Distinct patterns of functional brain connectivity correlate with objective performance and subjective beliefs

Pablo Barttfeld[a,b], Bruno Wicker[a,c], Phil McAleer[d], Pascal Belin[d], Yann Cojan[a], Martín Graziano[a], Ramón Leiguarda[e], and Mariano Sigman[a,f,1]

[a]Laboratory of Integrative Neuroscience, Physics Department, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and Instituto de Física de Buenos Aires, Conicet, Pabellón 1, Ciudad Universitaria, 1428 Buenos Aires, Argentina; [b]Cognitive Neuroimaging Unit, Institut National de la Santé et de la Recherche Médicale (INSERM), 91191 Gif sur Yvette, France; [c]Institut de Neurosciences de la Timone, Centre National de la Recherche Scientifique Unité Mixte de Recherche 7289 and Aix-Marseille University, 13385 Marseille, France; [d]Department of Psychology, Glasgow University, Glasgow G12 8QB, United Kingdom; and [e]Fundación para la Lucha contra las Enfermedades Neurológicas de la Infancia, 1428 Buenos Aires, Argentina; [f]Universidad Torcuato Di Tella, Almirante Juan Saenz Valiente 1010, C1428BIJ Buenos Aires, Argentina

Edited by Michael I. Posner, University of Oregon, Eugene, OR, and approved June 3, 2013 (received for review January 21, 2013)

The degree of correspondence between objective performance and subjective beliefs varies widely across individuals. Here we demonstrate that functional brain network connectivity measured before exposure to a perceptual decision task covaries with individual objective (type-I performance) and subjective (type-II performance) accuracy. Increases in connectivity with type-II performance were observed in networks measured while participants directed attention inward (focus on respiration), but not in networks measured during states of neutral (resting state) or exogenous attention. Measures of type-I performance were less sensitive to the subjects' specific attentional states from which the networks were derived. These results suggest the existence of functional brain networks indexing objective performance and accuracy of subjective beliefs distinctively expressed in a set of stable mental states.

interoception | metacognition | resting-state | partial-report-paradigm

Decisions often bear upon other decisions, as when we seek a second medical opinion before undergoing a risky surgical intervention. These "metadecisions" are mediated by confidence judgments, the degree to which decision makers consider that their choices are likely to be correct. Confidence judgments can be severely distorted: People may lack confidence when responding correctly and reciprocally, be very confident of incorrect responses (1–6). In classic perceptual tasks followed by a confidence report, one can distinguish between (*i*) the ability to correctly discriminate between stimulus alternatives, referred to as type-I performance, and (*ii*) the ability of confidence judgments to discriminate between correct and incorrect responses, referred to as type-II performance (2, 7). The objective of this work is to investigate whether functional brain networks distinctively covary with type-I and type-II performance.

Network organization of resting state functional brain activity can account for individual differences in several cognitive functions (8–13). These studies rely on networks derived from the "resting state" (14, 15). Recently, Tang et al. (16) showed the formation of distinct brain networks in the maintenance of three well-defined mental states that vary the focus of attention: resting, alert, and meditation states (16). Here we capitalize on this idea, deriving functional brain networks for each individual, varying the focus of attention toward internal states (interoception, focus on respiration), external stimulus (exteroception), or remaining in a resting state of free thought.

Interoception (generically defined as the ability to detect subtle changes in bodily systems, including muscles, skin, joints, and viscera) (17), is closely related to metacognition of agency (18, 19). We reasoned that this may more generally reflect a partially overlapping system regulating attention to internal states, including interoception (focus on body systems) and metacognitive ability (focus on internal thoughts and feelings). Hence, our working hypothesis is that increases in functional connectivity with the quality of subjective judgments (type-II performance) should be more sensitive in networks expressed while attention is directed inward compared with networks obtained in other attentional states.

We measured functional networks in three different attentional states: exteroceptive attention (detecting an oddball within a sequence of sounds), interoceptive attention (focusing on respiration), and resting state (relaxing without falling asleep), while a sequence of tones was presented at a very low volume in all states. We then investigated the covariance of functional connectivity, measured in different attentional states, with type-I and type-II performance in a perceptual decision task.

Results

After the functional MRI (fMRI) recordings, participants performed a partial report (PR) experiment, identifying a letter in a cued location of a cluttered field (20) and indicating the degree of confidence in their response in a continuous scale (Fig. 1*A*) (3). Type-II performance can be quantified by measuring the area under the receiver operating characteristic curve (A_{ROC}) (2, 21). This nonparametric test estimates the degree of overlap between confidence distributions for correct and error trials. Type I and type II varied widely between subjects (Fig. 1 *B* and *C*) with a significant correlation ($r = 0.74$, $P < 0.001$) but with sufficient dispersion to allow a reliable simultaneous regression of the fMRI signal to both factors.

To test whether coherence in spontaneous activity between brain regions in different attentional states covary with type-I and type-II performance, we conducted a functional connectivity analysis, based on 141 standard previously defined cortical regions of interest (ROIs) (22). Following the work of Dosenbach and colleagues (22, 23), ROIs were grouped in five different functional systems: frontoparietal (FP), cinguloopercular (CO), default brain network (DBN), sensorimotor (SM), and occipital (OC) (Fig. S1).

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

Author contributions: P. Barttfeld, B.W., M.G., and M.S. designed research; P. Barttfeld, B.W., P.M., and P. Belin performed research; P. Barttfeld, Y.C., M.G., and M.S. analyzed data; and P. Barttfeld, B.W., Y.C., R.L., and M.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: sigman@df.uba.ar.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1301353110/-/DCSupplemental.

