

Published in final edited form as:

Science. 2010 September 17; 329(5998): 1541–1543. doi:10.1126/science.1191883.

Relating introspective accuracy to individual differences in brain structure

Stephen M. Fleming^{*,1}, Rimona S. Weil^{*,1,2}, Zoltan Nagy¹, Raymond J. Dolan¹, and Geraint Rees^{1,2}

¹Wellcome Trust Centre for Neuroimaging, University College London, 12 Queen Square, London WC1N 3BG

²Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR

Abstract

Our ability to introspect about self-performance is core to human subjective experience, but the neuroanatomical basis of this ability is unknown. Such accurate introspection requires discriminating correct from incorrect decisions, a capacity that varies substantially across individuals. We dissociated variation in introspective ability from objective performance in a simple perceptual decision task, allowing us to determine whether this inter-individual variability was associated with a distinct neural substrate. We show that introspective ability is correlated with gray matter volume in anterior prefrontal cortex, a region which shows striking evolutionary development in humans. Moreover, inter-individual variability in introspective ability also correlated with white matter microstructure connected with this area of prefrontal cortex. Our findings point to a focal neuroanatomical substrate for introspective ability, a substrate distinct from that supporting primary perception.

Our moment-to-moment judgments of the outside world are often subject to introspective interrogation. In this context, introspective or “metacognitive” sensitivity refers to the ability to discriminate correct from incorrect perceptual decisions (1), and its accuracy is essential for the appropriate guidance of decision making and action (2, 3). For example, low confidence that a recent decision was correct may prompt us to re-examine the evidence or seek a second opinion. Recently, behavioural studies have begun to quantify metacognitive accuracy following simple perceptual decisions, and isolate variations in this ability: a decision may be performed poorly, yet an individual may believe that his or her performance was good, or vice-versa (4-8). While previous work has investigated how confidence in perceptual decisions varies from trial to trial (9, 10), little is known about the biological basis of metacognitive ability, here defined as how well an individual’s confidence ratings discriminate correct from incorrect decisions over time. We hypothesised that individual differences in metacognitive ability would be reflected in the anatomy of brain regions responsible for this function, in line with similar associations between brain anatomy and performance in other cognitive domains (11-15).

We objectively quantified variability in metacognitive sensitivity between individuals, and then related these inter-individual differences to brain structure measured using magnetic

Correspondence: Stephen M. Fleming, Wellcome Trust Centre for Neuroimaging, University College London, 12 Queen Square, London WC1N 3BG T: +44-20-7833-7486 F: +44-20-7813-1420 s.fleming@fil.ion.ucl.ac.uk.

*These authors contributed equally to this work

resonance imaging (MRI). This approach was motivated by observations that individual differences in a range of skills, such as language (11), decision making (12) and memory (13), are consistently associated with variation in healthy brain anatomy. Our design dissociated a quantitative measure of metacognitive accuracy, A_{roc} (which is specific to an individual) from both objective task performance and subjective confidence (which both vary on a trial-by-trial basis). Earlier patient studies describe candidate brain regions where damage is associated with poor introspective ability, in particular, a prefrontal-parietal network (16-18). Theories of prefrontal function have emphasised a role for anterior (rostrolateral) prefrontal cortex in carrying out second-order operations on internally generated information (19, 20), a core feature of metacognition. We hypothesised that the local structure of these regions (both grey matter volume and white matter integrity) might reflect an individual's metacognitive ability.

