Supporting Information

Barttfeld et al. 10.1073/pnas.1301353110

SI Materials and Methods

Partial Report (Behavioral) Experiment. The partial report (PR) experiment was programmed in Python (www.python.org). Stimuli were presented on a 19-inch screen (resolution of 800 × 600 pixels) placed at a distance of 73 cm (Fig. 1*A*). Letter fonts were uppercase Times New Roman with a font size of 1.2°. Letters were chosen randomly from the alphabet (26 symbols), without repetition. Eight inter-stimulus intervals (ISIs, the period between the offset of array to onset of cue) were used (24, 71, 129, 200, 306, 506, 753, and 1,000 ms). Each observer first completed a practice block of 50 trials. Subjects completed four blocks (384 trials). In each block all positions (total of eight) and all ISIs (total of eight) were uniformly sampled in random order. In each trial, eight letters were presented simultaneously for 106 ms. The eight letters were arranged on a circle, around the fixation point (eccentricity 5.2°). A red dot on an array of blue dots indicated the position of the target. Participants were asked to report, using a standard keyboard, the letter presented in the position cued by the red circle, which remained on screen until subject's response. Random performance is 1/26, because for each trial, subjects had to choose 1 of 26 possible letters. Subsequently, participants had to report the confidence of their response with an ad hoc bar placed in the center of the screen and composed of 13 division marks and two labels: "0% confidence" at the extreme left of the bar, and "100% confidence" at the extreme right of the bar ("0% seguro" and "100% seguro," in Spanish) (Fig. 1*A*). The experiment lasted ∼45 min.

Estimation of Individual Metacognitive Ability. To estimate subjects' metacognitive ability, we calculated a type-II receiver operating characteristic (ROC) curve for each participant (1), categorizing as a hit (*H*) a high confidence response after a correct decision, and as a false alarm (*FA*) when subject reported high confidence after a wrong decision. ROC curves were anchored at [0, 0] and [1, 1]. Curves were plotted using the cumulative probabilities of $H = \mathrm{p}(\mathrm{confidence} = = \mathrm{i} | \mathrm{correct\ \ trial})$ and $FA = \mathrm{p}(\mathrm{confidence} = = \mathrm{i} | \mathrm{error\ \ trial})$, where *i* represents the bin size, set at 10, to categorize the continuous subjective responses. A ROC curve that bows sharply upward indicates that the probability of being correct rises rapidly with confidence; conversely, a flat ROC function indicates a weak link between confidence and accuracy. We calculated the area between the ROC curve and the *x* axis (possible values range from 0 to 1) as an estimate of a subject's introspective ability.

Functional MRI Preprocessing. Functional MRI (fMRI) data were preprocessed using statistical parametric mapping (SPM5) software (http://www.fil.ion.ucl.ac.uk/spm). The first four image acquisitions of the task-free functional time series were discarded to allow for stabilization of the MR signal. The remaining 220 volumes (360 for the Glasgow dataset) underwent the following preprocessing steps: slice timing, realignment to the first scan, normalization, and smoothing [8 mm full width at half maximum (FWHM) isotropic Gaussian kernel]. Normalization to the Montreal Neurological Institute (MNI) template was computed on the structural image and then applied on functional data. Following the procedure of Fox et al. (2), we removed by regression the six parameters resulting from rigid body correction for head motion.

