FUNCTIONAL NETWORK PLASTICITY OF THE HUMAN SOCIAL BRAIN

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 Sofie L Valk

Email: valk@cbs.mpg.de

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.

Classification: Psychological and Cognitive Sciences

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

SUMMARY

Despite the importance of our ability to understand the thoughts and feelings of others, the social brain remains incompletely understood. Here, we studied the plasticity of social brain function in healthy adults following the targeted training of attention-mindfulness, socio-affective, and socio-cognitive skills for 9 months. All participants were followed with repeated multimodal neuroimaging and behavioral testing. Longitudinal analyses of functional networks indicated marked and specific reorganization following mental training. Socio-cognitive training resulting in an increased integration of multiple demand and default mode regions whereas attention-mindfulness resulted in their segregation. Socio-affective training resulted in an increased functional integration of ventral attention network with these regions. Changes in functional network organization were robust after varying analysis parameters, and predictive of change in behavioral markers of compassion and perspective-taking. Our results advance the understanding of the social brain, describing its intrinsic functional organization and reorganization following mental training.

Significance

Our work provides new evidence for functional network reconfigurations following the targeted training of different social skills (attention-mindfulness, socio-affective competences, socio-cognitive skills). Studying a large sample of healthy adults who underwent functional network neuroimaging, we could observe robust evidence for a reorganization of functional brain networks. Each training intervention resulted in unique patterns of functional network reorganization, which were found to relate to behavioral markers indexing these social skills. Thus, our longitudinal findings show that social brain networks can be reconfigured following social mental training in adults and that these changes result in behavioral change. These findings could inform the development and implementation of targeted training interventions aiming at cultivating social understanding and prosocial behavior.

INTRODUCTION

Humans have unique social skills that enhance cooperation and survival (1, 2). These social capacities can largely be divided into several components (3, 4): Firstly, socio-affective (or emotional-motivational) abilities allow us to share feelings with others, and may give rise to compassion and prosocial motivation (5-7). Secondly, a socio-cognitive system gains access to beliefs and intentions of others, a skill also referred to as Theory of Mind (ToM) or mentalizing (3, 8, 9). Finally, there is evidence for a role of attention, action observation, and mindfulness as important auxiliary functions of these social aptitudes, which underlie self-other distinctions and awareness (10-12). Together, these human capacities combine externally and internally oriented cognitive and affective processes and reflect both focused and ongoing thought processes (13-17). However, to what extent these processes are malleable following targeted training is incompletely understood.

Previous functional neuroimaging studies have identified dissociable brain networks involved in these components of social cognition, which somewhat recapitulate large-scale functional networks also seen in the resting brain, notably the ventral attention network (VAN) and paralimbic networks in the case of socio-affective processing (4, 18, 19), the default mode network (DMN) in the case of ToM (4, 19), and multiple demand network (MDN) in the case of attention and action observation (4, 20). Indeed, tasks probing empathy consistently elicit functional activations in anterior insula, supramarginal gyrus, and midcingulate cortex (3, 21, 22), while ToM paradigms activate the DMN regions such as medial frontal cortex, temporo-parietal junction, and superior temporal sulcus (4, 9, 23). Conversely, attention and action observation activate inferior parietal and lateral frontal and anterior insular cortices (24-26). At the same time, it has been shown that the capacity to understand the feelings and thoughts of others in naturalistic settings are supported by multiple processes and consequently implicate combinations of distributed brain networks (19, 27-33). Thus, social cognitive processing can be understood as a hierarchical phenomenon, including automatic and controlled functional processes (3, 32, 34, 35). However, a causal relationship between brain function and differentiable processes underlying social behavior has not been investigated.

Expanding cross-sectional association studies, longitudinal investigations can reveal links between social skills and the human brain, for example by targeted mental training designs. Only few previous studies have addressed longitudinal mental training effects in the domain of social cognition. While these have generally suggested changes in behaviour and MRI-based measures of brain structure and function (36, 37), sample sizes were relatively modest and training intervals short. Moreover, few studies have compared different practices or focussed on different social skills specifically. Our work was realized in the context of the *ReSource* study (38), a large-scale study involved a targeted training of attention-mindfulness (*Presence* module), followed by socio-affective (*Affect* module) and socio-cognitive/ToM training (*Perspective* module) over the course of nine months. 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. Based on the *ReSource* sample, our group could previously show differentiable change in cortical thickness following these modules, illustrating structural plasticity of the social brain (39). However, it remains unknown how targeted training of interoceptive and

attentional processes, socio-affective competences, and socio-cognitive skills relative to changes in brain functional organization.

The overlap between functional activation patterns during specific social cognition paradigms and large-scale functional systems at rest implicates that these functional processes are underpinned by the intrinsic organization of the cerebral cortex (40). This organisation enables the spatial integrating and segregating of different cortical networks, and has been increasingly conceptualized to run along continuous dimensions, also referred to as connectivity gradients (16, 41-43). Reliably identifiable from resting-state functional connectivity data, such gradients can be used to compactly describe neural organization (43-47). For example, the principal gradient places transmodal association networks such as the DMN and paralimbic networks at a maximal distance from sensory and motor regions (16, 42, 43). This segregation may allow association cortices to take on functions only loosely constrained by the environment, which likely facilitates the formation of abstract representations and transformations required for socio-cognitive and socio-affective processing (48). Complementing this sensory-transmodal hierarchy, the second gradient reflects a visual-somatomotor differentiation and the third gradient differentiates the MDS from the rest of the brain (15). Social processes likely have a differential positioning along these gradients, with ToM possibly being associated with transmodal networks such as the DMN (43). Attention may reflect more external focus (16), whereas socio-affective processes may include both alerting, emotionally salient, stimuli from the environment, and more abstract forms of ongoing thought and emotions (32). Effective connectivity studies in social cognition have furthermore shown that, depending on the task, different transmodal brain networks are either integrated and segregated (4, 44, 49-51). For example, complex and demanding ToM tasks often involve an integration of distributed brain networks, such as the DMN and frontal parietal network (4, 52). Conversely, segregation of brain networks may relate to a temporally and hierarchically reduced depth of processing, and may potentially be at play during attentional processes (53, 54). However, the precise mapping of social brain functions to intrinsic organizational axes is not understood to date.

Capitalizing on the longitudinal *Resource* study (38), here we studied functional network plasticity of the social brain. In particular (38), we examined functional network reorganization following the three training modules *i.e., Presence* (promoting attention and interoceptive awareness, resembling mindfulness-based interventions (55)), *Affect* (focusing on socio-affective capacities, such as compassion, gratitude, dealing with difficult emotions), and *Perspective* (focusing on socio-cognitive skills, such as perspective-taking on self and others/ToM as well as metacognition). We sought to specifically assess plasticity of task-based functional networks contributing to attention, socio-affective, and socio-cognitive skills (24, 56) in an intrinsic coordinate system spanned by the first three functional connectivity gradients, together describing functional network integration and segregation. We hypothesized that the training of these social skills would result in dissociable functional reorganization. In addition to assessing the reorganization of brain functional architecture, we assessed associations to behavioral change in attention, compassion, and ToM markers using machine learning with cross-validation. A series of additional analyses evaluated the robustness of our findings with respect to analysis parameter choices and further contextualized findings.

RESULTS

Changes in functional eccentricity following the targeted training of social skills

We analyzed resting-state as well as task-based fMRI and behavioral data from a large-scale mental training intervention, the *ReSource Project* (38). For details, see http://resource-project.org and the preregistered trial https://clinicaltrials.gov/ct2/show/NCT01833104.

We first derived task-based fMRI activation patterns associated with (i) attention, based on a Cued-flanker task (24), (ii) socio-affective functions, such as empathy and compassion, based on the affect contrast in the EmpaTom task (56), and (iii) socio-cognitive functioning, based on the ToM contrast of the EmpaTom (56) in the current sample. To compactly describe large-scale functional brain organization, we reduced the dimensionality of the resting-state fMRI connectomes using previously utilized techniques (57). The identified eigenvectors are spatial gradients describing continuous axes of connectivity space (43, 57, 58). Gradients of each individual were aligned to a template based on the Human Connectome Project dataset (57, 59). For each vertex, we calculated its distance from the center of the gradient coordinate system formed by G1, G2, and G3 for each individual. This eccentricity captured vertex-wise intrinsic functional integration (low eccentricity) and segregation (high eccentricity) in a single scalar value (58, 60, 61). Eccentricity indeed correlated positively ($r>0.5$, $p<0.001$ when controlling for spatial autocorrelation) with network clustering coefficient and characteristic path length, supporting that it captured segregation and integration of large-scale networks (**Supplementary Results**). Highest segregation was observed in visual and sensory-motor networks, while ventral attention and limbic networks were closest to the center of the space (62). Calculating eccentricity scores of the activations at baseline (**Figure 1C**), attention networks had a lower eccentricity than both social networks (both $p<0.001$).

We then examined how functional organization of the different social networks changed following their targeted training. *Resource* participants were randomly assigned to two training cohorts (TC1, N=80; TC2, N=81) and underwent a 9-month training consisting of three sequential training modules with weekly group sessions and daily exercises, completed via cell-phone and internet platforms (**Figure 1, Table 1-3,** *Materials and Methods and Supplementary Materials for more details*). 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 1C**). Additionally, a matched test-retest control cohort did not undergo any training (RCC, N=90), but was followed 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**.**

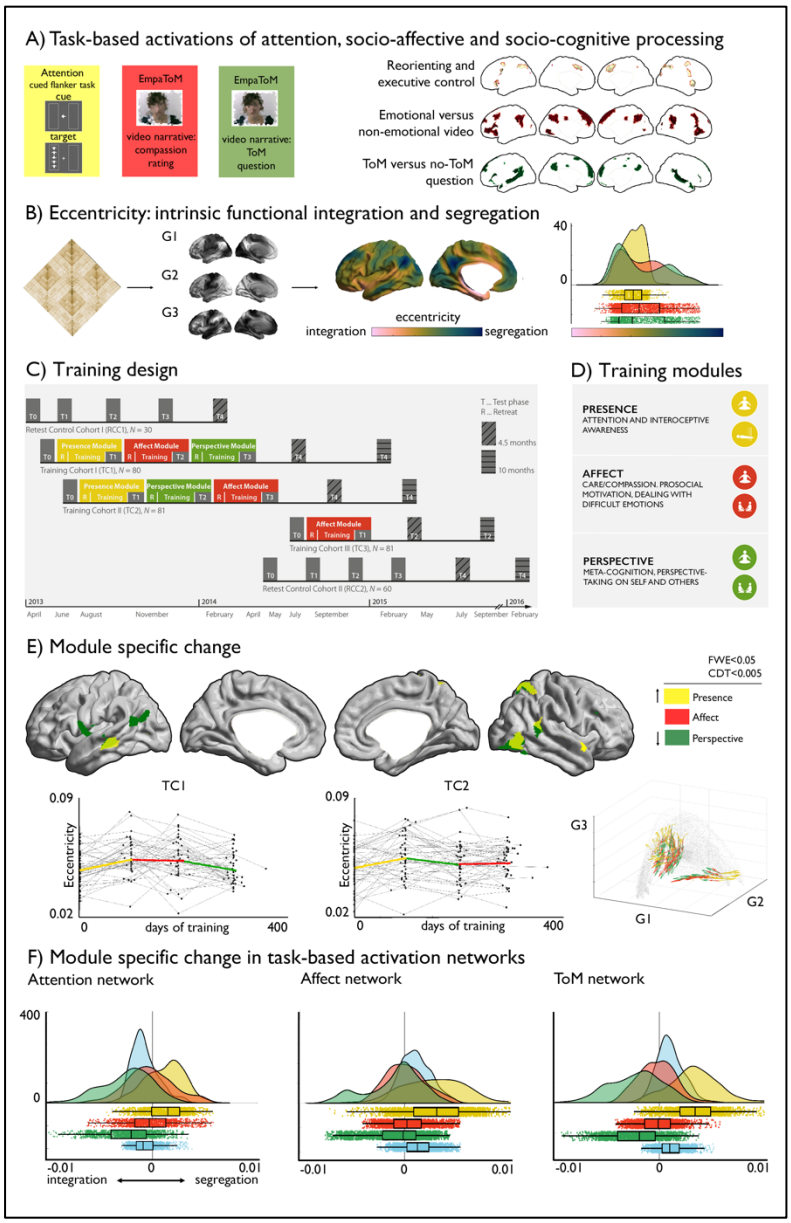

Figure 1. Functional eccentricity alterations of the social brain following mental training. A). Functional localizers (24, 56) and their mappings projected on the cortex**; B**) Vertex-level eccentricity from rs-fMRI by combining the first three gradients (G1-G3); *middle*: vertex-wise eccentricity and *right*: eccentricity of task-based social networks; **C**). Training design of the *ReSource* study; **D**). Training modules; **E**). *upper panel:* Differential change in 3D gradient distance following mental training, *lower panel left:* Average change in overlapping clusters associated with eccentricity 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 eccentricity changes as a function of 3d gradient space. Arrows indicate direction of change; **F**). Changes in average network eccentricity, colored by training: *Presence* (yellow), *Affect* (red), *Perspective* (green), and *retest controls* (light blue). Whole brain findings are corrected for multiple comparisons at $p_{FWE}<0.005$.