We studied 32 healthy human participants while they made a series of visual judgements (21). The difficulty of the visual judgement was varied on a per-participant basis in order to keep performance at a constant level (71%), near sensory threshold. As well as asking participants to make these objective perceptual judgements, we also asked them to provide ratings of confidence in their decisions following each trial (Fig. 1). These ratings were used to determine metacognitive ability at an individual level through the construction of Type II receiver operating characteristic (ROC) curves (21-23) (Fig. 2A). The ROC model provided an excellent fit to our data across participants (mean $R^2 = 0.97 \pm 0.023$). The area between the major diagonal and an individual's ROC curve is a measure of the ability to link confidence to perceptual performance (A_{roc}). We found considerable variation across individuals in metacognitive ability ($A_{roc} = 0.55 - 0.75$) despite underlying task performance being held constant (proportion correct: 70 - 74%); furthermore, these measures were uncorrelated (Pearson's $r = -0.11$, $P = 0.56$). To establish whether this variability was stable, we split data from each participant into two halves, and computed the test-retest reliability of the two sets. This analysis revealed intrasubject consistency in A_{roc} ($r = 0.69$, $P = 0.00001$; fig. S2).

Having quantified inter-individual variability in introspection, we now asked whether this variability in introspective judgements was predicted by variability in brain structure using two distinct measures: gray matter volume measured using MRI, and the fractional anisotropy of white matter measured using diffusion tensor imaging (DTI). Our analysis examined the possible relationship between brain structure and four different measures: the metacognitive ability (A_{roc}) of our participants, objective performance on the perceptual task (d' and c), and the tendency to use high or low confidence responses on individual trials (B_{roc}) (21). Having removed the potentially confounding factors (24) of overall brain size and gender (as regressors of no interest), we found that an individual's metacognitive ability (A_{roc}) was significantly correlated with gray matter volume in right anterior prefrontal cortex (Fig. 3A) (Brodmann area (BA) 10, peak voxel coordinates: [24, 65, 18], $t_{max} = 4.8$, $P < 0.05$, corrected for multiple comparisons). Furthermore, gray matter volume in this region did not correlate with task performance as indexed by d' (Fig. 3B) ($r = 0.15$, $P = 0.42$) or overall confidence (B_{roc} ; $r = -0.023$, $P = 0.90$). Gray matter volume in a homologous region in left anterior prefrontal cortex was also correlated with A_{roc} , but did not survive correction for multiple comparisons across the brain volume. Details of this and other clusters that did not survive a whole brain correction are listed in table S2. Thus, variability in introspective judgements of performance on a simple visual detection task was predicted by variability in the anatomical structure of anterior prefrontal cortex (BA 10) independently of both objective performance and level of confidence. Finally, while our primary question addressed positive dependence of gray matter on A_{roc} , we also found that left inferior temporal gyrus showed a negative correlation with metacognitive sensitivity (Fig. 3A) (coordinates: [-56 -30 -26], $t_{max} = 4.66$, $P < 0.05$, corrected for multiple comparisons),

accompanied by a similar region on the right that did not survive correction for multiple comparisons (see table S2 for full details and coordinates) (21).

Having established that gray matter volume was predictive of A_{roc} , we next analysed white matter microstructure. If the structure of anterior prefrontal cortex is functionally related to metacognitive performance, we hypothesised that white matter tracts connected with this region would also show a similar microstructural correlation with expression of this behavioural trait. In a whole-brain analysis of white matter microstructure (see Methods for details), we found that fractional anisotropy (a measure of white matter integrity) in the genu of the corpus callosum was positively dependent on A_{roc} (Fig. 4) ($P < 0.05$, corrected for multiple comparisons). This specific subdivision of the corpus callosum contains white matter fibres connected with the anterior and orbital prefrontal cortices in humans (25), consistent with metacognitive ability being dependent not only on anterior prefrontal grey matter but also on reciprocal projections to and from this area. Neither objective performance (stimulus contrast or d') nor overall confidence (B_{roc}) correlated with grey matter volume or white matter fractional anisotropy elsewhere in the brain ($P > 0.05$, corrected for multiple comparisons (see tables S2 and S3 for uncorrected correlations). We note that an absence of structural correlations with these parameters may have been due to our design deliberately minimising variability in both d' and B_{roc} in order to isolate the neural correlates of introspective ability (A_{roc}).