fMRI Analysis. Analyses were done using Matlab (MathWorks) and R software for statistics (3). To study the relation between functional connectivity and metacognitive ability, we used a previously defined set of regions of interest (ROIs) (4) composed of 141 ROIs comprising five functional systems [frontoparietal (FP), cinguloopercular (CO), default brain network (DBN), sensorimotor (SM), and occipital (OC)] (Fig. S1). Systems differ within a narrow range in the number of ROIs they contain, varying between 21 (FP) and 33 (SM) ROIs. We built ROIs containing the voxels of a 5-mm sphere around each ROI coordinate, as defined in ref. 4. For each ROI, a time series was extracted for each individual and each state, using the Marsbar software package (http://marsbar.sourceforge.net). These regional fMRI time series were then used to construct a 141-node functional connectivity network for each subject and attentional state. We used wavelet analysis to construct correlation matrices from the time series. We followed the procedures exactly as described by Supekar and collaborators (5): We applied a maximum overlap discrete wavelet transform (MODWT) to each of the time series to obtain the contributing signal in the following three frequency components: scale 1 (0.13–0.25 Hz), scale 2 (0.06–0.12 Hz), and scale 3 (0.01–0.05 Hz). Several studies have suggested that wavelet filtering might be better suited for fMRI time series than Fourier filtering (5, 6). All subsequent analysis was done based on the scale 3 component, whose frequency lies in the range of slow frequency correlations of the default network (2, 7). For each attentional state *s* (*interoceptive*, *exteroceptive*, or *resting*) and participant *p* we measured a 141 × 141 connectivity matrix $C_{s,p}$. The matrix entry $C_{s,p}(i,j)$ indicates the temporal correlation of the average fMRI signal of ROIs *i* and *j*, which henceforth is referred to as functional connectivity.

To study functional connectivity correlates of type-I and type-II performances (Fig. 2 *A* and *B*) we conducted an across-subjects bivariate linear regression, using the least squares method, between each entry *ij* of the connectivity matrix $C_{s,p}(i,j)$ and type-I and type-II performances in the PR task. This way we obtained a matrix $B_{s,r}(i,j)$ per attentional state *s* and regression *r* to type-I or type-II performance, in which each entry *ij* represents the dependence or beta (β) value for the connectivity between ROI(*i*) and ROI(*j*), and type-I or type-II performance. Fig. S2 shows the $B_{s,r}$ matrices. We also calculated the *R*-squared values, to measure the amount of total variance in the fMRI data explained by the linear model (Fig. S5). Visualizations of the $B_{s,r}(i,j)$ values in glass brains were done using custom software in Python for the Anatomist/BrainVisa software. We projected into a glass brain a link between ROI(*i*) and ROI(*j*) if the $B_{s,r}(i,j)$ value for that connection exceeded a certain threshold (Fig. 2 *A* and *B* and Tables 1 and 2 list the 15 ROIs with the highest rank in the number of connections whose β-value were positive and exceeded the threshold, for type-I and type-II performances). The threshold was set to a value of 3 SDs above the mean of the distribution of $B_{s,r}$ obtained from networks measured under exteroceptive state (for dependences to type-I performance) and interoceptive state (for dependences to type-II performance). We chose these particular states to calculate the threshold for each performance because, for networks derived from those attentional states, the average of $B_{s,r}$ reached the highest value. Thresholds are arbitrary, but they are used only for visualization purposes and play no role in statistical analysis.

To search for main effects and interactions of type-I and type-II performances on functional connectivity and attentional states, we conducted an analysis of covariance (ANCOVA). We measured the average connectivity matrix $\hat{C}_{s,p}$, a 5 × 5 matrix resulting from all possible pairings between FP, CO, DBN, SM,

and OC for each subject p and attentional state s. Each entry nm of this matrix represents the average connectivity between system n and system m. This matrix was submitted to a single ANCOVA as dependent variable, with type-I and type-II performance as continuous regressors, and attentional state (exteroceptive, resting, interoceptive) as within-subjects factor and subject identity as a random-effect factor. Previously to running the ANCOVA test, we assessed that our data satisfied assumptions of normality (Shapiro–Wilk normality test, $W = 0.99$, P value >0.05) and homogeneity of variances (Bartlett's K-squared = 1.62, df = 2, P value >0.1). As we report in the main text, type I and type II performances are not completely independent but their correlation is not strong enough to impair the ANCOVA analysis with these factors.

To study the effect of both types of performance on connectivity within each attentional state, we followed this analysis with three independent ANCOVA, one per attentional state. We measured the average connectivity $\hat{C}_{s,p}$ matrix between functional systems for each subject p and attentional state s, and submitted it to an ANCOVA, with type-I and type-II performances as continuous regressors, and subject identity as a random-effect factor.