Our main analysis tracked longitudinal change in functional eccentricity and each gradient separately (G1-G3) using mixed-effects models (63). We observed marked eccentricity changes following *Presence* and *Perspective*. *Presence* training resulted in increased eccentricity of bilateral temporal and right superior parietal areas (family-wise error corrected p-value, $p_{FWE}=0.002$), indicating increased segregation of these regions. *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 eccentricity change following *Affect* training. Trajectories were similar in TC1 and TC2 (**Figure S2**).

Following, we evaluated training-related alterations in the task-derived functional networks. We observed increases in eccentricity of the socio-cognitive network in following *Presence* training (t= 2.67 p=0.012), whereas both the attention (t=-2.41, p=0.024) and socio-cognitive network (t=-2.91, p=0.006) showed decreased eccentricity following *Perspective* training (**Figure 2D**). These findings indicated that functional connectivity within task-based attention network becomes more dissimilar from other networks following the training of mindfulness-attention during *Presence* training, whereas the training of socio-cognitive skills during *Perspective* leads to increased similarity of attention and socio-cognitive networks with the rest of the connectome. Post-hoc analysis confirmed that eccentricity changes in task-based functional networks were comparable within TC1 and TC2 separately (**Figure S2**), relative to RCC (**Figure S3**) and when controlling for global signal (**Figure S4**). Studying changes within large-scale functional communities (64), we found similar patterns indicating differentiable shifts in integration and segregation after *Presence* and *Perspective* training (**Supplementary Results**).

Relation to alterations in connectome gradient embedding

Having established changes in overall patterns of segregation and integration following *ReSource* training, we studied alterations within the first three large-scale gradients (**Figure 2**).

As in prior work (43), the first gradient (G1) differentiated sensory-motor from transmodal systems, such as the default mode network (DMN). At baseline, attention networks had a lower position relative to both socio-affective (p<0.001) and socio-cognitive (p<0.001) networks. The two latter networks both occupied the transmodal end of G1. Within G1, we found decreases after *Presence* in right inferior parietal cortex (p_{FWE}=0.005) and right superior frontal gyrus (p_{FWE}=0.007) and *Perspective* in left temporal parietal cortex (p_{FWE}=0.035). Conversely, *Affect* showed increases along G1 in right inferior parietal cortex (p_{FWE}=0.022) and right middle insula (p_{FWE}=0.010). Studying average G1 network change in task-based networks we found increase in socio-cognitive network following *Presence* (t=2.219, p=0.043).

The second gradient (G2) ran from sensory-motor to visual networks. For G2, attention networks were lower than socio-affective networks (p<0.003), but higher than socio-cognitive networks (p<0.001). Socio-affective networks were at a higher position on G2 relative to socio-cognitive networks (p<0.001). Studying alterations following mental training along G2, we found decreases in right posterior insula (p_{FWE}=0.004) following *Presence* and decreases in right orbitofrontal cortex (p_{FWE}=0.002) following *Perspective*. G2 change within task-based networks showed an increase of the attention (t=2.21, p=0.041) and socio-affective (t=2.515, p=0.018) networks following *Presence*.

The third gradient (G3) differentiated the multiple-demand network, MDN (65), from the rest of the brain. For G3, attention networks had a higher loading than socio-affective networks (p<0.001), which, in turn, had a higher loading than socio-cognitive networks (p<0.001), indicating a clear differentiation of all three networks along this gradient. Within G3, we observed increases in right inferior temporal cortex (p_{FWE}=0.038) following *Presence*, and decreases in bilateral temporal cortex (p_{FWE}=0.003) and left occipital (p_{FWE}<0.001) following

Affect. Repeating analysis in both training groups separately, indicated whole brain changes in *Affect* along G1 were consistently observed in both cohorts (**Supplementary Results**).

In sum, we could show consistent and divergent functional network plasticity of key regions in the MDN, VAN, and DMN following the training of attention-mindfulness, socio-affect and socio-cognition.

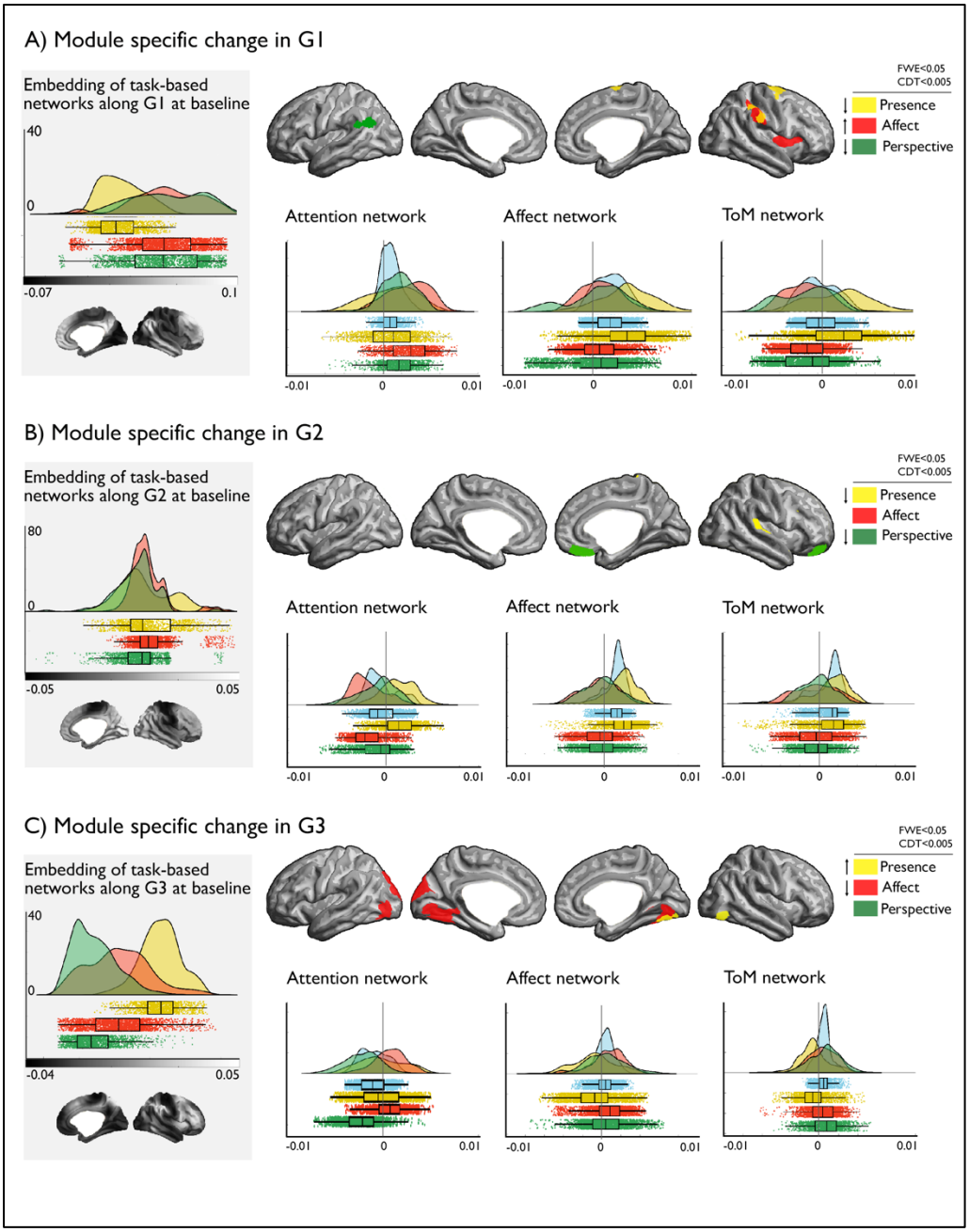

Figure 2. Functional gradient alterations of the social brain following mental training. A-C) *Left panel (grey):* Embedding of task-based networks along G1-G3. *upper panel*: Changes in G1-G3 respectively, arrows in the panel next to the color-boxed indicate increase of decrease; *lower panel*: visualization of task-based network-specific changes in each gradient. Colors indicate the training module (yellow=*Presence*; red=*Affect*; green=*Perspective*) and *retest controls* (light blue). Whole brain findings are corrected for multiple comparisons at $p_{FWE}<0.005$.

Associations between functional network reconfiguration and behavioral change

Last, we evaluated whether the above alterations in functional eccentricity following mental training could predict behavioral changes following the *Resource* training. We first replicated previous observations of module-specific behavioral change (66) in the current sample with fMRI data available. Attention scores increased during *Presence* relative to the other modules ($t=3.826$, $p<0.001$) and to RCC ($t=3.209$, $p<0.001$). Compassion scores increased following *Affect* relative to the other modules ($t=3.305$, $p<0.001$) and to RCC ($t=3.344$, $p<0.001$). Finally, ToM scores were marginally increased during *Perspective* relative to the other modules ($t=1.618$, $p=0.053$) and RCC ($t=1.562$, $p=0.059$).

Supervised learning (elastic net, 10 components, 5-fold cross validation, 5000 repetitions (28)) was utilized to predict behavioral change from the three gradients combined. G1-G3 combined could predict compassion (prediction – outcome $r=0.20$, mean absolute error (MAE): 2.44, $p=0.046$) and ToM scores ($r=0.29$, MAE:2.36, $p=0.005$) change. Binning change scores in 10 bins improved predictions relative to using the raw change scores in both compassion and ToM, yet change in attention was not predictable (**Supplementary Results**). Features were selected in anterior and inferior temporal, limbic, and visual cortex, and related predominantly to G2 for compassion, particularly with lower weights on attention and higher weights on socio-cognitive networks. Conversely, features predictive of ToM change related to a dissociable pattern between attention and socio-cognitive network weights in G3.

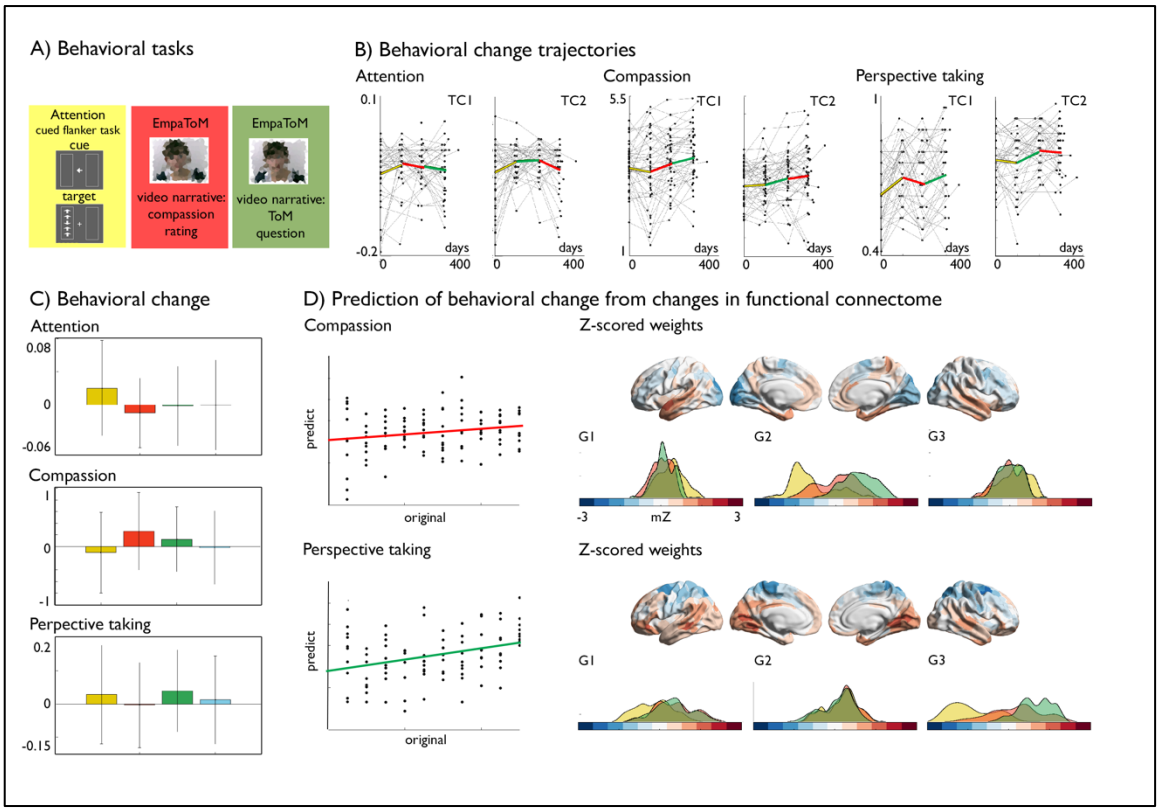

Figure 3. Association between changes in functional embedding and changes in behaviour. A). Behavioral tasks that measured attention, compassion, and perspective-taking; **B**). Behavioral measurements in TC1 and TC2 across time-points (66); **C**). Behavioral change across modules; **D**). Prediction behavior change across time-points. *left:* scatter of predicted versus original scores; *right upper:* average bootstrapped z-score; *right lower:* average bootstrapped z-score within task-based network per gradient dimension.

