◆ **Human Brain Mapping 00:00–00 (2016)** ◆

Substrates of Metacognition on Perception and Metacognition on Higher-order Cognition Relate to Different Subsystems of the Mentalizing Network

Sofie L. Valk,[1†*] **Boris C. Bernhardt,**[1,2†] **Anne Böckler,**[1,3] **Philipp Kanske,**[1] **and Tania Singer**[1]

[1]Department of Social Neuroscience, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[2]Neuroimaging of Epilepsy Lab and McConnell Brain Imaging Center, Department of Neurology and Neurosurgery, Montreal Neurological Institute and Hospital, McGill University, Montreal, QC, Canada
[3]Department of Psychology, Julius Maximilians University, Würzburg, Germany

◆

Abstract: Humans have the ability to reflect upon their perception, thoughts, and actions, known as metacognition (MC). The brain basis of MC is incompletely understood, and it is debated whether MC on different processes is subserved by common or divergent networks. We combined behavioral phenotyping with multi-modal neuroimaging to investigate whether structural substrates of individual differences in MC on higher-order cognition (MC-C) are dissociable from those underlying MC on perceptual accuracy (MC-P). Motivated by conceptual work suggesting a link between MC and cognitive perspective taking, we furthermore tested for overlaps between MC substrates and mentalizing networks. In a large sample of healthy adults, individual differences in MC-C and MC-P did not correlate. MRI-based cortical thickness mapping revealed a structural basis of this independence, by showing that individual differences in MC-P related to right prefrontal cortical thickness, while MC-C scores correlated with measures in lateral prefrontal, temporo-parietal, and posterior midline regions. Surface-based superficial white matter diffusivity analysis revealed substrates resembling those seen for cortical thickness, confirming the divergence of both MC faculties using an independent imaging marker. Despite their specificity, substrates of MC-C and MC-P fell clearly within networks known to participate in mentalizing, confirmed by task-based fMRI in the same subjects, previous meta-analytical findings, and *ad-hoc* Neurosynth-based meta-analyses. Our integrative multi-method approach indicates domain-specific substrates of MC; despite their divergence, these nevertheless likely rely on component processes mediated by circuits also involved in mentalizing. *Hum Brain Mapp 00:000–000, 2016.* © **2016 Wiley Periodicals, Inc.**

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: European Research Council under the European Community's Seventh Framework Program; Contract grant number: FP7/2007-2013/ERC Grant agreement number 205557 [EMPATHICBRAIN]; Contract grant sponsor: Canadian Institutes of Health Research

[†]These authors contributed equally to the study

*Correspondence to: Sofie Valk, Department of Social Neuroscience, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany. E-mail:valk@cbs.mpg.de

Received for publication 8 December 2015; Revised 15 April 2016; Accepted 24 April 2016.

DOI: 10.1002/hbm.23247

Key words: metacognition; structural MRI; individual differences; cognition; theory of mind; interoceptive accuracy

◆

INTRODUCTION

Metacognition (MC) is the process through which individuals reflect on and control their perceptions, memories, and actions [Metcalfe, 1994]. A key aspect of MC is the evaluation of one's own performance, and such metacognitive accuracy can be measured based on the fidelity of subjects' trial-by-trial confidence judgments with respect to their objective task performance [Clarke et al., 1959; Fleming et al., 2012a; Galvin et al., 2003; Koriat 2007; Maniscalco and Lau, 2012]. MC has been assessed by studying confidence ratings of signal detection task performance for perception [Fleming et al., 2010, 2012a] and with respect to memory accuracy [Koriat and Bjork, 2006; Oppenheimer, 2008]. It has been suggested that MC may be key to executive function [Koriat, 2007], social cognition [Frith, 2012], and mental health [Teasdale et al., 2002].

Despite its important role in guiding behavior and cognition, fundamental aspects of MC are still incompletely understood. Particularly, it is debated whether MC is a domain-specific skill (i.e., MC abilities are specific to a given domain, and are not mandatory to translate to others) or rather domain-general (i.e., MC is a general ability, independent of the specific domain involved). Behavioral studies addressing this question, mostly focusing MC on memory *vs* MC on perception (MC-P), have been rather inconclusive, with some work showing correlations between individual abilities across both MC faculties [McCurdy et al., 2013], while others suggested their independence [Baird et al., 2013; Fleming et al., 2014]. Similarly, the brain basis of MC is still unknown. On the one hand, functional and structural neuroimaging studies have frequently localized substrates of MC in prefrontal regions [Baird et al., 2013, 2013; Fleming et al., 2010, 2012b, 2014; McCaig et al., 2011; Yokoyama et al., 2010], possibly suggesting some domain-generality. In an influential theory, MC was related to mentalizing, also called Theory of Mind (ToM) [Frith, 2012], where taking the perspective on one's own actions may rely on processes and (largely prefrontal) networks similar to those involved in taking the perspective of others [Lombardo et al., 2010; Frith, 2012]. Yet, prefrontal substrates for MC have not always been reported [McCurdy et al., 2013], and there is overall only modest overlap across studies trying to localize MC [Baird et al., 2013; Cabeza, 2008; Fleck et al., 2006; Henson et al., 2000], possibly challenging domain-general accounts. Furthermore, no study to date has compared MC on perception to MC on higher-order cognition, involving high-level factual reasoning and mentalizing.

Based on previous work, one can thus test for two competing hypotheses. The first is that MC may be a rather domain-general ability, and that different types of MC may relate to mentalizing networks. Alternatively, MC on different processes may relate to different networks, possibly those involved in the basic process itself. This would then speak for a shared network of only MC on mentalizing/factual reasoning and mentalizing. To address this topic at a structural-anatomical level, we carried out a multimodal neuroimaging assessment in a large cohort of healthy participants and contrasted substrates of individual differences in well-established markers of MC-P with a novel marker of MC on the accuracy in a high-level cognitive task (henceforth, MC-C; [Kanske et al., 2015]. Our advanced MRI framework assessed cortical morphology and diffusion anisotropy of the superficial white matter (WM) running immediately below the cortical mantle, for the evaluation of multimethod consistency. Analyses were spatially unconstrained (i.e., no a-priori ROIs were chosen) to objectively evaluate whether substrates of MC-P and MC-C would overlap or diverge. However, based on previous conceptual suggestions indicating that MC abilities may rely on operations that also play a role in mentalizing [Frith, 2012], we tested specifically whether substrates of both might generally fall into subcomponents of mentalizing networks, such as the medial prefrontal, lateral temporal, temporo-parietal, and parietal midline regions [Bzdok et al., 2012; Mar, 2011]. We therefore assessed the overlap between our structural MRI findings and a map of task-based fMRI activations obtained during a mentalizing task in the same subjects from a previously published study [Kanske et al., 2015], meta-analytical data on imaging studies on mentalizing [Bzdok et al., 2012], and ad-hoc forward and reverse inference obtained from Neurosynth (http://www.neurosynth.org/).

MATERIALS AND METHODS

Participants

Following an extensive advertising campaign for a large-scale longitudinal study in 2013 (for details, see [Singer et al., in press]), we studied 191 consecutively enrolled healthy adults (116 women, age mean ± SD = 41 ± 10 years, 20–55 years). Participants were recruited as two matched subsamples from the cities of Berlin ($n = 93$; 55 women, age mean ± SD = 43 ± 8.4 years, 26–55 years) and Leipzig ($n = 98$; 61 women, age mean ± SD = 39.2 ± 10.1 years, 20–55 years). Participants had normal-to-high IQ (mean ± SD 115 ± 15 years, 78–152 years), an average of 18 ± 3 years of education, and normal or corrected-to-normal vision. Volunteers gave written and informed consent prior to participation. The study was approved by the Research Ethics Committees of the University of Leipzig and the Humboldt University in Berlin.