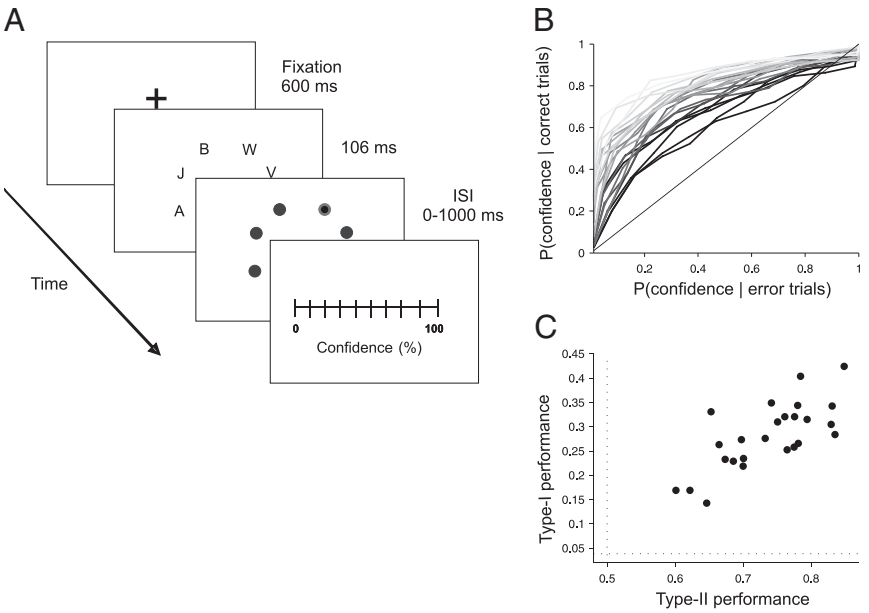

Fig. 1. (*A*) Subjects performed a partial report task experiment identifying a letter in a cued location of a cluttered field (3) and subsequently indicating the degree of confidence in their response in a continuous scale. (*B*) Individual ROC curves reflect a broad variability in metacognitive accuracy. (*C*) Type-I and type-II performance show a significant correlation across individuals, but with sufficient dispersion to perform a reliable bivariate regression to both factors. Dotted lines mark random type-I (1/26) and type-II (0.5) performance.

is referred as functional connectivity. To investigate connectivity changes associated with type-I and type-II performance, we conducted an across-subjects bivariate linear regression, between each entry of the correlation matrix $C_{s,p}$ and type-I and type-II performances of the subject in the PR task. This led to six matrices of beta (β) values, $B_{s,r}(i,j)$, one per attentional state s and regression r to type-I or type-II performance (Fig. S2). As an example, a positive value of $B_{resting,type\text{-}I}(i,j)$ indicates that connectivity between ROI(i) and ROI(j) increases with type-I performance for networks measured in the resting state.

For visualization purposes, we projected $B_{s,r}(i,j)$ values exceeding a threshold of 3 SDs into glass brains (Fig. 2 *A* and *B*). For type-I performance (Fig. 2*A*), the vast majority of values were positive, showing a marked tendency of increased connectivity with type-I performance. On the contrary, $B_{s,r}(i,j)$ values indexing how connectivity varies with type-II performance—$B_{s,type\text{-}II}(i,j)$—showed different patterns across states (Fig. 2*B*). For networks measured in the exteroceptive and resting states, the vast majority of values were negative, indicating a marked tendency of decrease in connectivity with type-II performance. This pattern was different for networks measured in the interoceptive state. The regression revealed a dense distribution of positive $B_{interoceptive,type\text{-}II}$ values for ROIs localized within the frontoparietal system and negative $B_{interoceptive,type\text{-}II}$ values within

