Relating Introspective Accuracy to Individual Differences in Brain Structure

Stephen M. Fleming, *et al.*

Science **329**, 1541 (2010);
DOI: 10.1126/science.1191883

increasing the maturation of miR-16 from its precursor pre/pri–miR-16. Raphe additionally responds to chronic fluoxetine treatment by releasing S100β, which in turn acts on the noradrenergic neurons of the locus coeruleus. By lowering miR-16 levels, S100β unlocks the expression of serotonergic functions in this noradrenergic brain area. Our pharmacological and behavioral data thus posit miR-16 as a central effector that regulates SERT expression and mediates the adaptive response of serotonergic and noradrenergic neurons to fluoxetine treatment.

References and Notes

1. G. E. Torres, R. R. Gainetdinov, M. G. Caron, *Nat. Rev. Neurosci.* **4**, 13 (2003).
2. E. C. Azmitia, *Int. Rev. Neurobiol.* **77**, 31 (2007).
3. O. Berton, E. J. Nestler, *Nat. Rev. Neurosci.* **7**, 137 (2006).
4. Y. Qian, H. E. Melikian, D. B. Rye, A. I. Levey, R. D. Blakely, *J. Neurosci.* **15**, 1261 (1995).
5. D. L. Murphy *et al.*, *Neuropharmacology* **55**, 932 (2008).
6. S. Benmansour, W. A. Owens, M. Cecchi, D. A. Morilak, A. Frazer, *J. Neurosci.* **22**, 6766 (2002).
7. D. P. Bartel, *Cell* **136**, 215 (2009).
8. C. M. Croce, *Nat. Rev. Genet.* **10**, 704 (2009).
9. K. S. Kosik, *Nat. Rev. Neurosci.* **7**, 911 (2006).
10. V. K. Gangaraju, H. Lin, *Nat. Rev. Mol. Cell Biol.* **10**, 116 (2009).
11. S. Mouillet-Richard *et al.*, *J. Biol. Chem.* **275**, 9186 (2000).
12. J. M. Launay, B. Schneider, S. Loric, M. Da Prada, O. Kellermann, *FASEB J.* **20**, 1843 (2006).
13. A. Frazer, *J. Clin. Psychopharmacol.* **17** (suppl. 1), 2S (1997).
14. E. J. Nestler, M. Alreja, G. K. Aghajanian, *Biol. Psychiatry* **46**, 1131 (1999).
15. G. Martello *et al.*, *Nature* **449**, 183 (2007).
16. M. A. Kim, H. S. Lee, B. Y. Lee, B. D. Waterhouse, *Brain Res.* **1026**, 56 (2004).
17. N. Shanmugam, M. A. Reddy, R. Natarajan, *J. Biol. Chem.* **283**, 36221 (2008).
18. R. Manev, T. Uz, H. Manev, *Eur. J. Pharmacol.* **420**, R1 (2001).
19. A. Surget *et al.*, *Neuropsychopharmacology* **34**, 1363 (2009).
20. P. Willner, *Psychopharmacology (Berlin)* **134**, 319 (1997).

21. Materials and methods and supporting data are available on *Science* Online.

22. We thank P. Weil-Malherbe, V. Mutel, F. d'Agostini, G. Zürcher, E. Borroni, J. L. Moreau, F. Jenck, M. Bühler, and N. Pieron for skillful methodological assistance, and S. Blanquet, M. Briley, and L. Aggerbeck for critical reading of the manuscript. O.K. is a professor at Paris XI University. This work was funded by CNRS, ANR, and INSERM.