Measuring Metacognition

In two separate sessions, participants completed a perceptual discrimination task and a high-level mentalizing/factual reasoning task (Figs. 1A and 2A). In both measures, signal-detection-theory [Green and Swets, 1974] was used to quantify individual differences in metacognitive ability ("type II sensitivity"), here defined as the ability to accurately link confidence with performance.

MC-P

A visual discrimination-task [Baird et al., 2013; Fleming et al., 2010; Song et al., 2011] was performed at an individually determined threshold. Each trial ($n = 120$) consisted of a display of six Gabor patches, followed by a blank screen, prior to a second display of six patches. In one of the two displays, one randomly selected Gabor patch was rotated using an adaptive staircase procedure. Participants were asked to assess whether the shift occurred in the first or second stimulus display, prior to them rating their confidence in the accuracy of their response on a trial-by-trial basis on a Likert scale [ranging from 0 (not confident) to 6 (confident)].

MC-C

The EmpaToM is a newly developed task measuring empathy, compassion, mentalizing, and MC based on naturalistic video stimuli [Kanske et al., 2015]. In this 30-min-long paradigm, people recount autobiographical episodes that are either emotionally negative (e.g., loss of a loved one) or neutral (e.g., commuting to work). Multiple-choice questions with three response options after each video assessed either mentalizing (24 questions about the mental states of people in the video) or factual reasoning (24 questions about the content of the story). After each trial (total of 48 trails), participants rated their confidence in performance accuracy after having performed the high-level cognition questions on a trial-by-trial basis on a continuous rating scale (ranging from "not confident" to "confident"). Agreement between confidence ratings and actual performance was then used as MC measure.

We chose the well-established receiver operator characteristic, *ROC*, to quantify meta-cognitive accuracy across both tasks using SPSS (Version 22, IBM, Armonck, NY). ROC analyses have been applied in several neuroimaging studies assessing MC [Baird et al., 2013; Fleming et al., 2010; Song et al., 2011]. Please note that the set-up of the MC-C task does not allow for separate computation of the so-called *meta d-prime* [Maniscalco and Lau, 2012] metric, an alternative index of MC. Kolmogorov–Smirnov-tests indicated that MC-P (% errors: mean ± SD = 25.8 ± 5.3, range = 15.9–49.0) and MC-C (% errors: mean ± SD = 40.6 ± 12.4, range = 12.5–68.7) were both normally distributed.

MRI Acquisition

MRI data from all participants, irrespective of recruitment site, were acquired on the same 3T Siemens Magnetom Verio (Siemens Healthcare, Erlangen, Germany) using a 32-channel head coil in Leipzig. Structural images were acquired using a T1-weighted 3D-MPRAGE sequence (176 sagittal slices, repetition time [TR] = 2,300 ms, echo time [TE] = 2.98 ms, flip angle = 7°, field-of-view [FOV] = 240×256 mm^2, matrix = 240×256, voxel size = $1 \times 1 \times 1$ mm^3). Diffusion-weighted images (DWI) were obtained using twice-refocused EPI sequence (TR = 8,900 ms, TE = 90 ms, flip angle = 90°, FOV = 210×210 mm^2, matrix = 110×110, no gap, 63 axial slices, voxel size = $1.9 \times 1.9 \times 1.9$ mm^3, 64 diffusion directions with $b = 1{,}000$ s mm^{-2} along with seven interspersed non-diffusion weighted volumes). Functional masks were derived from a task-based fMRI paradigm described in more detail by Kanske et al. [2015].

MRI-based Cortical Thickness Measurements

We used FreeSurfer to generate cortical surface models and to measure cortical thickness from T1-weigthed MRI (Version 5.1.0; http://surfer.nmr.mgh.harvard.edu). Previous work has cross-validated FreeSurfer with histological analysis [Cardinale et al., 2014; Rosas et al., 2002] and manual measurements [Kuperberg et al., 2003]. Processing steps have been outlined in detail elsewhere [Dale et al., 1999; Fischl et al., 1999; Han et al., 2006]. Following surface extraction, sulcal and gyral features of an individual were warped to an average spherical representation, fsaverage5, which allows for accurate matching of measurement locations across participants. Surfaces were visually inspected and inaccuracies manually corrected (SLV, BCB). The current analysis employed a 20-mm full-width-at-half-maximum (FWHM) Gaussian smoothing kernel, following previous recommendations [Lerch and Evans, 2005] and previous studies of our group and others [Bermudez et al., 2009; Bernhardt et al., 2010; Doyle-Thomas et al., 2013; Lerch et al., 2005; Shaw et al., 2006, 2015; Valk et al., 2016]. Smoothing was carried out along cortical surface topology, to minimize partial volume effects and to offer high anatomical sensitivity and specificity.

Assessment of the Superficial White Matter

DWI preprocessing

Preprocessing, based on FSL (Version 5.0; http://www.fmrib.ox.ac.uk/fsl), involved motion correction, eddy current correction, and estimation of the diffusion tensor and fractional anisotropy (FA), a measure of directionality of water diffusion. A boundary-based registration [Greve and

Fischl, 2009] aligned FA images with T1-weighted MRI by maximizing the intensity gradient across tissue boundaries, using the surfaces that separate brain structures and tissue types of the T1-weighted reference image, and the tissue intensity of diffusion image.

White matter surface generation

For each participant, we generated a surface to systematically sample diffusion anisotropy of the superficial WM immediately below the cortical mantle, similar to previous work [Fjell et al., 2008; Kang et al., 2011, 2012]. We estimated a Laplacian deformation field running from the cortical interface toward the ventricles, which guided subsequent placement of a surface running ∼2 mm below the cortical interface. A Laplacian field guarantees point-wise correspondence between cortical and superficial WM surfaces, a requirement for meaningful integration of cortical thickness and WM diffusion measures. The correspondence guarantees implicit between-subject alignment of WM parameters via the surface-registration estimated at the level of the neocortex. In agreement with previous findings [Kang et al., 2011, 2012], high anisotropy-values were detected in regions proximal to the corpus callosum, bilateral central cortices, and insula; conversely, low anisotropy was observed in occipital regions, temporo-parietal regions, and medial prefrontal cortices. Similar to thickness measures, superficial WM anisotropy data were surface-smoothed at 20-mm FWHM.

Quality control and case selection

A total of 155/191 (91 women, mean ± SD age = 40.1 ± 9.5 years; Berlin: 78, Leipzig: 77) participants had complete metacognitive and performance measures (Supporting Information Fig. 1). Quality controlled thickness measures were available in all remaining 155 participants, DWI data in 151 (Berlin: 77, Leipzig: 74).

Statistical Analyses

As in previous work [Bernhardt et al., 2014b; Bernhardt et al., 2015; Valk et al., 2015], analyses were performed using SurfStat [Worsley et al., 2009] for Matlab (Version 2013b; The Mathworks, Natick, MA).

Behavioral analysis

Correlational analyses tested for pair-wise associations between individual differences in MC-C, MC-P, performance measures, and IQ.

Cortical substrates: Cortical thickness mapping

Linear models at each cortical surface-point i assessed the relationship between thickness T and metacognitive capacities:

$$T_i = \beta_0 + \beta_1^* Sex + \beta_2^* Age + \beta_3^* Performance + \beta_4^* MC$$

T_i is the thickness at surface-point i. In the formula above, we corrected for Sex and Age, given their marked effects on brain structure [Salat et al., 2004; Sowell, et al., 2003, 2007]; the *MC* effects of interest were either MC-P or MC-C scores. In each domain, we controlled for *Performance*, given a possible relationship between task-performance and MC [Fleming et al., 2012a; Galvin et al., 2003; Maniscalco and Lau, 2012]. To account for shared variance between MC-P and MC-C, we also analyzed models that controlled for MC-C score when calculating brain correlates of MC-P, and including MC-P score when calculating brain correlates of MC-C. We also analyzed the model including recruitment site (Berlin, Leipzig) as a covariate to rule out possible recruitment effects, and a model including IQ score to rule out confounds of general intelligence.