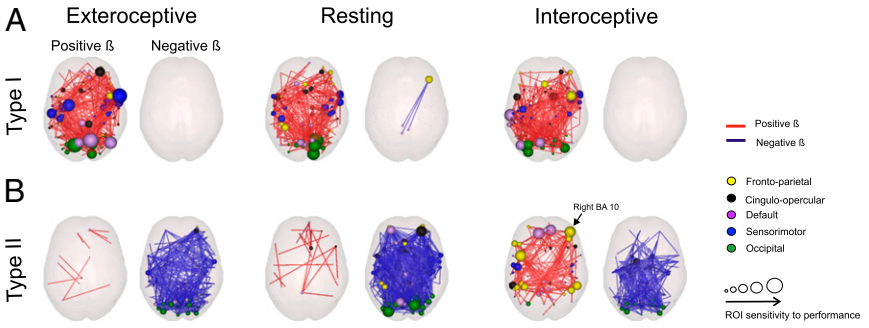

Fig. 2. Dependence of functional brain connectivity with type-I and type-II performance measured by a bivariate regression of connectivity to both factors. Red links denote positive beta (β) values (connectivity increases with performance). Blue links negative β-values (connectivity decreases with performance). For visualization purposes only β whose absolute value exceeded 3 SDs were depicted. For each ROI, the size of the sphere denotes the number of connections exceeding 3 SDs. Color indicates the functional system to which the ROI belongs. (*A*) β-Values for type-I performance. ROIs whose connectivity varies positively with type-I performance were mostly located in medial and posterior brain regions. The four ROIs with the highest rank in the number of connections exceeding a threshold (for networks measured under exteroceptive state, the state with the highest average β-value) are in dorsal frontal cortex ([60, 8, 34]), occipital cortex ([−16, −76, 33]), the precuneus ([11, −68, 42]) and parietal cortex ([−26, −8, 54]) (see Table 1 for the top 15 ROIs), which is consistent with previous findings of connectivity in the resting state predicting visual performance (11). (*B*) β-Values for type-II performance. The spatial distribution of ROIs whose connectivity increases for increasing type-II performance was mostly located in frontal regions in the FP and DBN. The 4 ROIs with the highest rank in the number of connections exceeding a threshold (for networks measured under interoceptive state, the state with the highest average β-value) are in the ventromedial prefrontal cortex ([−11, 45, 17] and [9, 51, 16]), the dorsal frontal cortex [−42, 7, 36] and the anterior prefrontal cortex [42, 48, −3] (see Table 2 for the top 15 ROIs).

medial and occipital regions, involving nodes from the occipital and cinguloopercular systems (Fig. 2*B*, *Right*).

To quantify these observations, we first reduced the dimensionality of the connectivity matrix by collapsing all of the connections between ROIs belonging to each pair of functional systems to a single scalar value. For each state s and subject p, we derived the average (reduced) connectivity matrix $\hat{C}_{s,p}$, a 5 × 5 matrix resulting from all possible pairings between FP, OP, DBN, SM, and OC. Each entry (n,m) of this matrix represents the average connectivity between system n and system m. Note that the interaction between a functional system with itself reflects the interaction between all of the ROIs within this functional system and hence the diagonal elements of $\hat{C}_{s,p}$ are not trivially and maximally correlated. The entries of these matrices were submitted as dependent variables to an analysis of covariance (ANCOVA), with type-I and type-II performance as continuous regressors, attentional state (exteroceptive, resting, and interoceptive) as within-subjects factor and subject identity as a random-effect factor. The ANCOVA revealed a main effect for type-I performance [$F(1, 21) = 6.26$, $P < 0$ 0.05] but not for type-II performance [$F(1, 21) = 1.38$, $P > 0.1$]. In Fig. 2 it can be observed that there is a main trend of type-II effect (when regressed together with type I) to decrease connectivity, but this effect does not reach significance when analyzed in an ANCOVA as a main effect. Conversely, we observed a significant interaction between attentional state and type-II performance [$F(2, 42) = 3.72$, $P < 0.05$] and a nonsignificant interaction between type-I performance and attentional state [$F(2, 42) = 0.37$, $P > 0.5$]. These results show that functional connectivity relates to type-I performance in a way that is independent of the attentional state, whereas the relation between connectivity and type-II performance varies with the attentional state of the subjects.

Next, to investigate the sensitivity of functional networks, measured under different attentional states, to type-I and type-II performance, we submitted independently for each state s, the entries of the reduced connectivity matrix $\hat{C}_{s,p}$ to three independent intrastate ANCOVAs with type-I and type-II performance as continuous regressors, and subject identity as a random variable. The ANCOVA tests revealed a main effect of type-I performance on networks measured under exteroceptive and interoceptive states [exteroceptive: $F(1,21) = 4.95$, $P < 0.05$; interoceptive: $F(1,21) = 4.72$, $P < 0.05$]. We observed a main effect of type-II performance on connectivity only in the interoceptive state [$F(1,21) = 6.48$, $P < 0.02$], reaching a higher level of significance than all other comparisons. The effect of type-II performance was not significant in the exteroceptive [$F(1,21) = 0.88$, $P > 0.1$] or resting [$F(1,21) = 0.05$, $P > 0.1$] states.

The ANCOVA analyses described above show the global trends of how connectivity (accumulated over all pairs of functional systems) depends on performance and attentional state. They show that connectivity increases for more accurate type-I performers for all attentional states. Instead, the effect of type-II performance shows a more complex pattern that depends on the attentional state. However, the ANCOVA cannot describe which specific connections contribute to these effects. To this aim we performed a one-sample t test following the ANCOVA reported above. For each attentional state s, regressor (r, type-I or type-II performance), and pair of functional systems S_n S_m, we considered all of the $B_{s,r}(i,j)$ where ROI(i) belongs to system n and ROI(j) belongs to system m. Note that connectivity (and hence $B_{s,r}$) is symmetric. This analysis naturally extends to connections of a system with itself by considering the $B_{s,r}(i,j)$ values of all of the ROIs within one system (excluding of course the connectivity of a ROI with itself). Then, for each condition and pair of systems (n,m) the statistical significance of the dependence of this specific connection with performance was assessed comparing whether the distribution of dependences [i.e., $B_{s,r}(i,j)$ values for all ROI(i) - ROI(j) connections] differed from zero, by means of a one-sample t test and correcting for multiple comparisons. A significant negative t value indicates that the distribution of $B_{s,r}(i,j)$ values for that system pair (n,m) is shifted toward negative values (indicating a tendency to decrease connectivity between systems n and m as performance increases). Instead, a significantly positive t value indicates that the distributions of $B_{s,r}(i,j)$ for that system pair (n,m) is shifted toward positive values (indicating that connectivity between systems n and m increases as performance increases).