One concern is that the structural covariation we observed here may have been potentially confounded by differences in perceptual ability per se. Good perceptual ability may be reflected in the staircase procedure converging on consistently low values for stimulus contrast for a given individual. We therefore carried out control analyses (21) (table S4) to rule out this alternative explanation. These results demonstrated significant correlations of gray matter and fractional anisotropy with A_{roc} in anterior prefrontal cortex when controlling for changes in task parameters, and an absence of correlations with task parameters themselves. Thus the structure-behaviour correlations we observed here are unlikely to be due to low-level differences in performance but instead relate to underlying differences in individual metacognitive ability.

How might these regions contribute to metacognition? Anterior subdivisions of prefrontal cortex have been implicated in high-level control of cognition (19, 20, 26, 27), and are well placed to integrate supramodal perceptual information with decision output (28), a process thought to be key for metacognitive sensitivity (1, 3). Dorsolateral prefrontal activity increases under conditions in which subjective reports match objective perceptual performance (29), suggesting a computational role in linking performance to confidence. Consistent with prefrontal grey matter volume playing a causal role in metacognition, patients with lesions to anterior prefrontal cortex show deficits in subjective reports compared to controls, after factoring out differences in objective performance (16). Furthermore, impairing dorsolateral prefrontal cortex function with theta-burst transcranial magnetic stimulation compromises the metacognitive sensitivity of subjective reports of awareness, while leaving underlying task performance intact (30). These findings, together with the present work, suggest a central role for anterior and dorsolateral PFC in metacognitive sensitivity. Our present findings may reflect innate differences in anatomy, or alternatively reflect the effects of experience and learning, as has been found in the sensorimotor domain (14, 15). This raises the tantalising possibility of being able to “train” metacognitive ability by harnessing underlying neural plasticity in the regions we identify here (31).

Our key finding is a delineation of a strikingly focal anatomical substrate that predicts inter-individual variability in metacognitive ability. As with any correlational method, we cannot

establish whether the covariation we observed here between brain structure and metacognition reflects a causal role. However, given a wealth of evidence for changes in grey matter volume within and between individuals associated with a range of skills, we propose that underlying differences in metacognitive ability are similarly dependent on large-scale brain anatomy. Our data provide an initial window onto the biological substrates of the ability to link objective performance to subjective confidence. The demonstration that this ability may be dependent on local and phylogenetically recent prefrontal anatomy is consistent with a conjecture that metacognitive function has been selected for during evolution (32), facilitating computations that allow us to introspect about self-performance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was funded by the Wellcome Trust (G.R., R.J.D., Z.N.), the UCL 4-Year PhD in Neuroscience (S.M.F.) and the Medical Research Council (R.W.). We thank T. Sharot, G. Ridgway and C. Frith for comments on earlier drafts of this manuscript.