Analysis of the superficial WM

We followed the same modeling approach as in (b) to evaluate the relationship between individual differences in MC and diffusion FA sampled from the superficial WM.

Multi-modal overlap analysis

Within each MC domain (i.e., MC-P, MC-C), we intersected findings from cortical thickness and superficial WM analysis (i.e., (b) and (c)) to assess common substrates across MRI modalities. Within intersections, we evaluated effects of the other MC domain, as above, to test for specificity.

Multi-method overlaps with mentalizing network

We overlaid results generated by (d) with task-based activations during ToM questions in the same subjects from a previously published study [Kanske et al., 2015], a previously published meta-analysis on mentalizing [Bzdok et al., 2012], and an automated *ad-hoc* meta-analysis using both forward and reverse inference masks of mentalizing based on a total of 124 studies at the time of study in Neurosynth (October 2015; http://www.neurosynth.org/analyses/terms/mentalizing/). Furthermore, to test the association of the MC-C and MC-P correlates with ROIs based on the fMRI ToM localizer in the same subjects [Kanske et al., 2015], we performed a post-hoc analysis correlating thickness in the 10 largest regions of this activation mask (cluster size > 250 mm^2) with MC-P and MC-C score.

Correction for multiple comparisons

Surface-based findings were adjusted using random field theory for nonisotropic images [Worsley et al., 1999], controlling the probability of reporting a family-wise error to $P_{\mathrm{FWE}} < 0.05$. *Post-hoc* analyses within overlaps were corrected at a family-wise error level of $P_{\mathrm{FWE}} < 0.05$ using Bonferroni-adjustment.

RESULTS

Behavioral Findings

Neither MC-P nor MC-C showed significant correlations with task accuracy in their respective domain ($r = -0.11$, $P > 0.1$). Importantly, although task-accuracies marginally correlated across both domains ($r = 0.17$, $P < 0.04$), MC-P and MC-C did not correlate ($r = -0.08$, $P > 0.1$). No correlation was consistently seen when additionally controlling for age and sex. Power analysis (G*power 3.1; [Faul et al., 2009]) based on the observed effect size indicated that >1600 participants would have been necessary to detect a positive correlation, even at uncorrected thresholds (set 1-beta = 0.9, alpha = 0.05, two-tailed). For a single inter-correlation between two behavioral variables, our study in general had a high power to detect small-to-medium effects (for an $r = 0.25$, we would have achieved a power of 1-beta = 0.94 with 155 subjects at alpha = 0.05). On the other hand, for very small effects, such as the correlation of $r = -0.08$ between MC-C and MC-P, our large sample was not offering sufficient power (1-beta = 0.25).

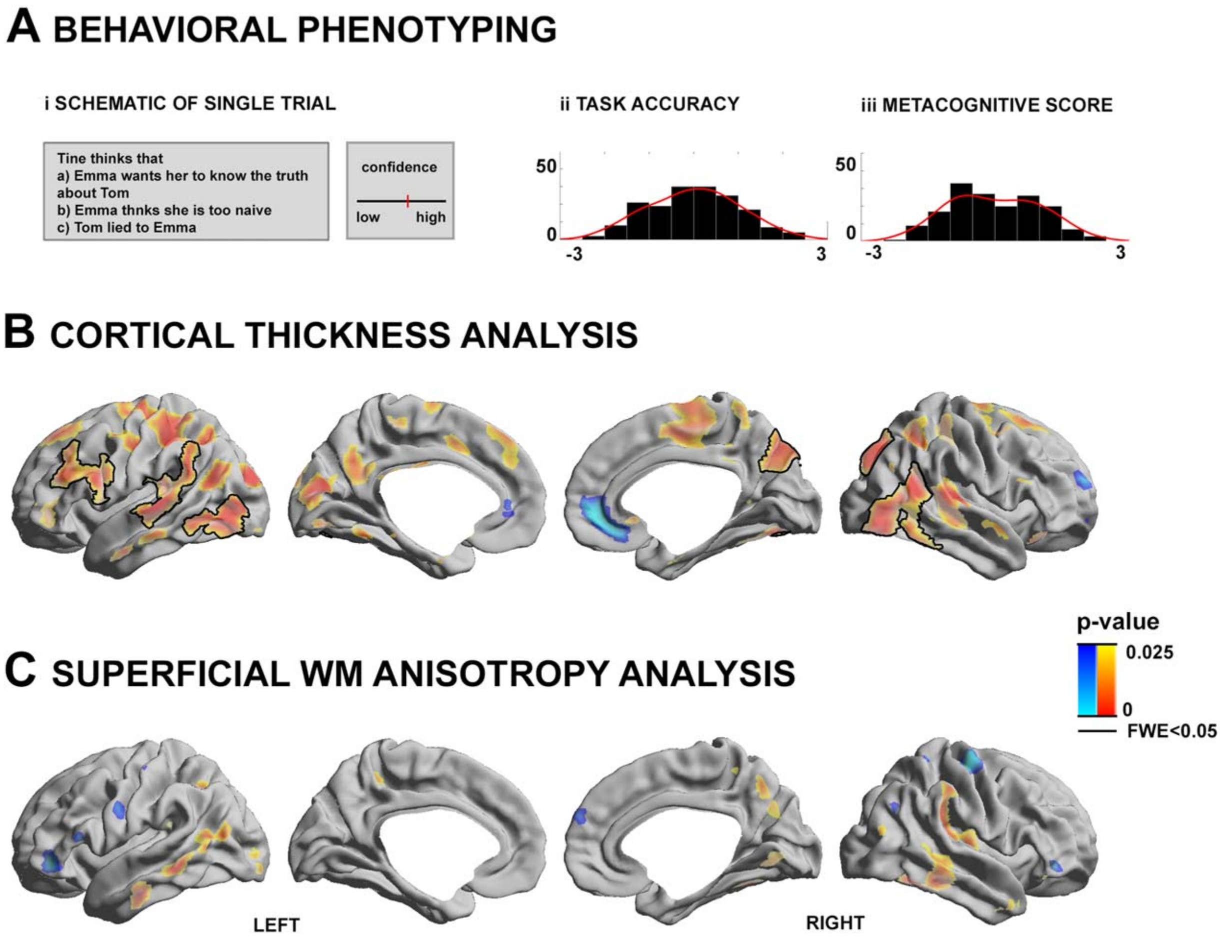

Figure 1.

Substrates of metacognition on higher-order cognition (MC-C). (**A**) i Behavioral assessment: all participants underwent the EmpaTom-task ($n = 48$ trials) [Kanske et al., 2015], where they were asked to answer three-way multiple choice questions concerning details of 15-sec video-stories they just saw, followed by confidence ratings; ii Distribution of task-performance (left) and respective MC-C scores (right) (z-scored); (**B**) Cortical thickness correlates of individual differences in MC-C, where *warm/cold* colors indicate cortical thickness *increases/decreases* in individuals with higher MC-C scores; (**C**) Superficial WM anisotropy (FA) correlates with individual differences in MC-C. To correct for multiple comparisons in B and C, findings were thresholded at $P_{\mathrm{FWE}} < 0.05$, using random field theory for non-isotropic images (black outlines), superimposed on trends (semi-transparent, no black outlines).