Robustness analyses

Various additional analysis indicated robustness of our findings (**Supplementary Results**);

- To validate our vertex-level approach, we additionally computed network-level index of segregation and integration; the *dispersion* within as well as *distance* between task-based fMRI activations in 3D gradient space (60), respectively reflective of segregation within and between networks. Using these network-level metrics we observed similar patterns of network integration and segregation following the mental training modules.
- Taking a graph theoretical approach to integration and segregation, using clustering coefficient and path length (67), we could confirm differentiable patterns of integration and segregation in task-based functional networks after *Presence* and *Perspective* modules.
- We observed similar results when performing global signal regression, indicating the observed patterns of intrinsic functional integration and segregation did not relate to global signal changes as a function of training.
- We observed no significant association with change in task-unrelated thoughts when evaluating change in post-scan thought probes after each mental training block, suggesting the contents of thought during resting state assessment did not systematically vary as a function of mental training.
- As the task-based activations within the current sample were key to our analysis, we performed extensive decoding to leverage the dissociability of the functional activations associated with attention, socio-affect, and socio-cognitive tasks in the present sample with canonical functional-networks, transcriptomic expression, and meta-analytical task space. These measures provide further contextualization of our sample specific networks, and as such help to interpret the sample and metric-specific networks. These additional analyses confirmed the observed task-based networks were consistently different across samples and approaches.

DISCUSSION

Assessing longitudinal change following mental training provided evidence for the interrelation of intrinsic functional brain organization and social and attentional functions. Particularly, our findings show functional segregation of regions in the multiple demand network (MDN) and default mode network (DMN) following *Presence* training, indicating that these networks differentiated from the rest of the cortex. Conversely, *Perspective* training resulted in their functional integration. *Affect* resulted in marked alterations of functional organization of regions in the ventral attention network (VAN) along the first gradient specifically, reflecting an increased integration of these regions with MDN and DMN networks. Second, we could predict changes in compassion and ToM scores based on co-occurring changes in large-scale functional organization, suggesting behavioral relevance of functional network alterations following mental training. In sum, our study showed that the concerted training of social skills results in dissociable changes in the functional connectome. Although our work focused on healthy adults ranging from 20 to 55 years of age, our findings overall support the possibility that targeted mental training can enhance social skills and lead to co-occurring functional network reconfigurations. These findings could inform future efforts to develop interventions aimed at cultivating the understanding of thoughts and feelings of others.

Higher cognition reflects a balance between parallel processing in an integrated architecture versus more segregated hierarchical processing (4, 32, 40, 68). Analyzing task-derived fMRI networks involved in social and attentional processes at baseline, we could show that these networks were positioned differently in a coordinate system spanned by the first three functional gradients. Echoing prior work in young adults (43), the principal gradient ran from unimodal to transmodal systems such as the DMN. This axes aligns with classic notions of cortical hierarchy (43), axes of microstructural differentiation (41, 69, 70) and cortical evolution, with heteromodal regions undergoing recent expansions in the human lineage (71, 72). In our findings, socio-affective and socio-cognitive functional activations were located at the higher end of this functional hierarchy stretching from perception/action to flexible cognition. Functions key to social cognition such as ToM are often co-localized with, if not equal to, heteromodal association systems such as the DMN (19, 73-75). The DMN is thought to integrate multiple streams of information (15, 43, 76), coordinating internal and external information (16, 32, 77, 78), and supporting predictive processing (79). Conversely, attention was associated with the middle of this axes, indicating its function to bridge sensory and integrative processing. Notably, task-based networks involved in attention, socio-affective, and socio-cognitive skills were linearly ordered along the third gradient, which dissociates the MDN from rest of the brain, with attention closest to the MDN and socio-cognitive skills closest to the DMN. Unlike the DMN, the MDN engages in preferentially in externally oriented tasks (43, 80). Recent studies have shown that the third gradient furthermore differentiates socio-episodic memory processes in DMN from task-focused processing taking place in the MDN (15, 81). The MDN, including fronto-parietal and dorsal attention networks, has been implicated in attention/executive function (19, 20). Indeed, whereas task-based activations in perspective-taking showed overlap with the DMN, task-based activation associated with attention related to MDN. Conversely, socio-affective skills were found to be in the middle of the organizational axis dissociating MDN from DMN. This positioning of socio-affective task-based activations along the third gradient between attention and ToM may be reflective of a regulative role of

functional activity of MDN and DMN (62). As such, we could show that the task-based functional activations associated with attention, socio-affect and socio-cognition each were placed at unique locations along cardinal axes of functional organization, particularly those dissociating MDN from DMN and other cortical networks.

Studying functional network plasticity following *ReSource* training, we found that regions in both the MDN and DMN integrated further following *Perspective* training, whereas they segregated following *Presence*. In the context of social cognition, integrated states might reflect active thought processes such as those required when engaging in ToM, while more segregated states may reflect automation and domain-specific function (53). Mindfulness cultivated during *Presence* may reduce habitual thought patterns and enhance momentary awareness (82) – possibly captured by functional network segregation. Corroborating these findings, *Presence* training shifted the socio-cognitive network along the principal gradient G1, indicating a segregation of this network from more perceptual/action driven modes of processing. Our observations in the domain of social skill training align with previous mental training studies in other domains; working memory task performance has been associated with decreased modularity (increased integration) whereas automation following working memory training was associated with increased modularity (segregation) of MDN and DMN (53). Second, mental training has been reported to result in reduced intra-network connectivity in the DMN, VAN and somatomotor networks, reflecting integration (83). More generally, our findings may be in line with the Global Workspace Theory, which poses that automated tasks can be performed within segregated clusters of regions, whereas those that are novel and challenging require integration (84). Importantly, previous literature has indicated that functional brain networks are dominated by stable group and individual factors (85). At the same time, functional activity is inherently dynamic, reflecting different cognitive states associated with faciliatory and inhibitory processing (54, 86). Thus, it is likely that though each functional network has its place within large-scale functional organization, changes enabled by mental training reflect more dynamic elements of functional network organization. As such, our work expands pervious findings on region-specific effects (36) by providing a system-level perspective on how the social brain adaptively reconfigures following targeted, and multifaceted interventions.

We did not observe changes in eccentricity of task-based networks following the *Affect* training module. However, core nodes of the VAN, also referred to as saliency network (87), such as the supramarginal gyrus and insula, shifted upwards along the principal gradient G1, indicating that this network may more closely affiliate with the MDN and the DMN following this module. The VAN has been implicated in empathy (18, 88), and may also contribute to the regulation of high-level cognitive processes and attention. Previous work has furthermore suggested VAN are part of the diverse club, a set of regions with high participation coefficients reflecting their diverse connectivity patterns supporting global integration (89). Its change may reflect a processing shift of the VAN towards a more integrative functional role. In addition, these regions are suggested to mediate the activity within fronto-parietal and DMN (62, 90). It is possible that the lack of change in eccentricity and increases of the VAN along G1 following socio-affective training reflects such a coordinating role (whereas its decreases following mindfulness attention result in lack of coordination and increased 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 (13, 14). Notably, though the current work focused on cortical networks, it is likely also subcortical regions contribute to the functional organization of the social brain (56, 91). Follow-up work that studies plasticity of sub-cortical function and structure in the context of the social brain will provide additional system-level insights on the social brain.

Using supervised learning with cross-validation, we found behavioral markers of social functions to be associated with training-related reorganization in functional connectome axes. In particular, we could predict both compassion and perspective-taking change during the *Affect* and *Perspective* blocks. *Affect* training nurtured compassion, prosocial motivation, and emotional regulation, as reflected in the increases in compassion scores following the *Affect* module in particular. We observed that increased weights in temporal/limbic regions following *Affect* could predict compassion change. Previous work suggests compassion/positive affect is likely emotion-motivational systems (92) but also interoceptive information (14), as supported by (para)limbic regions such as ventral anterior insula and parahippocampa regions (14, 15, 88). On the other hand, *Perspective* training focused on training perspective-taking of self and others and was associated with increases in ToM scores (66). Behavioral change in perspective-taking following *Perspective* could be predicted by positive weights in the DMN (which overlaps with the task-based ToM network), whereas the MDN (which overlaps with the task-based attention network) had decreased weights. This may reflect that increases in perspective-taking are supported by a dissociation of focussed, externally-oriented processes mediated by MDN function and DMN functions that increasingly integrate these with internally-oriented processing (15). These observations underscore the differentiation between socio-emotional and socio-cognitive processes, and provide evidence that change in compassion and perspective-taking is captured by change intrinsic functional reorganization at the level of the individual.

In sum, by combining cross-sectional with longitudinal approaches, we could show how different modules of social mental training resulted in changes in functional network organization. In line with prior work on the relationship between grey matter structure and the *ReSource* training in the same sample (39), our work shows intrinsic separability of processes underlying social functions. Notably, our work, in part, relied on within-sample task-based activations associated with attention, socio-affective, and socio-cognitive domains and previous work indicates that functional activations in tasks probing social processes are variable as a function of task operationalization (4, 32, 93). Indeed, though meta-analytical decoding showed all three tasks were associated with the probed functions based on the task-based neuroimaging literature, they additionally involved other processes such as inhibition, cognitive control and memory (4, 32, 74, 93). Moreover, task-based activations used in the current study could be associated with different transcriptomic patterns, replicating and expanding prior work linking task-based activations with transcriptomic organization (94), suggesting they may reflect differentiable neurogenetic pathways. Taken together, the current work underscores the differentiation of processes underlying our ability of understand the thoughts and feelings of ourselves and others within the functional organization of the human brain, providing a system-level perspective on social functioning.

MATERIALS AND 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, see Table 1 for more details. 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 (38)]. 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 (95) and a personal interview, SCID-II, for Axis-II disorders (96, 97). 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 (38).

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 (38).

Final sample. We excluded individuals with missing structural and/or functional MRI data and/or a framewise-displacement of >0.3mm (<5%) (98) and individuals with gradients that showed less than r=0.5 correspondence to the average gradient patterns were disregarded for analysis (<1%), as recommended (46). Further details of sample size per time-point and exclusion criteria are in **Table 1-3**.

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 (38); *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 (38); *no MRT*: due to illness / scheduling issues / discomfort in scanner; *other*: non-disclosed.

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 (24, 56). 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 center of a black screen.

Task-free functional MRI analysis. Processing was based on DPARSF/REST for Matlab [http://www.restfmri.net (99)]. 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 (98) 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 evaluating functional gradients (41, 43, 57, 60, 100, 101) the functional connectivity matrix was proportionally thresholded at 90% per row and converted into a normalised angle matrix using the BrainSpace toolbox for MATLAB (57). Diffusion map embedding, a nonlinear manifold learning technique (58), 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 (α = 0, maximal influence; α = 1, no influence). Based on previous work (43), we set α = 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 (59) via Procrustes rotations (57). The Procrustes alignment enables comparison across individual

embedding solutions, provided the original data is equivalent enough to produce comparable Euclidean spaces.

3D gradient metric: eccentricity. 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.

3D gradient metric: dispersion. To investigate multi-dimensional differences in cortical organization, we quantified dispersion of task-based networks (24, 56) and functional networks (64) 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 (60).

Graph theoretical metrics. To further validate our gradient-based metric of integration and segregation, we also computed clustering coefficient and path length using the Brain Connectivity Toolbox (67). Both measures were computed in the parcel-level, by averaging the individual connectomes within the 400 Schaefer parcels (102) and with a threshold at the top 10% of connections, analogue to our gradient approach.

Task-based functional networks. Paradigms identifying functional localizers are previously described (24, 56). 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 to evaluate their embedding within the intrinsic gradient space constructed from the resting-state fMRI data.

<u>*Behavioral markers*</u>

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 (66). The measure for compassion was based on the EmpaToM task, a developed and validated naturalistic video paradigm in the current subjects (56, 103). 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 (56) 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 (56).

