A LOW-DIMENSIONAL CONNECTOME MANIFOLD GOVERNS THE ORGANIZATION AND PLASTICITY OF SOCIAL BRAIN FUNCTIONS IN HUMANS

Sofie L Valk[1,2Δ]; Philipp Kanske[3,4]; Bo-yong Park[5]; Seok-Jun Hong[6-8]; Anne Böckler-Raettig[9]; Fynn-Mathis Trautwein[10]; Boris C. Bernhardt[5]*, Tania Singer[11]*

* joint co-authors

1. Otto Hahn Group Cognitive Neurogenetics, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; 2. INM-7, FZ Jülich, Jülich, Germany; 3. Clinical Psychology and Behavioral Neuroscience, Faculty of Psychology, Technische Universität Dresden, Dresden, Germany; 4. Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; 5. Multimodal Imaging and Connectome Analysis Lab, McConnell Brain Imaging Centre, Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec, Canada; 6. Center for the Developing Brain, Child Mind Institute, NY, USA; 7. Center for Neuroscience Imaging Research, Institute for Basic Science, Suwon, South Korea; 8. Department of Biomedical Engineering, Sungkyunkwan University, Suwon, South Korea; 9. Department of Psychology, Leibniz University Hannover, Germany; 10. Klinik für Psychosomatische Medizin und Psychotherapie, Universitäts Klinikum Freiburg, Freiburg, Germany; 11. Social Neuroscience Lab, Max Planck Society, Berlin, Germany

[Δ] Correspondence to valk@cbs.mpg.de

SUMMARY

Social skills such as our abilities to understand feelings and thoughts are at the core of what makes us human. Here, we combined a unique longitudinal intervention study with cutting-edge connectomics to study the organization and plasticity of brain networks associated with different social skills (socio-affective, socio-cognitive, and attention-mindfulness). Baseline analysis in our cohort (n=332) showed that social brain networks have *(i)* compact and dissociable signatures in a low-dimensional manifold governed by gradients of brain connectivity, *(ii)* specific neurobiological underpinnings, and *(iii)* reflect inter-individual variations in social behavior. Furthermore, longitudinal brain network analyses following a 9-month training intervention indicated *(iv)* domain-specific reorganization of these signatures that was furthermore predictive of behavioral change in social functions. Multiple sensitivity analyses supported robustness. Our findings, thus, provide novel evidence on macroscale network organization and plasticity underlying human social cognition and behavior, and suggest connectome-reconfigurations as a mechanism of adult skill learning.

Keywords: social cognition; ToM; empathy; compassion; social emotions; attention, mental training; plasticity; functional connectivity; connectome; gradients

INTRODUCTION

Humans have unique social skills that enhance cooperation and survival (Dunbar, 1998; Ochsner and Lieberman, 2001). These social capacities can largely be divided into at least two subsystems (Schurz et al., 2020a; Singer, 2006): *(i)* a socio-affective (or emotional-motivational) system that entails the ability to empathize with another, that is, to share feelings with others, and that may give rise to compassion and prosocial motivation (Batson, 2009; de Vignemont and Singer, 2006; Eisenberg and Fabes, 1990) and *(ii)* a socio-cognitive system that represents the ability to understand the believes and intentions of others, referred to as Theory of Mind (ToM) or mentalizing (Frith and Frith, 2006; Saxe and Kanwisher, 2003; Singer, 2006). In addition, *(iii)* attention, interoception, and mindfulness are important auxiliary functions of these social aptitudes, which underlie self-other distinctions and awareness (Craig, 2009; Kleckner et al., 2017; Tomasello, 1995). Together, these complex human capacities combine externally and internally oriented cognitive and affective processes across the domains of perception, reflective meta-cognition as well as interoception and allostatic regulation (Barrett, 2017; Chun et al., 2011).

Previous functional neuroimaging studies have identified the brain networks involved in these differentiable aspects of social emotion and cognition. Tasks probing empathy have studied how affective states of others are shared (Singer, 2006; Singer and Lamm, 2009; Singer et al., 2004), whereas tasks assessing ToM have evaluated the functional network involved in processing the cognitive inference of others' mental states (Bzdok et al., 2012; Saxe and Kanwisher, 2003; Schurz et al., 2020a). Recruitment of these networks overlaps with established resting-state networks, such as the default-mode, attention, saliency, and control networks (Alcala-Lopez et al., 2019; Kanske et al., 2016; Schurz et al., 2020a). These intrinsic functional networks support processes that are not strictly social in nature, and recent evidence suggests that 'social' brain networks overlap considerably with networks involved in memory, language, attention, interoception, and meta-cognition, as well as sensory processes (Alcala-Lopez et al., 2018). Thus, higher-level social capacities such as our capacity to understand feelings and thoughts of others are supported by multiple processes and consequently implicate distributed brain networks rather than single regions (Alcala-Lopez et al., 2018; Chang et al., 2015; Pessoa, 2014; Raz et al., 2016; Schaafsma et al., 2015).

An increasing body of work has indicated that specific social and cognitive functions are facilitated by the segregation and integration of different brain systems depending on the task at hand (Barrett and Satpute, 2013; Betzel et al., 2016; Dajani and Uddin, 2015; Schurz et al., 2020a; Shine et al., 2019a). Performance of complex, cognitively demanding, tasks, such as working memory-tasks, is supported by integration of relevant brain networks. However, the training of such tasks, and associated automation, results in segregation of relevant networks and a temporally and hierarchically reduced depth of processing (Finc et al., 2020; Shine and Poldrack, 2018). A combination of both functions is essential for successful social cognition and interaction, combining specialized and flexible processing (Schurz et al., 2020a; Schurz et al., 2020b). A recent stream of neuroscientific work suggests that integrative and segregated properties of higher order functional processes can be conceptualized by studying their associations to macroscale motifs of brain organization, also referred to as gradients or manifolds of connectivity (Huntenburg et al., 2018; Margulies et al., 2016; Murphy et al.,

2019; Paquola et al., 2019). Such large-scale functional gradients can be reliably identified from *in vivo* functional connectivity information, by applying techniques that identify principal dimensions of connectivity variations from one region to the next. As such, they provide a low-dimensional coordinate system reflecting macroscale neural organization, specifically accounting for both functional segregation across regions as well as their integration (Haak and Beckmann, 2020; Haak et al., 2018; Hong et al., 2020; Margulies et al., 2016; Shine et al., 2019a).

The macroscale manifolds that can be identified by these connectome-wide decomposition techniques have been argued to reflect evolutionary processes, and relate to both genetic expression profiles and established functional hierarchies, underscoring the close link between the physical layout of the brain, innate factors, and function (Fornito et al., 2019; Krienen et al., 2016; Margulies et al., 2016; Mesulam, 1998; Murphy et al., 2019; Paquola et al., 2019; Valk et al., 2020; Xu et al., 2020). For example, the principal gradient of intrinsic functional connectivity (derived from resting-state fMRI) places transmodal association cortices at a maximal distance from primary regions that are specialized for perceiving and acting in the here and now (Huntenburg et al., 2018; Margulies et al., 2016; Murphy et al., 2019). This increased spatial distance from primary cortex may allow association cortices to take on functions that are only loosely constrained by the immediate environment, facilitating abstract representations and transformations in cognitive, emotional, and social processing (Murphy et al., 2019). Indeed, ToM relies on inner reflective processes mostly based on self-generated internal thoughts only loosely constrained by direct input from the environment, and is associated with transmodal networks such as the default mode system, showing a high distance from primary cortices (Margulies et al., 2016). Conversely empathy may be more closely coupled to primary sensory and attention networks, as it processes potentially alerting, emotionally salient, stimuli from the environment. Variations in functional topology, reflecting the relative integration and segregation of different brain networks, have been previously associated with social and cognitive skills in health and disease (Hong et al., 2019; Marquand et al., 2017; Murphy et al., 2019; Schmalzle et al., 2017; Shine et al., 2019a). However, though previous longitudinal mental training research has indicated observable changes in behaviour and MRI-based measures of local brain structure following the training of social skills (Klimecki et al., 2014; Tang et al., 2015; Trautwein et al., 2020; Valk et al., 2017), it remains unknown whether targeted training of socio-affective competences (such as empathy, compassion) and socio-cognitive skills (such as ToM and cognitive perspective taking) can have domain specific effects on the functional connectome.

Taken together, current evidence suggests that *(i)* the neural basis of social capacities can be divided into multiple sub-systems which relate to various functional networks located in heteromodal association cortices and paralimbic areas; *(ii)* specific social and cognitive functions are likely facilitated by the segregation and integration of different brain systems depending on the task at hand; *(iii)* large-scale gradients reflect overarching organizational motifs of brain function, particularly functional integration across cortical regions as well as functional segregation within specific regions. However, the exact relationship between specific social capacities and large-scale functional organization remains unclear. Our study sought to directly examine the intrinsic relationship between social skills and functional topology and to study whether training attention, socio-emotional, and socio-cognitive skills

would elicit dissociable alterations in large-scale functional organization using a combination of cross-sectional and longitudinal approaches.

Our work was realized in the context of the *ReSource* study (Singer et al., 2016), a large-scale 9-month study involves the targeted training of attention-mindfulness (*Presence* module), socio-affective skills (*Affect* module), and socio-cognitive/ToM abilities (*Perspective* module) over 3-months-long training modules. We tracked training related change with a wide variety of markers including behavioral measures of social emotions and cognition and multimodal imaging. Whereas *Presence* aimed at initially stabilizing the mind and nurturing introspective abilities, the *Affect* and *Perspective* modules focussed on nurturing social skills such as empathy, compassion and perspective taking on self and others. Our analysis approach combined cross-sectional and longitudinal assessments with extensive decoding to comprehensively study the relationships between social capacities and large-scale functional organization. We utilized connectome-based manifold learning to identify low-dimensional connectivity gradients, which synoptically capture the balance of integration and segregation at the local and network level.

The baseline data (before any training) was used to assess the functional organization of the human social brain. Networks associated with attentional, socio-affective and socio-cognitive functions at baseline were identified via previously validated functional imaging tasks (Kanske et al., 2015; Trautwein et al., 2016) and projected into this new connectivity-determined space derived from parallel resting-state fMRI acquisition. We then calculated individualized metrics of global and local network integration and segregation. We hypothesized that task-based networks probing attention, empathy, and perspective-taking would have unique manifold representations in low-dimensional connectivity space and that individual variation of these signatures reflect domain-specific social capacities at baseline. To further validate the dissociability of neurobiological processes underlying attention, socio-affective, and socio-cognitive processing in the current sample, we performed an extensive decoding of task-based functional networks involved in attention, socio-affective, and socio-cognitive processing with respect to meta-analytical task-based activation libraries and gene expression profiles.

Capitalizing on the unique longitudinal design of the ReSource study, we then examined functional network reorganization following the training of the three 3-month training modules of *(i)* attention-based mindfulness (*Presence*), *(ii)* socio-affective abilities (*Affect*), and *(iii)* socio-cognitive competences (*Perspective*). We hypothesized that the training of these skills would result in dissociable alterations of functional topology reflecting network integration and segregation. Accordingly, the *Presence* module cultivates states in which attention is focused in a non-elaborative way on aspects of sensory and interoceptive inputs. An example is attending to the breath, bodily signals, and sounds. Such mindfulness-related skills are sometimes described as "bare attention"(Slagter et al., 2007). On the other hand, the *Perspective* module supports states of more integrated awareness and cognition including rather abstract processing of information that are decoupled from environmental inputs. Examples include reflecting on one's own and others' mental states. Finally, the *Affect module* aimed to cultivate compassion, emotional regulation, and acceptance of difficult emotions. An additional aim was to nurture feelings of care and prosocial motivation. Such socio-emotional

skills likely require a balance of automated and active processes (Barrett, 2017; Chun et al., 2011). Indeed, affective processing require on the one hand fast and automatic analyses of external alerting and emotionally salient stimuli (Barrett, 2017). On the other hand, they also require complex interoceptive regulatory processes involving the integration of bodily signals and multiple context information to form conscious emotions and plan for adaptive responses (Barrett, 2017). Mixed effects models operating on the manifold signatures of the task-relevant functional networks tracked longitudinal trajectories of their reorganization in functional integration and segregation subsequent to training. We further explored associations between the plasticity of their manifold representation and behavioral change in attention, compassion, and ToM using machine learning with cross-validation. Multiple sensitivity analyses verified the robustness of our findings.