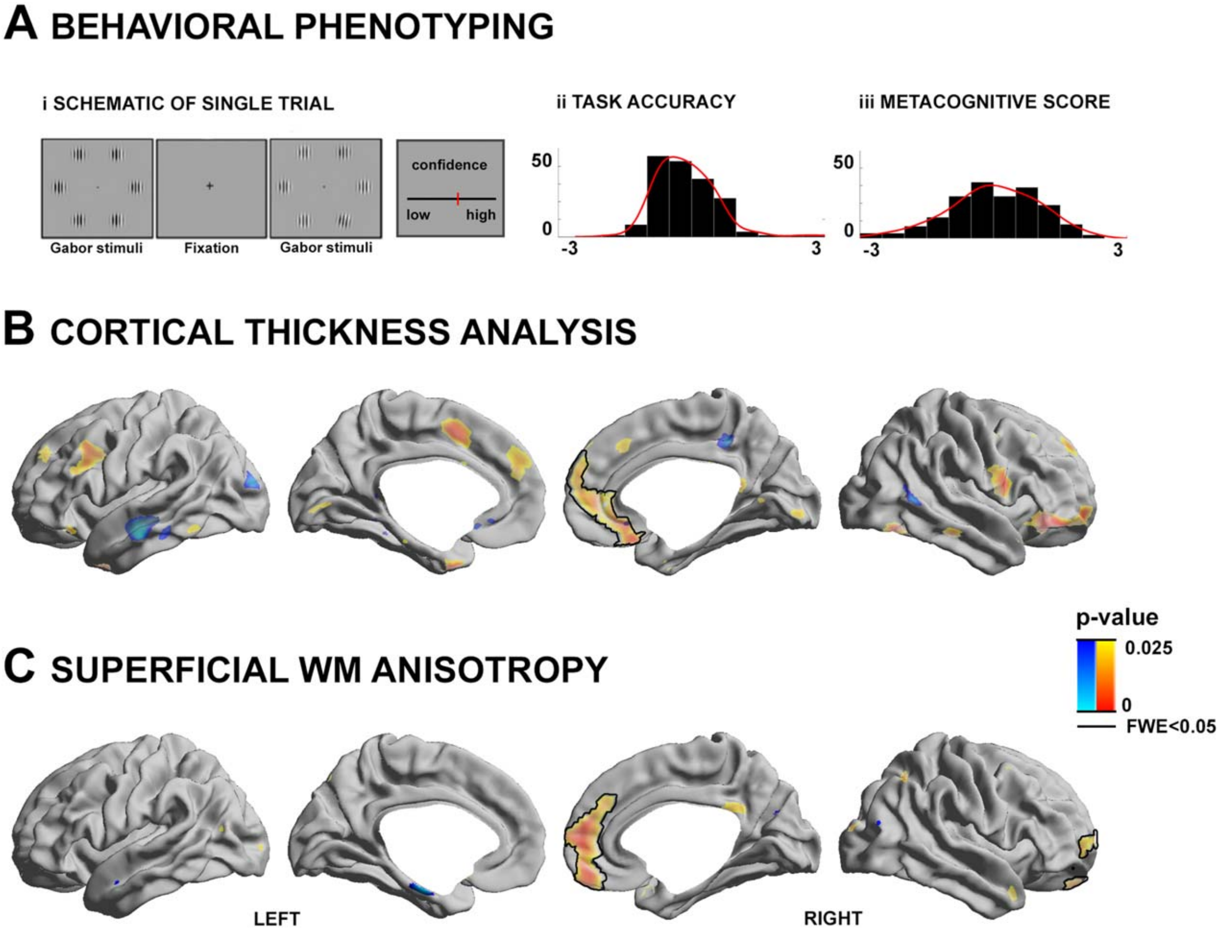

Figure 2.

MC-P, metacognition on perception. (**A**) **i** *Behavioral assessment*: participants underwent a perceptual discrimination task, where they were asked whether the first or second Gabor-patch included titled lines, followed by confidence ratings ($n = 120$ trials); **ii** Distribution of z-scored behavioral score distribution accuracy and respective MC-P scores (z-scored); (**B**) Cortical thickness correlates of individual differences in MC-P; (**C**) Superficial WM anisotropy (FA) correlates with individual differences in perceptual metacognition. Please see Figure 1 for further details.

Individual difference analysis, thus, suggests independence between the abilities to accurately evaluate one's own performance during perceptual processing as opposed to higher-level cognitive tasks.

Cortical Substrates of Metacognitive Abilities

Cortical thickness analysis complemented behavioral findings and revealed specific, non-overlapping structural substrates of individual differences in both MC domains.

Higher MC-C scores correlated positively with thickness increases in a large bilateral cluster encompassing lateral frontal, superior and inferior temporal, temporo-parietal, and posterior midline regions ($P_{\mathrm{FWE}} < 0.05$; Fig. 1). Separate assessment of WM diffusion parameters revealed overlapping, yet more restricted and only marginally significant positive correlations between MC-C and FA of right middle temporal regions ($P_{\mathrm{FWE}} < 0.07$), please see Supporting Information Table I for further details on the clusters reported.

Considering MC-P (Fig. 2), we observed a positive correlation with thickness in right medial prefrontal and anterior cingulate cortices ($P_{\mathrm{FWE}} < 0.01$), as well as in right lateral prefrontal regions at trend levels ($P_{\mathrm{FWE}} < 0.09$). Individuals with a more accurate assessment of their visual perceptual sensitivity, thus, showed increased thickness in these regions relative to those with low accuracy. Additionally, diffusion anisotropy findings in the superficial WM converged with those at the level of cortical thickness, by showing colocalized, albeit more restricted substrates in medial prefrontal cortices ($P_{\mathrm{FWE}} < 0.03$). Please see

Supporting Information Table II for further details on the clusters reported. Findings were consistent when controlling for the other MC scores (i.e., MC-C score when assessing MC-P correlates, MC-P score when assessing MC-C correlates; Supporting Information Fig. 2), when controlling for recruitment site (Supporting Information Fig. 3), and when controlling for IQ (Supporting Information Fig. 4). Complementing the partial correlation analysis, we assessed differences in correlations in clusters of findings using a framework that accounts for shared variables [Steiger, 1980]. For cortical thickness, this analysis confirmed marked interactions in mPFC ($Z = -3.3$, $P < 0.001$) and trends in right STG/STS ($Z = 1.74$, $P < 0.1$). For FA, it confirmed marked interactions in mPFC ($Z = -2.9$, $P < 0.005$) and STG/STS ($Z = 2.3$, $P < 0.02$). Interactions in right inferior temporal regions were pointing in consistent directions, but were of lower significance (CT: $Z = 1.65$, $P = 0.1$, FA: $Z = 0.9$. n.s.).

Multimethod Overlaps

Cortical thickness and WM diffusion anisotropy findings (at $P_{\mathrm{FWE}} < 0.1$) overlapped within a given MC domain, providing cross-validation across two imaging modalities and reemphasizing divergence of MC-P and MC-C (Fig. 3, Supporting Information Table II). Indeed, correlations of individual differences in MC-P with cortical thickness and FA of the superficial WM consistently intersected in medial prefrontal regions, while cortical thickness and diffusion substrates of MC-C showed overlaps in right lateral temporal, and superior temporal sulcus/gyrus. *Post-hoc* analysis showed that mean thickness and FA in the overlap cluster relating to individual differences of MC-C did not correlate with MC-P individual difference scores, and vice versa, further supporting their independence.

Overlap with Mentalizing Network

Despite their regional divergence, substrates of MC-P and MC-P were both falling into a network known to play a role in mentalizing (Fig. 4, Supporting Information Table III). First, overlaps were observed with task-based fMRI activations (during a mentalizing *vs* control condition) carried out in the same subjects [Kanske et al., 2015], suggesting sample-specific relevance (see panel A). A subset of ROIs (cluster size > 250 mm^2) based on this functional contrast correlated with MC scores. Please see Supporting Information Table IV. Second, intersections also overlapped with previous meta-analytical findings in the domain of mentalizing (please see Bzdok et al. [2012]; panel B). Last, they also fell into regions highlighted by *ad-hoc* meta-analytical forward/reverse inference based on 124 studies using *Neurosynth* (see panel C and D).