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 (104). We employed linear mixed-effects models, a flexible statistical technique that allows for inclusion of multiple measurements per subjects and irregular measurement intervals (105). In all models, we controlled for age and sex, and subject. Inference was performed on subject-specific eccentricity/gradient change maps, Δeccentricity/gradient, which were generated by subtracting vertex-wise eccentricity/gradient maps of subsequent time points for a given participant.

a) 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 one training module against another (for example *Affect* versus *Perspective*). To compare a given module against RCC, we estimated contrasts for training cohort change relative to RCC (*Presence, Perspective, Affect*).

b) Correction for multiple comparisons. Whole brain findings were corrected for multiple comparisons using random field theory for non-isotropic images (106) 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 (107). Post-hoc whole cortex eccentricity/gradient change in TC1 and TC2 separately were thresholded with cluster-forming threshold at p=0.01.

c) 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 (102) and captured change within task-based 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 (98).

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 4**). We regressed out age and sex, and computed a binned change score to account for outliers. 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 5000 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 (28).

Meta-analytical decoding

We used the NeuroSynth meta-analytic database (http://www.neurosynth.org) (108) 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 (109, 110). 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)(111). 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 ($r>0.5$). 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) (112). 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.

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. **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 can be requested from the authors.

REFERENCES

1. K. N. Ochsner, M. D. Lieberman, The emergence of social cognitive neuroscience. *Am Psychol* **56**, 717-734 (2001).
2. R. I. M. Dunbar, The Social Brain hypothesis. *Evol. Anthropol.* **6**, 178-190 (1998).
3. T. Singer, The neuronal basis and ontogeny of empathy and mind reading: review of literature and implications for future research. *Neurosci Biobehav Rev* **30**, 855-863 (2006).
4. M. Schurz, L. Maliske, P. Kanske, Cross-network interactions in social cognition: A review of findings on task related brain activation and connectivity. *Cortex* https://doi.org/10.1016/j.cortex.2020.05.006 (2020).
5. C. D. Batson, "These things called empathy." in The Social Neuroscience of Empathy, J. Decety, W. Ickes, Eds. (MIT press, Cambridge, MA, 2009), pp. 16-31.
6. N. Eisenberg, R. A. Fabes, Empathy: conceptualization, measurement, and relation to prosocial behavior. . *Motivation and Emotion* **14**, 131-149 (1990).
7. F. de Vignemont, T. Singer, The empathic brain: how, when and why? *Trends Cogn Sci* **10**, 435-441 (2006).
8. C. D. Frith, U. Frith, The neural basis of mentalizing. *Neuron* **50**, 531-534 (2006).
9. R. Saxe, N. Kanwisher, People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage* **19**, 1835-1842 (2003).
10. M. Tomasello, "Joint atttention as social cognition" in Joint attentioon: Its origins and role in development, C. Moore, P. J. Dunham, Eds. (LawrenceErlbaum Associates, Inc., 1995), pp. 103-130.
11. A. D. Craig, How do you feel--now? The anterior insula and human awareness. *Nat Rev Neurosci* **10**, 59-70 (2009).
12. I. R. Kleckner *et al.*, Evidence for a Large-Scale Brain System Supporting Allostasis and Interoception in Humans. *Nat Hum Behav* **1** (2017).
13. M. M. Chun, J. D. Golomb, N. B. Turk-Browne, A taxonomy of external and internal attention. *Annu Rev Psychol* **62**, 73-101 (2011).
14. L. F. Barrett, The theory of constructed emotion: an active inference account of interoception and categorization. *Soc Cogn Affect Neurosci* **12**, 1-23 (2017).
15. A. Turnbull *et al.*, Reductions in task positive neural systems occur with the passage of time and are associated with changes in ongoing thought. *Sci Rep* **10**, 9912 (2020).
16. C. Murphy *et al.*, Modes of operation: A topographic neural gradient supporting stimulus dependent and independent cognition. *Neuroimage* **186**, 487-496 (2019).
17. M. Sormaz *et al.*, Default mode network can support the level of detail in experience during active task states. *Proc Natl Acad Sci U S A* **115**, 9318-9323 (2018).
18. C. Lamm, J. Decety, T. Singer, Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* **54**, 2492-2502 (2011).
19. D. Alcala-Lopez *et al.*, Computing the Social Brain Connectome Across Systems and States. *Cereb Cortex* **28**, 2207-2232 (2018).
20. M. Assem, M. F. Glasser, D. C. Van Essen, J. Duncan, A Domain-General Cognitive Core Defined in Multimodally Parcellated Human Cortex. *Cereb Cortex* **30**, 4361-4380 (2020).
21. T. Singer, C. Lamm, The social neuroscience of empathy. *Ann N Y Acad Sci* **1156**, 81-96 (2009).
22. T. Singer *et al.*, Empathy for pain involves the affective but not sensory components of pain. *Science* **303**, 1157-1162 (2004).
23. D. Bzdok *et al.*, Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Struct Funct* **217**, 783-796 (2012).

24. F. M. Trautwein, T. Singer, P. Kanske, Stimulus-Driven Reorienting Impairs Executive Control of Attention: Evidence for a Common Bottleneck in Anterior Insula. *Cereb Cortex* 10.1093/cercor/bhw225 (2016).
25. M. Corbetta, G. Patel, G. L. Shulman, The reorienting system of the human brain: from environment to theory of mind. *Neuron* **58**, 306-324 (2008).
26. M. Corbetta, G. L. Shulman, Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* **3**, 201-215 (2002).
27. G. Raz *et al.*, Functional connectivity dynamics during film viewing reveal common networks for different emotional experiences. *Cogn Affect Behav Neurosci* **16**, 709-723 (2016).
28. L. J. Chang, P. J. Gianaros, S. B. Manuck, A. Krishnan, T. D. Wager, A Sensitive and Specific Neural Signature for Picture-Induced Negative Affect. *PLoS Biol* **13**, e1002180 (2015).
29. S. M. Schaafsma, D. W. Pfaff, R. P. Spunt, R. Adolphs, Deconstructing and reconstructing theory of mind. *Trends Cogn Sci* **19**, 65-72 (2015).
30. L. Pessoa, Understanding brain networks and brain organization. *Phys Life Rev* **11**, 400-435 (2014).
31. J. Zaki, J. Weber, N. Bolger, K. Ochsner, The neural bases of empathic accuracy. *Proc Natl Acad Sci U S A* **106**, 11382-11387 (2009).
32. M. Schurz *et al.*, Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychol Bull* 10.1037/bul0000303 (2020).
33. L. Schilbach *et al.*, Toward a second-person neuroscience. *Behav Brain Sci* **36**, 393-414 (2013).
34. F. B. de Waal, The antiquity of empathy. *Science* **336**, 874-876 (2012).
35. S. D. Preston, F. B. de Waal, Empathy: Its ultimate and proximate bases. *Behav Brain Sci* **25**, 1-20; discussion 20-71 (2002).
36. Y. Y. Tang, B. K. Holzel, M. I. Posner, The neuroscience of mindfulness meditation. *Nat Rev Neurosci* **16**, 213-225 (2015).
37. C. R. Pernet, N. Belov, A. Delorme, A. Zammit, Mindfulness related changes in grey matter: a systematic review and meta-analysis. *Brain Imaging Behav* 10.1007/s11682-021-00453-4 (2021).
38. T. Singer *et al.*, *The ReSource Project. Background, design, samples and measurements (2nd ed)*. L. Max Planck Institute for Human Cognitive and Brain Sciences, Germany, Ed., MPI special issue in Human Cognitive and Brain Sciences (2016).
39. S. L. Valk *et al.*, Structural plasticity of the social brain: Differential change after socio-affective and cognitive mental training. *Sci Adv* **3**, e1700489 (2017).
40. M. M. Mesulam, From sensation to cognition. *Brain* **121 (Pt 6)**, 1013-1052 (1998).
41. C. Paquola *et al.*, Microstructural and functional gradients are increasingly dissociated in transmodal cortices. *PLoS Biol* **17**, e3000284 (2019).
42. J. M. Huntenburg, P. L. Bazin, D. S. Margulies, Large-Scale Gradients in Human Cortical Organization. *Trends Cogn Sci* **22**, 21-31 (2018).
43. D. S. Margulies *et al.*, Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc Natl Acad Sci U S A* **113**, 12574-12579 (2016).
44. J. M. Shine *et al.*, Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. *Nat Neurosci* **22**, 289-296 (2019).
45. K. V. Haak, A. F. Marquand, C. F. Beckmann, Connectopic mapping with resting-state fMRI. *Neuroimage* **170**, 83-94 (2018).
46. S. J. Hong *et al.*, Towards a connectivity gradient-based frameworrk for reproducible biomarker discovery. *bioRxiv* (2020).

47. K. V. Haak, C. F. Beckmann, Understanding brain organisation in the face of functional heterogeneity and functional multiplicity. *Neuroimage* **220**, 117061 (2020).
48. J. Smallwood *et al.*, The neural correlates of ongoing conscious thought. *iScience* **24**, 102132 (2021).
49. R. F. Betzel, M. Fukushima, Y. He, X. N. Zuo, O. Sporns, Dynamic fluctuations coincide with periods of high and low modularity in resting-state functional brain networks. *Neuroimage* **127**, 287-297 (2016).
50. L. F. Barrett, A. B. Satpute, Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain. *Current opinion in neurobiology* **23**, 361-372 (2013).
51. D. R. Dajani, L. Q. Uddin, Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends Neurosci* **38**, 571-578 (2015).
52. J. M. Shine *et al.*, The Dynamics of Functional Brain Networks: Integrated Network States during Cognitive Task Performance. *Neuron* **92**, 544-554 (2016).
53. K. Finc *et al.*, Dynamic reconfiguration of functional brain networks during working memory training. *Nat Commun* **11**, 2435 (2020).
54. J. M. Shine, R. A. Poldrack, Principles of dynamic network reconfiguration across diverse brain states. *Neuroimage* **180**, 396-405 (2018).
55. J. Kabat-Zinn, *Full catastrophe living: using the wisdom of your body and mind to face stress, pain, and illness.* (Delacorte, New York, NY, 1990).
56. P. Kanske, A. Bockler, F. M. Trautwein, T. Singer, 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 (2015).
57. R. Vos de Wael *et al.*, BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. *Nat Commun Biology* (2020).
58. R. R. Coifman *et al.*, 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 (2005).
59. D. C. Van Essen *et al.*, The WU-Minn Human Connectome Project: an overview. *Neuroimage* **80**, 62-79 (2013).
60. R. A. I. Bethlehem *et al.*, Dispersion of functional gradients across the lifespan. *NeuroImage* (2020).
61. B. Park *et al.*, Macroscale connectome manifold expansion in adolescence. *bioRxiv* (2020).
62. V. Menon, L. Q. Uddin, Saliency, switching, attention and control: a network model of insula function. *Brain Struct Funct* **214**, 655-667 (2010).
63. A. M. Dale, B. Fischl, M. I. Sereno, Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179-194 (1999).
64. B. T. Yeo *et al.*, The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol* **106**, 1125-1165 (2011).
65. J. Duncan, The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn Sci* **14**, 172-179 (2010).
66. F. M. Trautwein, P. Kanske, A. Bockler, T. Singer, Differential benefits of mental training types for attention, compassion, and theory of mind. *Cognition* **194**, 104039 (2020).
67. M. Rubinov, O. Sporns, Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**, 1059-1069 (2010).
68. M. A. Bertolero, B. T. Yeo, M. D'Esposito, The modular and integrative functional architecture of the human brain. *Proc Natl Acad Sci U S A* **112**, E6798-6807 (2015).
69. J. M. Huntenburg *et al.*, A Systematic Relationship Between Functional Connectivity and Intracortical Myelin in the Human Cerebral Cortex. *Cereb Cortex* **27**, 981-997 (2017).