Supporting Online Material

www.sciencemag.org/cgi/content/full/329/5998/1537/DC1
Materials and Methods
Figs. S1 to S10
References

14 June 2010; accepted 4 August 2010
10.1126/science.1193692

Relating Introspective Accuracy to Individual Differences in Brain Structure

Stephen M. Fleming,[1]*† Rimona S. Weil,[1,2]* Zoltan Nagy,[1] Raymond J. Dolan,[1] Geraint Rees[1,2]

The ability to introspect about self-performance is key to human subjective experience, but the neuroanatomical basis of this ability is unknown. Such accurate introspection requires discriminating correct decisions from incorrect ones, 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 interindividual variability was associated with a distinct neural basis. We show that introspective ability is correlated with gray matter volume in the anterior prefrontal cortex, a region that shows marked evolutionary development in humans. Moreover, interindividual variation in introspective ability is also correlated with white-matter microstructure connected with this area of the 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 reexamine the evidence or seek a second opinion. Recently, behavioral studies have begun to quantify metacognitive accuracy following simple perceptual decisions and to isolate variations in this ability: A decision may be made poorly, yet an individual may believe that his or her performance was good, or vice versa (*4*–*8*). Whereas 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, defined here as how well an individual's confidence ratings discriminate correct from incorrect decisions over time. We hypothesized 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 interindividual differences to brain structure measured with magnetic 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 experimental 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 in which damage is associated with poor introspective ability: in particular, a prefrontal-parietal network (*16*–*18*). Theories of prefrontal

[1]Wellcome Trust Centre for Neuroimaging, University College London, 12 Queen Square, London WC1N 3BG, UK. [2]Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR, UK.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: s.fleming@fil.ion.ucl.ac.uk

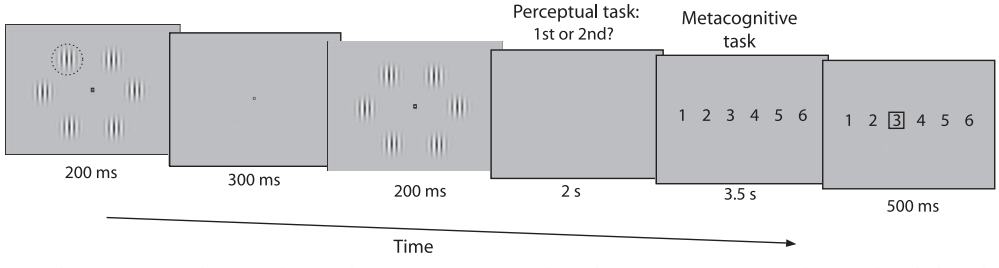

Fig. 1. Behavioral task. Participants completed a two-alternative forced-choice task that required two judgments per trial: a perceptual response followed by an estimate of relative confidence in their decision. The perceptual response indicated whether the first or second temporal interval contained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

function have emphasized a role for anterior (rostrolateral) prefrontal cortex (PFC) in carrying out second-order operations on internally generated information (*19*, *20*), a process necessary for metacognition. We hypothesized that the local structure of these regions (both gray-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 to keep performance at a constant level (71%), near sensory threshold. In addition to asking participants to make these objective perceptual judgements, we also asked them to provide ratings of confidence in their decisions after each trial (Fig. 1). We then used these ratings to determine metacognitive ability at an individual level through the construction of type II receiver operating characteristic (ROC) curves (Fig. 2A) (*21*–*23*). The ROC model provided an excellent fit to our data across participants (mean explained variance $R^2 = 0.97 \pm 0.023$). 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 to 0.75), despite underlying task performance being held constant (proportion correct: 70 to 74%); furthermore, these measures were uncorrelated (Pearson's correlation coefficient $r = -0.21$, $P = 0.24$). 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 intraparticipant consistency in A_{roc} ($r = 0.69$, $P = 0.00001$) (fig. S2).

Having quantified interindividual variability in introspection, we then asked whether this variability in introspective judgements was predicted by variability in brain structure using two distinct measures: gray-matter volume measured from T1-weighted anatomical images and the fractional anisotropy (FA) of white matter measured from diffusion tensor images. Our analysis examined the possible relation between brain structure and four different measures: the metacognitive ability (A_{roc}) of our participants, objective performance on the perceptual task (sensitivity, d', and criterion, c), and the tendency to use high or low confidence responses on individual trials (B_{roc}) [see supporting online material (SOM) methods section for details]. 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 the right anterior PFC (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})($r = -0.023$, $P = 0.90$).