DISCUSSION

To study the brain basis of meta-cognition, the current work tested whether meta-cognitive abilities in different domains share common or distinct structural-anatomical substrates. We contrasted a widely studied measure of meta-cognitive accuracy on perception, MC-P [Fleming et al., 2010], with a novel marker of meta-cognitive accuracy on higher-level cognition, MC-C [Kanske et al., 2015]. Behavioral correlation analyses in a large sample of healthy adults revealed no evidence for a significant association between individual differences in MC-P and MC-C; moreover, multimodal MRI analyses indicated that individual variations in both domains related to specific and non-overlapping substrates at the level of cortical morphology and diffusion anisotropy of the superficial WM (MC-P: prefrontal regions, MC-C: parietal and posterior temporal cortices). The lack of behavioral inter-correlation together with the divergence of structural substrates speaks against a domain-general substrate of MC. Despite the regional divergence, however, overlaying findings with a task-based fMRI localizer from the same sample [Kanske et al., 2015] and meta-analytical data across previous studies demonstrated that each MC substrate clearly fell within a different large-scale functional network known to be involved in mentalizing. Our multi-method structural and diffusion MRI findings, thus, suggest that different components in this overarching network may subserve both the reflection about self and others, allowing for meta-cognitive processing in general as well as introspection.

At a behavioral level, the observed lack of correlations between MC-C and MC-P supports claims that metacognitive capacities may be domain-selective [Baird et al., 2013; Kelemen et al., 2000; Pannu et al., 2005; Schnyer et al., 2004]. As our findings are based on one of the largest cohorts published in the MC literature to date, our sample has sufficiently high power for even small-to-medium effects. Indeed, for the observed correlations as small as $r = -0.08$, samples of 1,600 and more participants would have been necessary to obtain even marginal significance levels at a power of 0.9. Multimodal neuroimaging complemented behavioral results by showing non-overlapping structural substrates of each meta-cognitive faculty. Indeed, while MC-C ability correlated positively with thickness of lateral and midline parietal cortices, as well as lateral temporal and dorsolateral prefrontal regions, MC-P related to anterior and medial prefrontal grey matter. Superficial WM analysis provided further support for their distinction, by highlighting effects in similar regions, albeit of a more restricted extent. Our confidence in the validity of our findings is high, as results were consistent across different imaging modalities and when controlling for various possible confounds such as recruitment site and IQ. On a histological level, macroscopic cortical thickness changes may reflect alterations in neuronal and synaptic markers [Giedd et al., 1999; Sowell et al., 2003], while

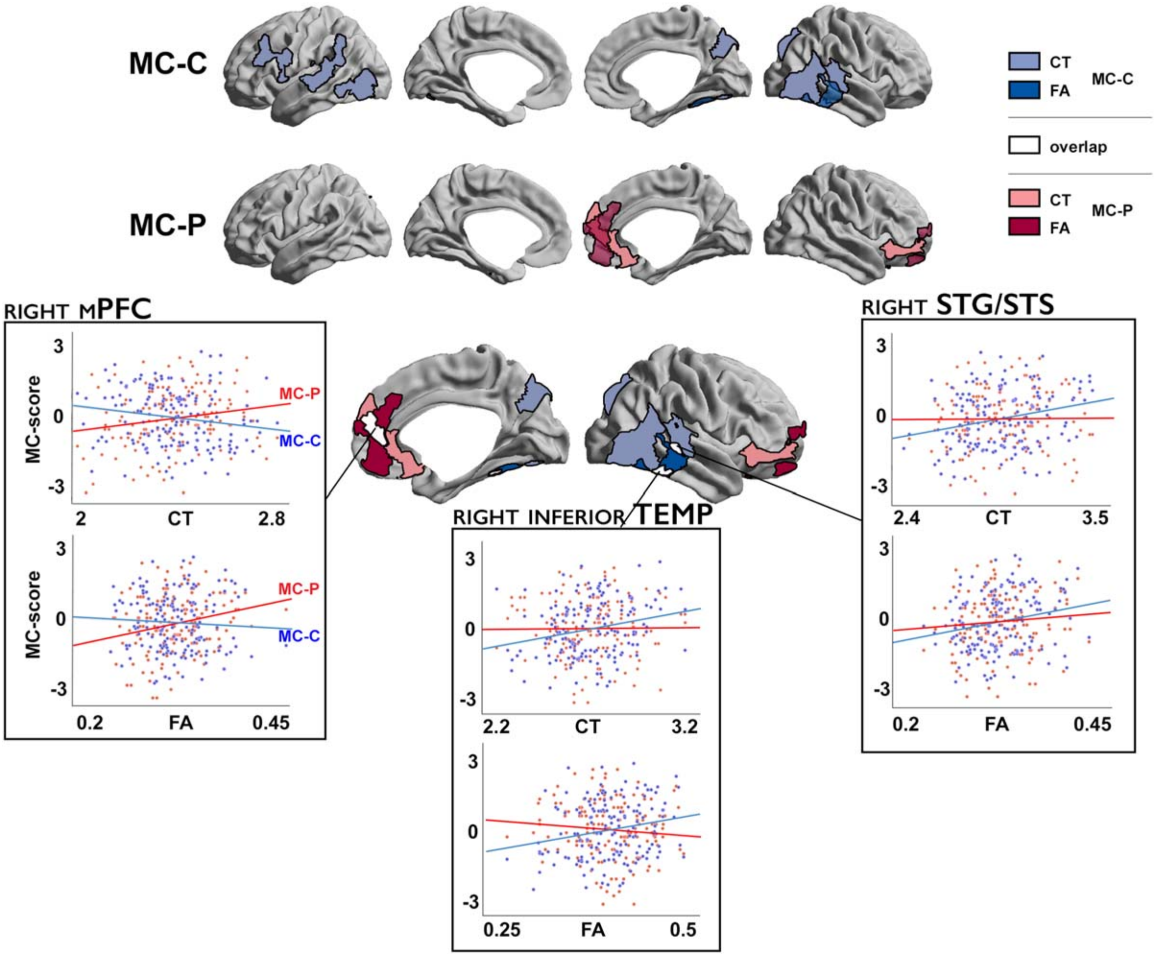

Figure 3.

Multi-modal overlap analysis. The small brains in top panels show cortical thickness (CT) (light blue, $P_{\mathrm{FWE}} < 0.1$) and superficial WM anisotropy (FA) (dark blue, $P_{\mathrm{FWE}} < 0.1$) substrates of MC-C and MC-P, respectively. Overlaps of substrates across both modalities are delineated in white. The larger surface rendering display MC-P and MC-C on the same brain, illustrating their topographic divergence. Scatterplots are shown for multi-modal overlaps within each MC domain (left two scatter plots: FA/CT overlap for MC-P in medial prefrontal cortex; middle two scatter plots: FA/CT overlap in inferior temporal regions; right two scatter plots: FA/CT overlap in superior temporal regions). MC-P/MC-C scores are marked in red/blue.

superficial WM anisotropy may relate to membrane integrity, myelination, and fiber arrangement [Concha et al., 2010; Fjell et al., 2008; Kang et al., 2012]. Given their overlap, our findings may ultimately reflect to a more efficient architecture of anatomical connections within and below these regions in individuals with high meta-cognitive ability, possibly allowing for more integrated functional processing within each domain.