RESULTS

Connectome manifolds provide compact signatures to dissociate human social brain networks

To examine the intrinsic relationship between social skills and functional topology, we analyzed behavioral as well as task-based and task-free fMRI data from a large-scale mental training intervention, the *ReSource Project* (Singer et al., 2016). For further details, please see http://resource-project.org and the preregistered randomized controlled trial https://clinicaltrials.gov/ct2/show/NCT01833104 at http://ClinicalTrials.gov.

The first series of analyses studied baseline (T_0) data prior to any training to profile human social brain networks (**Figure 1A, Supplementary Materials**). Specifically, we derived task-based fMRI activation patterns associated with (i) attention based on a Cued-flanker task (Trautwein et al., 2016), (ii) socio-affective functions, such as empathy and compassion, based on the affect contrast in the EmpaTom task (Kanske et al., 2015) and (iii) socio-cognitive functioning, based on the ToM contrast of the EmpaTom task (Kanske et al., 2015). To identify the manifold signatures of these networks, we applied non-linear dimensionality reduction to resting-state functional MRI connectomes based on our open access software (Vos De Wael et al., 2020). These techniques identify a series of eigenmodes that describe spatial variations in connectivity, so called connectivity gradients (Coifman et al., 2005; Margulies et al., 2016; Vos De Wael et al., 2020). Gradients of each individual were aligned to a template based on the Human Connectome Project dataset (Van Essen et al., 2013; Vos De Wael et al., 2020) (**Figure 1B**). In line with previous work (Margulies et al., 2016), the first gradient (G1) differentiated sensory-motor networks from transmodal systems, such as the fronto-parietal and default mode network. The second gradient (G2) reflected a segregation of sensory-motor and visual networks. The third gradient (G3) differentiated the multiple demand network (Duncan, 2010) from transmodal and sensory-motor systems. Projecting the fMRI activations from the three social brain profiling tasks into the gradient-based manifold revealed a unique spatial signature for each task-based network, particularly with respect to G1 and G3 (**Figure 1B-C**). Task-derived attention networks were located in the middle of G1, between unimodal and default mode regions, and at the upper multiple demand end of G3. Conversely, task-derived socio-cognitive activations were placed at the default mode end of G1 and G3, maximally distant from the sensory and multiple-demand networks. In contrast, task-derived socio-affective networks resided in the upper default mode

half of G1 and the middle of G3, between default and multiple-demand networks, in line with the role of socio-affective processes in balancing externally and internally oriented functions (Barrett, 2017; Chun et al., 2011).

For each vertex, we calculated the Euclidian distance from the centre of the low-dimensional gradient manifold for each individual – creating a single *eccentricity* score, reflecting local network integration/segregation – as done recently (Bethlehem et al., 2020; Park et al., 2020). This score quantifies vertex-specific offsets in the low-dimensional connectivity gradient manifold (Coifman et al., 2005). It enables the study of localized changes within the cortical connectome capturing multivariate associations of the first three gradients combined. Here, higher scores indicate stronger segregation from an integrated functional core. Such segregated functions are for example those mediated by visual and sensory-motor networks(Bethlehem et al., 2020; Margulies et al., 2016), whereas saliency processing is a more integrative function (Menon and Uddin, 2010). We projected the activation maps from the fMRI tasks at baseline (T_0, **Figure1C**) into the intrinsic gradient manifold, and calculated average vertex-wise eccentricity scores for each activation map. The conjunction network of socio-affective and socio-cognitive activation had highest eccentricity, whereas the overlap of the attention and socio-affective network had lowest eccentricity. In addition, we computed network-level index of segregation and integration; the *dispersion* within as well as *distance* between task-based fMRI activations in 3D gradient space (Bethlehem et al., 2020) (**Figure 1D**). Attention networks had higher dispersion relative to social task-based networks, and there was a larger distance between attention and social networks rather than within social networks, with highest distance between attention and socio-cognitive networks. The embedding of task-based activation networks in the intrinsic gradient manifold supports that processes shared by attention and socio-affective functioning are integrative functions (*e.g.*, integrating perceptual and abstract as well as interoceptive and exteroceptive information). This aligns well with the putative functional role of the dorsal anterior insula and anterior-mid cingulate (Uddin, 2015). Conversely, shared processes underlying both social domains might be segregated, decoupled, functions supported by the default mode (*e.g.*, a perceptually decoupled mode of processing, most distant from regions that directly govern action and perception).

The manifold representation provided dissociable signatures of the three task-based activation networks implicated in social function. In the next step, we used three complementary approaches to further support dissociability. First, capitalizing on intrinsic functional communities (Yeo et al., 2011) as an alternative and well-established approach to characterize macroscale functional organization (**Figure S1, Supplementary Results**), we confirmed that the three task-activation networks indeed reflected differentiable combinations of large-scale communities, underlining different mixtures of the fronto-parietal and default mode network. Secondly, we carried out meta-analytic decoding against 20 topic terms associated with social, emotional, and cognitive processing. This analysis confirmed that fMRI activation of the three capacities related to differentiable functional taxonomies ascribed by the prior task-based fMRI literature (**Figure S1, Supplementary Results**). Thirdly, as functional gradients are thought to reflect spatiotemporal trends in gene expression, we also performed transcriptomic association analysis with *post-mortem* gene expression data from the Allen Institute for Brain Sciences (AIBS) (Arnatkeviciute et al., 2019; Gorgolewski et al., 2014;

Hawrylycz et al., 2012). Significant gene lists after multiple comparisons correction were fed into a developmental gene expression analysis, which highlights spatio-temporal trajectories for these genes are strongly expressed. This analysis implicated differentiable regional enrichment patterns of genes spatially co-localized with attention, socio-affective, and socio-cognitive functional activations. Notably, whereas functional activations associated with socio-emotional processing were associated with early fetal gene expression patterns, task-based activation related to attention and ToM was linked to gene expression in early adulthood (**Figure S1, Supplementary Results**).

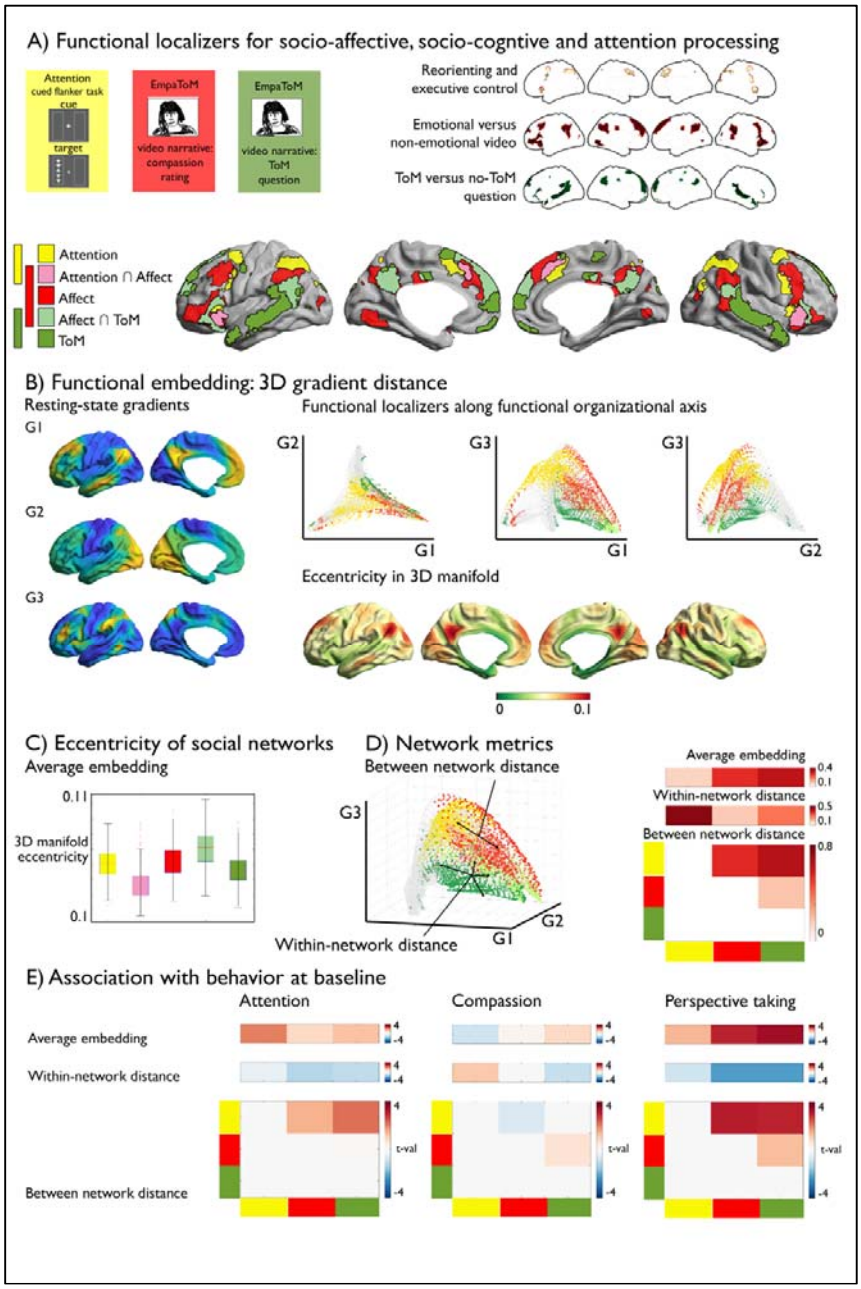

Figure 1. Manifold embedding of functional activation during attention, socio-affective, and socio-cognitive tasks. A). Functional localizers (Kanske et al., 2015; Trautwein et al., 2016) and their mappings projected on the cortex: yellow: attention; pink: overlap between attention and socio-affective activation; red: socio-affective activation; light green: overlap between socio-affective and socio-cognitive activation; dark green: socio-cognitive activation; **B**). *left*: The first three gradients of based on resting-state functional

connectivity; *right*: the embedding of task-based activations; *bottom*: vertex-level eccentricity by combining the first three gradients; **C**) Manifold embedding (vertex-wise eccentricity in 3D manifold) of unique and overlapping task-based activations; **D**). *left*: Embedding of functional networks in the three gradients combined, and visualization of network-based dispersion metrics (within network distance and between network distance) and *right:* average vertex-wise manifold embedding and network-based dispersion of task-based functional activations (yellow=attention; red=socio-affect; green=socio-cognition); **E**). Correlation between vertex-level and network-level embeddings and behavioural markers.

To then assess behavioral relevance of the manifold signature representation, we evaluated network-based manifold associations to behavioral markers of attention, compassion, and ToM (**Figure 1E**) (FDR-corrected). ToM ability was positively associated with manifold eccentricity of socio-cognitive ($t=3.15$, $p<0.001$) and socio-affective networks ($t=2.74$, $p<0.01$), but not attention networks ($p>0.1$). Both socio-cognitive and socio-affective networks showed decreased within-network dispersion (socio-cognitive: $t=-2.34$, $p<0.05$; socio-affective: $t=-2.34$, $p<0.05$) and increased distance to attention networks in individuals with better ToM (socio-affective: $t=2.47$, $p<0.01$; socio-cognitive: $t=2.66$, $p<0.02$). Conversely, attention performance correlated with eccentricity of attention networks ($t=2.16$, $p<0.05$), but not social networks ($p>0.1$). We observed no significant correlation for compassion. Our baseline evaluations, thus, point to a dissociation of social brain networks, at the level of topological signatures, neurobiological correlates, and their relation to behavior. Biggest differences were seen between attentional and perspective-taking functions. Socio-affective processes seemed to be placed in between these extremes, which may point towards both internally and externally oriented processes playing a role for empathic responding, compassion and emotion regulation.