70. C. Paquola, R. A. I. Bethlehem, B. C. Bernhardt, A moment of change: shifts in myeloarchitecture characterise adolescent development of cortical gradients. *eLife* (2020).
71. T. Xu *et al.*, Cross-species Functional Alignment Reveals Evolutionary Hierarchy Within the Connectome. *Neuroimage* (2020).
72. S. L. Valk *et al.*, Shaping Brain Structure: Genetic and Phylogenetic Axes of Macro Scale Organization of Cortical Thickness. *Sci Adv* (2020).
73. P. N. Alves *et al.*, An improved neuroanatomical model of the default-mode network reconciles previous neuroimaging and neuropathological findings. *Commun Biol* **2**, 370 (2019).
74. D. Alcala-Lopez, K. Vogeley, F. Binkofski, D. Bzdok, Building blocks of social cognition: Mirror, mentalize, share? *Cortex* **118**, 4-18 (2019).
75. J. M. Kernbach *et al.*, Subspecialization within default mode nodes characterized in 10,000 UK Biobank participants. *Proc Natl Acad Sci U S A* **115**, 12295-12300 (2018).
76. C. Paquola *et al.*, Convergence of cortical types and functional motifs in the human mesiotemporal lobe. *Elife* **9** (2020).
77. U. Frith, C. D. Frith, Development and neurophysiology of mentalizing. *Philos Trans R Soc Lond B Biol Sci* **358**, 459-473 (2003).
78. C. Murphy *et al.*, Distant from input: Evidence of regions within the default mode network supporting perceptually-decoupled and conceptually-guided cognition. *Neuroimage* **171**, 393-401 (2018).
79. L. Chanes, L. F. Barrett, Redefining the Role of Limbic Areas in Cortical Processing. *Trends Cogn Sci* **20**, 96-106 (2016).
80. R. L. Buckner, J. R. Andrews-Hanna, D. L. Schacter, The brain's default network: anatomy, function, and relevance to disease. *Ann N Y Acad Sci* **1124**, 1-38 (2008).
81. A. Turnbull *et al.*, Left dorsolateral prefrontal cortex supports context-dependent prioritisation of off-task thought. *Nat Commun* **10**, 3816 (2019).
82. A. Lutz, J. Mattout, G. Pagnoni, The epistemic and pragmatic value of non-action: a predictive coding perspective on meditation. *Curr Opin Psychol* **28**, 166-171 (2019).
83. F. A. Cotier, R. Zhang, T. M. C. Lee, A longitudinal study of the effect of short-term meditation training on functional network organization of the aging brain. *Sci Rep* **7**, 598 (2017).
84. S. Dehaene, M. Kerszberg, J. P. Changeux, A neuronal model of a global workspace in effortful cognitive tasks. *Proc Natl Acad Sci U S A* **95**, 14529-14534 (1998).
85. C. Gratton *et al.*, Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation. *Neuron* **98**, 439-452 e435 (2018).
86. J. M. Shine, M. J. Aburn, M. Breakspear, R. A. Poldrack, The modulation of neural gain facilitates a transition between functional segregation and integration in the brain. *Elife* **7** (2018).
87. L. Q. Uddin, Salience processing and insular cortical function and dysfunction. *Nat Rev Neurosci* **16**, 55-61 (2015).
88. A. Turnbull *et al.*, The ebb and flow of attention: Between-subject variation in intrinsic connectivity and cognition associated with the dynamics of ongoing experience. *Neuroimage* **185**, 286-299 (2019).
89. M. A. Bertolero, B. T. T. Yeo, M. D'Esposito, The diverse club. *Nat Commun* **8**, 1277 (2017).
90. V. Menon, "Salience Network" in Brain Mapping: An Encyclopedic Reference, A. W. Toga, Ed. (Academic Press: Elsevier, 2015), vol. 2, pp. 597-611.
91. J. M. Shine, The thalamus integrates the macrosystems of the brain to facilitate complex, adaptive brain network dynamics. *Prog Neurobiol* 10.1016/j.pneurobio.2020.101951, 101951 (2020).

92. O. M. Klimecki, S. Leiberg, M. Ricard, T. Singer, Differential pattern of functional brain plasticity after compassion and empathy training. *Soc Cogn Affect Neurosci* **9**, 873-879 (2014).
93. M. Schurz, J. Radua, M. Aichhorn, F. Richlan, J. Perner, Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neurosci Biobehav Rev* **42**, 9-34 (2014).
94. J. Y. Hansen *et al.*, Mapping gene transcription and neurocognition across human neocortex. *Nat Hum Behav* 10.1038/s41562-021-01082-z (2021).
95. H. U. Wittchen, H. Pfister, *Diagnostisches Expertensystem für psychische Störungen (DIA-X)* (Swets & Zeitlinger, Frankfurt, 1997).
96. H. U. Wittchen, M. Zaudig, T. Fydrich, *SKID-Strukturiertes Klinisches Interview für DSM-IV. Achse I und II* (Hogrefe, Göttingen, 1997).
97. M. B. First, M. Gibbon, R. L. Spitzer, J. B. W. Williams, L. S. Benjamin, *Structured Clinical Interview for DSM-IV Axis II Personality Disorders, (SCID-II).* (American psychiatric Press, Inc., Washington, D.C., 1997).
98. J. D. Power, K. A. Barnes, A. Z. Snyder, B. L. Schlaggar, S. E. Petersen, Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **59**, 2142-2152 (2012).
99. C. Song *et al.*, Relating inter-individual differences in metacognitive performance on different perceptual tasks. *Conscious Cogn* **20**, 1787-1792 (2011).
100. S. J. Hong *et al.*, Atypical functional connectome hierarchy in autism. *Nat Commun* **10**, 1022 (2019).
101. R. Vos de Wael *et al.*, Anatomical and microstructural determinants of hippocampal subfield functional connectome embedding. *Proc Natl Acad Sci U S A* **115**, 10154-10159 (2018).
102. A. Schaefer *et al.*, Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb Cortex* **28**, 3095-3114 (2018).
103. M. G. Tholen, F. M. Trautwein, A. Böckler, T. Singer, P. Kanske, Functional magnetic resonance imaging (fMRI) item analysis of empathy and theory of mind. *Hum Brain Mapp* (2020).
104. K. Worsley *et al.*, 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** (2009).
105. J. C. Pinheiro, D. Bates, *Mixed-effect models in S and S-PLUS* (Springer, 2000).
106. K. J. Worsley, M. Andermann, T. Koulis, D. MacDonald, A. C. Evans, Detecting changes in nonisotropic images. *Hum Brain Mapp* **8**, 98-101 (1999).
107. Y. Benjamini, Y. Hochberg, 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 (1995).
108. T. Yarkoni, R. A. Poldrack, T. E. Nichols, D. C. Van Essen, T. D. Wager, Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods* **8**, 665-670 (2011).
109. M. J. Hawrylycz *et al.*, An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391-399 (2012).
110. K. J. Gorgolewski *et al.* (2014) Tight fitting genes: finding relations between statistical maps and gene expression patterns. in *OHBM 2014*.
111. A. Arnatkeviciute, B. D. Fulcher, A. Fornito, A practical guide to linking brain-wide gene expression and neuroimaging data. *Neuroimage* **189**, 353-367 (2019).
112. J. D. Dougherty, E. F. Schmidt, M. Nakajima, N. Heintz, Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. *Nucleic Acids Res* **38**, 4218-4230 (2010).

FUNCTIONAL NETWORK PLASTICITY OF THE HUMAN SOCIAL BRAIN

SUPPLEMENT

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

SUPPLEMENTARY MATERIALS AND METHODS

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 (**Figure 1**) (*1*). 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 modules were always first in order to stabilize the mind and prepare the participants for the socio-cognitive and socio-affective training modules.

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 (*2*). 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 (*3*) and the 'Affect Dyad' (*1, 4*). 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 (*5*), forgiveness meditation, and development of self-compassion (*6*).

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 (*7, 8*) and on theoretical accounts distinguishing between affective (*e.g.,* compassion and empathy) and cognitive (*e.g.,* Theory of Mind, ToM) routes of social cognition (*9, 10*). 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 (*7*). 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 (*1, 11*).

SUPPLEMENTARY RESULTS

Task-based activation patterns in attention, socio-affect, and socio-cognitive function.

Assessing task-based activation patterns in attention, socio-affect, and socio-cognition. (**Figure 1**), we observed task-specific and task-general activations between attention and social functioning. Attention processing (*i.e.,* the overlap between stimulus driven reorienting and executive control in the cued-flanker task) was associated with functional activations in bilateral inferior parietal and superior frontal regions. On the other hand, socio-affective processing (*i.e.,* the contrast of viewing emotional and neutral videos in the EmpaTom task), implicated bilateral supramarginal, dorsolateral prefrontal and bilateral inferior frontal regions. ToM (*i.e.,* the contrast of ToM versus no-ToM questions in the EmpaToM) related to activations in bilateral temporal and temporo-parietal, as well as medial prefrontal regions. Attention and socio-affective processing both implicated the bilateral anterior insula and anterior midline regions, and socio-affective and socio-cognitive processing both related to temporal parietal, midline parietal and prefrontal, as well as inferior frontal regions, although the peaks of the activation pattern in these broader brain areas clearly differ for the socio-affective and socio-cognitive domains (see also (*12*)). Our findings are thus in line with previous work and demonstrate differentiable, but overlapping networks for attention, socio-emotional, and socio-cognitive activations (*13*).

Leveraging eccentricity against graph theoretical measures

To further validate our gradient-based model of eccentricity (segregation and integration) we also created metrics of regional integration and segregation using graph theoretical approaches (**Figure S1**). We found a positive correspondence between gradient-based eccentricity and graph theory-based clustering coefficient ($r=0.641$, $p<0.001$) and path length ($r=0.510$, $p<0.001$). Evaluating network-specific alterations following mental training these measures, we found increases of clustering coefficient (attention network: $t=2.489$, $p=0.021$) and path length (socio-cognitive network: $t=2.374$, $p=0.027$) following *Presence* training and decreases in both measures (clustering coefficient: socio-cognitive network: $t=-2.629$, $p=0.015$ and path length attention network: $t=-2.43$, $p=0.03$, socio-affective network: $t=-2.843$, $p=0.01$, socio-cognitive network: $t=-3.335$, $p=0.003$).

Eccentricity change in each training cohort

Evaluating the alterations in each training cohort separately we observed similar effects (**Figure S2**). Specifically, in TC1 we found increases in eccentricity of left TPJ (FWEp=0.039) following *Presence* and decreases following *Perspective* (FWEp=0.005). Studying the changes in task-based networks, we observed decreases in eccentricity following *Perspective* in the socio-cognitive task-based network ($t=-2.711$, $p=0.015$). For TC2, we found increases in eccentricity in right opercular regions following *Presence* (FWEp=0.006) and decreases in right inferior temporal/occipital regions following *Perspective* (FWEp=0.01).

Evaluating change against retest controls (**Figure S3**), we found an increase of distance between attention and socio-cognitive networks following *Presence* ($t=2.04$, $p=0.02$) and a decrease of eccentricity following *Perspective* in task-based socio-cognitive network ($t=-2.084$,

p=0.02) at uncorrected thresholds. Last, evaluating change between two modules, we observed predominantly difference between *Presence* and *Perspective* particularly in attention networks (t=2.725, p=0.015).

Global signal corrected analyses

Correcting the functional connectivity matrices for global signal (GSR) before computing gradient and eccentricity measures gave similar results as without GSR (**Figure S4**). Specifically, along G1 we found a difference between attention and socio-affect network (t=-5.228, p<0.001) and between the attention and socio-cognitive network (t=-5.63, p<0.001). Along G2 we observed a difference between socio-affect and socio-cognitive network (t=-2.813, p=0.015). Along G3 we observed a difference between all networks; attention versus socio-affect (t=-6.353, p<0.001), attention versus socio-cognition (t=-9.283, p<0.001), and socio-affect and socio-cognition (t=-2.96, p=0.01). Testing alterations following training, we observed that eccentricity increases in attention task-based networks following *Presence* (t=2.232, p=0.04), and decreased in attention task-based network (t=-2.50, p=0.015) following *Perspective.* In individual gradients we found that, along G1, the attention network decreased following *Presence* (t=-2.157, p=0.048). For G2, *Presence* resulted in a decrease of the socio-cognitive network (t=-2.918, p=0.006).

Social connections with eccentricity of functional communities.

Evaluating longitudinal change within large-scale functional communities (*14*) (**Figure S5**), we found that *Presence* training resulted in increased eccentricity of dorsal attention networks (t=3.01, p=0.01). Conversely, *Perspective* training resulted in decreased embedding of dorsal attention networks (t=-3.18, p<0.001) and ventral attention network (t=-2.69, p=0.028). Studying change in the individual large-scale gradients (G1-G3), we observed decreases in the ventral attention network following *Presence* (t=-3.18, p=0.006) but increased following *Affect* (t=2.983, p=0.015) along G1. Last we observed increase of the visual network following *Presence* (t=3.51, p=0.0015) but decrease following *Affect* (t=-4.0193, p<0.001) along G3.

Change in first three large-scale gradients in each training cohort

Reiterating analysis of **Figure 3** in each training cohort separately (**Figure S6** and **Figure S7**), we observed decreases in G1 following *Presence* in right superior parietal cortex extending to supramarginal gyrus (FWEp<0.001) and right superior frontal cortex (FWEp=0.026) in TC1. Conversely, *Affect* resulted in increases in G1 in right superior parietal cortex extending to supramarginal gyrus (FWEp=0.01) in TC1. In TC2, we observed decreases along G1 in left opercular regions (FWEp=0.04) following *Presence* and increases in bilateral mid/anterior insula (left: FWEp=0.004, right: FWEp=0.002) and right mid cingulate (FWEp<0.001) following *Affect*. No changes in task-based networks were significant. We observed no whole cortex alterations along G2 in TC1 or TC2, however we found increases in G2 within socio-affect task-based network (t=2.518 p=0.02) following *Presence* and decreases (t=-2.719 p=0.01) following *Affect* in TC2. Along G3 we found decreases in TC2 in left occipital regions following *Affect* (FWEp<0.001) and decreases in left superior frontal regions following

Perspective (FWEp=0.01) in TC2. Studying change along G1-G3 by comparing only two training modules, we found a difference between *Presence* and *Affect* along G2 in the attention network (t=2.198, p=0.03) and socio-affective network (t=2.339, p=0.03).