Motivated by conceptual accounts suggesting a link between MC and mentalizing [Frith, 2012], we overlaid the location of structural-anatomical substrates with meta-analytical as well as task-based fMRI findings on mentalizing. Despite them being non-overlapping, substrates of MC-P and MC-C clearly fell into components of the mentalizing circuitry, supporting the notion that similar brain regions may be essential for understanding both the self and others [Amodio and Frith, 2006; Frith, 2012; Lombardo et al., 2010; Mitchell et al., 2006; Saxe et al., 2006; Singer et al., 2004]. The mentalizing network has been related to processes sub-serving simulating an alternate viewpoint to the present, a process by which mental models are used adaptively to imaging events beyond those that emerge from the current setting [Buckner et al., 2008]. These findings may also relate to the substantial similarity between the mentalizing circuitry and the default mode network [Buckner and Vincent, 2007; Fox et al., 2005; Greicius et al., 2003; Raichle et al., 2001]; a network active in the absence of a specific task, and related to self-generated, internally

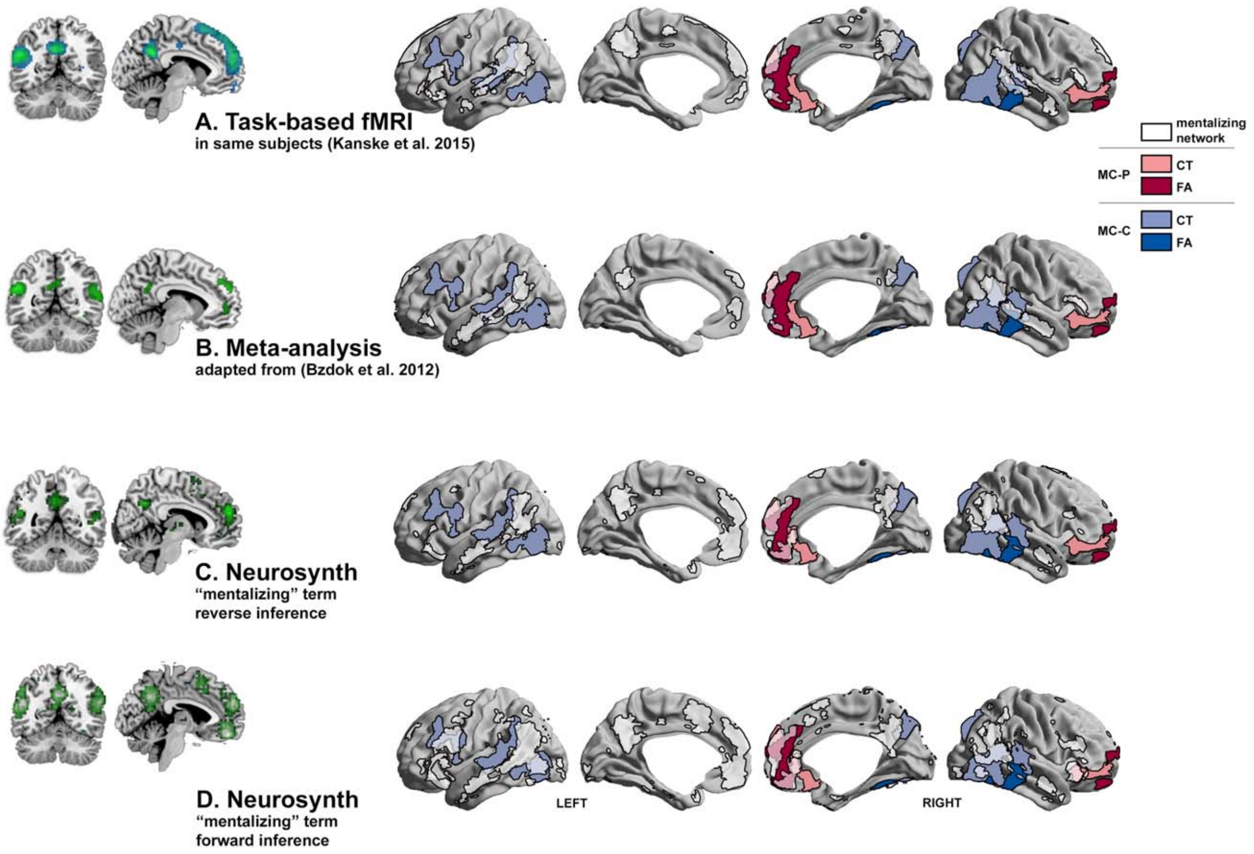

Figure 4.

Relation of MC substrates the mentalizing network, derived from three different methods. (**A**) Correspondence to fMRI activations derived from a mentalizing task in the same subjects [Kanske et al., 2015]; (**B**) Correspondence to a previously published meta-analysis on mentalizing and social cognition [Bzdok et al., 2012]; (**C**) Correspondence with Neurosynth reverse inference map of the term "mentalizing"; (**D**) Correspondence with Neurosynth forward inference map of the term "mentalizing". (Masks are depicted in semi-transparent white with black delineation).

focused, and self-referential thought [Amodio and Frith, 2006; Bernhardt et al., 2014a; Botvinick et al., 2004; Christoff et al., 2009; Smallwood and Schooler, 2006]. The default mode network has been associated with "decoupling," a class of cognitive processes possibly underlying both taking the cognitive perspective of others and by MC [Andrews-Hanna et al., 2010; Buckner et al., 2008; Buckner and Vincent, 2007; Vincent et al., 2006].

The focus on structural substrates underlying individual differences in the current work may not have targeted all brain regions functionally involved in MC-P and MC-C. Future studies based on fMRI may provide further insights, and also address dynamic and connectional properties of different network components. Nevertheless, the specific role of prefrontal regions in MC-P may relate to their involvement in both decoupled self-generated thought and the evaluation of incoming sensory information [Gilbert et al., 2005; Sommer et al., 2007], possibly mediated by pathways from and to higher visual areas [Fleming et al., 2012b; Ramnani and Owen, 2004; Young, 1992]. This might, in turn, subserve action monitoring and the planning of possible future actions [D'Argembeau et al., 2007; Gusnard et al., 2001; Mitchell et al., 2005; Northoff et al., 2006; Vogeley et al., 2004], likely requiring accurate representations of sensory stimuli and perceptual performance. Contrary, MC-C substrates fell more specifically into posterior midline and lateral temporal nodes of the mentalizing and default mode networks. Posterior nodes are heavily involved in episodic memory processing, and strongly interconnected with the medial temporal lobe system [Andrews-Hanna et al., 2010; Buckner et al., 2008; Buckner and Carroll, 2007]. Episodic memory retrieval may indeed have been relevant for MC-C, as the employed paradigm queried participants' confidence on given answers to questions relating to the content of others' autobiographical narratives. At trend level, we found that high MC-C ability related to decreases in mPFC. In previous literature not only increases but also decreases in thickness have been related to expert knowledge and enhanced behavioral performance [Hyde et al., 2007; Shaw et al., 2006]. Though beyond the scope of the current project, future studies might follow up on this interesting pattern of results. Also, it is worth pointing out that mentalizing relies on multiple inferences (e.g., infer actor's goals; infer most likely action given a goal), which individually will have a certain amount of associated uncertainty, all of which would finally be integrated when making confidence judgments. On this basis, one might have expected a more widespread substrate involved in MC for high-level cognition than for MC about perception, as shown by our findings.

To conclude, our findings robustly suggest divergent structural–anatomical substrates for MC in different domains that nevertheless colocalize with different components of the mentalizing circuitry. Our result could inform training studies that aim at enhancing MC, ToM, and social cognition in general. Specifically, a crucial question for future intervention research will be to investigate whether improvement in MC transfers to mentalizing skills, and vice versa.

ACKNOWLEDGMENTS

The authors are thankful to the Department of Social Neurosciences for their support with the *ReSource* project.

REFERENCES

Amodio DM, Frith CD (2006): Meeting of minds: The medial frontal cortex and social cognition. Nat Rev Neurosci 7:268–277.

Andrews-Hanna JR, Reidler JS, Sepulcre J, Poulin R, Buckner RL (2010): Functional-anatomic fractionation of the brain's default network. Neuron 65:550–562.

Baird B, Smallwood J, Gorgolewski KJ, Margulies DS (2013): Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception. J Neurosci 33:16657–16665.

Baird B, Cieslak M, Smallwood J, Grafton ST, Schooler JW (2013): Regional white matter variation associated with domain-specific metacognitive accuracy. J Cogn Neurosci 33:16657–16665.