Macroscale network plasticity following the targeted training of social skills

Having established the above network motifs of attentional, socio-affective, and socio-cognitive functions, we examined change patterns of these different networks in the gradient manifold following the targeted training of these core domains. In the *Resource* project, participants were randomly assigned to two training cohorts (TC1, N=80; TC2, N=81) who underwent three distinct and subsequent 3-months training modules with weekly-instructed group sessions and daily exercises, completed via cell-phone and internet platforms (**Figure 2, Table 1-3,** *Methods for more details*). The first module, *Presence*, promotes attention and interoceptive awareness, resembling mindfulness-based interventions (Kabat-Zinn, 1990). This module was followed by *Affect* focusing on socio-affective capacities, such as compassion, gratitude, dealing with difficult emotions, and prosocial motivation as well as by *Perspective* focusing on socio-cognitive skills, such as perspective-taking on self and others as well as metacognition. TC1 and TC2 underwent the latter two modules in different order (TC1: *Affect→Perspective*; TC2 *Perspective→Affect*) to serve as active control groups for each other (**Figure 2A**). Additionally, a matched test-retest control cohort did not undergo any training (RCC, n=90), but was assessed with the same measures as TC1 and TC2, as well as an active control group (TC3; N=81 completed three months of *Affect* training only). All participants were measured at the end of each three-month module (T_1, T_2, T_3) using 3T MRI and behavioral measures that were identical to the baseline (T_0) measures**.**

Longitudinal changes in manifold eccentricity were assessed using mixed-effects analysis (Dale et al., 1999) (**Figure 2B**). We observed marked eccentricity changes following the

Presence and *Perspective* modules. *Presence* training resulted in increased eccentricity of bilateral temporal and right superior parietal areas (family-wise error corrected p-value, $p_{FWE}<0.002$). This indicates that after training *Presence* these regions became more segregated. Conversely, *Perspective* training resulted in decreased eccentricity of temporal, temporo-parietal and parietal regions, together with left insular cortices ($p_{FWE}<0.001$). We observed no significant whole brain alterations following *Affect*. Effects were similar when assessing TC1 and TC2 separately (**Figure S2**) and when controlling for global signal effects (**Figure S2**), supporting robustness. This indicates that the integration and segregation of functional organization shows unique alterations following different mental training regimes, with regions shifting their functional embedding, depending on the functions trained.

Following, we evaluated whether we could observe training-related alterations in integration/segregation of the task derived functional networks. We observed a largely inverse relationship between *Presence* and *Perspective* in eccentricity change of task-based networks of attention, social affect, and social cognition (**Figure 2D**; **Table S1**). Whereas the attention and social networks became more segregated following *Presence*, they became more integrated following *Perspective*. Post-hoc analysis confirmed that eccentricity changes in task-based functional networks were comparable, but weaker, when comparing TC1 and TC2 against the RCC (**Figure S2**). When studying the changes within cortex-wide functional communities (Yeo et al., 2011), we found similar patterns indicating differentiable patterns of integration and segregation in functional networks of the association cortex after *Presence* and *Perspective*. Notably, *Affect* resulted in increased segregation of visual cortices from attention networks, and integration of ventral attention and fronto-parietal networks (**Figure S3**). Follow-up analysis confirmed differential alterations following each training in the first three functional gradients individually (**Figure S4, Supplementary Results**). Evaluations of sub-cortical and cerebellar embeddings suggested increased segregation of bilateral thalamus and left hippocampus following *Affect* (**Figure S5, Supplementary Results**). The contents of thought during resting state assessment did not systematically vary as a function of mental training, as we observed no significant association with change in task-unrelated thoughts when evaluating change in post-scan thought probes after each mental training block (**Supplementary Results**).

Collectively, our findings implicate an increase in connectome manifold eccentricity of transmodal association networks following *Presence* training, reflecting increased segregation of task-based networks, whereas *Perspective* training shows an integration of these networks. The results in our longitudinal assessments enhance baseline observations that juxtapose externally-oriented attention from internally-oriented cognitive functioning, such as perspective taking. *Affect* training altered the embedding of the thalamus and hippocampus, possibly underscoring their role in coordinating affective functions (Lindquist et al., 2012). Whereas our cross-sectional analyses highlighted different manifold signatures of task-based activation networks, longitudinal analysis thus indicated differentiable network plasticity specifically between networks implicated in focused attention (*Presence*) versus integrative social cognition (*Perspective*).

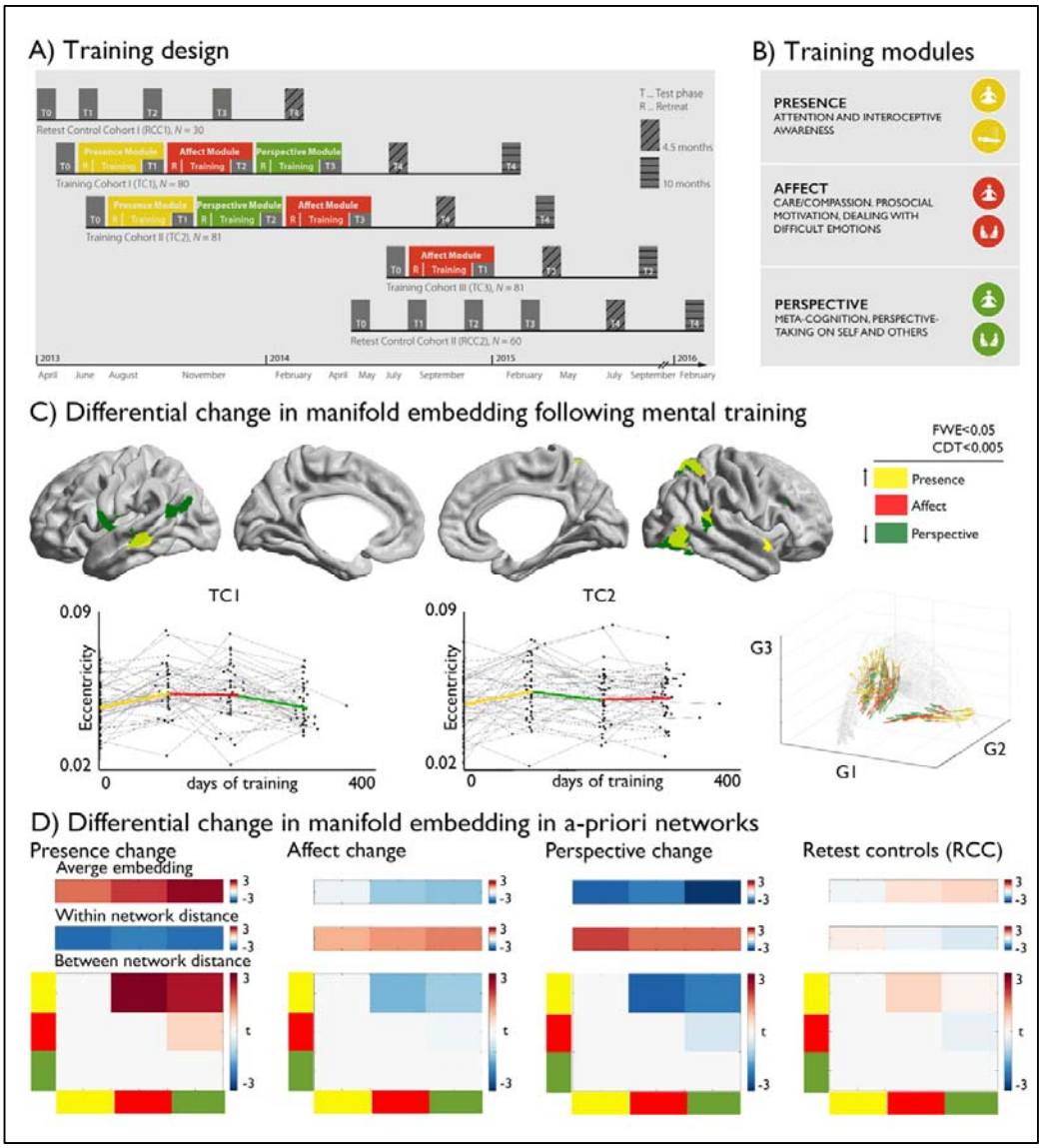

Figure 2. Embedding changes of the social brain following mental training. A). Training design of the *ReSource* study; **B**). Training modules; **C**). *upper panel:* Differential change in 3D gradient distance following mental training, *lower panel left:* Average change in overlapping clusters associated with manifold change in Presence and *Perspective* (light green) following *Presence* (yellow), *Affect* (red), and *Perspective* (green) in TC1 and TC2; *lower panel right*: visualization of local manifold changes as a function of 3d gradient space. Arrows indicate direction of change; **D**). Change in average eccentricity, as well as change in network-level dispersion in a-priori defined task-based networks. Whole brain findings are corrected for multiple comparisons at FWEp<0.005.

Associations between functional network reconfiguration and behavioral change

Last, we evaluated whether the above alterations in manifold eccentricity following mental training could predict behavioral changes in the three domains *i.e.*, attention, compassion, and ToM. Specifically, we leveraged supervised machine learning with 5-fold cross validation (Chang et al., 2015) to predict behavioral change across time-points from the change scores of the vertex-wise connectome manifold eccentricity metric. Manifold eccentricity predicted behavioral change in attention ($r = 0.39$; mean absolute error (MAE) = 0.041; $p < 0.001$), compassion ($r = 0.26$; MAE = 0.583; $p < 0.001$), and ToM ($r = 0.40$, MAE=0.117, $p<0.001$). Notably, behavioral change prediction was intervention-specific, and change could not be

predicted from weight-maps training on the other behavioral change scores. Post-hoc assessment in a-priori task-based networks indicated that positive manifold features, reflecting increased segregation, were selected in attention and socio-affective networks, rather than socio-cognitive networks, when predicting changes in attentional scores from the Cued-Flanker task. Change in compassion ratings using the EmpaToM task was predicted by negative weights in the attention-network, reflecting increased integration, whereas changes in ToM scores from the EmpaToM task were associated with positive weights in the attention network and negative weights in the socio-cognitive network. These results indicate that distinct, multivariate reconfigurations of local functional topology relate to training-induced behavioral changes in the three domains of attention, compassion and ToM. All predictor maps show both positive and negative weights across the cortex, suggesting that behavioral change in attention, compassion, and ToM relates to a combination of segregation and integration of local manifold embeddings.

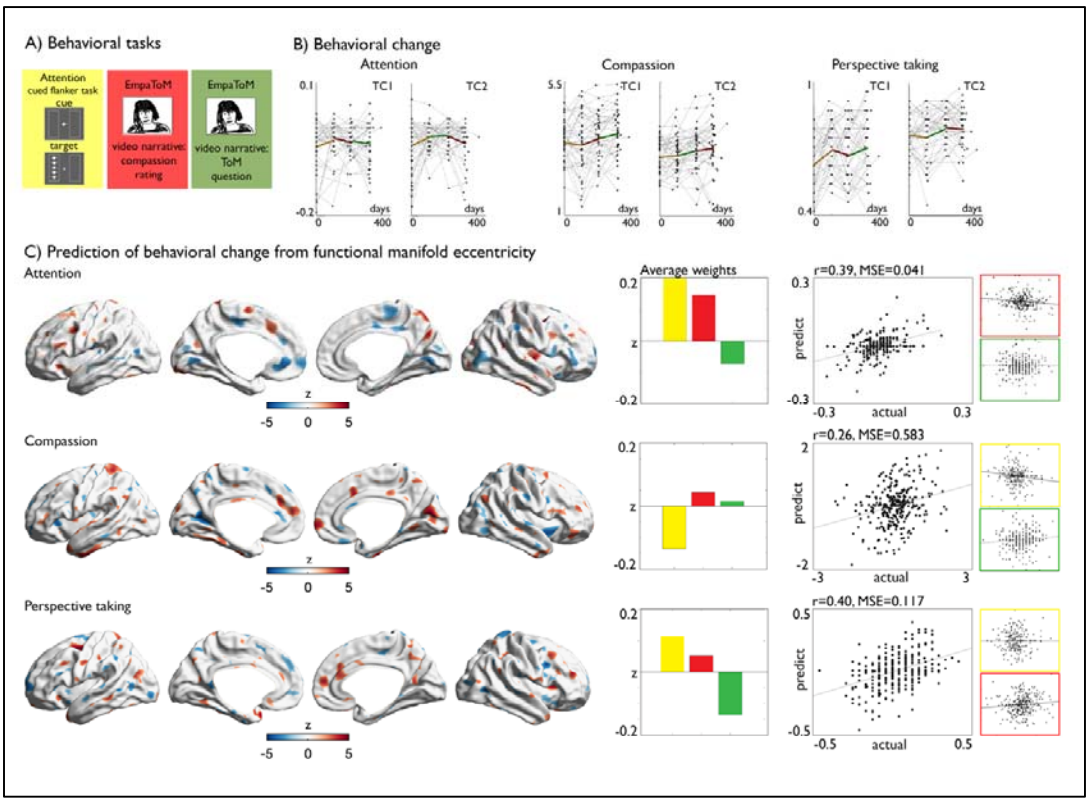

Figure 3. Association between changes in functional embedding and changes in behaviour. A). Behavioral tasks that measured attention, compassion, and perspective-taking; **B**). Behavioral measurements in TC1 and TC2 across time-points (Trautwein et al., 2020); **C**). Prediction of behavioral change based on functional manifold change across time-points. *Left:* bootstrapped loadings of weights ($p<0.05$); *Middle:* average bootstrapped z-score within task-based network; *Right*: Prediction of behavioral change within domain, as well as using the predictor of the other domains (yellow=attention, red=compassion, green=perspective-taking).