Behavior change prediction

Attention was not predictable using a multi-gradient approach (raw: r=-0.21, p=0.06; bin: r=-0.04, p>0.1). Compassion and perspective-taking were less predictable using raw scores (compassion raw: r=0.19, p=0.06, MAE=0.504; perspective-taking raw: r=0.20, p=0.05, MAE=0.099), yet distribution of weight maps was similar (**Figure S8**).

Functional decoding of localizer networks

Decoding task-based activations using large-scale functional communities (*14*) we found that task-based activations associated with attention were predominantly in dorsal and ventral attention networks and fronto-parietal networks (**Figure S5**). Conversely, task-based activations associated with socio-affective processing related to frontal parietal networks as well as default mode network and ventral attention network. ToM-based activations were associated with default mode network, and ventral attention and somato-sensory network.

To further decode the association between task-based activations of social processing in the current sample we took an ad-hoc a meta-analytic approach using meta-analytic maps of social, cognitive, affective, and attentional terms in the Neurosynth database (**Figure S9**). We found differential relationships between each of the functional localizers and meta-analytical maps (FDRq<0.05). Task-based activation of attention was associated with inhibition, executive functioning, cognitive control, attention, and working memory. Conversely, task-based activation of socio-affective functioning was associated with affect, goal-directed behavior, episodic memory, executive functioning, emotion, social-interaction, cognitive control, emotion regulation, inhibition, self, empathy, social, ToM, and mentalizing. Socio-cognitive task-based activations related to episodic memory, semantic memory, emotion, emotion regulation, empathy, social interaction, self, mentalizing, social, and ToM – meta-analytical maps.

Genetic decoding of the social brain

Having functionally decoded task-based activations associated with attention and social processing, we sought further understand the factors that contribute to the dissociable activation patterns of social functioning (**Figure S10, Tables S1-3**)

Specifically, we correlated the task-based activation during attention, socio-affective and socio-cognitive tasks with post-mortem gene expression data from the Allen Institute of Brain Sciences (AIBS) (*15*). Among significantly associated genes, we selected only those consistently expressed (r>0.5) among donors. We then used developmental gene set enrichment analysis (*16*), to identify spatiotemporal time windows in which these genes are most frequently expressed. The analysis suggested a differential patterning between attention, socio-affective, and socio-cognitive processes. Activations associated with reorienting and executive control of

attention were predominantly in cerebellum and thalamus during childhood and adolescence, and showed a positive increase along development (r=0.797, p=0.006), whereas socio-cognitive activations were associated with striatal expression in adolescence and early adulthood, and did not show association with development (r=-0.497, p>0.1). Socio-affective activations were associated with both thalamus in early fetal and adolescent time windows, as well as cerebellum across development, and had a negative relationship with developmental expression (r=-0.86, p=0.002).

Network-based metrics of eccentricity change

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 (*17*) (**Supplementary Figure 11**). Using these metrics to evaluate change we found a decrease in within-network distance in the attention network (t=-2.31, p=0.045) and socio-cognitive network (t=-2.25, p=0.03), and increased between network distance between attention and socio-affective (t=2.81, p=0.015) and attention and socio-cognitive (t=2.46, p=0.03) networks following *Presence.* Conversely, *Perspective* resulted in a decrease in between network distance between attention and socio-affective network (t=-2.42, p=0.03).

SUPPLEMENTARY FIGURES

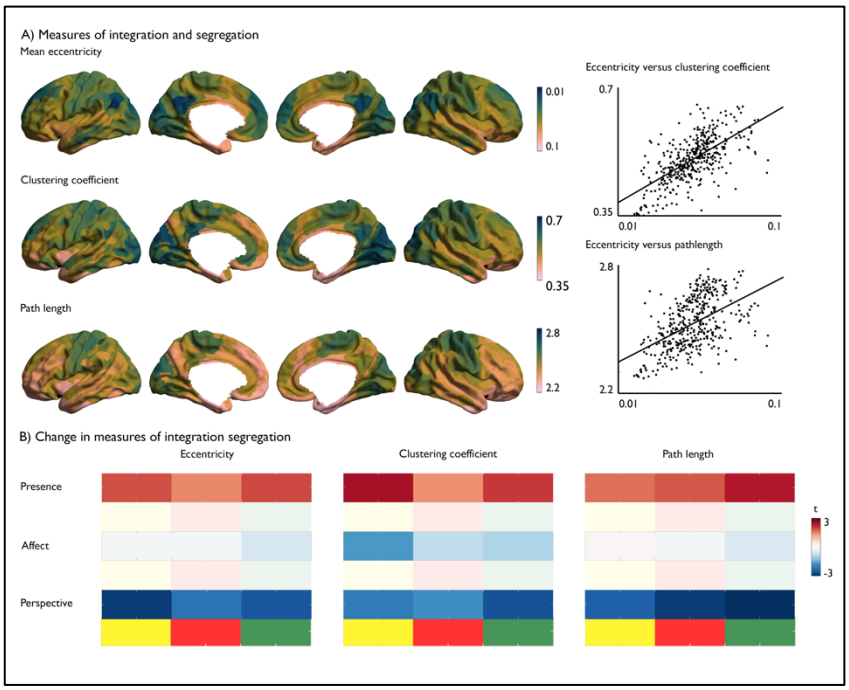

Figure S1. Comparison of gradient-based eccentricity measure with graph theory-based measures of integration and segregation. A) Measures of integration and segregation; B) Change in difference measures of integration and segregation (clustering coefficient and path length). Colors reflect the relevant task-based network (yellow is attention, red is socio-affect and green is socio-cognition).

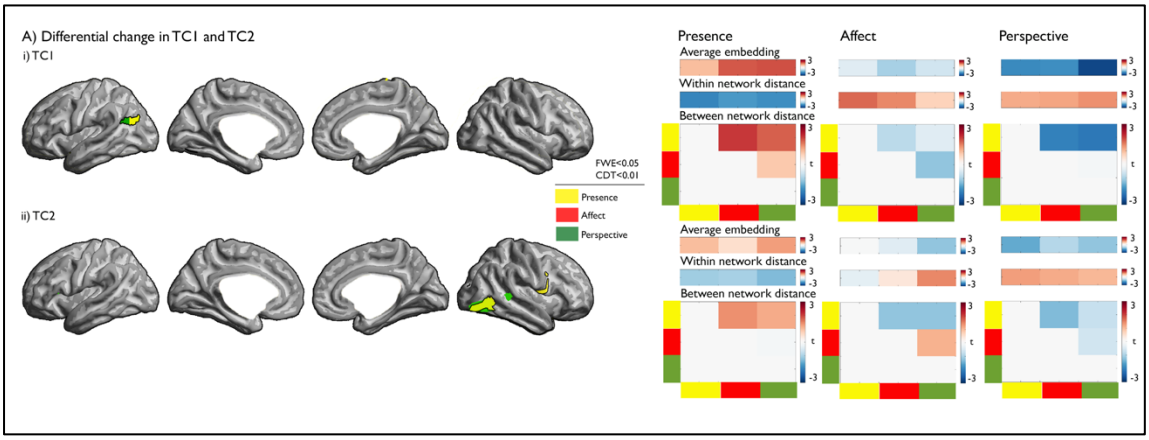

Figure S2. Module specific change in eccentricity in TC1 and TC2 separately. Whole brain differential change of manifold embedding in training cohort 1 (TC1) (i) and training cohort 2 (TC2) (ii), FWEp<0.01, as well as the respective change with task-based functional modules in following each training versus the other two trainings

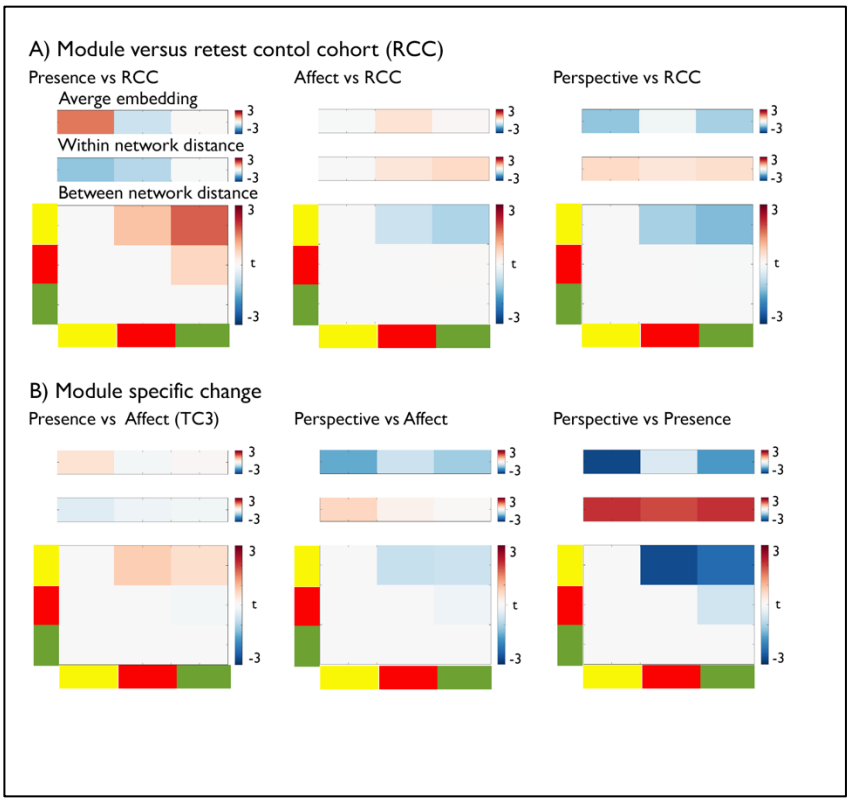

Figure S3. Change versus RCC and across modules. Colors reflect the task-based networks, for distribution of data in raincloud plots see Figure 1.

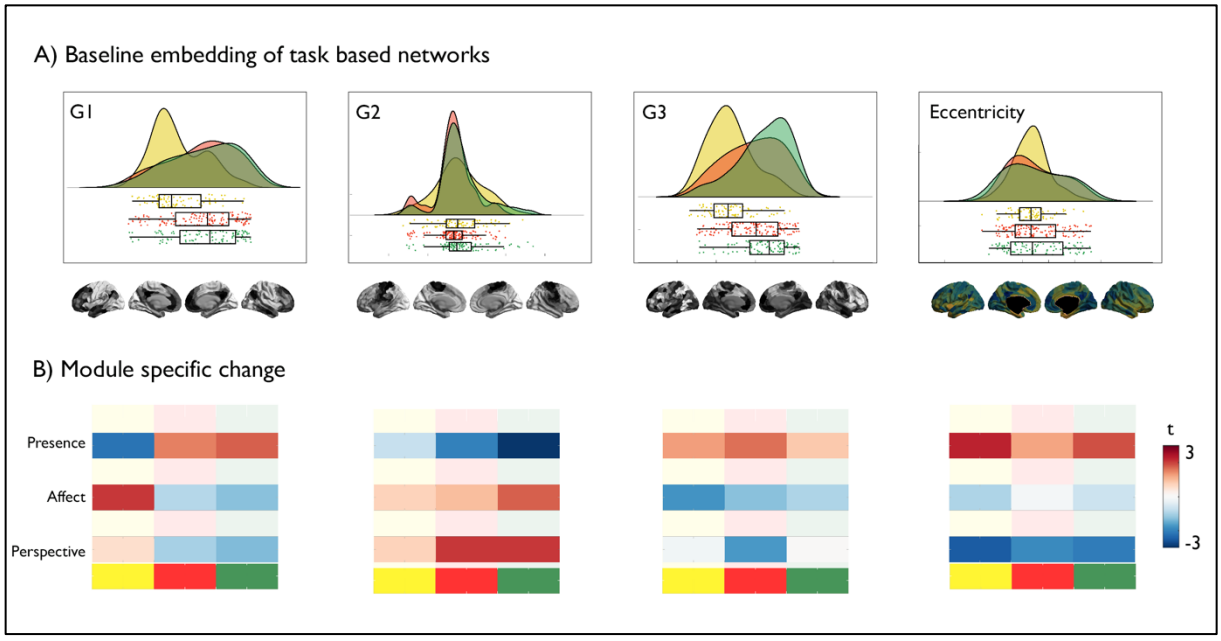

Figure S4. Large-scale organization of social brain and plasticity following training when controlling for GSR. Colors reflect the task-based networks.