Bermudez P, Lerch JP, Evans AC, Zatorre RJ (2009): Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. Cereb Cortex 19:1583–1596.

Bernhardt BC, Bernasconi N, Concha L, Bernasconi A (2010): Cortical thickness analysis in temporal lobe epilepsy: Reproducibility and relation to outcome. Neurology 74:1776–1784.

Bernhardt BC, Smallwood J, Tusche A, Ruby FJ, Engen HG, Steinbeis N, Singer T (2014a): Medial prefrontal and anterior cingulate cortical thickness predicts shared individual differences in self-generated thought and temporal discounting. NeuroImage 90:290–297.

Bernhardt BC, Valk SL, Silani G, Bird G, Frith U, Singer T (2014b): Selective disruption of sociocognitive structural brain networks in autism and alexithymia. Cereb Cortex 24:3258–3267.

Bernhardt BC, Hong S, Bernasconi A, Bernasconi N (2015): MRI pattern learning in temporal lobe epilepsy: patient classification and prognostics. Annals of Neurology 77:436–446

Botvinick MM, Cohen JD, Carter CS (2004): Conflict monitoring and anterior cingulate cortex: An update. Trends Cogn Sci 8: 539–546.

Buckner RL, Carroll DC (2007): Self-projection and the brain. Trends Cogn Sci 11:49–57.

Buckner RL, Vincent JL (2007): Unrest at rest: Default activity and spontaneous network correlations. NeuroImage 37:1091–1096; discussion 1097–1099.

Buckner RL, Andrews-Hanna JR, Schacter DL (2008): The brain's default network: Anatomy, function, and relevance to disease. Ann N Y Acad Sci 1124:1–38.

Bzdok D, Schilbach L, Vogeley K, Schneider K, Laird AR, Langner R, Eickhoff SB (2012): Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. Brain Struct Funct 217:783–796.

Cabeza R (2008): Role of parietal regions in episodic memory retrieval: The dual attentional processes hypothesis. Neuropsychologia 46:1813–1827.

Cardinale F, Chinnici G, Bramerio M, Mai R, Sartori I, Cossu M, Lo Russo G, Castana L, Colombo N, Caborni C, De Momi E, Ferrigno G (2014): Validation of FreeSurfer-estimated brain cortical thickness: Comparison with histologic measurements. Neuroinformatics 12:535–542.

Christoff K, Gordon AM, Smallwood J, Smith R, Schooler JW (2009): Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. Proc Natl Acad Sci USA 106:8719–8724.

Clarke F, Birdsall T, Tanner W (1959): Two types of ROC curves and definition of parameters. J Acoust Soc Am 31:629–630.

Concha L, Livy DJ, Beaulieu C, Wheatley BM, Gross DW (2010): In vivo diffusion tensor imaging and histopathology of the fimbria-fornix in temporal lobe epilepsy. J Neurosci 30:996–1002.

D'Argembeau A, Ruby P, Collette F, Degueldre C, Balteau E, Luxen A, Maquet P, Salmon E (2007): Distinct regions of the medial prefrontal cortex are associated with self-referential processing and perspective taking. J Cogn Neurosci 19:935–944.

Dale AM, Fischl B, Sereno MI (1999): Cortical surface-based analysis. I. Segmentation and surface reconstruction. NeuroImage 9: 179–194.

Doyle-Thomas KA, Duerden EG, Taylor MJ, Lerch JP, Soorya LV, Wang AT, Fan J, Hollander E, Anagnostou E (2013): Effects of age and symptomatology on cortical thickness in autism spectrum disorders. Res Autism Spectr Disord 7:141–150.

Faul F, Erdfelder E, Buchner A, Lang AG (2009): Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behav Res Methods 41:1149–1160.

Fischl B, Sereno MI, Dale AM (1999): Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. NeuroImage 9:195–207.

Fjell AM, Westlye LT, Greve DN, Fischl B, Benner T, van der Kouwe AJ, Salat D, Bjornerud A, Due-Tonnessen P, Walhovd KB (2008): The relationship between diffusion tensor imaging and volumetry as measures of white matter properties. NeuroImage 42:1654–1668.

Fleck MS, Daselaar SM, Dobbins IG, Cabeza R (2006): Role of prefrontal and anterior cingulate regions in decision-making processes shared by memory and nonmemory tasks. Cereb Cortex 16:1623–1630.

Fleming SM, Dolan RJ, Frith CD (2012a): Metacognition: Computation, biology and function. Philos Trans R Soc Lond B Biol Sci 367:1280–1286.

Fleming SM, Huijgen J, Dolan RJ (2012b): Prefrontal contributions to metacognition in perceptual decision making. J Neurosci 32: 6117–6125.

Fleming SM, Weil RS, Nagy Z, Dolan RJ, Rees G (2010): Relating introspective accuracy to individual differences in brain structure. Science 329:1541–1543.

Fleming SM, Ryu J, Golfinos JG, Blackmon KE (2014): Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. Brain 137:2811–2822.

Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME (2005): The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci USA 102:9673–9678.

Frith CD (2012): The role of metacognition in human social interactions. Philos Trans R Soc Lond B Biol Sci 367:2213–2223.

Galvin SJ, Podd JV, Drga V, Whitmore J (2003): Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. Psychon Bull Rev 10:843–876.

Giedd JN, Blumenthal J, Jeffries NO, Castellanos FX, Liu H, Zijdenbos A, Paus T, Evans AC, Rapoport JL (1999): Brain development during childhood and adolescence: A longitudinal MRI study. Nat Neurosci 2:861–863.

Gilbert SJ, Frith CD, Burgess PW (2005): Involvement of rostral prefrontal cortex in selection between stimulus-oriented and stimulus-independent thought. Eur J Neurosci 21:1423–1431.

Green DM, Swets JA (1974): Signal Detection Theory and Psychophysics (A Reprint, With Corrections of the Original 1966 ed.). Huntington, NY: Robert E. Krieger Publishing Co.

Greicius MD, Krasnow B, Reiss AL, Menon V (2003): Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. Proc Natl Acad Sci USA 100: 253–258.

Greve DN, Fischl B (2009): Accurate and robust brain image alignment using boundary-based registration. NeuroImage 48: 63–72.

Gusnard DA, Akbudak E, Shulman GL, Raichle ME (2001): Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function. Proc Natl Acad Sci USA 98:4259–4264.

Han X, Jovicich J, Salat D, van der Kouwe A, Quinn B, Czanner S, Busa E, Pacheco J, Albert M, Killiany R, Maguire P, Rosas D, Makris N, Dale A, Dickerson B, Fischl B (2006): Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. NeuroImage 32:180–194.

Henson RN, Rugg MD, Shallice T, Dolan RJ (2000): Confidence in recognition memory for words: Dissociating right prefrontal roles in episodic retrieval. J Cogn Neurosci 12:913–923.

Hyde KL, Lerch JP, Zatorre RJ, Griffiths TD, Evans AC, Peretz I (2007): Cortical thickness in congenital amusia: When less is better than more. J Neurosci 27:13028–13032.

Kang X, Herron TJ, Woods DL (2011): Regional variation, hemispheric asymmetries and gender differences in pericortical white matter. NeuroImage 56:2011–2023.

Kang X, Herron TJ, Turken AU, Woods DL (2012): Diffusion properties of cortical and pericortical tissue: Regional variations, reliability and methodological issues. Magn Reson Imaging 30: 1111–1122.

Kanske P, Bockler A, Trautwein FM, Singer T (2015): Dissecting the social brain: Introducing the EmpaToM to reveal distinct neural networks and brain-behavior relations for empathy and theory of mind. NeuroImage 122:6–19.

Kelemen WL, Frost PJ, Weaver CA III (2000): Individual differences in metacognition: Evidence against a general metacognitive ability. Mem Cogn 28:92–107.