DISCUSSION

This study explored the functional organization and plasticity of brain networks involved in attention, social emotions such as empathy and compassion, and social cognition such as our ability for cognitive perspective taking (ToM). We studied a large cohort of healthy adults between 20 and 55 years of age that took part in a 9-month longitudinal mental training study, the *Resource* project (Singer et al., 2016). Our main goals were *(i)* to evaluate the intrinsic relationship between social processes and macroscale network motifs, and *(ii)* to examine whether the training of such skills with three targeted mental training modules would result in differentiable and behaviorally relevant alterations in functional network topology. Our multi-level approach integrated task-based functional activation mapping across core domains of social functioning with novel analysis concepts to profile macroscale organization along continuous connectivity-governed manifolds. This new approach placed the distinct, yet interacting components of the social brain into a novel intrinsic functional coordinate system, which synoptically visualized their functional organization and which was found to relate to inter-individual differences in behavioral measures of attentional and social skills. In the scope of the *Resource* project (Singer et al., 2016), we established differential longitudinal training effects on functional network architecture and behavioral markers in a training specific manner resulting from serially administered and targeted 3-months intervention modules: *Presence* training aiming at enhancing attention; *Affect* training aiming at enhancing socio-affective competences; and *Perspective* training aiming at enhancing meta-cognitive capacities including perspective taking on self and others (ToM). Multiple sensitivity analyses supported robustness of observed effects. Collectively, our findings offer a novel conceptualization of the distributed functional architecture underlying attention and social cognition in humans, and shine a new light onto network topology that adapts to targeted training of distinct social skills in adults.

Our study builds on a growing body of work suggesting an interplay between large-scale human brain organization and complex behavior (Bethlehem et al., 2020; Hong et al., 2019; Marquand et al., 2017; Murphy et al., 2019; Park et al., 2020; Schmalzle et al., 2017; Shine et al., 2019a). Converging evidence shows that the cortex displays consistent transitions in molecular, structural and functional properties, an architecture that balances integration and segregation according to internal and environmental demands (Bertolero et al., 2018; Finc et al., 2020; Murphy et al., 2020; Shine and Poldrack, 2018). At baseline i.e., before any training, we assessed whether different task-derived fMRI networks involved in social and attentional processes would be associated to differentiable signatures in the lower dimensional manifold governed by connectome gradients. The dimensionality reduction techniques applied here are particularly suitable, as they provide an intrinsic and continuous coordinate system that situates individual regions within the global landscape of the cortical connectome. Gradients in our sample of middle old healthy individuals who underwent extensive screening before study enrollment (Singer et al., 2016), recapitulated prior work in healthy young adults (Hong et al., 2020; Margulies et al., 2016). Indeed, in accordance to prior studies, we observed that G1 traversed a sensory-fugal axis, running from unimodal to transmodal systems such as the default mode and fronto-parietal networks. Parallel explorations have demonstrated that this gradient aligns with the putative cortical hierarchy (Margulies 2016), measures of microstructural differentiation (Huntenburg et al., 2017; Paquola et al., 2020b;

Paquola et al., 2019), and main axes in cortical evolution (Valk et al., 2020; Xu et al., 2020). In our findings, the overlap between socio-affective and socio-cognitive functional activations was found to be located at the apex of this G1, linking social behavior to integrative systems developed late in evolution. Networks implicated in social cognition such as ToM are often co-localized with, if not equal to, heteromodal association systems such as the default mode network (Alcala-Lopez et al., 2018; Alcala-Lopez et al., 2019; Alves et al., 2019; Kernbach et al., 2018). The default mode network is placed at the apex of the functional hierarchy (Margulies et al., 2016). Its broad functional role might be to integrate multiple streams of information (Margulies et al., 2016; Paquola et al., 2020a) and to orchestrate cognitive processes decoupled from external information (Frith and Frith, 2003; Murphy et al., 2018; Murphy et al., 2019; Schurz et al., 2020b) and of its potential ability to engage in high level predictive processing (Chanes and Barrett, 2016). While G2 commonly dissociates visual and somatosensory systems, G3 juxtaposes the multiple demand network known to engage in externally oriented tasks on the one hand, and sensory-motor and transmodal systems on the other hand. Task-based networks involved in attention, socio-affective, and socio-cognitive skills were linearly ordered along G3, with attention on one end of this gradient and socio-cognitive skills on the other. As such, our work suggests a topological dissociation of externally oriented attentional functions and internally oriented socio-cognitive skills. Socio-affective skills were found to be largely at the middle of these organizational gradients, providing topological evidence of the role of affect in allocentric processing and coordination of external and internal cognition (Barrett, 2017; Chun et al., 2011). Indeed, emotional-motivational processes are neither entirely externally oriented nor do they encompass only de-coupled abstract reasoning. By contrast, they need to integrate external salient cues with internal processes of autonomic body regulation and interoception, which could indeed explain their intermediary position in the gradient based coordinate system. Beyond the specific location of functional systems in manifold space, the use of gradients as a suitable coordinate system has been supported by parallel explorations in non-human animals, structural MRI studies, and computational modelling experiments. Indeed, experimental work in non-human animals and *in silico* studies have implicated similar large-scale gradients in region-to-region variations in micro-circuitry (Burt et al., 2018; Chaudhuri et al., 2015; Fulcher, 2019; Park et al., 2020; Wang et al., 2019) alongside the regions' role in processing hierarchies (Mesulam, 1998); in recent theoretical work, this architecture has been suggested to contribute to predictive coding in affect and cognition (Chanes and Barrett, 2016).

In addition to situating different components of the social brain in the intrinsic coordinate system spanned by G1-G3, a major goal of the study was to evaluate the functional brain reorganization following the training of socio-affective, socio-cognitive, and attention-mindfulness skills. Again, our approach enabled us to represent functional network plasticity in the compact manifold space governed by axes of brain connectivity (Hong et al., 2019; Margulies et al., 2016). We found increased segregation of temporal and parietal structures following a three-month attention-based *Presence* training, whereas the *Perspective* module resulted in an increased integration of these regions. Assessments of the network-based dispersion of task-based networks associated with attention, socio-affective and socio-cognitive functioning indicated a segregation of attention and default networks following *Presence* training, while *Perspective* training resulted in increased integration of these

networks, suggesting integration of attentional and cognitive processes. Both integrative and segregating functional processes are competing constraints on higher-level functioning, reflecting a balance between parallel processing and efficient, integrated, architecture versus the ability to exploit learned associations through automation (Shine and Poldrack, 2018). A decrease of network integration has also been reported in motor and sustained vigilance tasks (Wang et al., 2016) and it might contribute to automatic processes by potentially releasing association hubs for flexible processing (Shine and Poldrack, 2018). Thus, integrated states might reflect active thought processes such as those required when engaging in mind-reading of others (ToM), while more segregated states may reflect to a less integrated mode of information processing, less tethered with conceptual, motivational and emotional systems, as trained in the Presence module (Tang et al., 2015). This view is in line with accounts that mindfulness-based practices reduce habitual trajectories of (pre)motor and autonomic reactions, as well as the pull of distracting spontaneous thought (Lutz et al., 2019). In line with this hypothesis, the mindfulness-attention intervention resulted in functional segregation, whereas the training of perspective-taking resulted in an overall increase of network integration. *Affect* was associated with increased functional embedding of the hippocampus and thalamus, whereas *Presence* or *Perspective* did not result in significant alterations of these regions. Hippocampus and thalamus contribute to various aspects of emotional processing (Kober et al., 2008; Lindquist et al., 2012; Wager et al., 2008). Specifically, the hippocampus is associated with the encoding of salient stimuli (Lindquist et al., 2012) and its anatomy is at the nexus of various processing streams integrating the multiple demand and default mode cortical networks (Paquola et al., 2020a; Vos de Wael et al., 2018). Conversely, the thalamus is an integrative hub region that is known to contribute to overall cortico-cortical network organization (Hwang et al., 2017; Shine et al., 2019b), cortical excitability (Wang et al., 2019; Weng et al., 2020), as well as homeostatic processes and the endocrine system (Kober et al., 2008; Muller et al., 2020; Shine et al., 2019b). Thus, it is possible that functional reorganisation of integrative subcortical nodes and the hippocampus contributes to an altered coordination of cortical network integration and segregation. Such an interpretation aligns with theories of emotional allostasis, suggesting that affective processing may balance integration and segregation of brain function to regulate resources dynamically (Barrett, 2017; Chun et al., 2011).

Through availability of behavioral markers of attentional, affective, and socio-cognitive functions, we could relate differential patterns of local functional integration and segregation to behavioral outcomes. Specifically, increased segregation of the attention network related to behavioral improvements in attention, whereas its integration predicted behavioral improvement in compassion. Increased integration of socio-cognitive networks, and to all lesser extend segregation of task-based attention networks, predicted increased ToM ratings. Affective processes trained in the *Affect* module include compassion, prosocial motivation but also emotional regulation. Such processes may be supported by an increased integration of perceptual and abstract as well as interoceptive and exteroceptive information (Barrett, 2017). Conversely, attention-mindfulness training aimed at increased decoupling and untethering from such processes (Lutz et al., 2019). On the other hand, socio-cognitive training, and improvement ToM, may rather relate to complex, cognitively demanding, processing, such as perspective-taking, and therefor relates to the integration of functionally relevant, ToM, brain

networks. These observations are broadly in line with our findings at baseline in the functional manifold and underscore the dissociation between attentional, socio-emotional, and socio-cognitive processes, as well as the dissociable functional profiles underlying network segregation and integration (Finc et al., 2020). Indeed, in a previous *ReSource* study focusing on the analyses of autonomic measurements during engagement in the different types of mental practices of each training module, the *Presence* module was reported to be more relaxing and requiring less efforts, whereas the two social modules, *Affect* and *Perspective*, were associated with increased effort and arousal (Lumma et al., 2015). Our observations that such patterns are specific depending on the functional domain corroborate this notion, and support the notion of a functional gradient contrasting externally-oriented from internally-oriented functions (Chun et al., 2011).

Decoding the task-based activations during attention, socio-affective, and socio-cognitive processing with meta-analytical and *post-mortem* gene expression data provided functional and neurobiological context to our findings. Task-based activations associated with attention, socio-affective, and socio-cognitive domains implicated networks involved in both social and non-social processes (Alcala-Lopez et al., 2019). These results likely underscore that social processes are embedded in a broader functional landscape, integrating sensory and bodily processes with both internally- and externally-oriented cognitive functions. As expected, meta-analytical functional decoding could show that attention-related brain activation related to attention terms, while the social task-based activations related to terms describing elements of social cognition. Whereas social emotion could be linked to affective but also social and control processes, we found that socio-cognitive processing related to social and mnemonic functions. These results are in line with the functions probed by the task, and again suggest socio-affective processes might integrate external and internal cognitive processing. Second, we observed different compositions of developmental genetic expression patterns in cortical but also subcortical and cerebellar regions for each of the three task-based activation maps. The developmental time-windows of transcription patterns associated with socio-cognitive and attentional activations were primarily linked to genes expressed in late childhood and adolescence, functional activations associated with affective processing were linked with genes expressed in early fetal time-windows. This pattern potentially points to a differentiable neurodevelopmental origin of each of these functions and is in line with theories that suggest that socio-affective processes such as empathy and compassion have different developmental time-courses than higher-level cognitive processes such as our ability for ToM developing much later in child development (Singer, 2006). It will be important for future work to explore the development and location of genetic expression associated with socio-emotional and socio-cognitive functioning to better understand the phylogenetic and ontogenetic determinants of our social brain.

As our work demonstrates, the application of advanced resting-state fMRI analytics can effectively visualize modular and integrative functional signatures across the cortical landscape, and specifically capture the contribution of transmodal association networks in higher order social cognition (Bertolero et al., 2018; Yeo et al., 2011). Different functional processes likely reflect prominent locations in intrinsic functional organization, in which external attention may differ from internally oriented cognition. Integrating the current observations with previous work on the relationship between grey matter structure and social

and cognitive skill in the same sample (Valk et al., 2017), we could again show separability of component processes underlying higher-order social functioning. Whereas the current work focusses on functional organization, earlier work from our group assessed regional changes in cortical thickness. While both brain function and structure relate to behavior (Grasby et al., 2020; Shine et al., 2019a), it is likely that the measured brain changes follow different time-windows and underlying biological mechanisms. Further work integrating both structural and functional measures will need to explore the interrelation between anatomy, connectivity, and brain function in both cross-sectional and longitudinal contexts.