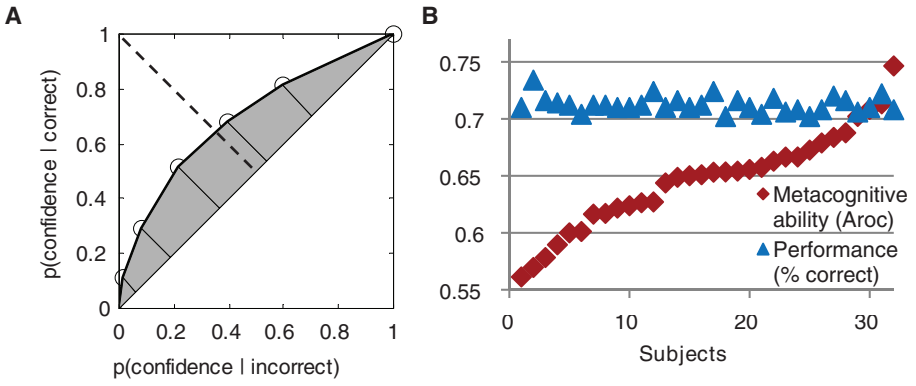

Fig. 2. ROC calculation and behavioral performance. (**A**) Participants' confidence ratings were used to construct a type II ROC function that quantifies the ability to discriminate between correct and incorrect responses cumulated across levels of confidence. A_{roc} was calculated as the shaded area between the ROC curve and the major diagonal (*21*). Mutually perpendicular dotted and solid lines represent the minor and major diagonals, respectively. (**B**) Plot of the relation between task performance (percentage correct) and A_{roc}, with participants ordered by increasing A_{roc} value.

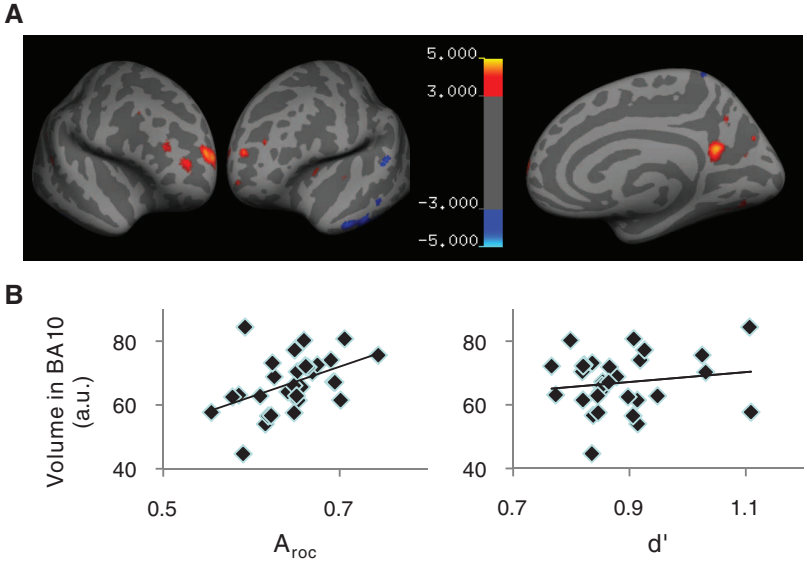

Fig. 3. Gray-matter volume correlated with introspective ability. (**A**) Projection of statistical (T) maps for positive (hot color map: red, orange, yellow) and negative (cool color map: blue) correlations with A_{roc} onto an inflated cortical surface (T1-weighted template, thresholded at $T > 3$ for display purposes). Significant clusters ($P < 0.05$, corrected for multiple comparisons) where metacognitive ability correlated with gray-matter volume (see SOM methods) were found in right anterior PFC (BA 10; positive correlation) and the left inferior temporal gyrus (negative correlation), accompanied by contralateral homologous clusters at $P < 0.001$, uncorrected. (**B**) Plot of gray-matter volume in the right BA 10 cluster against both A_{roc} and d' (see SOM methods for full details), indicating that the correlation with metacognitive ability was independent of task performance. a.u., arbitrary units.

Gray-matter volume in a homologous region in the left anterior PFC 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 the anterior PFC (BA 10), independently of both objective performance and level of confidence. Finally, whereas our primary question addressed positive dependence of gray matter on A_{roc}, we also found that the 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).