To create Fig. 3 *A* and *B*, we conducted a one-sample t test analysis comparing β-value changes across attentional states at a functional network level. Unlike the ANCOVA analysis, this analysis is performed directly on the β-values ($B_{s,r}$ matrices), not on the functional connectivity (temporal similarity between time series of ROI pairs) values ($C_{s,p}$ matrices). For each pair of functional systems (n,m) we consider all of the $B_{s,r}(i,j)$ where ROI(i) belongs to system n and ROI(j) belongs to system m. For each pair of systems (n,m) we obtained a distribution of β-values (i.e., all ROIs i and j dependences). For example, for the SM–FP system pair, because SM is composed of 33 ROIs and FP is composed of 21 ROIs, the β-distribution is composed of 33 × 21 β-values. To obtain the distribution of β for a within-network connectivity (for example SM–SM), only the upper diagonal of the β-matrix is averaged because connectivity between pairs of ROIs is symmetric and excluding the triangular part because the correlation between a ROI and itself is trivially 1. For each pair of systems (n,m) the statistical significance of the dependence of this specific connection with performance is assessed comparing the mean value of the distribution of dependences against zero, by means of a one-sample t test and correcting for multiple comparisons. We display a link between two functional systems if the t value for that connection is higher than 5.35, corresponding to a P value of 10^{-5}, Bonferroni corrected for multiple comparisons (two-tailed one-sample t test, 15 pairs of systems (n, m) × 3 attentional states × 2 types of performance).

To create Fig. S3, we followed the same procedure as that for Fig. 2 *A* and *C*. Because this work includes two datasets, obtained with different scanners, we conducted the bivariate regression between functional connectivity and type-I and type-II performances separately for the two datasets. The objective was to investigate whether our main finding, the interaction of the effects between attentional state and type-II performance on connectivity, was observed in each dataset. Fig. S3 shows $B_{s,r}$ matrices (similar to the ones shown in Fig. S2) for type-II performance for all attentional states for all subjects in the study (Fig. S3, *Top* row), for the Buenos Aires dataset (Fig. S3, *Middle* row), and for the Glasgow dataset (Fig. S3, *Bottom* row). Despite having fewer subjects, in $B_{s,r}$ matrices from both datasets we observe the increment in β-values for networks measured under interoceptive state, visually depicted in this figure as an increase of β-values involving frontoparietal, sensorimotor, and occipital systems.

To create Fig. S4, we conducted the same analyses described above, using time series of 3-, 4-, 5-, 6-, and 7.22-min length, which is equivalent to 90, 120, 150, 180, and 220 scans. For the Glasgow dataset, we extended this analysis to 12 min. For each block duration, we measured a connectivity matrix ($C_{s,p}$) per subject p and attentional state s, and linearized each connectivity matrix to obtain a vector of length N = 19,881 (141 × 141). We then calculated the dot product $a.b = \sum_{i=1}^{N} a_i.b_i$, between the vector, corresponding to the linearized $C_{s,p}$ for the time series of different length, and vector b, the linearized $C_{s,p}$ obtained from the full-length time series. This way we quantified the similarity between a $C_{s,p}$ matrix obtained from the full-length time series and the $C_{s,p}$ matrices obtained from shorter time series. A value of 1 means perfect concordance of values, whereas a value of 0 implies full orthogonality. Similarity values approach monotonically to 1 as the time series increases (Fig. S4). Even using time series of 5 min, the projection into the full-length matrix yields a similarity above 0.95, showing that time series length hardly affected the results obtained.

Fig. S5 was generated to explore how much of the total variance in the fMRI data was explained by the two behavioral regressors (type-I and type-II performances). We calculated the R-squared value, quantifying the proportion of variance that the model explains. We collapsed R-squared values across attentional states (variations due to attentional state were minimal) to obtain a single distribution of R-squared values. Fig. S5 shows that only a minor portion of the total variance is explained by the model, including both behavioral regressors, as expected by the noisy nature of the fMRI data and the complexity of the sources of variation in brain activation during resting state fMRI, showing that the connectivity between any pair of ROIs cannot be strongly related to another variable.