To close, our work demonstrates a tight inter-relation between social functions and functional network organization and plasticity, capitalizing on a 9-month mental training study in a large group of adults focusing on the training of attention, social emotions such as empathy as well as compassion, and social cognition such as our ability for ToM. Projecting functional network information into compact lower dimensional spaces governed by intrinsic connectivity, we were able to show that cortical networks implicated in cognitive and social functions co-localized yet segregated mainly in transmodal association cortices. These systems are thought to be at the top of the putative cortical hierarchy, have undergone considerably evolutionary expansion, show protracted neurodevelopment, and harbor the potential for extensive plasticity and increased inter-individual variation (Baum et al., 2020; Mueller et al., 2013; Wei et al., 2019; Xu et al., 2020). Our study has shown that the topology of such systems at rest is closely related to social emotions and social cognition, and that the concerted training of such skills implicates dissociable changes in their network signatures in connectome gradient space. Although the current work focused on middle-aged healthy adults ranging from 20 to 55 years of age, our framework may be useful to explore dynamic changes during brain development in children and adolescence, and be a promising approach to understand social and cognitive deficits and challenges to mental health in individuals suffering from neuropsychiatric disorders.

Methods

Experimental design

Participants

A total of 332 healthy adults (197 women, mean±SD=40.7±9.2 years, 20-55 years), recruited in 2012-2014 participated in the study. More than 95% of our sample was Caucasian, with catchment areas balanced across two German municipalities (Berlin and Leipzig). Participant eligibility was determined through a multi-stage procedure that involved several screening and mental health questionnaires, together with a phone interview [for details, see (Singer et al., 2016)]. Next, a face-to-face mental health diagnostic interview with a trained clinical psychologist was scheduled. The interview included a computer-assisted German version of the Structured Clinical Interview for DSM-IV Axis-I disorders, SCID-I DIA-X (Wittchen and Pfister, 1997) and a personal interview, SCID-II, for Axis-II disorders (First et al., 1997; Wittchen et al., 1997). Participants were excluded if they fulfilled criteria for: i) an Axis-I disorder within the past two years; ii) Schizophrenia, psychotic disorders, bipolar disorder, substance dependency, or an Axis-II disorder at any time in their life. No participant had a history of neurological disorders or head trauma, based on an in-house self-report questionnaire used to screen all volunteers prior to imaging investigations. In addition, participants underwent a diagnostic radiological evaluation to rule out the presence of mass lesions (e.g., tumors, vascular malformations). All participants gave written informed consent and the study was approved by the Research Ethics Committees of the University of Leipzig (#376/12-ff) and Humboldt University in Berlin (#2013-02, 2013-29, 2014-10). The study was registered at ClinicalTrials.gov under the title “Plasticity of the Compassionate Brain” (#NCT01833104). For details on recruitment and sample selection, see the full cohort and study descriptor (Singer et al., 2016).

Study design

Our study focused on two training groups: training cohort 1 (TC1, n=80 at enrolment) and training cohort 2 (TC2, n=81 at enrolment), as well as a retest control cohort (RCC) that was partly measured prior to (n=30 at enrolment) and partly after (n=60 at enrolment) TC1 and TC2. A third training cohort (TC3, n=81 at enrolment) underwent an independent training program, and was included as an additional active control for the *Presence* module. Participants were selected from a larger pool of potential volunteers by bootstrapping without replacement, creating cohorts not differing significantly with respect to several demographic and self-report traits. Total training duration of TC1 and TC2 was 39 weeks (~nine months), divided into three modules: *Presence*, *Affect*, and *Perspective* (see below, for details), each lasting for 13 weeks (**Figure 1**). TC3 only participated in one 13-week *Affect* training, and is only included in supplementary robustness analyses, so that the main analysis of functional plasticity focusses on TC1 and TC2. Our main cohorts of interest, TC1 and TC2, underwent *Affect* and *Perspective* modules in different order to act as active control cohorts for each other. Specifically, TC1 underwent “*Presence-Affect-Perspective*”, whereas TC2 underwent “*Presence-Perspective-Affect*”. TC1, TC2, and RCC underwent four testing phases. The baseline-testing phase is called T0; testing phases at the end of the xth module are called Tx

(*i.e.*, T1, T2, T3). In RCC, testing was carried out at similarly spaced intervals. The study had a slightly time-shifted design, where different groups started at different time points to simultaneously accommodate scanner and teacher availability. As we focused on training-related effects, we did not include analysis of a follow-up measurement T4 that was carried out 4, 5, or 10 months after the official training had ended. For details on training and practice set-up, timeline, and measures, see (Singer et al., 2016).

Final sample

We excluded individuals with missing structural and/or functional MRI data and/or a framewise-displacement of >0.3mm (<5%) (Power et al., 2012) and individuals with gradients that showed less than r=0.5 correspondence to the average gradient patterns were disregarded for analysis (<1%). Further details of sample size per time-point and exclusion criteria are in **Table 1, Table 2** and **Table 3**.

Recruited (N, mean age, % female)	T_0 (N)	T_1 (N)	T_2 (N)	T_3 (N)
Total (N = 332)	284	264	185	188
TC1 (N = 80; 41.3; 58.8)	72	66	56	55
TC2 (N = 81; 41.2; 59.3)	69	60	63	66
RCC (N =90; 40.0; 58.9)	72	71	66	67
TC3 (N = 81; 40.4; 60.5)	71	67		

Table 1. Participant inclusion in resting-state analysis.

Reason for dropout (TC1, TC2, RCC)	T_0	T_1	T_2	T_3
Structural MR incidental finding	5	(5 based on T_0)	(5 based on T_0)	(5 based on T_0)
Structural MRI quality control	7	6	4	2
Dropout	2	7(2 based on T_0)	7(9 based on T_{01})	7 (16 based on T_{012})
Medical reasons	1	7(1 based on T_0)	7(8 based on T_{01})	(15 based on T_{012})
Other	4	10	7	7
Functional MRI missing	4	10	11	4
Functional MRI quality	14 (26)	5(23)	8(14)	3(5)

Table 2. Reason for missing data across the study duration. *MR incidental findings* are based on T_0 radiological evaluations; participants who did not survive *MRI quality control* refers to movement and/or artefacts in the T1-weighted MRI; dropout details can be found in (Singer et al., 2016); *no MRT*: due to illness / scheduling issues / discomfort in scanner; *other*: non-disclosed; *functional MRI missing:* no complete functional MRI; *functional MRI quality:* >0.3mm movement (low quality in volume + surface)

Reason for dropout (TC3)	T_0	T_1
MR incidental finding	3	(3 based on T_0)
MRI quality control	0	0
Dropout	0	3
Medical reasons	1	2
Other	5	3
Functional MRI missing	2	0
Functional MRI quality	1(1)	3(4)

Table 3. Reason for missing data across the study duration. *MR incidental findings* are based on T_0 radiological evaluations; participants who did not survive *MRI quality control* refers to movement and/or artefacts in the T1-weighted MRI; dropout details can be found in (Singer et al., 2016); *no MRT*: due to illness / scheduling issues / discomfort in scanner; *other*: non-disclosed.

Training modules

Each module started off with a three-day intensive retreat, followed by weekly group sessions with the teachers and daily home practice supported by a custom-made internet platform and smartphone applications providing: i) audio streams for guided meditations and ii) an interface for dyadic exercises (**Figure2b**) (Singer et al., 2016). During the retreat, participants were introduced to topics and associated core exercises of the upcoming module. Training during the subsequent eight weeks included weekly two-hour-long sessions with instructors that included discussion of training challenges and effects, practice of the core exercises, and introduction to new contemplative practices. The last five weeks of each module were used to consolidate previous topics, with no new topics being introduced.

Presence: Core exercises, practiced repeatedly during the retreat, in the weekly sessions, and at home (instruction was to practice at least five times per week), were Breathing Meditation and Body Scan (Kabat-Zinn, 1990). The instruction for the Breathing Meditation was to focus attention on sensations of breathing and to refocus attention whenever it drifted. The Body Scan involved focusing on various parts of the body in a systematic fashion (*e.g.*, from toes to head), while paying close attention to perceptions occurring in these body parts. Additional exercises of Presence, also practiced during the retreat and weekly sessions, were walking meditation, meditations on vision, sound, and taste, as well as an open presence meditation. These practices require a deliberate focus of attention on certain facets of present moment-to-moment experience, monitoring of distractions, and reorienting towards the object of attention in the meditation, be it the breath, a sound or a visual object.

Affect: Core exercises were Loving-kindness Meditation (Salzberg, 1995) and the 'Affect Dyad' (Kok and Singer, 2016; Singer et al., 2016). For Loving-kindness Meditation, participants were first familiarised to ways of connecting with the feeling and motivation of love and care, such as imagining a baby, a cute animal, a close benevolent other, a place of safety and comfort, or focusing on sensations of warmth in the body. These feelings can then be directed towards oneself and others. The typical instruction for the Loving-kindness Meditation was to start with imagining oneself and then a benefactor, where such feelings might arise naturally, and then to extend feelings of Loving-kindness and good wishes to self and then the benefactor. Over the course of several sessions, participants were asked to successively extent these feelings to others to whom one feels neutral, people one has

difficulties with, and ultimately all humans and sentient beings. To foster experiences of Loving-kindness, participants were instructed to mentally repeat phrases such as “May you be happy”, “May you be healthy”, “May you be safe”, and “May you live with ease”.

The Affect Dyad is a partner practise done face-to-face during the retreat, during the weekly sessions, and via the web- or smartphone-based application during daily practice at home. During this exercise, participants contemplated situations from their last day that they experienced as difficult and situations which they were grateful for. Partner A was instructed to listen attentively to what the speaker (partner B) has to say without giving verbal or non-verbal feedback, promoting empathic listening. The speaker remembered the situation and how it felt like and focused on the immediate subjective affective and bodily experience without engaging in abstract reasoning or interpretation. After a first run, roles were switched. This contemplative dialogue allows cultivating empathic listening in the listener and observing difficult emotions and their effect on the body as well as developing gratitude and positive affect in the speaker.

Additional features of *Affect* were exploration of emotions in an attitude of acceptance and care, a guided mental training that contrasts empathy and compassion and teaches participants how to transform an empathic into a loving and compassionate response when confronted with the suffering of others (Klimecki et al., 2014), forgiveness meditation, and development of self-compassion (Neff and Germer, 2013).

Perspective: Core exercises were an Observing-thoughts Meditation and a ‘Perspective Dyad’. In the former, the objective was to observe thoughts as mental events or natural phenomena and not as direct representations of reality. In the starting phase of the practice, this was supported by identifying thoughts using opposite poles such as me/other, past/future, positive/negative, or more generic labels such as “judging” and “thinking”. Later in the program, participants were instructed to observe the coming and going of thoughts without getting involved in them.

The Perspective Dyad is a partner exercise with a structure comparable to the Affect Dyad. This exercise was partly based on the Internal Family System approach by Schwartz and colleagues (Holmes, 2007; Schwarz, 1997) and on theoretical accounts distinguishing between affective (*e.g.*, compassion and empathy) and cognitive (*e.g.*, Theory of Mind, ToM) routes of social cognition (Singer, 2006, 2012). For this training of perspective-taking on self and others, participants were first introduced to the concept of inner parts, personality-trait-like patterns of cognition, emotion, and behavioral tendencies that dominate in certain situations and shape experience as well as behavior (Holmes, 2007). During the retreat and throughout the course, participants were supported in detecting inner parts. In the Perspective Dyad, the speaker described a situation of the last day from the perspective of one of his/her inner parts, that is, how the experience might have been if a certain inner part had been dominant. The other participant listened attentively without giving verbal or non-verbal feedback and tried to find out from which inner part the speaker was recounting the situation. The listener, thus, had to engage in cognitive perspective-taking on the other to find out “who is speaking” and to infer the needs, desires, and beliefs of the other. The speaker, in turn, needed to take a meta-perspective onto its own self-related aspects and to decouple from a lived and experienced reality.

Additional elements of *Perspective* were exercises in which participants needed to take the perspective of people they encountered whom they have difficulties with in their daily lives, reflections on the central role that thoughts play in our lives, how these might differ for thoughts of others, and why understanding them differs from approving their behavior. This description of the training protocol was adapted from previous work (Singer et al., 2016; Valk et al., 2017).