After we established that gray-matter volume was predictive of A_{roc}, we next analyzed white-matter microstructure. If the structure of the anterior PFC is functionally related to metacognitive performance, we hypothesized that white-matter tracts connected with this region would also show a similar microstructural correlation with

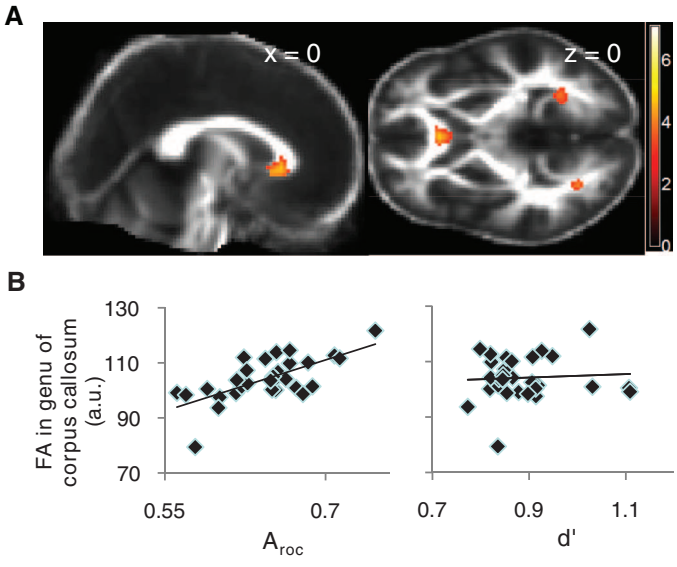

Fig. 4. White-matter microstructure correlated with introspective ability. (**A**) Statistical (T) map of voxelwise correlations between FA and A_{roc}, thresholded at $T > 3$ for display purposes and overlaid on sagittal (left) and axial (right) slices of the average FA image across participants, at the x and z coordinates indicated. A region within the genu of the anterior corpus callosum showed a correlation between FA and metacognitive ability 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.

expression of this behavioral trait. In a whole-brain analysis of white-matter microstructure (*21*), we found that FA (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 fibers connected with the anterior and orbital PFCs in humans (*25*), consistent with metacognitive ability being dependent not only on anterior prefrontal gray 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 gray-matter volume or white-matter FA 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 minimizing variability in both d' and B_{roc} to isolate the neural correlates of introspective ability (A_{roc}).

One concern is that the structural covariation that we observed may have been potentially confounded by differences in perceptual ability. Good perceptual ability may be reflected in the staircase procedure converging on consistently low values for stimulus contrast for a given individual. Therefore, we carried out control analyses (table S4) (*21*) to rule out this alternative explanation. These results demonstrated significant correlations of gray matter and FA with A_{roc} in the anterior PFC when controlling for changes in task parameters and an absence of correlations with task parameters themselves. Thus, the structure-behavior 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 the PFC 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*). 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 gray-matter volume playing a causal role in metacognition, patients with lesions to the anterior PFC show deficits in subjective reports as compared with controls, after factoring out differences in objective performance (*16*). Furthermore, impairing dorsolateral PFC function with theta-burst transcranial magnetic stimulation compromises the metacognitive sensitivity of subjective reports of awareness but leaves underlying task performance intact (*30*). Together with the present work, these findings suggest a central role for anterior and dorsolateral PFC in metacognitive sensitivity. Our present findings may reflect innate differences in anatomy or, alternatively, may reflect the effects of experience and learning, as has been found in the sensorimotor domain (*14*, *15*). This raises the tantalizing possibility of being able to "train" metacognitive ability by harnessing underlying neural plasticity in the regions that we identify here (*31*).

Our main finding is a delineation of a noticeably focal anatomical substrate that predicts interindividual variability in metacognitive ability. As with any correlational method, we cannot establish whether the covariation we observed between brain structure and metacognition reflects a causal relation. However, given a wealth of evidence for changes in gray-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 to the biological basis 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.