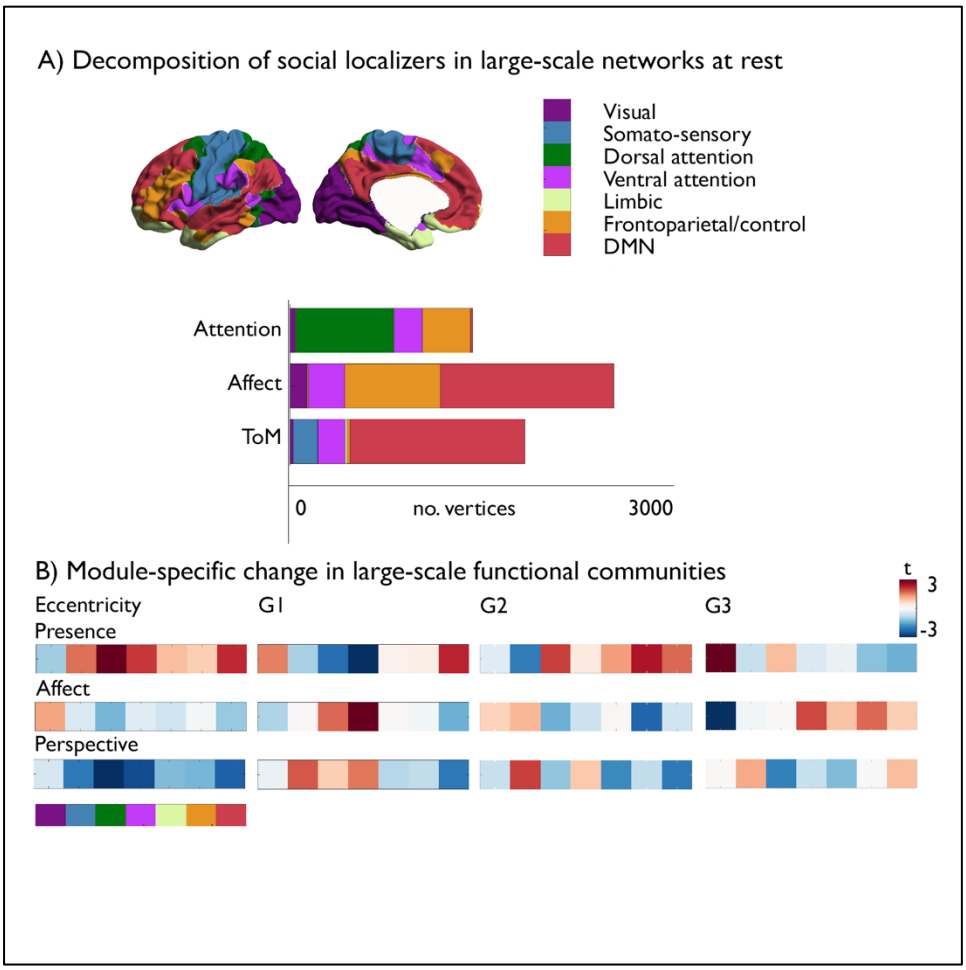

Figure S5. Functional community decoding. A) Decomposition of task-based functional activations into large-scale functional community's at rest (*14*); **B**) Association between the change in embedding, and the first three gradients, of large-scale functional communities (*14*) following mental training.

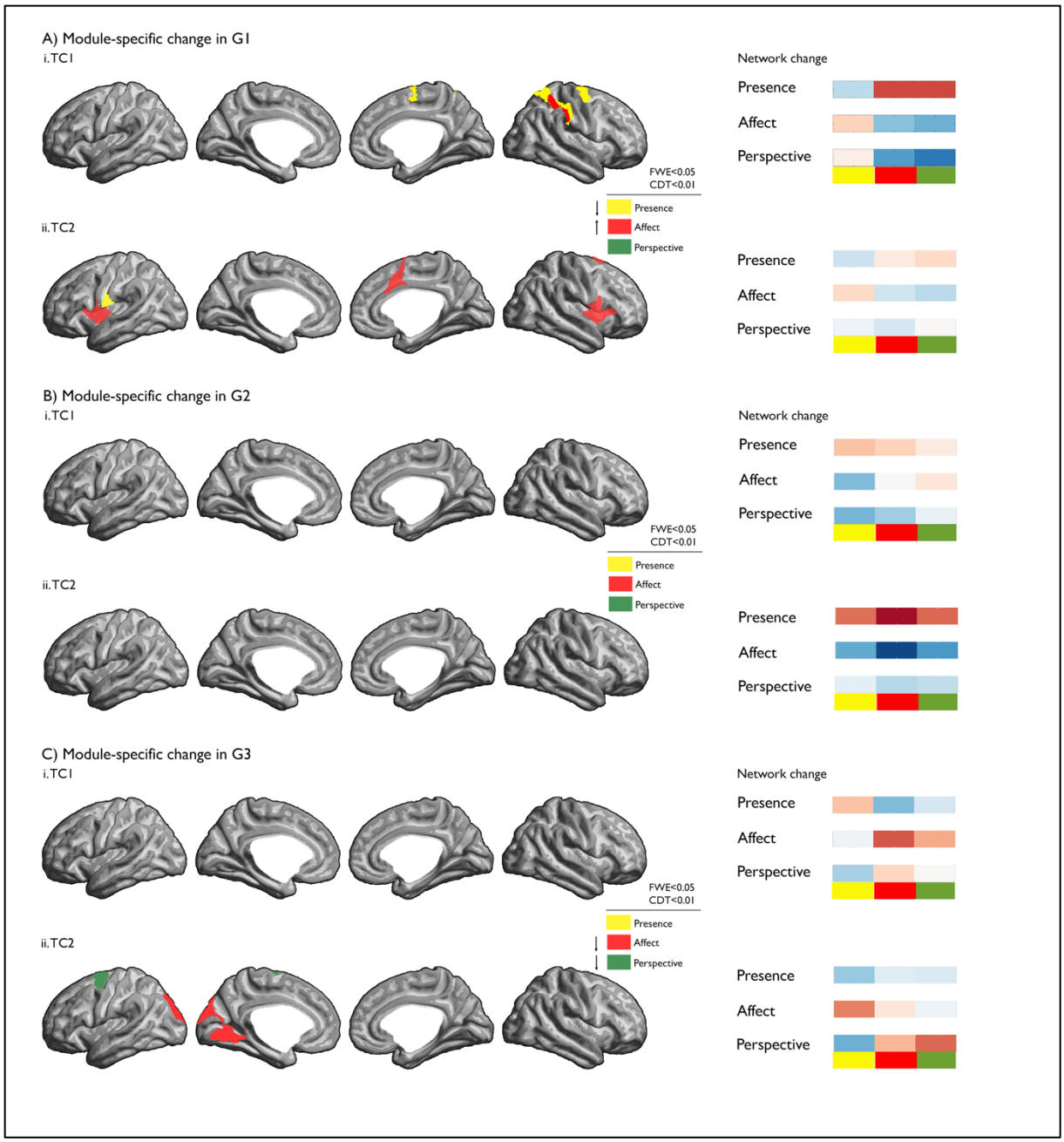

Figure S6. Gradient changes in TC1 and TC2 separately. Change in G1-G3 in training cohort 1 and 2, evaluated using a cortex-wide analysis, as well as within task-based networks.

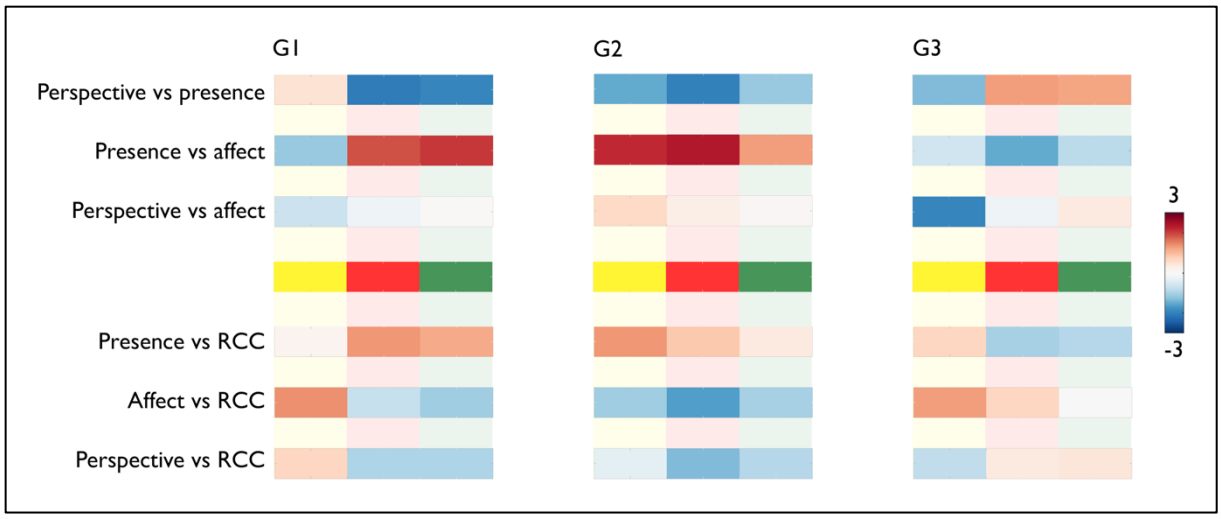

Figure S7 Change in G1-G3 vs other module and retest controls. Colors (yellow, red, green) reflect task-based network (attention, socio-affect and socio-cognition respectively), whereas red-blue ([-3 3]) indicates the t-value associated with change in each task-based network.

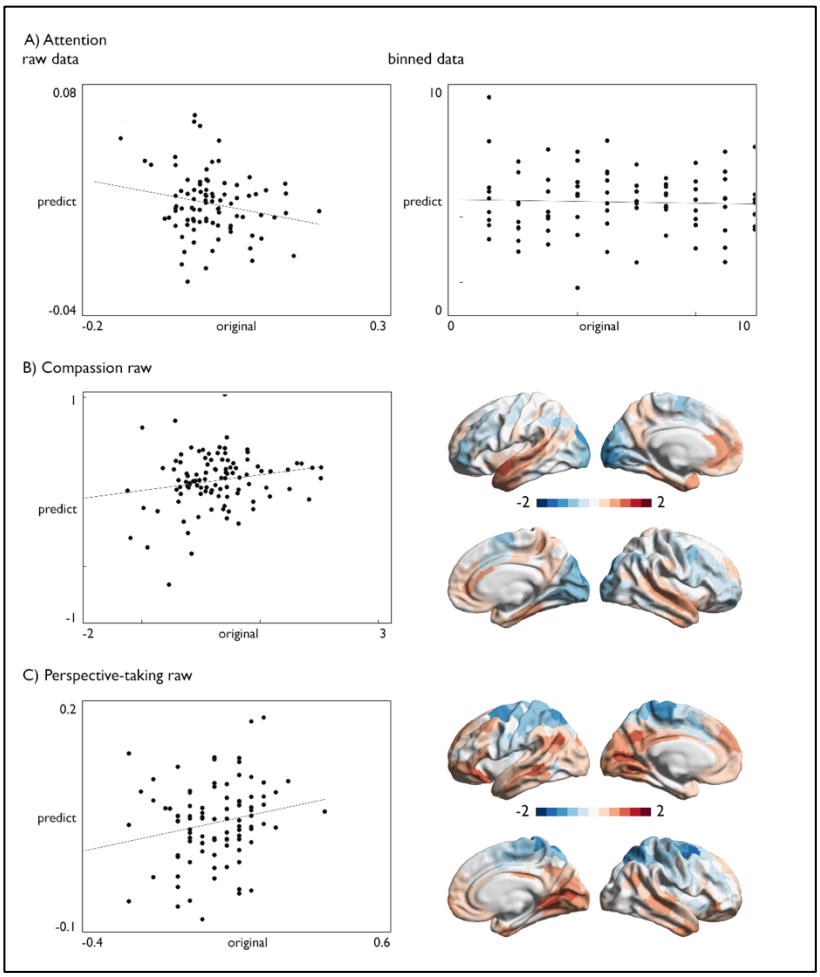

Figure S8. Predicting attention, compassion and ToM. A) Prediction of attention using raw and binned scores; **B**) Prediction of compassion and perspective-taking using raw change scores, as well as the associated weight maps (z-scored weights between -2 2 are visualized).