Neuroimaging acquisition and analysis

MRI acquisition

MRI data were acquired on a 3T Siemens Magnetom Verio (Siemens Healthcare, Erlangen, Germany) using a 32-channel head coil. We recorded task-free functional MRI using a T2*-weighted gradient 2D-EPI sequence (repetition time [TR]=2000ms, echo time [TE]=27ms, flip angle=90°; 37 slices tilted at approximately 30° with 3 mm slice thickness, field of view [FOV]=210×210mm2, matrix=70×70, 3×3×3 mm3 voxels, 1 mm gap; 210 volumes per session). We also acquired a T1-weighted 3D-MPRAGE sequence (176 sagittal slices, TR=2300 ms, TE=2.98 ms, inversion time [TI]=900 ms, flip angle=7°, FOV=240×256 mm^2, matrix=240×256, 1×1×1 mm^3 voxels). Throughout the duration of our longitudinal study, imaging hardware and console software (Syngo B17) were held constant. Participants also underwent multiple functional tasks, which had a variable number of volumes due to the different lengths of the tasks, but otherwise identical sequence parameters analogous to the task-free data (Kanske et al., 2015; Trautwein et al., 2016). During the functional session, participants were instructed to lie still in the scanner, think of nothing in particular, and fixate a white cross in the centre of a black screen.

Task-free functional MRI analysis

Processing was based on DPARSF/REST for Matlab [http://www.restfmri.net (Song et al., 2011)]. We discarded the first 5 volumes to ensure steady-state magnetization, performed slice-time correction, motion correction and realignment, and co-registered functional time series of a given subject to the corresponding T1-weighted MRI. Images underwent unified segmentation and registration to MNI152, followed by nuisance covariate regression to remove effects of average WM and CSF signal, as well as 6 motion parameters (3 translations, 3 rotations). We included a scrubbing (Power et al., 2012)that modelled time points with a frame-wise displacement of ≥0.5 mm, together with the preceding and subsequent time points as separate regressors during nuisance covariate correction. Volume-based timeseries were mapped to the fsaverage5 surface using bbregister.

Gradient construction

In line with previous studies (Bethlehem et al., 2020; Hong et al., 2019; Margulies et al., 2016; Paquola et al., 2019; Vos De Wael et al., 2020; Vos de Wael et al., 2018) the functional connectivity matrix was proportionally thresholded at 90% per row and converted into a

normalised angle matrix using the BrainSpace toolbox for MATLAB (Vos De Wael et al., 2020). Diffusion map embedding, a nonlinear manifold learning technique (Coifman et al., 2005), identified principal gradient components, explaining functional connectivity variance in descending order. In brief, the algorithm estimates a low-dimensional embedding from a high-dimensional affinity matrix. In this space, cortical nodes that are strongly interconnected by either many suprathreshold edges or few very strong edges are closer together, whereas nodes with little or no functional connectivity are farther apart. The name of this approach, which belongs to the family of graph Laplacians, derives from the equivalence of the Euclidean distance between points in the embedded space and the diffusion distance between probability distributions centred at those points. It is controlled by the parameter α, which controls the influence of the density of sampling points on the manifold ($\alpha = 0$, maximal influence; $\alpha = 1$, no influence). Based on previous work (Margulies et al., 2016), we set $\alpha = 0.5$, a choice that retains the global relations between data points in the embedded space and has been suggested to be relatively robust to noise in the functional connectivity matrix. The diffusion time (t), which controls the scale of eigenvalues of the diffusion operator was set at t=0 (default). Individual embedding solutions were aligned to the group-level embedding based on the Human Connectome Project S1200 sample (Van Essen et al., 2013) via Procrustes rotations (Vos De Wael et al., 2020). The Procrustes alignment enables comparison across individual embedding solutions, provided the original data is equivalent enough to produce comparable Euclidean spaces.

3D gradients

To construct the combined gradient, we computed the Euclidean distance to the individual center for gradient 1-3. Next, to evaluate change within and between individuals, we computed the difference between gradient scores between different time-points. To investigate multi-dimensional differences in cortical organization, we quantified dispersion of task-based networks (Kanske et al., 2015; Trautwein et al., 2016) and functional networks (Yeo et al., 2011) in 3D space. Each axis of this 3D space was defined by the values along the first three gradients. Within network dispersion was quantified as sum squared Euclidean distance of network nodes to the network centroid at individual level. Between network dispersion was calculated as the Euclidean distance between network centroids. These metrics were calculated for each subject within the individualized, aligned gradient space similar to previous work (Bethlehem et al., 2020).

Functional Networks

Paradigms identifying functional localizers are previously described (Kanske et al., 2015; Trautwein et al., 2016). In short, the attention network was defined based on the conjunction of activation following selective attention and executive control during a cued-flanker task. The socio-affective network was based on the contrast between emotional and non-emotional videos and the socio-cognitive network was based on the contrast between ToM and non-ToM questions. Networks were mapped to the fsaverage5 surface-template, and following

function-specific and function-unspecific networks were constructed by assessing the overlap between the attention, socio-affective, and socio-cognitive network.

Behavioral markers

We assessed a battery of behavioral markers developed and adapted to target the main aims of the *Presence, Perspective,* and *Affect* modules: selective attention, compassion, and ToM. Behavioral changes of these markers elicited by the different modules are reported elsewhere (Trautwein et al., 2020). The measure for compassion was based on the EmpaToM task, a developed and validated naturalistic video paradigm in the current subjects (Kanske et al., 2015; Tholen et al., 2020). Videos showed people recounting autobiographical episodes that were either emotionally negative (e.g., loss of a loved one) or neutral (e.g., commuting to work), followed by Likert-scale ratings of experienced valence and compassion. Since the conceptual understanding of compassion might change due to the training, we ensured a consistent understanding by defining it prior to each measurement as experiencing feelings of care, warmth, and benevolence. Compassion was quantified as mean of compassion ratings across all experimental conditions. The EmpaToM task (Kanske et al., 2015) also allowed for measurement of ToM performance. After the ratings, multiple-choice questions requiring inference of mental states (thoughts, intentions, beliefs) of the person in the video or factual reasoning on the video's content (control condition) were asked. Questions had three response options and only one correct answer, which had been validated during pre-study piloting (Kanske et al., 2015). Here, we calculated participants' error rates during the ToM questions after the video, collapsed across neutral and negative conditions.

Statistical analysis

Analysis was performed using SurfStat for Matlab (Worsley et al., 2009). We employed linear mixed-effects models, a flexible statistical technique that allows for inclusion of multiple measurements per subjects and irregular measurement intervals (Pinheiro and Bates, 2000). In all models, we controlled for a random effect term of Subject. Inference was performed on subject-specific manifold change maps, Δeccentricity, which were generated by subtracting vertex -wise manifold maps of subsequent time points for a given participant.

a) Cross-sectional association between functional embedding and behaviour: To evaluate the relationship between behavior and functional networks, we selected participants at baseline and used a linear model controlling for age and sex to evaluate the association between network-specific functional embedding and behavior.

b) Assessing module-specific change. To compare the modules against each other, we contrasted change of one given module against the average of the other two modules. To compare two modules, we compared module A to module B. To compare a given module against RCC, we estimated contrasts for training cohort change relative to RCC (*Presence, Perspective, Affect*).

c) Correction for multiple comparisons. Whole brain findings were corrected for multiple comparisons using random field theory for non-isotropic images (Worsley et al., 1999) with cluster-forming thresholds of p=0.005. Post-hoc network-based assessments were corrected

for number of tests within the analysis step using FDR correction (Benjamini and Hochberg, 1995).

d) Global signal regression. To assess the effects of global signal on our main longitudinal results we created connectivity matrixes while regressing out global signal within a 400 parcel solution (Schaefer et al., 2018) and captured change within functional community (Yeo et al., 2011) as well as task-based networks.

e) Gradient-specific assessments. Associations between behavior, training, and behavioral change and individual gradients were computed using linear models. Analysis were performed both by investigating whole-brain associations as well as within the functional networks.

Theoretically, the cross-over design of the study and the inclusion of number of scans since baseline as covariance controlled for test-retest effects on motion (as participants may become calmer in scanner after repeated sessions). Nevertheless, to control for outliers, we removed all individuals with >0.3 mm/degree movement (Power et al., 2012).

Behavioral prediction

We adopted a supervised machine learning framework with cross-validation to predict behavioral change based on change in functional connectivity organization. We aimed at predicting attention, compassion, and perspective-taking (**Figure 3**). First, we regressed out age and sex, as well as performed mixed modelling to control for repeated measures. We utilized 5-fold cross-validation separating training and test data. Feature selection procedure was conducted using the training data (4/5 segments) and it was repeated 5 times with different segments of training data. Based on our vertex wise manifold summary, a set of features that could predict each behavioral score was identified using Lasso regularization, i.e. an L1 penalization on the parameters. Linear regression for predicting behavioral scores was constructed using the selected features as independent variables within the training data (4/5 segments) and it was applied to the test data (1/5 segment) to predict their behavioral scores. Prediction procedure was repeated 100 times with different set of training and test data to avoid bias for separating subjects. The prediction accuracy was assessed by calculating Pearson's correlation between the actual and predicted behavioral scores as well as their mean absolute error, MAE (Chang et al., 2015). To evaluate model specificity, we used the output weights of each prediction model to predict the other behavioral domain.

Meta-analytical decoding

We used the NeuroSynth meta-analytic database (http://www.neurosynth.org) (Yarkoni et al., 2011) to assess topic terms associated with social, affective, interoceptive, and cognitive functioning. (*"affect", "autonomic", "emotion", "emotion regulation", "empathy", "episodic memory", "goal directed", "interoceptive", "mentalizing", "reward", "self", "semantic memory", "social", "social interaction", "theory mind", "inhibition", "executive", "cognitive control", "attention", "working memory"*). We used a t-test to

compare there was significantly activation associated with a certain term relative to the activity outside the task-based activation. Following, after correcting for number of contrasts tested, we plotted the functions significantly associated with the respective task-based activation.

Transcriptomic association analysis

To provide additional neurobiological context of our findings, we assessed spatial correlations of task-based functional activations in attention and socio-affective, and socio-cognitive processing and gene expression patterns. Initially, we correlated the t-statistics map of the functional activations and the post-mortem gene expression maps provided by Allen Institute for Brain Sciences (AIBS) using the Neurovault gene decoding tool (Gorgolewski et al., 2014; Hawrylycz et al., 2012). Neurovault implements mixed-effect analysis to estimate associations between the input t-statistic map and the genes of AIBS donor brains yielding the gene symbols associated with the input t-statistic map. Gene symbols that passed a significance level of FDR-corrected $p < 0.05$ and showed a positive association with the task-based were further tested whether they are consistently expressed across the donors using abagen toolbox (https://github.com/rmarkello/abagen)(Arnatkeviciute et al., 2019). For each gene, we estimated whole-brain gene expression map and correlated it between all pair of different donors. Genes showing consistent whole-brain expression pattern across donors (FDR < 0.05). In a second stage, gene lists that were significant were fed into enrichment analysis, which involved comparison against developmental expression profiles from the BrainSpan dataset (http://www.brainspan.org) using the cell-type specific expression analysis (CSEA) developmental expression tool (http://genetics.wustl.edu/jdlab/csea- tool-2) (Dougherty et al., 2010). As the AIBS repository is composed of adult post-mortem datasets, it should be noted that the associated gene symbols represent indirect associations with the input t-statistic map derived from the developmental data.

Subcortical and cerebellar analysis

To assess the cross-sectional and longitudinal association between social behavior and functional embedding of sub-cortical regions we computed the embedding of sub-cortical regions defined by the Harvard-Oxford sub-cortical regions (50%) threshold using their functional connectivity maps with cortical vertices, similar to previous work (Vos de Wael et al., 2018). Following we computed the gradient scores of each subcortical region and calculated the multidimensional manifold embedding by evaluating the Euclidian distance of the first two gradients. Using this information, we computed the 3D manifold embedding for sub-cortical regions. We used an analogous approach for post-hoc assessment of Harvard-Oxford cerebellar regions, resulting in three gradients along which cerebellar sub-regions were embedded in the cortical functional connectome.