Koriat A (2007): Metacognition and conciousness. In: Zelalazo PD, Moscovitch M, Thompson E, editors. Cambridge Handbook of Conciousness. New York, USA: Cambridge University Press.

Koriat A, Bjork RA (2006): Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. J Exp Psychol Learn Mem Cogn 32:1133–1145.

Kuperberg GR, Broome MR, McGuire PK, David AS, Eddy M, Ozawa F, Goff D, West WC, Williams SC, van der Kouwe AJ, Salat DH, Dale AM, Fischl B (2003): Regionally localized thinning of the cerebral cortex in schizophrenia. Arch Gen Psychiatry 60:878–888.

Lerch JP, Evans AC (2005): Cortical thickness analysis examined through power analysis and a population simulation. NeuroImage 24:163–173.

Lerch JP, Pruessner JC, Zijdenbos A, Hampel H, Teipel SJ, Evans AC (2005): Focal decline of cortical thickness in Alzheimer's disease identified by computational neuroanatomy. Cereb Cortex 15:995–1001.

Lombardo MV, Chakrabarti B, Bullmore ET, Wheelwright SJ, Sadek SA, Suckling J, Consortium MA, Baron-Cohen S (2010): Shared neural circuits for mentalizing about the self and others. J Cogn Neurosci 22:1623–1635.

Maniscalco B, Lau H (2012): A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Conscious Cogn 21:422–430.

Mar RA (2011): The neural bases of social cognition and story comprehension. Annu Rev Psychol 62:103–134.

McCaig RG, Dixon M, Keramatian K, Liu I, Christoff K (2011): Improved modulation of rostrolateral prefrontal cortex using real-time fMRI training and meta-cognitive awareness. NeuroImage 55:1298–1305.

McCurdy LY, Maniscalco B, Metcalfe J, Liu KY, de Lange FP, Lau H (2013): Anatomical coupling between distinct metacognitive systems for memory and visual perception. J Neurosci 33: 1897–1906.

Metcalfe J (1994): Metacognition: Knowing About Knowing. Cambridge, MA: Bradford Books.

Mitchell JP, Banaji MR, Macrae CN (2005): The link between social cognition and self-referential thought in the medial prefrontal cortex. J Cogn Neurosci 17:1306–1315.

Mitchell JP, Macrae CN, Banaji MR (2006): Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. Neuron 50:655–663.

Northoff G, Heinzel A, de Greck M, Bermpohl F, Dobrowolny H, Panksepp J (2006): Self-referential processing in our brain—A meta-analysis of imaging studies on the self. NeuroImage 31: 440–457.

Oppenheimer DM (2008): The secret life of fluency. Trends Cogn Sci 12:237–241.

Pannu JK, Kaszniak AW, Rapcsak SZ (2005): Metamemory for faces following frontal lobe damage. J Int Neuropsychol Soc 11:668–676.

Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL (2001): A default mode of brain function. Proc Natl Acad Sci USA 98:676–682.

Ramnani N, Owen AM (2004): Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging. Nat Rev Neurosci 5:184–194.

Rosas HD, Liu AK, Hersch S, Glessner M, Ferrante RJ, Salat DH, van der Kouwe A, Jenkins BG, Dale AM, Fischl B (2002): Regional and progressive thinning of the cortical ribbon in Huntington's disease. Neurology 58:695–701.

Salat DH, Buckner RL, Snyder AZ, Greve DN, Desikan RS, Busa E, Morris JC, Dale AM, Fischl B (2004): Thinning of the cerebral cortex in aging. Cereb Cortex 14:721–730.

Saxe R, Moran JM, Scholz J, Gabrieli J (2006): Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. Social Cogn Affect Neurosci 1: 229–234.

Schnyer DM, Verfaellie M, Alexander MP, LaFleche G, Nicholls L, Kaszniak AW (2004): A role for right medial prefontal cortex in accurate feeling-of-knowing judgements: Evidence from patients with lesions to frontal cortex. Neuropsychologia 42: 957–966.

Shaw P, Greenstein D, Lerch J, Clasen L, Lenroot R, Gogtay N, Evans A, Rapoport J, Giedd J (2006): Intellectual ability and cortical development in children and adolescents. Nature 440: 676–679.

Shaw P, Sharp W, Sudre G, Wharton A, Greenstein D, Raznahan A, Evans A, Chakravarty MM, Lerch JP, Rapoport J (2015): Subcortical and cortical morphological anomalies as an endophenotype in obsessive-compulsive disorder. Mol Psychiatry 20:224–231.

Singer T, Kok BE, Bornemann B, Zurborg S, Bolz M, Bochow CA. (in press) The ReSource Project. Background, Design, Samples and Measurements, 2nd ed.

Singer T, Seymour B, O'Doherty J, Kaube H, Dolan RJ, Frith CD (2004): Empathy for pain involves the affective but not sensory components of pain. Science 303:1157–1162.

Smallwood J, Schooler JW (2006): The restless mind. Psychol Bull 132:946–958.

Sommer M, Dohnel K, Sodian B, Meinhardt J, Thoermer C, Hajak G (2007): Neural correlates of true and false belief reasoning. NeuroImage 35:1378–1384.

Song C, Kanai R, Fleming SM, Weil RS, Schwarzkopf DS, Rees G (2011): Relating inter-individual differences in metacognitive performance on different perceptual tasks. Conscious Cogn 20: 1787–1792.

Sowell ER, Peterson BS, Thompson PM, Welcome SE, Henkenius AL, Toga AW (2003): Mapping cortical change across the human life span. Nat Neurosci 6:309–315.

Sowell ER, Peterson BS, Kan E, Woods RP, Yoshii J, Bansal R, Xu D, Zhu H, Thompson PM, Toga AW (2007): Sex differences in cortical thickness mapped in 176 healthy individuals between 7 and 87 years of age. Cereb Cortex 17:1550–1560.

Steiger JH (1980): Tests for comparing elements of a correlation matrix. Psychol Bull 87:245–251.

Teasdale JD, Moore RG, Hayhurst H, Pope M, Williams S, Segal ZV (2002): Metacognitive awareness and prevention of relapse in depression: Empirical evidence. J Consult Clin Psychol 70: 275–287.

Valk SL, Di Martino A, Milham MP, Bernhardt BC (2015): Multicenter mapping of structural network alterations in autism. Hum Brain Mapp 36:2364–2373.

Valk SL, Bernhardt BC, Bockler A, Trautwein FM, Kanske P, Singer T (2016): Socio-cognitive phenotypes differentially modulate large-scale structural covariance networks. Cereb Cortex. doi: 10.1093/cercor/bhv319.

Vincent JL, Snyder AZ, Fox MD, Shannon BJ, Andrews JR, Raichle ME, Buckner RL (2006): Coherent spontaneous activity identifies a hippocampal-parietal memory network. J Neurophysiol 96:3517–3531.

Vogeley K, May M, Ritzl A, Falkai P, Zilles K, Fink GR (2004): Neural correlates of first-person perspective as one constituent of human self-consciousness. J Cogn Neurosci 16:817–827.

Worsley KJ, Andermann M, Koulis T, MacDonald D, Evans AC (1999): Detecting changes in nonisotropic images. Human brain Mapping 8:98–101.

Worsley K, Taylor JE, Carbonell F, Chung MK, Duerden E, Bernhardt BC, Lyttelton OC, Boucher M, Evans A (2009): SurfStat: A Matlab toolbox for the statistical analysis of univariate and multivariate surface and volumetric data using linear mixed effect models and random field theory. NeuroImage S102.

Yokoyama O, Miura N, Watanabe J, Takemoto A, Uchida S, Sugiura M, Horie K, Sato S, Kawashima R, Nakamura K (2010): Right frontopolar cortex activity correlates with reliability of retrospective rating of confidence in short-term recognition memory performance. Neuroscience Research 68:199–206.

Young MP (1992): Objective analysis of the topological organization of the primate cortical visual system. Nature 358:152–155.