References and Notes

1. A. Cleeremans, B. Timmermans, A. Pasquali, *Neural Networks* **20**, 1032 (2007).
2. J. Metcalfe, *Metacognition: Knowing About Knowing* (MIT Press, Cambridge, MA, 1996).
3. H. C. Lau, *Prog. Brain Res.* **168**, 35 (2007).
4. D. A. Washburn, J. D. Smith, L. A. Taglialatela, *J. Gen. Psychol.* **132**, 446 (2005).
5. C. Kunimoto, J. Miller, H. Pashler, *Conscious. Cogn.* **10**, 294 (2001).
6. R. Szczepanowski, L. Pessoa, *J. Vis.* **7**, 1 (2007).
7. M. Graziano, M. Sigman, A. Rustichini, *PLoS ONE* **4**, e4909 (2009).
8. S. M. Fleming, R. J. Dolan, *Conscious. Cogn.* **19**, 352 (2010).
9. R. Kiani, M. N. Shadlen, *Science* **324**, 759 (2009).
10. A. Kepecs, N. Uchida, H. A. Zariwala, Z. F. Mainen, *Nature* **455**, 227 (2008).
11. M. Carreiras *et al.*, *Nature* **461**, 983 (2009).
12. D. S. Tuch *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12212 (2005).
13. L. Fuentemilla *et al.*, *J. Neurosci.* **29**, 8698 (2009).
14. J. Scholz, M. C. Klein, T. E. J. Behrens, H. Johansen-Berg, *Nat. Neurosci.* **12**, 1370 (2009).
15. B. Draganski *et al.*, *Nature* **427**, 311 (2004).
16. A. Del Cul, S. Dehaene, P. Reyes, E. Bravo, A. Slachevsky, *Brain* **132**, 2531 (2009).
17. A. P. Shimamura, *Conscious. Cogn.* **9**, 313 (2000).
18. J. S. Simons, P. V. Peers, Y. S. Mazuz, M. E. Berryhill, I. R. Olson, *Cereb. Cortex* **20**, 479 (2010).
19. P. C. Fletcher, R. N. A. Henson, *Brain* **124**, 849 (2001).
20. K. Christoff, J. D. E. Gabrieli, *Psychobiology* **28**, 168 (2000).
21. Materials, methods, discussion of ROC model fits, and details of control analyses are available as supporting material on *Science* Online.
22. S. J. Galvin, J. V. Podd, V. Drga, J. Whitmore, *Psychon. Bull. Rev.* **10**, 843 (2003).
23. D. E. Kornbrot, *Percept. Psychophys.* **68**, 393 (2006).
24. C. D. Smith, H. Chebrolu, D. R. Wekstein, F. A. Schmitt, W. R. Markesbery, *Neurobiol. Aging* **28**, 1075 (2007).
25. H. J. Park *et al.*, *Hum. Brain Mapp.* **29**, 503 (2008).
26. N. D. Daw, J. P. O'Doherty, P. Dayan, B. Seymour, R. J. Dolan, *Nature* **441**, 876 (2006).
27. P. W. Burgess, I. Dumontheil, S. J. Gilbert, *Trends Cogn. Sci.* **11**, 290 (2007).
28. N. Ramnani, A. M. Owen, *Nat. Rev. Neurosci.* **5**, 184 (2004).
29. H. C. Lau, R. E. Passingham, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18763 (2006).
30. E. Rounis *et al.*, *Cognit. Neurosci.* **1**, 165 (2010).
31. E. B. Titchener, *Lectures on the Experimental Psychology of the Thought-Processes* (Macmillan, New York, 1909).
32. J. Metcalfe, in *Handbook of Metamemory and Memory*, J. Dunlosky, R. A. Bjork, Eds. (Psychology Press, New York, 2008), pp. 27–46.

33. This research was funded by the Wellcome Trust (G.R., R.J.D., Z.N.), the University College London four-year Ph.D. program in neuroscience (S.M.F.), and the Medical Research Council (R.S.W.). We thank T. Sharot, G. Ridgway, and C. Frith for comments on earlier drafts of this manuscript.

Supporting Online Material

www.sciencemag.org/cgi/content/full/329/5998/1541/DC1
Materials and Methods
SOM Text
Figs. S1 to S4
Tables S1 to S4
References

5 May 2010; accepted 23 July 2010
10.1126/science.1191883