Acknowledgements

Data for the *ReSource* project were collected between 2013 and 2016 at the Department of Social Neuroscience at the Max Planck Institute for Human Cognitive and Brain Sciences Leipzig. Tania Singer (Principal Investigator) received funding for the ReSource Project from the European Research Council (ERC) under the European Community's Seventh Framework Program (FP7/2007–2013) ERC grant agreement number 205557. Sofie Valk received support from the Max Planck Society (Otto Hahn Award). Boris Bernhardt acknowledges research support from the NSERC (Discovery-1304413), the Canadian Institutes of Health Research (CIHR FDN-154298), SickKids Foundation (NI17-039), Azrieli Center for Autism Research (ACAR-TACC), and the Tier-2 Canada Research Chairs program. Bo-yong Park was funded by Molson Neuro-Engineering fellowship by Montreal Neurological Institute and Hospital (MNI) and the Fonds de la Recherche du Quebec – Santé. We thank Lisa Feldman Barrett for fruitful discussions during the conception of this manuscript. **Author contributions:** SLV and BCB were involved in data acquisition and processing, and conceived and designed the resting-state computational experiments. SH and BP contributed to development of gradient and genetic enrichment analysis. F.-M.T., A.B., and P.K. designed and analyzed the functional and behavioral data used in this study. T.S. initiated and developed the *ReSource* Project and model, as well as the training protocol. All authors discussed, wrote, and approved the final version of the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

References

Alcala-Lopez, D., Smallwood, J., Jefferies, E., Van Overwalle, F., Vogeley, K., Mars, R.B., Turetsky, B.I., Laird, A.R., Fox, P.T., Eickhoff, S.B., and Bzdok, D. (2018). Computing the Social Brain Connectome Across Systems and States. Cereb Cortex 28, 2207-2232.

Alcala-Lopez, D., Vogeley, K., Binkofski, F., and Bzdok, D. (2019). Building blocks of social cognition: Mirror, mentalize, share? Cortex 118, 4-18.

Alves, P.N., Foulon, C., Karolis, V., Bzdok, D., Margulies, D.S., Volle, E., and Thiebaut de Schotten, M. (2019). An improved neuroanatomical model of the default-mode network reconciles previous neuroimaging and neuropathological findings. Commun Biol 2, 370.

Arnatkeviciute, A., Fulcher, B.D., and Fornito, A. (2019). A practical guide to linking brain-wide gene expression and neuroimaging data. Neuroimage 189, 353-367.

Barrett, L.F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. Soc Cogn Affect Neurosci 12, 1-23.

Barrett, L.F., Quigley, K.S., and Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. Philos Trans R Soc Lond B Biol Sci 371.

Barrett, L.F., and Satpute, A.B. (2013). Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain. Current opinion in neurobiology 23, 361-372.

Batson, C.D. (2009). These things called empathy. In The Social Neuroscience of Empathy, J. Decety, and W. Ickes, eds. (Cambridge, MA: MIT press), pp. 16-31.

Baum, G.L., Cui, Z., Roalf, D.R., Ciric, R., Betzel, R.F., Larsen, B., Cieslak, M., Cook, P.A., Xia, C.H., Moore, T.M., *et al.* (2020). Development of structure-function coupling in human brain networks during youth. Proc Natl Acad Sci U S A 117, 771-778.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological) 57, 289-300.

Bertolero, M.A., Yeo, B.T.T., Bassett, D.S., and D'Esposito, M. (2018). A mechanistic model of connector hubs, modularity and cognition. Nat Hum Behav 2, 765-777.

Bethlehem, R.A.I., Paquola, C., Seidlitz, J., Ronan, L., Bernhardt, B.C., Consortium, C.-C., and Tsvetanov, K.A. (2020). Dispersion of functional gradients across the lifespan. NeuroImage

Betzel, R.F., Fukushima, M., He, Y., Zuo, X.N., and Sporns, O. (2016). Dynamic fluctuations coincide with periods of high and low modularity in resting-state functional brain networks. Neuroimage 127, 287-297.

Burt, J.B., Demirtas, M., Eckner, W.J., Navejar, N.M., Ji, J.L., Martin, W.J., Bernacchia, A., Anticevic, A., and Murray, J.D. (2018). Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. Nat Neurosci 21, 1251-1259.

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

Chanes, L., and Barrett, L.F. (2016). Redefining the Role of Limbic Areas in Cortical Processing. Trends Cogn Sci 20, 96-106.

Chang, L.J., Gianaros, P.J., Manuck, S.B., Krishnan, A., and Wager, T.D. (2015). A Sensitive and Specific Neural Signature for Picture-Induced Negative Affect. PLoS Biol 13, e1002180.

Chaudhuri, R., Knoblauch, K., Gariel, M.A., Kennedy, H., and Wang, X.J. (2015). A Large-Scale Circuit Mechanism for Hierarchical Dynamical Processing in the Primate Cortex. Neuron 88, 419-431.

Chun, M.M., Golomb, J.D., and Turk-Browne, N.B. (2011). A taxonomy of external and internal attention. Annu Rev Psychol 62, 73-101.

Coifman, R.R., Lafon, S., Lee, A.B., Maggioni, M., Nadler, B., Warner, F., and Zucker, S.W. (2005). Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. Proc Natl Acad Sci U S A 102, 7426-7431.
Craig, A.D. (2009). How do you feel--now? The anterior insula and human awareness. Nat Rev Neurosci 10, 59-70.
Dajani, D.R., and Uddin, L.Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. Trends Neurosci 38, 571-578.
Dale, A.M., Fischl, B., and Sereno, M.I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9, 179-194.
de Vignemont, F., and Singer, T. (2006). The empathic brain: how, when and why? Trends Cogn Sci 10, 435-441.
Dougherty, J.D., Schmidt, E.F., Nakajima, M., and Heintz, N. (2010). Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. Nucleic Acids Res 38, 4218-4230.
Dunbar, R.I.M. (1998). The Social Brain hypothesis. Evol Anthropol 6, 178-190.
Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. Trends Cogn Sci 14, 172-179.
Eisenberg, N., and Fabes, R.A. (1990). Empathy: conceptualization, measurement, and relation to prosocial behavior. . Motivation and Emotion 14, 131-149.
Finc, K., Bonna, K., He, X., Lydon-Staley, D.M., Kuhn, S., Duch, W., and Bassett, D.S. (2020). Dynamic reconfiguration of functional brain networks during working memory training. Nat Commun 11, 2435.
First, M.B., Gibbon, M., Spitzer, R.L., Williams, J.B.W., and Benjamin, L.S. (1997). Structured Clinical Interview for DSM-IV Axis II Personality Disorders, (SCID-II). (Washington, D.C.: American psychiatric Press, Inc.).
Fornito, A., Arnatkeviciute, A., and Fulcher, B.D. (2019). Bridging the Gap between Connectome and Transcriptome. Trends Cogn Sci 23, 34-50.
Frith, C.D., and Frith, U. (2006). The neural basis of mentalizing. Neuron 50, 531-534.
Frith, U., and Frith, C.D. (2003). Development and neurophysiology of mentalizing. Philos Trans R Soc Lond B Biol Sci 358, 459-473.
Fulcher, B.D. (2019). Discovering Conserved Properties of Brain Organization Through Multimodal Integration and Interspecies Comparison. J Exp Neurosci 13, 1179069519862047.
Gorgolewski, K.J., Fox, A.S., Chang, L.J., Schaefer, A., Arelin, K., Burmann, I., Sacher, J., and Margulies, D.S. (2014). Tight fitting genes: finding relations between statistical maps and gene expression patterns. In OHBM 2014.
Grasby, K.L., Jahanshad, N., Painter, J.N., Colodro-Conde, L., Bralten, J., Hibar, D.P., Lind, P.A., Pizzagalli, F., Ching, C.R.K., McMahon, M.A.B., *et al.* (2020). The genetic architecture of the human cerebral cortex. Science 367.
Haak, K.V., and Beckmann, C.F. (2020). Understanding brain organisation in the face of functional heterogeneity and functional multiplicity. Neuroimage 220, 117061.
Haak, K.V., Marquand, A.F., and Beckmann, C.F. (2018). Connectopic mapping with resting-state fMRI. Neuroimage 170, 83-94.
Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., Shen, E.H., Ng, L., Miller, J.A., van de Lagemaat, L.N., Smith, K.A., Ebbert, A., Riley, Z.L., *et al.* (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. Nature 489, 391-399.
Holmes, T. (2007). Parts word: An illustrated guide to your inner life (Kalamazoo: Winged Heart Press).
Hong, S.J., Vos de Wael, R., Bethlehem, R.A.I., Lariviere, S., Paquola, C., Valk, S.L., Milham, M.P., Di Martino, A., Margulies, D.S., Smallwood, J., and Bernhardt, B.C. (2019). Atypical functional connectome hierarchy in autism. Nat Commun 10, 1022.

Hong, S.J., Xu, T., Nikolaidis, A., Smallwood, J., Margulies, D.S., Bernhardt, B.C., Vogelstein, J.T., and Milham, M.P. (2020). Towards a connectivity gradient-based frameworrk for reproducible biomarker discovery. bioRxiv

Huntenburg, J.M., Bazin, P.L., Goulas, A., Tardif, C.L., Villringer, A., and Margulies, D.S. (2017). A Systematic Relationship Between Functional Connectivity and Intracortical Myelin in the Human Cerebral Cortex. Cereb Cortex 27, 981-997.

Huntenburg, J.M., Bazin, P.L., and Margulies, D.S. (2018). Large-Scale Gradients in Human Cortical Organization. Trends Cogn Sci 22, 21-31.

Hwang, K., Bertolero, M.A., Liu, W.B., and D'Esposito, M. (2017). The Human Thalamus Is an Integrative Hub for Functional Brain Networks. J Neurosci 37, 5594-5607.

Kabat-Zinn, J. (1990). Full catastrophe living: using the wisdom of your body and mind to face stress, pain, and illness. (New York, NY: Delacorte).

Kanske, P., Bockler, A., Trautwein, F.M., Parianen Lesemann, F.H., and Singer, T. (2016). Are strong empathizers better mentalizers? Evidence for independence and interaction between the routes of social cognition. Soc Cogn Affect Neurosci.

Kanske, P., Bockler, A., Trautwein, F.M., and 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.

Kernbach, J.M., Yeo, B.T.T., Smallwood, J., Margulies, D.S., Thiebaut de Schotten, M., Walter, H., Sabuncu, M.R., Holmes, A.J., Gramfort, A., Varoquaux, G., *et al.* (2018). Subspecialization within default mode nodes characterized in 10,000 UK Biobank participants. Proc Natl Acad Sci U S A 115, 12295-12300.

Kleckner, I.R., Zhang, J., Touroutoglou, A., Chanes, L., Xia, C., Simmons, W.K., Quigley, K.S., Dickerson, B.C., and Barrett, L.F. (2017). Evidence for a Large-Scale Brain System Supporting Allostasis and Interoception in Humans. Nat Hum Behav 1.

Klimecki, O.M., Leiberg, S., Ricard, M., and Singer, T. (2014). Differential pattern of functional brain plasticity after compassion and empathy training. Soc Cogn Affect Neurosci 9, 873-879.

Kober, H., Barrett, L.F., Joseph, J., Bliss-Moreau, E., Lindquist, K., and Wager, T.D. (2008). Functional grouping and cortical-subcortical interactions in emotion: a meta-analysis of neuroimaging studies. Neuroimage 42, 998-1031.

Kok, B.E., and Singer, T. (2016). Effects of Contemplative Dyads on Engagement and Perceived Social Connectedness Over 9 Months of Mental Training: A Randomized Clinical Trial. JAMA Psychiatry.

Krienen, F.M., Yeo, B.T., Ge, T., Buckner, R.L., and Sherwood, C.C. (2016). Transcriptional profiles of supragranular-enriched genes associate with corticocortical network architecture in the human brain. Proc Natl Acad Sci U S A 113, E469-478.

Lindquist, K.A., Wager, T.D., Kober, H., Bliss-Moreau, E., and Barrett, L.F. (2012). The brain basis of emotion: a meta-analytic review. Behav Brain Sci 35, 121-143.

Lumma, A.L., Kok, B.E., and Singer, T. (2015). Is meditation always relaxing? Investigating heart rate, heart rate variability, experienced effort and likeability during training of three types of meditation. Int J Psychophysiol 97, 38-45.

Lutz, A., Mattout, J., and Pagnoni, G. (2019). The epistemic and pragmatic value of non-action: a predictive coding perspective on meditation. Curr Opin Psychol 28, 166-171.