Fig. 3 represents these results displaying a link for each pair of functional systems when the t value of the dependence in connectivity is higher than 5.35, corresponding to a P value of 10^{-5}, Bonferroni corrected for multiple comparisons (Fig. 3 *A* and *B*). First, we observe that connectivity between many pairs varied with type I and type II. This shows that the effect of the ANCOVA described above does not result from a sparse distribution of changes but instead from broad and distributed changes in the connectivity pattern. Functional connectivity showed an overall increase in connectivity with type-I performance. A major core, including interactions between OC–CO–SM systems and the SM–DBN interaction, reached the highest levels of significance (Fig. 3*A*). This indicates that subregions within this core coherently increased their connectivity with increasing type-I performance. For type-II performance, connectivity between functional systems showed an overall decrease with performance for the exteroceptive and resting states. Only for networks measured under the interoceptive state, we observed increases in connectivity with

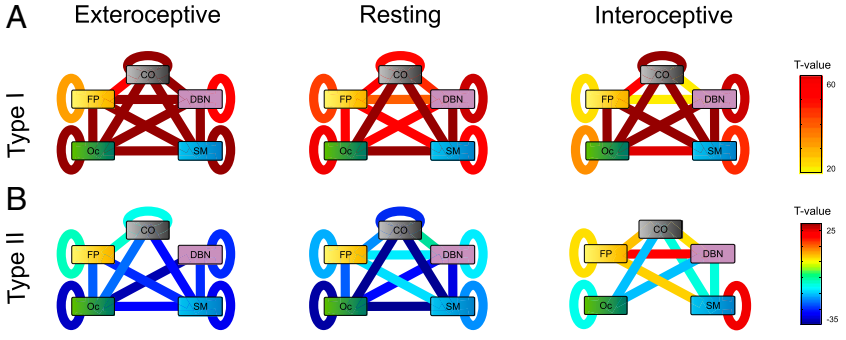

Fig. 3. Connectivity between functional systems and its relation with type-I and type-II performance. (*A*) T values from a one-sample t test analysis quantifying the magnitude of β-values for type-I performance averaged across system pairs, for all attentional states. Negative t values indicate a decrease in connectivity between systems n and m as type-I performance increases, whereas a significantly positive value indicates a connectivity increase between systems n and m increases as type-I performance increases (P value of 10^{-5}, Bonferroni corrected for multiple comparisons). (*B*) T values from a one-sample t test analysis quantifying the magnitude of β-values for type-II performance averaged across system pairs, for all attentional states.

Table 1. Regression to type-I performance, top 15 ROIs

MNI coordinates				
x	*y*	*z*	ROI label	Network
60	8	34	Dorsal frontal cortex	Sensorimotor
−16	−76	33	Occipital	Occipital
11	−68	42	Precuneus	Default
−26	−8	54	Parietal	Sensorimotor
45	−72	29	Occipital	Default
43	1	12	Ventral frontal cortex	Sensorimotor
17	−68	20	Postoccipital	Occipital
−55	−22	38	Parietal	Occipital
−36	−12	15	Mid insula	Sensorimotor
58	−3	17	Precentral gyrus	Sensorimotor
−9	−72	41	Occipital	Sensorimotor
27	49	26	Anterior PFC	Default
−41	−31	48	Postparietal	Cinguloopercular
33	−12	16	Mid insula	Sensorimotor
60	8	34	Dorsal frontal cortex	Sensorimotor

ROIs with the highest rank in the number of connections whose β-value for type-I performance exceeded a threshold of 3 SDs, for networks measured under exteroceptive state. MNI, Montreal Neurological Institute.

increasing type-II performance, specifically involving systems FP (to SM, DBN, OC, and FP itself), SM (to itself), and CO (to DBN). This reveals a core formed by reciprocal connections between ROIs belonging to the FP–DBN–CO systems whose connectivity shows distinct patterns of dependence with type-II performance in the interoceptive state compared with resting and exteroceptive states.

Discussion

We combined measures of objective performance, fluctuations of brain activity in different states, and subjective estimates of performance (24) to investigate which aspects of functional connectivity correlate with the wide variability observed in objective (type I) performance and metacognitive (type II) ability. Specifically, we examined whether increases in functional connectivity with type-II performance are distinctively manifested when attention is directed inward (focus on respiration). We found that connectivity in states of neutral (resting state) or exogenous attention globally decreased with increasing type-II performance. Instead, connectivity measured in the interoceptive state showed a more complex dependence with type-II performance: connectivity within a core formed by FP, SM, and DBN systems increases with type-II performance and connectivity between OC and CO systems decreases with type-II performance. Contrary to this state dependence observed in the covariation of connectivity with type-II performance, the relation between type-I performance and functional connectivity was less sensitive to the specific subjects' mental states from which the functional networks were derived.

We emphasize that these analyses are only correlational and do not imply any causality or directionality. Our hypothesis is that connectivity between brain regions should have an effect on (type I and type II) performance in a task. On the other hand, the ANCOVA analyses test how connectivity varies as a function of task performance and attentional state. We have tried not to use semantic descriptions involving causal relations (such as "predict" and "explain") but as a note of caution here we explicitly mention that all our analyses can only show a correlational and nondirectional relation between connectivity and performance.

As expected from previous anatomical (2), lesion (25), and transcranial magnetic stimulation (TMS) (7) studies, we found that the prefrontal cortex (PFC) was one of the regions whose connectivity was more sensitive to type II performance. This includes Brodmann area 10 and dorsolateral prefrontal cortex, known to play an important role in linking objective performance to subjective beliefs (2, 21). However, beyond these specific nodes, we also observed an increase in connectivity with an individual's type-II performance in a much broader network including ventrolateral PFC, bilateral inferior parietal lobules, angular gyrus, and bilateral precentral gyrus. This observation is in line with theories of conscious perception relying on long-distance brain networks linking prefrontal cortex with other brain regions, including the parietal cortex (26–28).