1. Fleming SM, Weil RS, Nagy Z, Dolan RJ, Rees G (2010) Relating introspective accuracy to individual differences in brain structure. *Science* 329(5998):1541–1543.
2. Fox MD, et al. (2005) The human brain is intrinsically organized into dynamic, anti-correlated functional networks. *Proc Natl Acad Sci USA* 102(27):9673–9678.
3. R Development Core Team (2008) R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna). Available at www.R-project.org.
4. Dosenbach NU, et al. (2010) Prediction of individual brain maturity using fMRI. *Science* 329(5997):1358–1361.
5. Supekar K, Menon V, Rubin D, Musen M, Greicius MD (2008) Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. *PLOS Comput Biol* 4(6): e1000100.
6. Achard S, Salvador R, Whitcher B, Suckling J, Bullmore E (2006) A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J Neurosci* 26(1):63–72.
7. Raichle ME (2009) A paradigm shift in functional brain imaging. *J Neurosci* 29(41): 12729–12734.

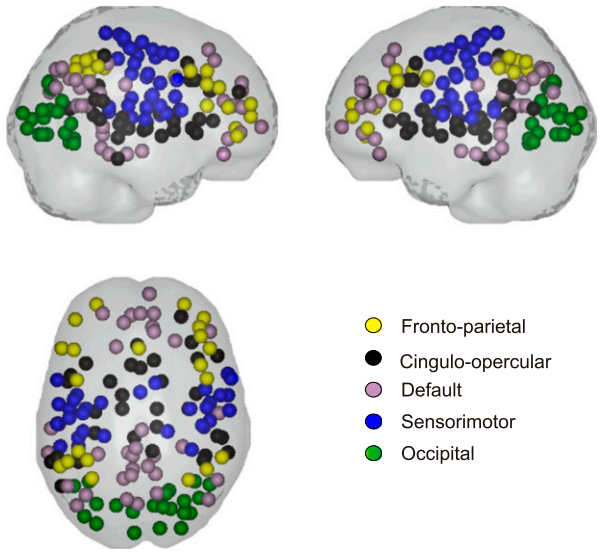

Fig. S1. A total of 141 regions of interest (ROIs) used in the analysis.

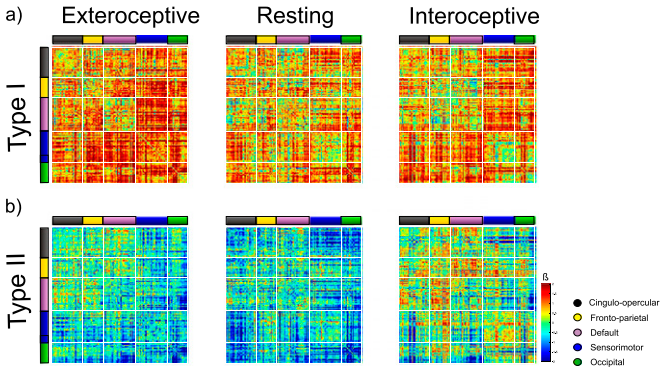

Fig. S2. Organization of functional brain networks according to subjects' individual metacognitive ability and performance. (*A*) Slope (β) values for type-I performance of the bivariate regression between functional connectivity and both types of performance. (*B*) Slopes for type-II performance of the bivariate regression between functional connectivity and both types of performance.