Margulies, D.S., Ghosh, S.S., Goulas, A., Falkiewicz, M., Huntenburg, J.M., Langs, G., Bezgin, G., Eickhoff, S.B., Castellanos, F.X., Petrides, M., *et al.* (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. Proc Natl Acad Sci U S A 113, 12574-12579.

Marquand, A.F., Haak, K.V., and Beckmann, C.F. (2017). Functional corticostriatal connection topographies predict goal directed behaviour in humans. Nat Hum Behav 1, 0146.

Menon, V., and Uddin, L.Q. (2010). Saliency, switching, attention and control: a network model of insula function. Brain Struct Funct 214, 655-667.
Mesulam, M.M. (1998). From sensation to cognition. Brain 121 (Pt 6), 1013-1052.
Mueller, S., Wang, D., Fox, M.D., Yeo, B.T., Sepulcre, J., Sabuncu, M.R., Shafee, R., Lu, J., and Liu, H. (2013). Individual variability in functional connectivity architecture of the human brain. Neuron 77, 586-595.
Muller, E.J., Munn, B., Hearne, L.J., Smith, J.B., Fulcher, B., Arnatkeviciute, A., Lurie, D.J., Cocchi, L., and Shine, J.M. (2020). Core and matrix thalamic sub-populations relate to spatio-temporal cortical connectivity gradients. Neuroimage 222, 117224.
Murphy, A.C., Bertolero, M.A., Papadopoulos, L., Lydon-Staley, D.M., and Bassett, D.S. (2020). Multimodal network dynamics underpinning working memory. Nat Commun 11, 3035.
Murphy, C., Jefferies, E., Rueschemeyer, S.A., Sormaz, M., Wang, H.T., Margulies, D.S., and Smallwood, J. (2018). Distant from input: Evidence of regions within the default mode network supporting perceptually-decoupled and conceptually-guided cognition. Neuroimage 171, 393-401.
Murphy, C., Wang, H.T., Konu, D., Lowndes, R., Margulies, D.S., Jefferies, E., and Smallwood, J. (2019). Modes of operation: A topographic neural gradient supporting stimulus dependent and independent cognition. Neuroimage 186, 487-496.
Neff, K.D., and Germer, C.K. (2013). A pilot study and randomized controlled trial of the mindful self-compassion program. J Clin Psychol 69, 28-44.
Ochsner, K.N., and Lieberman, M.D. (2001). The emergence of social cognitive neuroscience. Am Psychol 56, 717-734.
Paquola, C., Benkarim, O., DeKraker, J., Lariviere, S., Frassle, S., Royer, J., Tavakol, S., Valk, S., Bernasconi, A., Bernasconi, N., *et al.* (2020a). Convergence of cortical types and functional motifs in the human mesiotemporal lobe. Elife 9.
Paquola, C., Bethlehem, R.A.I., and Bernhardt, B.C. (2020b). A moment of change: shifts in myeloarchitecture characterise adolescent development of cortical gradients. eLife.
Paquola, C., Vos De Wael, R., Wagstyl, K., Bethlehem, R.A.I., Hong, S.J., Seidlitz, J., Bullmore, E.T., Evans, A.C., Misic, B., Margulies, D.S., *et al.* (2019). Microstructural and functional gradients are increasingly dissociated in transmodal cortices. PLoS Biol 17, e3000284.
Park, B., Bethlehem, R.A.I., Paquola, C., Lariviere, S., Cruces, R.R., Vos De Wael, R., Consortium, N.i.P.N., Bullmore, E.T., and Bernhardt, B.C. (2020). Macroscale connectome manifold expansion in adolescence. bioRxiv
Pessoa, L. (2014). Understanding brain networks and brain organization. Phys Life Rev 11, 400-435.
Pinheiro, J.C., and Bates, D. (2000). Mixed-effect models in S and S-PLUS (Springer).
Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., and Petersen, S.E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59, 2142-2152.
Raz, G., Touroutoglou, A., Wilson-Mendenhall, C., Gilam, G., Lin, T., Gonen, T., Jacob, Y., Atzil, S., Admon, R., Bleich-Cohen, M., *et al.* (2016). Functional connectivity dynamics during film viewing reveal common networks for different emotional experiences. Cogn Affect Behav Neurosci 16, 709-723.
Salzberg, D. (1995). Lovingkindness. The Revolutionary Art of Happiness. (Boston, MA: Shambala).
Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". Neuroimage 19, 1835-1842.
Schaafsma, S.M., Pfaff, D.W., Spunt, R.P., and Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. Trends Cogn Sci 19, 65-72.

Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.N., Holmes, A.J., Eickhoff, S.B., and Yeo, B.T.T. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. Cereb Cortex 28, 3095-3114.
Schmalzle, R., Brook O'Donnell, M., Garcia, J.O., Cascio, C.N., Bayer, J., Bassett, D.S., Vettel, J.M., and Falk, E.B. (2017). Brain connectivity dynamics during social interaction reflect social network structure. Proc Natl Acad Sci U S A 114, 5153-5158.
Schurz, M., Maliske, L., and Kanske, P. (2020a). Cross-network interactions in social cognition: A review of findings on task related brain activation and connectivity. Cortex.
Schurz, M., Radua, J., Tholen, M.G., Maliske, L., Margulies, D.S., Mars, R.B., Sallet, J., and Kanske, P. (2020b). Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. Psychol Bull.
Schwarz, R.C. (1997). Internal familly systems therapy. (New York: Guilford).
Shine, J.M., Breakspear, M., Bell, P.T., Ehgoetz Martens, K.A., Shine, R., Koyejo, O., Sporns, O., and Poldrack, R.A. (2019a). Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. Nat Neurosci 22, 289-296.
Shine, J.M., Hearne, L.J., Breakspear, M., Hwang, K., Muller, E.J., Sporns, O., Poldrack, R.A., Mattingley, J.B., and Cocchi, L. (2019b). The Low-Dimensional Neural Architecture of Cognitive Complexity Is Related to Activity in Medial Thalamic Nuclei. Neuron 104, 849-855 e843.
Shine, J.M., and Poldrack, R.A. (2018). Principles of dynamic network reconfiguration across diverse brain states. Neuroimage 180, 396-405.
Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: review of literature and implications for future research. Neurosci Biobehav Rev 30, 855-863.
Singer, T. (2012). The past, present and future of social neuroscience: a European perspective. Neuroimage 61, 437-449.
Singer, T., Kok, B.E., Bornemann, B., Zurborg, S., Bolz, M., and Bochow, C.A. (2016). The ReSource Project. Background, design, samples and measurements (2nd ed).
Singer, T., and Lamm, C. (2009). The social neuroscience of empathy. Ann N Y Acad Sci 1156, 81-96.
Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R.J., and Frith, C.D. (2004). Empathy for pain involves the affective but not sensory components of pain. Science 303, 1157-1162.
Slagter, H.A., Lutz, A., Greischar, L.L., Francis, A.D., Nieuwenhuis, S., Davis, J.M., and Davidson, R.J. (2007). Mental training affects distribution of limited brain resources. PLoS Biol 5, e138.
Song, C., Kanai, R., Fleming, S.M., Weil, R.S., Schwarzkopf, D.S., and Rees, G. (2011). Relating inter-individual differences in metacognitive performance on different perceptual tasks. Conscious Cogn 20, 1787-1792.
Tang, Y.Y., Holzel, B.K., and Posner, M.I. (2015). The neuroscience of mindfulness meditation. Nat Rev Neurosci 16, 213-225.
Tholen, M.G., Trautwein, F.M., Böckler, A., Singer, T., and Kanske, P. (2020). Functional magnetic resonance imaging (fMRI) item analysis of empathy and theory of mind. Hum Brain Mapp.
Tomasello, M. (1995). Joint atttention as social cognition. In Joint attentioon: Its origins and role in development, C. Moore, and P.J. Dunham, eds. (LawrenceErlbaum Associates, Inc.), pp. 103-130.
Trautwein, F.M., Kanske, P., Bockler, A., and Singer, T. (2020). Differential benefits of mental training types for attention, compassion, and theory of mind. Cognition 194, 104039.
Trautwein, F.M., Singer, T., and Kanske, P. (2016). Stimulus-Driven Reorienting Impairs Executive Control of Attention: Evidence for a Common Bottleneck in Anterior Insula. Cereb Cortex.

Uddin, L.Q. (2015). Salience processing and insular cortical function and dysfunction. Nat Rev Neurosci 16, 55-61.

Valk, S.L., Bernhardt, B.C., Trautwein, F.M., Bockler, A., Kanske, P., Guizard, N., Collins, D.L., and Singer, T. (2017). Structural plasticity of the social brain: Differential change after socio-affective and cognitive mental training. Sci Adv 3, e1700489.

Valk, S.L., Xu, T., Margulies, D.S., Kharabian Masouleh, S., Paquola, C., Goulas, A., Kochunov, P., Smallwood, J., Yeo, B.T., Bernhardt, B.C., and Eickhoff, S.B. (2020). Shaping Brain Structure: Genetic and Phylogenetic Axes of Macro Scale Organization of Cortical Thickness. Sci Adv.

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., and Consortium, W.U.-M.H. (2013). The WU-Minn Human Connectome Project: an overview. Neuroimage 80, 62-79.

Vos De Wael, R., Benkarim, O., Paquola, C., Lariviere, S., Royer, J., Tavakol, S., Xu, T., Hong, S.J., Valk, S.L., Misic, B., *et al.* (2020). BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. Nat Commun Biology.

Vos de Wael, R., Lariviere, S., Caldairou, B., Hong, S.J., Margulies, D.S., Jefferies, E., Bernasconi, A., Smallwood, J., Bernasconi, N., and Bernhardt, B.C. (2018). Anatomical and microstructural determinants of hippocampal subfield functional connectome embedding. Proc Natl Acad Sci U S A 115, 10154-10159.

Wager, T.D., Davidson, M.L., Hughes, B.L., Lindquist, M.A., and Ochsner, K.N. (2008). Prefrontal-subcortical pathways mediating successful emotion regulation. Neuron 59, 1037-1050.

Wang, C., Ong, J.L., Patanaik, A., Zhou, J., and Chee, M.W. (2016). Spontaneous eyelid closures link vigilance fluctuation with fMRI dynamic connectivity states. Proc Natl Acad Sci U S A 113, 9653-9658.

Wang, P., Kong, R., Kong, X., Liegeois, R., Orban, C., Deco, G., van den Heuvel, M.P., and Thomas Yeo, B.T. (2019). Inversion of a large-scale circuit model reveals a cortical hierarchy in the dynamic resting human brain. Sci Adv 5, eaat7854.

Wei, Y., de Lange, S.C., Scholtens, L.H., Watanabe, K., Ardesch, D.J., Jansen, P.R., Savage, J.E., Li, L., Preuss, T.M., Rilling, J.K., *et al.* (2019). Genetic mapping and evolutionary analysis of human-expanded cognitive networks. Nat Commun 10, 4839.

Weng, Y., Lariviere, S., Caciagli, L., Vos de Wael, R., Rodriguez-Cruces, R., Royer, J., Xu, Q., Bernasconi, N., Bernasconi, A., Thomas Yeo, B.T., *et al.* (2020). Macroscale and microcircuit dissociation of focal and generalized human epilepsies. Commun Biol 3, 244.

Wittchen, H.U., and Pfister, H. (1997). Diagnostisches Expertensystem für psychische Störungen (DIA-X) (Frankfurt: Swets & Zeitlinger).

Wittchen, H.U., Zaudig, M., and Fydrich, T. (1997). SKID-Strukturiertes Klinisches Interview für DSM-IV. Achse I und II (Göttingen: Hogrefe).

Worsley, K., Taylor, J.E., Carbonell, F., Chung, M.K., Duerden, E., Bernhardt, B.C., Lyttelton, O.C., Boucher, M., and 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.

Worsley, K.J., Andermann, M., Koulis, T., MacDonald, D., and Evans, A.C. (1999). Detecting changes in nonisotropic images. Hum Brain Mapp 8, 98-101.

Xu, T., Nenning, K., Schwartz, E., Hong, S.J., Vogelstein, J.T., Fair, D.A., Schroeder, C.E., Margulies, D.S., Smallwood, J., Milham, M.P., and Langs, G. (2020). Cross-species Functional Alignment Reveals Evolutionary Hierarchy Within the Connectome. Neuroimage.

Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., and Wager, T.D. (2011). Large-scale automated synthesis of human functional neuroimaging data. Nat Methods 8, 665-670.

Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zollei, L., Polimeni, J.R., *et al.* (2011). The organization of the

human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol 106, 1125-1165.