as a measure of visual memory.

In conclusion, we calculated four covariates, whose variance we removed from the original executive test scores: (i) verbal abilities; (ii) spatial abilities; (iii) verbal memory; and (iv) visual memory.

Replacing missing data in the covariates. Our dataset was missing data for some of our executive tests and covariate scales. Table S1 lists the sample sizes for each neuropsychological test. Because the executive tests were our target scales, we used a subject-specific multiple linear regression for replacing data in the covariates only.

Thus, a missing data point for subject i on one of the covariates c (c_i) was predicted by using the following general linear model (GLM):

\[ c_i = \beta_0X_{i,-i} + \beta_1X_{1,i} + \beta_2X_{2,i} + \ldots + \beta_nX_{n,i}, \]

where \( \beta_0 \ldots \beta_n \) are parameter estimates and \( X_{1,i} \ldots X_{n,i} \) are n covariates other than c. The explanatory variables \( X_{1,i} \ldots X_{n,i} \) are composed of all other subjects except subject i (-i) who have valid data on these other covariates. In essence, we are using the data from all other subjects with valid data to predict the missing data point of subject i on the covariate c. This procedure ensures
that we have obtained the best estimates in an ordinary least-squares sense. This GLM is redefined for every subject based on that subject’s profile of available and missing data on the covariates so that we obtain the most precise estimate for missing covariate data on a subject-by-subject basis.

**Standardization of neuropsychological tests.** For the neuropsychological tests, we used normative data to convert the original scores to z-scores by using the mean and SDs from the respective age range of the normative sample. Where applicable, we also applied a correction for educational level. The following published norms were used for our cognitive control tests: (i) TMT, Mitrushina et al. (7); (ii) WCST, Heaton et al. (8); (iii) Stroop Test, Golden et al. (9); and (iv) COWA, Benton (10).

The Net Score (advantageous – disadvantageous) choices of the IGT was used in its original (unstandardized) form.

**Inverting scale direction for errors and reaction-time scales.** Some scales of our executive tests are “reversed” such that a better performance would be indicated by a lower score (WCST perseverative errors, TMT scores). Thus, to provide a uniform score interpretation, we inverted the direction of these scales with the following transformation:

\[ y_i = -(y_i - \bar{y}) \]

where \( y_i \) is a data point of subject \( i \) on any of the above mentioned scales. Thus, all scores used in the study were quantified so that higher scores indicate better performance.

**Removing variance explained by the covariates.** Using the covariates from above, we removed their variance from the five target scores \( y \) by submitting the residuals of the following GLM to the lesion overlap analysis:

\[ \hat{y} = \beta_1 \text{ (verbal abilities)} + \beta_2 \text{ (spatial abilities)} + \beta_3 \text{ (verbal memory)} + \beta_4 \text{ (spatial memory)}. \]

Residuals \( r \) of this GLM were calculated as

\[ r = y - \hat{y}. \]

These residuals were then analyzed for the main non-parametric VLSM analysis.

**Preprocessing of Neuroanatomical Data.** The visible lesion of each patient’s MRI or CT scan (CT only when MRI was contra-indicated) was traced manually slice by slice on corresponding regions of a reference brain (11) using MAP-3 (12) by a neuroanatomical expert (H.D.) who has demonstrated high reliability (13). Tracing was only carried out when the matching between corresponding slices in the lesioned brain and the reference brain was achieved with confidence. Therefore, lesions were only mapped if (i) they were clearly distinguishable from the ventricular system, which might be dilated following the neurological incident, (ii) there were no signs of cortical atrophy, and (iii) there were no imaging artifacts in the MRI/CT scans. Because of the manual tracing technique, no automated spatial normalization was necessary. Lesion maps were resampled to a 1-mm isotropic voxel size, smoothed with a Gaussian kernel (4 mm full width at half maximum), and binarized by using a threshold of 0.2.