References

1. Cleeremans A, Timmermans B, Pasquali A. Consciousness and metarepresentation: A computational sketch. *Neural Networks*. 2007; 20:1032. [PubMed: 17904799]
2. Metcalfe, J. *Metacognition: Knowing About Knowing*. MIT Press; 1996.
3. Lau HC. A higher order Bayesian decision theory of consciousness. *Prog. Brain Res*. 2008; 168:35. [PubMed: 18166384]
4. Washburn DA, Smith JD, Tagliatela LA. Individual differences in metacognitive responsiveness: cognitive and personality correlates. *J. Gen. Psychol*. 2005; 132:446. [PubMed: 17607328]
5. Kunimoto C, Miller J, Pashler H. Confidence and accuracy of near-threshold discrimination responses. *Conscious. Cogn*. 2001; 10:294. [PubMed: 11697867]
6. Szczepanowski R, Pessoa L. Fear perception: can objective and subjective awareness measures be dissociated? *J. Vis*. 2007; 7:1.
7. Graziano M, Sigman M. The spatial and temporal construction of confidence in the visual scene. *PLoS ONE*. 2009; 4:e4909. [PubMed: 19290055]
8. Fleming SM, Dolan RJ. Effects of loss aversion on post-decision wagering: Implications for measures of awareness. *Conscious. Cogn*. 2010; 19:352. [PubMed: 20005133]
9. Kiani R, Shadlen MN. Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*. 2009; 324:759. [PubMed: 19423820]
10. Kepecs A, Uchida N, Zariwala H, Mainen Z. Neural correlates, computation and behavioural impact of decision confidence. *Nature*. 2008; 455:227. [PubMed: 18690210]
11. Carreiras M, et al. An anatomical signature for literacy. *Nature*. 2009; 461:983. [PubMed: 19829380]
12. Tuch DS, et al. Choice reaction time performance correlates with diffusion anisotropy in white matter pathways supporting visuospatial attention. *Proc. Natl. Acad. Sci. U.S.A*. 2005; 102:12212. [PubMed: 16103359]
13. Fuentemilla L, et al. Individual differences in true and false memory retrieval are related to white matter brain microstructure. *J. Neurosci*. 2009; 29:8698. [PubMed: 19587276]
14. Scholz J, Klein MC, Behrens TEJ, Johansen-Berg H. Training induces changes in white-matter architecture. *Nat. Neurosci*. 2009; 12:1370. [PubMed: 19820707]
15. Draganski B, et al. Neuroplasticity: changes in grey matter induced by training. *Nature*. 2004; 427:311. [PubMed: 14737157]
16. Del Cul A, Dehaene S, Reyes P, Bravo E, Slachevsky A. Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain*. 2009; 132:2531. [PubMed: 19433438]

17. Shimamura AP. Toward a cognitive neuroscience of metacognition. *Conscious. Cogn.* 2000; 9:313. [PubMed: 10924251]
18. Simons JS, Peers PV, Mazuz YS, Berryhill ME, Olson IR. Dissociation between memory accuracy and memory confidence following bilateral parietal lesions. *Cereb. Cortex.* 2010; 20:479. [PubMed: 19542474]
19. Fletcher PC, Henson RNA. Frontal lobes and human memory: insights from functional neuroimaging. *Brain.* 2001; 124:849. [PubMed: 11335690]
20. Christoff K, Gabrieli JDE. The frontopolar cortex and human cognition: evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology.* 2000; 28:168.
21. Materials, methods, discussion of ROC model fits and details of control analyses are available as supporting material on *Science* Online.
22. Galvin SJ, Podd JV, Drga V, Whitmore J. Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. *Psychon. Bull. Rev.* 2003; 10:843. [PubMed: 15000533]
23. Kornbrot DE. Signal detection theory, the approach of choice: model-based and distribution-free measures and evaluation. *Percept. Psychophys.* 2006; 68:393. [PubMed: 16900832]
24. Smith CD, Chebrolu H, Wekstein DR, Schmitt FA, Markesbery WR. Age and gender effects on human brain anatomy: A voxel-based morphometric study in healthy elderly. *Neurobiol. Aging.* 2007; 28:1075. [PubMed: 16774798]
25. Park HJ, et al. Corpus callosal connection mapping using cortical gray matter parcellation and DT-MRI. *Hum. Brain Mapp.* 2008; 29:503. [PubMed: 17133394]
26. Daw N, O'Doherty J, Dayan P, Seymour B, Dolan R. Cortical substrates for exploratory decisions in humans. *Nature.* 2006; 441:876. [PubMed: 16778890]
27. Burgess PW, Dumontheil I, Gilbert SJ. The gateway hypothesis of rostral prefrontal cortex (area 10) function. *Trends Cogn. Sci.* 2007; 11:290. [PubMed: 17548231]
28. Ramnani N, Owen AM. Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nat. Rev. Neurosci.* 2004; 5:184. [PubMed: 14976518]
29. Lau H, Passingham R. Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proc. Natl. Acad. Sci. U.S.A.* 2006; 103:18763. [PubMed: 17124173]
30. Rounis, E., et al. *Cognitive Neurosci.* in press (available at <http://dx.doi.org/10.1080/17588921003632529>)
31. Titchener, EB. *Lectures on the experimental psychology of the thought-processes.* Macmillan; 1909.
32. Metcalfe, J. *Handbook of Metamemory and Memory.* Dunlosky, J.; Bjork, RA., editors. Psychology Press; 2008. p. 27-46.