Our work builds on previous studies showing that variability in several cognitive functions, including reading (11), executive control (12), intelligence (13), insight (10), working memory (8), perceptual learning (29), and attention (9) can be accounted for by the organization of resting state networks. The uniqueness of our work is twofold: First, it distinctively identifies networks that covary with objective (type I) performance on a visual task and metacognitive accuracy (type II). Second, it measures functional networks in different attentional states to examine whether a relatively narrow library of networks of stable mental states may be better indicators of individual traits than measures of resting state per se (16, 30).

Relative to this second specific aim, the most important result of this study is that networks measured during a state of interoception show a distinct pattern of dependence on type-II performance compared with networks obtained in the resting state or state of exteroceptive attention. Only networks measured in the interoceptive state show connections increasing with type-II performance. One of the regions showing increased connectivity to other nodes with type-II performance is the angular gyrus (AG), typically associated with awareness of action authorship (31). However, we did not find a change in connectivity with type II performance for other brain structures involved in interoception, such as insula and anterior cingulate (32, 33). Hence, interoception and metacognitive ability show partial overlap on brain circuitry.

These findings build on Garfinkel and colleagues' behavioral study (34) of recall and confidence of stimuli presented at different moments of the heart cycle; systole (bursts indicating heart contractions) and diastole (heart relaxations). A deficit in recall was observed specifically for targets perceived with low confidence during the systole. Participants with high interoceptive accuracy were more immune to this deficit in recall and as a consequence

Table 2. Regression to type-II performance, top 15 ROIs

MNI coordinates				
x	*y*	*z*	ROI label	Network
−11	45	17	Ventromedial PFC	Default
9	51	16	Ventromedial PFC	Default
−42	7	36	Dorsal frontal cortex	Frontoparietal
42	48	−3	Ventral anterior PFC	Frontoparietal
54	−44	43	Inferior parietal lobule	Frontoparietal
−47	−12	36	Parietal	Sensorimotor
−55	−44	30	Parietal	Cinguloopercular
44	−52	47	Inferior parietal lobule	Frontoparietal
−35	−46	48	Postparietal cortex	Frontoparietal
−5	−52	17	Postcingulate cortex	Default
−52	28	17	Ventral PFC	Frontoparietal
40	36	29	Dorsolateral PFC	Frontoparietal
−54	−9	23	Precentral gyrus	Sensorimotor
−11	−58	17	Postcingulate	Default
−11	45	17	Inferior parietal lobule	Frontoparietal

ROIs with the highest rank in the number of connections whose β-value for type-II performance exceeded a threshold of 3 SDs, for networks measured under interoceptive state.

confidence becomes a worse indicator of future recall (because both high and low confidence elements are recalled). Thus, participants with high interoceptive accuracy have worse metacognitive accuracy of future recall during the systole. This leads to a negative correlation between type-II performance and introspective ability, which may seem at odds with our finding that only networks measured in the interoceptive state show connections increasing with type-II performance. However, there is no intrinsic contradiction between these results: a partial overlap on brain circuitry of interoception and metacognitive ability may reveal itself as a competition between these processes, yielding results similar to those found by Garfinkel and colleagues (34). In the following paragraphs we argue how these arguments can be sketched for specific predictions of interactions between the systems of metacognitive ability and interoception. More generally, our work on functional brain networks and Garfinkel et al.'s behavioral studies (34), are only the first steps to understanding what may be a complex pattern of interactions between the systems of metacognitive ability and interoception. This may help bridge the fertile but largely disconnected literature of metacognitive ability (2, 35–37) and interoception (17, 32, 33, 38).

Beyond the results described in this study, other predictions derive from the hypothesis of partially overlapping systems of metacognitive ability and interoception: (*i*) Training interoception—for instance with interventions of mindfulness—may be a vehicle to partially improve metacognitive ability in a broad and nonspecific manner. (*ii*) Psychiatric disorders with deficits of interoception (such as depersonalization disorders; ref. 39), should reflect a specific deficit in type-II performance without affecting type-I performance. (*iii*) Synchronic expression of introspective and interoceptive tasks may reflect a bottleneck and hence interference. As in ref. 2, concurrent performance with an interoceptive task may impair (or delay, or interact with) type-II, but not type-I performance. (*iv*) Finally, a more speculative and theoretically provoking thought is that metacognitive ability and interoception may share a fundamental role in cementing conscious experience. In several psychological theories, metacognition is considered a process of second-order (meta) representation of first-order processes, which is constitutive of consciousness (see ref. 36 for a review). Similarly, interoceptive sensitivity has often been identified as a precursor of consciousness, although of a different kind: awareness of one's body, which is intimately linked to self-identity and self-consciousness (40). Hence a further prediction that can be examined empirically is that manipulations affecting conscious state (through sleep or mild sedation for instance) should show a correlated fade out of interoceptive and metacognitive abilities.

Our results extend the reach of the covariations of connectivity of previous studies to the domain of metacognition and highlight that mental states of interoceptive, resting, and exteroceptive attention convey different information about future objective performance and accuracy of subjective beliefs. Beyond the specific consequences for the domain of metacognition, our results indicate how information about individual traits may be enriched when based on a set of functional brain networks obtained from different mental states.