**Characterizing Lesion Location.** Of the entire sample of 344 patients, 174 had left hemispheric lesions, 122 had right hemispheric lesions, and 48 had bilateral lesions. Given the neuroanatomical focus of this study on the PFC, it is relevant to specify how many patients had damage to the PFC and to present lesion density maps for these patients. Fig. S3A shows a histogram of sample sizes for patients with different degrees of PFC overlap. The PFC mask (Fig. S3B) was defined as all regions anterior to the precentral gyrus on the lateral surface and all regions anterior to and including the ACC and SMA on the medial wall (the insular cortex was not included in the PFC mask). Fig. S3C shows the lesion density for all patients with no overlap with the PFC mask, and Fig. S3D shows the lesion density for subjects with at least 40% overlap with the PFC mask. Fig. S1 shows the lesion density map for all patients on each of the target scores used for the statistical analysis below.

For the scatter plot of IGT Net Score vs. extent of vmPFC damage (Fig. 2C) we defined a region of interest (ROI) based on the VLSM results for the IGT. All voxels with BM test values of 4 or greater were included. Overlap was then computed as the number of voxels of each patient’s lesion covering this ROI divided by the number of the voxels in the ROI.

**Statistical Lesion Analysis.** Nonparametric voxel-based lesion symptom mapping (15) uses the nonparametric Brunner–Munzel test, a nonparametric variant of the two-sample t test that allows for heteroscedasticity between the groups (16). This procedure is implemented in Non-Parametric Mapping (NPM), which is part of the MRcron software package (www.mccauslandcenter.sc.edu/mrcron/mrcron/index.html). We used a statistical threshold of 5% false discovery rate (FDR) (17) to control for false positives in the context of multiple comparisons.

Statistical power maps (18) are also implemented in NPM and were computed by using the nonparametric Wilcoxon–Mann–Whitney probability to estimate a power threshold. As an example, had our sample size been 10 patients of whom (at a particular voxel) only 3 had a lesion, then the most extreme ranking would be \( W = 6 \) (sum of the rank 1, 2, and 3), which corresponds to a \( P \) value of 0.01667 or a Z value of 2.13. Therefore, if our statistical threshold corresponding to a 5% FDR threshold had been \( Z = 2.56 \), we would not expect to detect this voxel no matter how large the effect size actually is.

**Difference Images.** Difference images are suited to highlight differences between two tasks. In these images, the mean difference between individual z-scores of two different tasks are color coded and projected onto a template brain in a voxel-wise manner. We computed the z-score of the cognitive control factor scores (Fig. 3A) based on the mean and SD of the entire sample and computed the pair-wise difference scores between the IGT and the cognitive control factor (Fig. 3C). A positive mean difference value between IGT and the executive factor (red areas in Fig. 3C) maps out reduced cognitive control, whereas a negative value (blue areas in Fig. 3C) delineates a reduction in decision-making performance on the IGT.

**Lesion Overlap Between Cognitive Tasks.** We computed the lesion overlap by counting the number of overlapping voxels with a significant lesion effect in all pairs of neuropsychological scores (e.g., STROOP ∩ COWA) divided by the number of significant voxels in either score (e.g., STROOP or COWA). This ratio normalizes the overlap measure by different “base rates” yielding a potentially asymmetrical overlap matrix (Fig. 2B). Identical entries in corresponding fields in the overlap matrix (e.g., IGT and TMT) indicate that the overlap between the two tests (if any) covers a nearly identical percentage of each test. If, however, as in the example of STROOP and COWA, the overlap measure is grossly asymmetrical, then this asymmetry indicates that the area covered by one test (STROOP) is “included” in the other (COWA). In summary, the overlap matrix reveals in a condensed display the mutual exclusivity of the neural correlates of each test (overlap measure near zero) and potential “inclusion” phenomena between two tests by using test-specific base rates in the calculation of the percent overlap measures.