One-sentence summary

The local anatomical structure of human prefrontal cortex predicts an individual's ability to introspect.

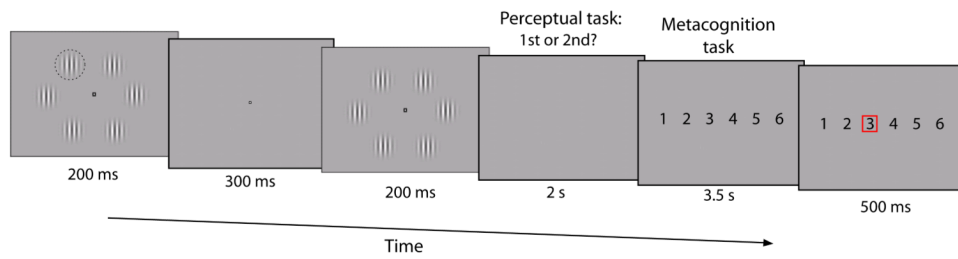


Figure 1.

Behavioral task.

Subjects completed a two-alternative forced choice task that required two judgments per trial: a perceptual response followed by an estimate of relative confidence in their decision. The perceptual response indicated whether the first or second temporal interval contained the higher contrast (pop-out) Gabor patch (highlighted here with a dashed circle which was not present in the actual display), which could appear at any one of 6 locations around a central fixation point. Pop-out Gabor contrast was continually adjusted using a staircase procedure to maintain ~71% performance. Confidence ratings were made using a 1-6 scale, with participants encouraged to use the whole scale from 1 = low relative confidence to 6 = high relative confidence.

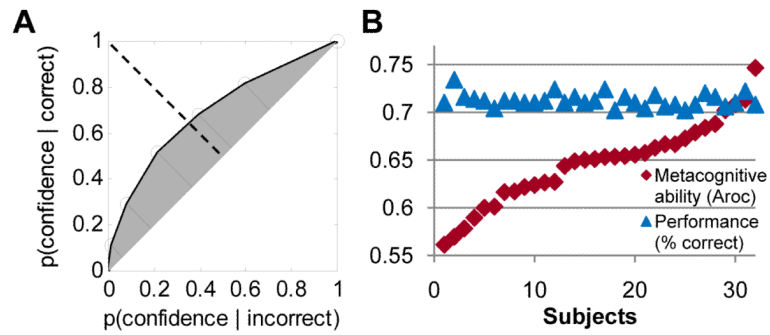


Figure 2.

ROC calculation and behavioural performance

(A) Participants' confidence ratings were used to construct a Type II ROC function that quantifies the ability to discriminate between correct and incorrect responses cumulated across levels of confidence. A_{ROC} was calculated as the shaded area between the ROC curve and major diagonal (21). Mutually perpendicular dotted and solid lines represent the minor and major diagonals respectively. (B) Plot of the relationship between task performance (% correct) and A_{ROC} , with subjects ordered by increasing A_{ROC} .