Materials and Methods

Participants. Twenty-five subjects (12 male; mean age = 25.06 y) with normal or corrected-to-normal vision, no report of history of psychiatric or neurological disorders, and no current use of any psychoactive medications, gave their written consent to participate in the experiment. The study was conducted in accordance with the Declaration of Helsinki and approved by the institutional ethics committees of the Fundación para la Lucha contra las Enfermedades Neurológicas de la Infancia (Argentina) and Glasgow university (UK). Sixteen subjects were from Buenos Aires and 9 from Glasgow, Scotland. Both groups showed a very similar pattern of results in the main observations of this study (Fig. S3) and hence were pooled together to increase the statistical power.

Partial Report (Behavioral) Experiment. Several days after the fMRI recordings, participants performed a PR experiment, identifying a letter in a cued location of a cluttered field (3). 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 the subject's response. Random performance is 1/26, because for each trial, subjects had to choose 1 of 26 possible letters. Subsequently, participants indicated the degree of confidence in their response in a continuous scale (Fig. 1*A*). Performance in the objective task (reporting the letter, or type-I performance) and performance in the subjective task (reporting confidence on response, or type-II performance) were used as linear regressors for functional MRI connectivity. To explore how much of the total variance in the fMRI data was explained by the two behavioral regressors, we calculated the R-square value (Fig. S5).

fMRI Recordings and Analysis. Functional images from Buenos Aires were acquired on a GE HDx 3.0T MR system with a conventional eight-channel head coil. Twenty-four axial slices (5 mm thick) were acquired parallel to the plane connecting the anterior and posterior commissures and covering the whole brain [repetition time (TR) = 2,000 ms, echo time (TE) = 35 ms, flip angle = 90]. To aid in the localization of functional data, high-resolution structural T1 image [3D Fast inversion recovery spoiled gradient echo (SPGR-IR), inversion time 700 mm; flip angle (FA) = 15, field of view (FOV) = 192 × 256 × 256 mm; matrix 512 × 512 × 168; slice thickness 1.1 mm] was also acquired. Images from Glasgow were acquired on a 3-T Siemens MRI system (Magnetom Vision; Siemens Electric) with the same parameters.

Subjects underwent three functional runs lasting 7 min 22 s each for the Buenos Aires dataset and 12 min for the Glasgow dataset. We ran the experiment in Glasgow with longer time series to assure that the functional networks measured with 7 min 22 s were stable and close to convergence to a stationary value (Fig. S4). Subjects were instructed to keep their eyes closed without falling asleep. Random sequences of tones with the same distribution of duration (200 ms), pitch (400 Hz), and oddball frequency (pitch 410 Hz every 15 tones) were presented every 400 ms during the three blocks at very low volume. In the interoceptive attention run, participants were instructed to focus on their respiration cycle, perceiving the air flowing in and out. In the exteroceptive attention, participants were informed that they would hear a series of sounds and should focus on it. In the resting block, subjects were instructed to relax, not to do any mental effort and not to fall asleep. After the recordings we asked subjects whether they heard the beeps in the other runs. None of the subjects reported noticing the tones in the resting state or interoceptive attention, indicating that in absence of directed attention the tones were camouflaged within the noise of the scanner. Conversely, all subjects reported a consistent approximate number of odd tones during the exteroceptive attention run. As the goal of our study was simply to direct subjects' attention to different states, we did not measure auditability but sounds were presented with exactly the same parameters in all three states.

ACKNOWLEDGMENTS. The authors thank Alejo Salles, Ariel Zylberberg and Simon van Gall for useful suggestions for the manuscript. This work is funded by the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), the Secretaría de Ciencia y Técnica de la Universidad de Buenos Aires (UBACYT), and by the Human Frontiers Science Program. M.S. is sponsored by the James McDonnell Foundation 21st Century Science Initiative in Understanding Human Cognition – Scholar Award. B.W. is supported by the Centre National de la Recherche Scientifique. P. Barttfeld was supported by a fellowship from the Consejo Nacional de Investigaciones Científicas y Técnicas and a Human Frontiers Science Program research grant.

1. Dienes Z, Seth A (2010) Gambling on the unconscious: A comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task. *Conscious Cogn* 19(2):674–681.
2. 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.
3. Graziano M, Sigman M (2009) The spatial and temporal construction of confidence in the visual scene. *PLoS ONE* 4(3):e4909.
4. Kunimoto C, Miller J, Pashler H (2001) Confidence and accuracy of near-threshold discrimination responses. *Conscious Cogn* 10(3):294–340.
5. Persaud N, McLeod P, Cowey A (2007) Post-decision wagering objectively measures awareness. *Nat Neurosci* 10(2):257–261.
6. Wilimzig C, Tsuchiya N, Fahle M, Einhauser W, Koch C (2008) Spatial attention increases performance but not subjective confidence in a discrimination task. *J Vis* 8(5):7.1–7.10.