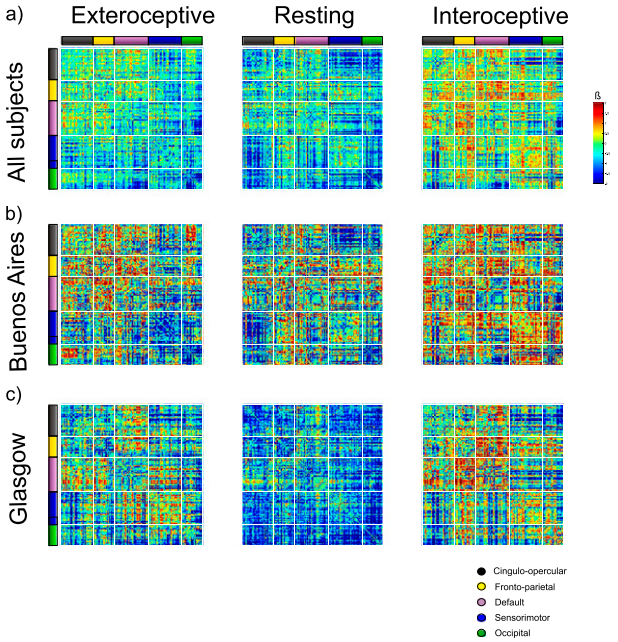

Fig. S3. Dependence of functional brain connectivity with type-I and type-II performances for the two datasets. (*A*) Matrix of slope (β) values for type-II performance of the bivariate regression between functional connectivity and both types of performance, for all subjects in the experiment. (*B*) Matrix of β-values for type-II performance of the bivariate regression between functional connectivity and both types of performance for the 16 subjects from Buenos Aires. (*C*) Matrix of β-values for type-II performance of the bivariate regression between functional connectivity and both types of performance, for the 9 subjects from Glasgow. Despite the low number of subjects, there is still a visible effect of positive β-values involving frontoparietal connectivity systems in the interoceptive state.

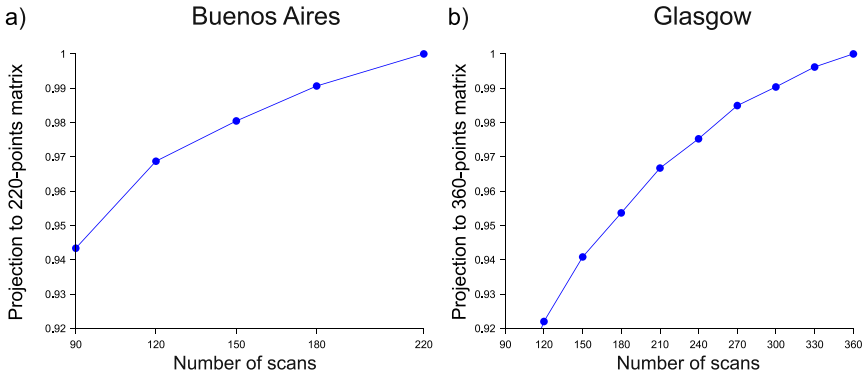

Fig. S4. Connectivity analysis using shorter time series. (*A*) Projection of each matrix (3, 4, 5, 6, and 7.22 min or 90, 120, 150, 180, and 220 scans) into the 220-time points matrix for the subjects from Buenos Aires. A value of 1 means perfect concordance of values, whereas a value of 0 implies full orthogonality. Matrices approached monotonically to the matrix calculated with full-length time series, as the time series length used to calculated them increased. Even using time series of 5 min, the projection into a the full-length matrix yields a similarity above 95% (*B*). (*B*) Projection of each matrix (3–12 min, or 90, 120, 150, 180, 210, 240, 300, 330, and 360 scans) into the 360-time points matrix for the subjects from Glasgow.

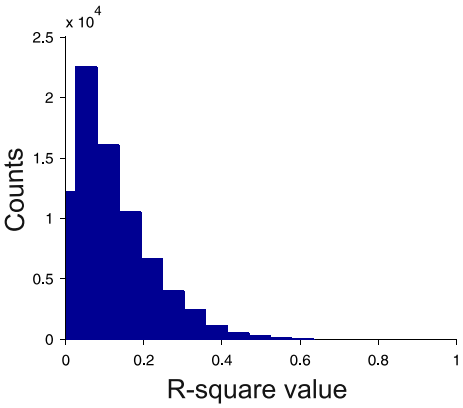

Fig. S5. Variance in the fMRI data explained by the regressors of interest. Distribution of all *R*-squared values of the bivariate regression between connectivity and type-I and type-II performances, collapsed across all attentional states and regressors. Almost all values are lower than 0.5, showing that the amount of variance explained by the two behavioral regressors is moderate.