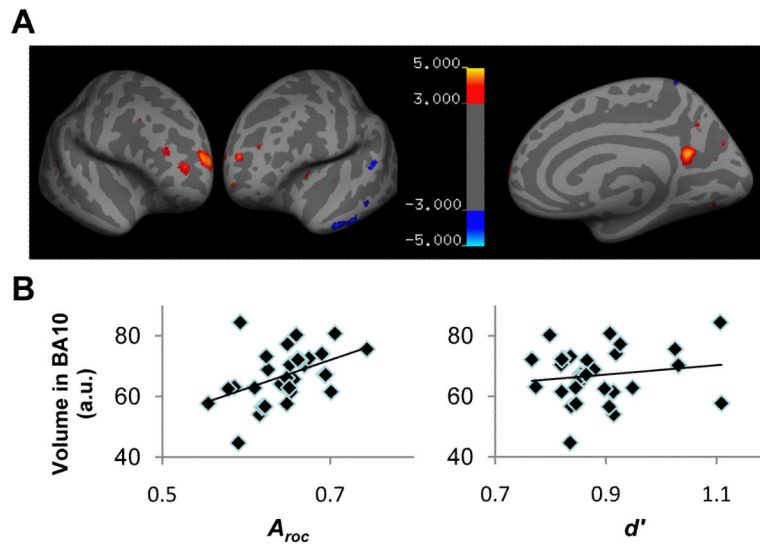


Figure 3.

Gray matter volume correlated with introspective ability

(A) Projection of statistical (T) maps for positive (hot colormap) and negative (cool colormap) correlations with A_{roc} onto an inflated cortical surface T1-weighted template, thresholded at $T > 3$ for display purposes. Significant clusters ($P < 0.05$, corrected for multiple comparisons) where metacognitive ability correlated with gray matter volume (see Methods) were found in right anterior prefrontal cortex (Brodmann area 10; positive correlation) and left inferior temporal gyrus (negative correlation), accompanied by contralateral homologous clusters at $P < 0.001$, uncorrected. (B) Plot of grey matter volume in the right BA10 cluster against both A_{roc} and d' (see Methods for full details) indicating that the correlation with metacognitive ability was independent of task performance.

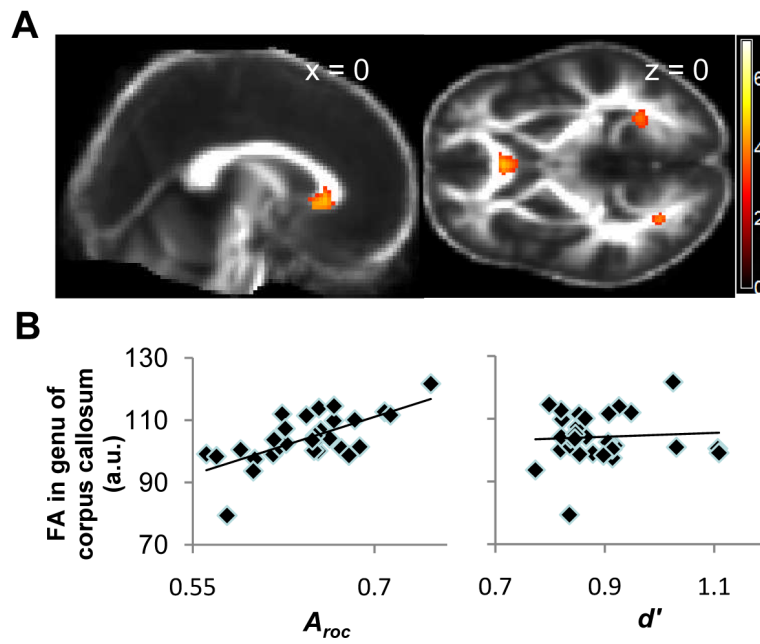


Figure 4.

White matter microstructure correlated with introspective ability

(A) Statistical (T) map of voxel-wise correlations between fractional anisotropy (FA) and A_{roc} , thresholded at $T > 3$ for display purposes and overlaid on sagittal (left) and axial (right) slices of the average FA image across subjects, at the x and z co-ordinates indicated. A region within the genu of the anterior corpus callosum showed a correlation between FA and metacognitive accuracy that was statistically significant after correcting for multiple comparisons ($P < 0.05$). (B) Plot of FA in the anterior corpus callosum cluster against both A_{roc} and d' indicating that the correlation with metacognitive ability was independent of task performance.