7. Rounis E, Maniscalco B, Rothwell J, Passingham R, Lau H (2010) Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience* 1(3):165–175.
8. Hampson M, Driesen NR, Skudlarski P, Gore JC, Constable RT (2006) Brain connectivity related to working memory performance. *J Neurosci* 26(51):13338–13343.
9. He BJ, et al. (2007) Breakdown of functional connectivity in frontoparietal networks underlies behavioral deficits in spatial neglect. *Neuron* 53(6):905–918.
10. Kounios J, et al. (2008) The origins of insight in resting-state brain activity. *Neuropsychologia* 46(1):281–291.
11. Koyama MS, et al. (2011) Resting-state functional connectivity indexes reading competence in children and adults. *J Neurosci* 31(23):8617–8624.
12. Seeley WW, et al. (2007) Dissociable intrinsic connectivity networks for salience processing and executive control. *J Neurosci* 27(9):2349–2356.
13. van den Heuvel MP, Stam CJ, Kahn RS, Hulshoff Pol HE (2009) Efficiency of functional brain networks and intellectual performance. *J Neurosci* 29(23):7619–7624.
14. Fox MD, et al. (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci USA* 102(27):9673–9678.
15. Raichle ME (2009) A paradigm shift in functional brain imaging. *J Neurosci* 29(41): 12729–12734.
16. Tang YY, Rothbart MK, Posner MI (2012) Neural correlates of establishing, maintaining, and switching brain states. *Trends Cogn Sci* 16(6):330–337.
17. Dunn BD, et al. (2010) Listening to your heart. How interoception shapes emotion experience and intuitive decision making. *Psychol Sci* 21(12):1835–1844.
18. Metcalfe J, Greene MJ (2007) Metacognition of agency. *J Exp Psychol Gen* 136(2): 184–199.
19. Seth AK, Suzuki K, Critchley HD (2011) An interoceptive predictive coding model of conscious presence. *Front Psychol* 2:395.
20. Graziano M, Sigman M (2008) The dynamics of sensory buffers: Geometric, spatial, and experience-dependent shaping of iconic memory. *J Vis* 8(5):9.1–9.13.
21. Lau H, Maniscalco B (2010) Neuroscience. Should confidence be trusted? *Science* 329(5998):1478–1479.
22. Dosenbach NU, et al. (2010) Prediction of individual brain maturity using fMRI. *Science* 329(5997):1358–1361.
23. Dosenbach NU, et al. (2007) Distinct brain networks for adaptive and stable task control in humans. *Proc Natl Acad Sci USA* 104(26):11073–11078.
24. Jack AI, Roepstorff A (2002) Introspection and cognitive brain mapping: From stimulus-response to script-report. *Trends Cogn Sci* 6(8):333–339.
25. Del Cul A, Dehaene S, Reyes P, Bravo E, Slachevsky A (2009) Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain* 132(Pt 9):2531–2540.
26. Crick F, Koch C (1995) Are we aware of neural activity in primary visual cortex? *Nature* 375(6527):121–123.
27. Dehaene S, Naccache L (2001) Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition* 79(1-2):1–37.
28. Dehaene S, Sergent C, Changeux JP (2003) A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc Natl Acad Sci USA* 100(14):8520–8525.
29. Baldassarre A, et al. (2012) Individual variability in functional connectivity predicts performance of a perceptual task. *Proc Natl Acad Sci USA* 109(9):3516–3521.
30. Barttfeld P, et al. (2012) State-dependent changes of connectivity patterns and functional brain network topology in autism spectrum disorder. *Neuropsychologia* 50(14):3653–3662.
31. Farrer C, et al. (2008) The angular gyrus computes action awareness representations. *Cereb Cortex* 18(2):254–261.
32. Craig AD (2009) How do you feel—now? The anterior insula and human awareness. *Nat Rev Neurosci* 10(1):59–70.
33. Craig AD (2003) Interoception: The sense of the physiological condition of the body. *Curr Opin Neurobiol* 13(4):500–505.
34. Garfinkel S, et al. (2013) What the heart forgets: Cardiac timing influences memory for words and is modulated by metacognition and interoceptive sensitivity. *Psychophysiology* 50(6):505–512.
35. Fleming SM, Huijgen J, Dolan RJ (2012) Prefrontal contributions to metacognition in perceptual decision making. *J Neurosci* 32(18):6117–6125.
36. Lau H, Rosenthal D (2011) Empirical support for higher-order theories of conscious awareness. *Trends Cogn Sci* 15(8):365–373.
37. Zylberberg A, Barttfeld P, Sigman M (2012) The construction of confidence in a perceptual decision. *Front Integr Neurosci* 6:79.
38. Critchley HD, Wiens S, Rotshtein P, Ohman A, Dolan RJ (2004) Neural systems supporting interoceptive awareness. *Nat Neurosci* 7(2):189–195.
39. Simeon D (2004) Depersonalisation disorder: A contemporary overview. *CNS Drugs* 18(6):343–354.
40. Tsakiris M, Tajadura-Jimenez A, Costantini M (2011) Just a heartbeat away from one's body: Interoceptive sensitivity predicts malleability of body-representations. *Proc Biol Sci* 278(1717):2470–2476.

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.

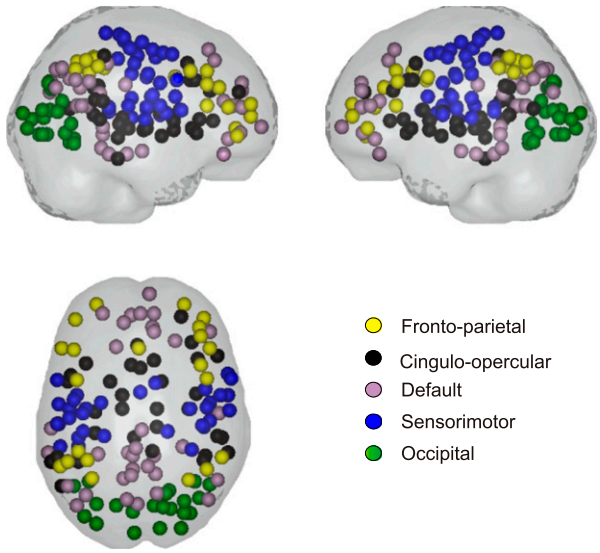

Fig. S1. A total of 141 regions of interest (ROIs) used in the analysis.