Fig. S1. Lesion density maps for each of the five target scores. The color bar encodes the number of subjects with a lesion in each voxel in the brain. Neurological convention (R = right) is used in all of the images.
Fig. S2. Maps of statistical power for each of the five target scores. Areas in red have sufficient statistical power to detect a significant lesion deficit effect at $P < 0.05$ (FDR). These nonparametric power calculations are implemented in Non-Parametric Mapping, which was used to compute nonparametric voxel-based lesion symptom maps. (see SI Materials and Methods for details). The areas in yellow show the overlap of statistical VLSM results at $P < 0.05$ (FDR) with the power map.
Fig. S3. (A) Histogram of number of patients with different degrees of overlap with an anatomical PFC mask shown in B. The green and red bars show the number of patients with complete data on all cognitive control tasks (green) and all tasks (red) and different degrees of PFC overlap (blue). (C) Lesion density map for all patients without any overlap with the PFC mask. (D) Lesion density maps for all patients with at least 40% of overlap with the PFC mask separated for each target score. In all images, neurological convention (R = right) is used.
Fig. S4. Lesion overlap analysis (VLSM) in the smaller sample with complete neuropsychological data (n = 62) and comparison with primary findings displayed in neurological convention (R = right). (A) The follow-up analysis was computed in those regions that exhibited a significant effect in the primary analysis (BM test, \( p < 0.05 \), FDR) (grayed out areas showing the disjunction mask of impairment on any one of the tasks in the primary analysis). Lesion density in the smaller sample was focused in the PFC. (B) Overlap (yellow) between primary findings (green) and the follow-up findings in the smaller sample (red) revealed a substantial amount of overlap for TMT, WCST, IGT and, to a lesser extent, for COWA and STROOP.
Sex differences for the five target neuropsychological tasks. The same VLSM analysis as in Fig. 1 was conducted for both genders (BM test, $P < 0.05$, FDR). Significant lesion impairment effects are shown in red for male and in green for female participants (overlap in yellow) in neurological convention ($R =$ right). The male and female sample sizes were as follows: TMT $f = 108$, $m = 127$; WCST $f = 86$, $m = 129$; STROOP $f = 50$, $m = 65$, COWA $f = 148$, $m = 173$; IGT $f = 43$, $m = 67$.

**Table S1. Pearson correlations between target scores and covariates.**

<table>
<thead>
<tr>
<th></th>
<th>WCST</th>
<th>STROOP</th>
<th>TMT</th>
<th>COWA</th>
<th>IGT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r$</td>
<td>$n$</td>
<td>$r$</td>
<td>$n$</td>
<td>$r$</td>
</tr>
<tr>
<td>WCST</td>
<td>1</td>
<td>215</td>
<td>0.22</td>
<td>110</td>
<td>0.43</td>
</tr>
<tr>
<td>STROOP</td>
<td>1</td>
<td>115</td>
<td>0.24</td>
<td>107</td>
<td>0.39</td>
</tr>
<tr>
<td>TMT</td>
<td>1</td>
<td>236</td>
<td>0.28</td>
<td>232</td>
<td>0.20</td>
</tr>
<tr>
<td>COWA</td>
<td>1</td>
<td>326</td>
<td>0.08</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>IGT</td>
<td>1</td>
<td>111</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Intercorrelation of target scores after removal of covariates**

<table>
<thead>
<tr>
<th></th>
<th>WCST</th>
<th>STROOP</th>
<th>TMT</th>
<th>COWA</th>
<th>IGT</th>
</tr>
</thead>
<tbody>
<tr>
<td>WCST</td>
<td>1</td>
<td>215</td>
<td>0.10</td>
<td>110</td>
<td>0.43</td>
</tr>
<tr>
<td>STROOP</td>
<td>1</td>
<td>115</td>
<td>0.11</td>
<td>107</td>
<td>0.58</td>
</tr>
<tr>
<td>TMT</td>
<td>1</td>
<td>236</td>
<td>0.21</td>
<td>232</td>
<td>0.20</td>
</tr>
<tr>
<td>COWA</td>
<td>1</td>
<td>326</td>
<td>0.02</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>IGT</td>
<td>1</td>
<td>111</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The correlation coefficients ($r$) are computed on the listed sample sizes ($n$), which vary because not all subjects were administered all tasks. The total number of patients available for each test are listed in the main diagonal of the sample sizes ($n$) in bold type. Because a lower score indicates better performance on WCST and TMT, these scales were reversed for an easier comparison with other correlations (see Materials and Methods for details).