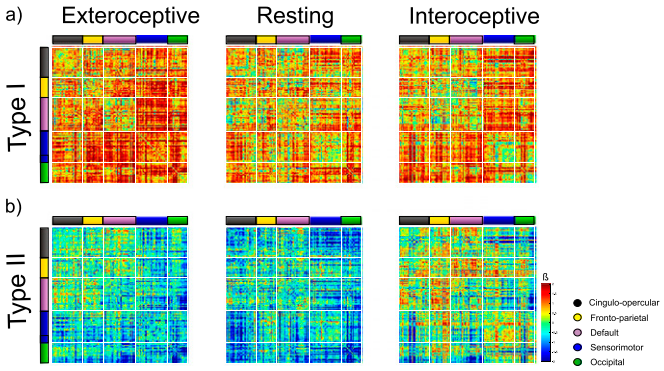

Fig. S2. Organization of functional brain networks according to subjects' individual metacognitive ability and performance. (*A*) Slope (β) values for type-I performance of the bivariate regression between functional connectivity and both types of performance. (*B*) Slopes for type-II performance of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

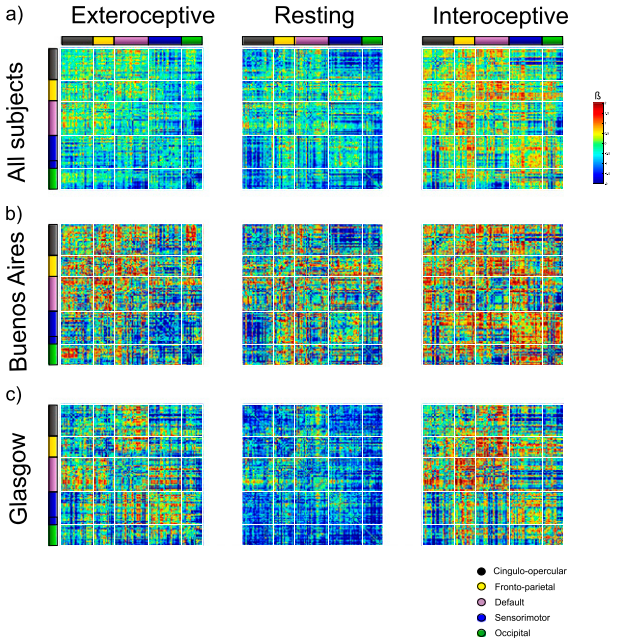

Fig. S3. Dependence of functional brain connectivity with type-I and type-II performances for the two datasets. (*A*) Matrix of slope (β) values for type-II performance of the bivariate regression between functional connectivity and both types of performance, for all subjects in the experiment. (*B*) 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.

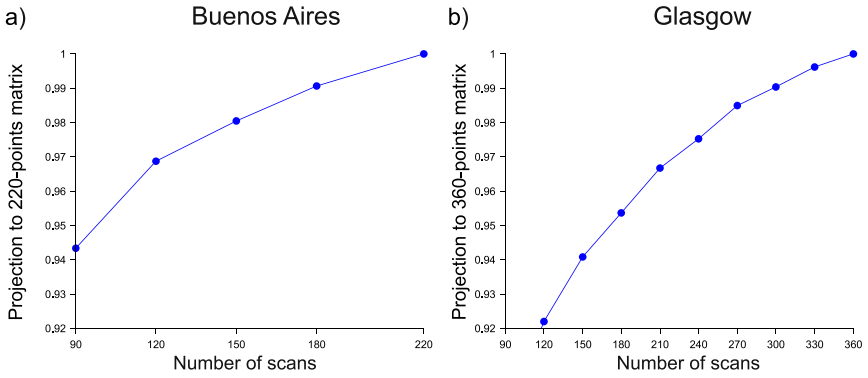

Fig. S4. Connectivity analysis using shorter time series. (*A*) Projection of each matrix (3, 4, 5, 6, and 7.22 min or 90, 120, 150, 180, and 220 scans) into the 220-time points matrix for the subjects from Buenos Aires. A value of 1 means perfect concordance of values, whereas a value of 0 implies full 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.

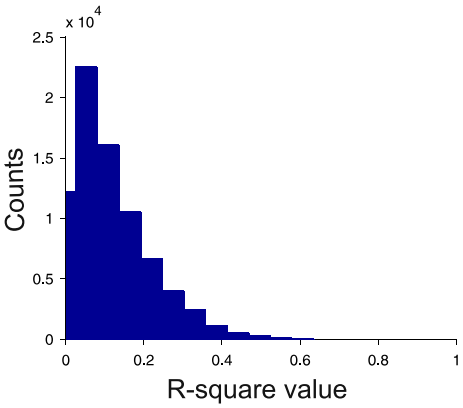

Fig. S5. Variance in the fMRI data explained by the regressors of interest. Distribution of all *R*-squared values of the bivariate regression between connectivity and type-I and type-II performances, collapsed across all attentional states and regressors. Almost all values are lower than 0.5, showing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