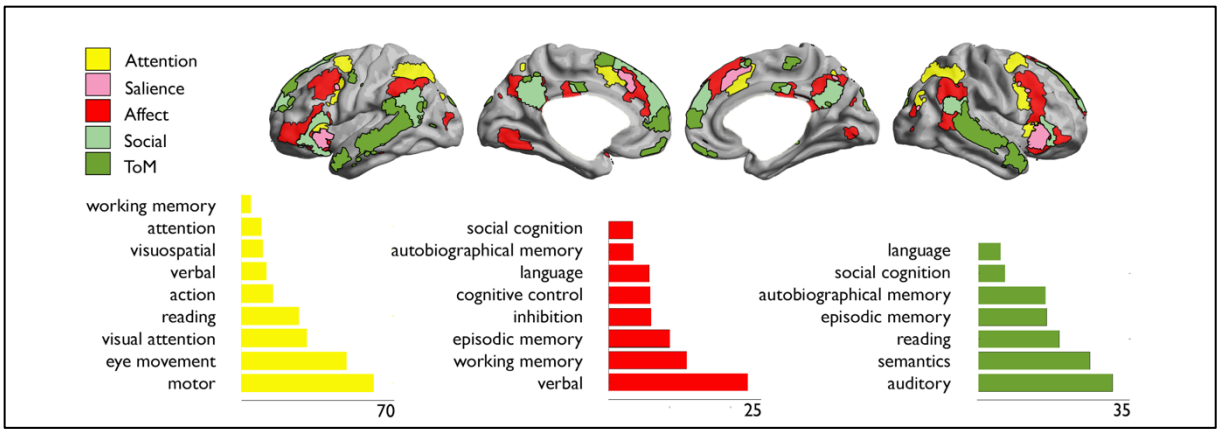

Figure S9. Functional decoding of task-based activations associated with attention, using 20 meta-analytical topic terms of Neurosynth reflecting social, affective, cognitive, and attentional functioning. Significant associations and associated t-values are represented in the bar plot, where yellow is attention, red is socio-affective, and green is socio-cognitive task-based activation maps.

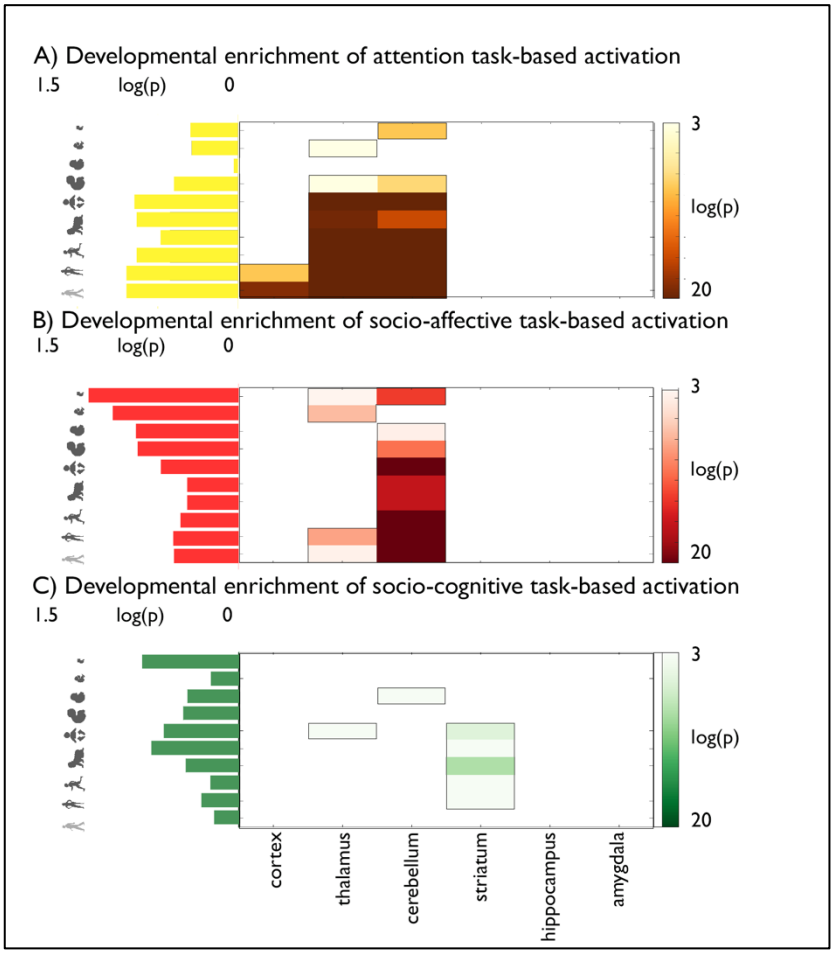

Figure S10. Developmental enrichment for each of the functional activations in attention, socio-affective, and socio-cognitive tasks at FDRp<0.05. The bar plot on the represents the log transformed p-values that averaged across all brain structures that reported in the enrichment analysis.

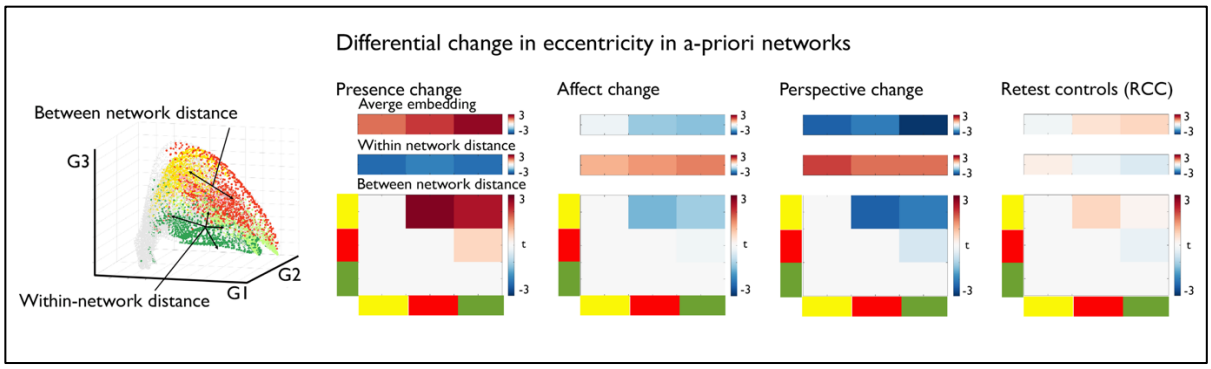

Figure S11. Differential change in eccentricity using network-based approaches. Average embedding recapitulates the average change score, within-network distance describes the change within each task-based network (yellow=attention, red=socio-affect, green=socio-cognition), between-network distance describes the change between task-based network. Effects are displayed as colored t-values [-3 3].

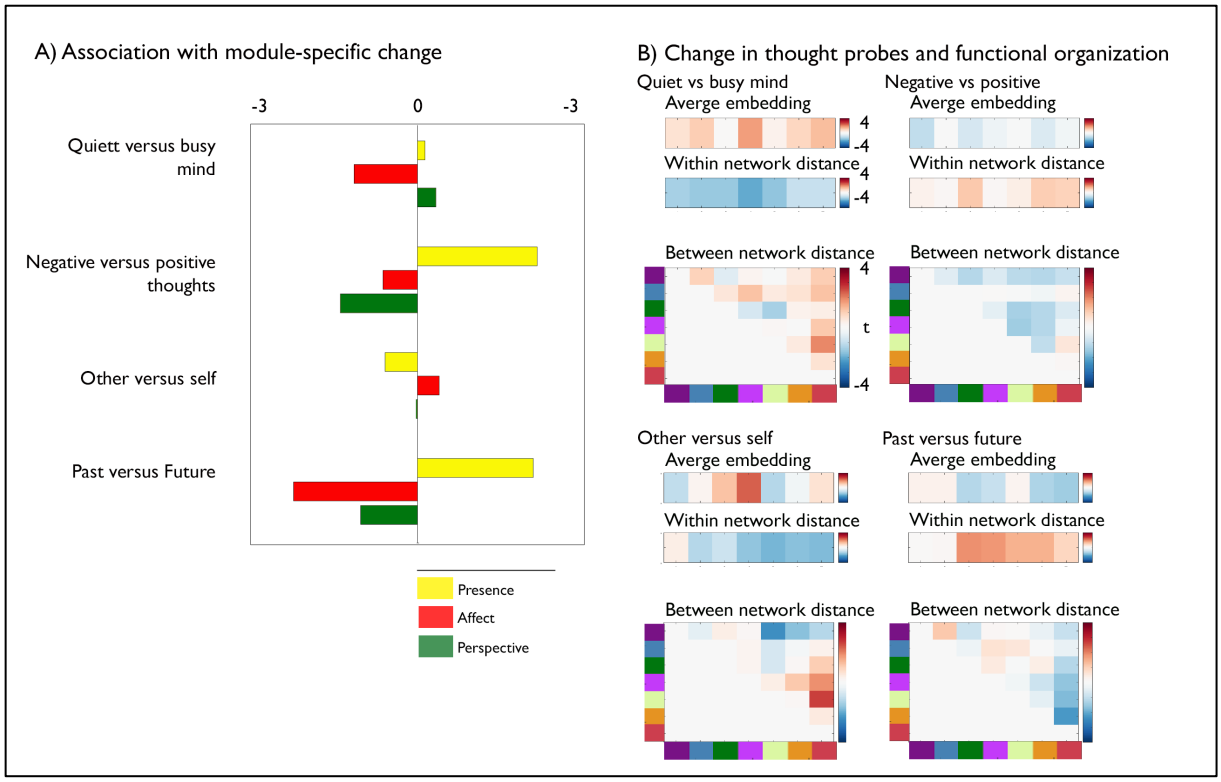

Figure S12 Association between change in task-unrelated thoughts and functional embedding. Change in task-related thought probes per module, and association with functional organization.

SUPPLEMENTARY TABLES

ACAN	DCUN1D2	GPX3	MAP3K13	PVALB	SULF1
ACYP2	DENND2D	GSTT1	MAP9	QRFPR	SV2C
ADM	DEXI	GUCA2B	MCF2	RAB37	SYCP2
ALDH1A3	DNAJC4	HAPLN4	MFGE8	RAD54B	SYT2
AMDHD1	DSCC1	HR	MGST2	RAMP3	TFAM
ANK1	ECSIT	HS3ST1	MIR31HG	RCAN2	THAP10
ANKRD34C	EFNA5	HS3ST5	MREG	RELL2	TIFA
AR	EIF5A2	IGDCC3	MX1	RET	TPTE2P6
ARHGAP9	ELMO3	INA	MYH7B	RFPL1	TRAK2
ASB13	EPB41	INTS8	NEFH	RHOBTB2	TRPC3
ATG4D	EPN3	IPW	NR1D2	RILP	TTC39A
AVPI1	ERRFI1	IRS1	NR3C1	RORA	TTC39B
BCAT1	ESRRG	ISCU	NRIP2	RPP25	UPP1
BTBD17	EXTL2	KCNA2	NT5M	RSRC1	VAV3
C17orf75	FAM20A	KCNAB3	NXPH3	SAP30L	
C3orf18	FAM71E1	KCNC1	ONECUT1	SCN1A	
C6orf106	FBLN7	KCNC3	ONECUT2	SCN1B	
CACNA2D2	FBXO9	KCTD9	OPRM1	SCRT1	
CAMK2D	FER1L4	KNG1	P2RX6	SDSL	
CAMK2G	FES	KRT31	PAG1	SEMA7A	
CCNI	FGF18	LAG3	PCDH12	SHD	
CDH7	FGF9	LAMA2	PCP4	SIX4	
CDR2L	FLT3	LAPTM4B	PGK1	SLC16A6	
CDS1	FNDC5	LCP2	PHLDA2	SLC17A6	
CERK	FRAT1	LGALS1	PHYH	SLC25A5	
CITED2	FSTL1	LINC00473	PLCB4	SLC38A1	
CLEC2L	FSTL5	LINC00515	PLCH1	SLC39A13	
CMYA5	GAL	LRCH1	PLEKHH3	SLC39A14	
CNTN6	GDPD1	LRRC38	PLXDC1	SMPX	
CPNE9	GLS2	LRRC3B	PNKD	SPAG4	
CRTAC1	GPLD1	LUZP1	PPARGC1A	SPTSSB	
CTXN3	GPR137C	LYPLA1	PPTC7	ST8SIA1	
CUX1	GPR161	MADCAM1	PSMD12	STARD10	

DCBLD2	GPR68	MAGI3	PTCHD1	STARD5	

Table S1. Genes associated with attention network activations.

ALDH1A3	CITED2	FSTL5	IPW	PAG1	SCN9A
AMDHD1	CNTN6	FZD1	KCTD12	PGAP1	SLC38A1
ANK1	DCUN1D2	GPLD1	KLK10	PLCB4	SPON2
ANKRD34C	DENND2D	GPR137C	LRCH1	PLEKHH3	ST8SIA1
ANKRD6	EPB41	GPR68	N4BP2	PLSCR4	TGFBI
CACNA2D2	FABP7	GPX3	NANOS1	PLXNC1	TMEM159
CADM1	FLT3	HIST1H1D	NGB	RAVER2	UPP1
CAMK2D	FNDC5	IGDCC3	ONECUT1	RORA	WTIP
CDH7	FRAT1	INTS8	OPRM1	SCGN	

Table S2. Genes associated with compassion network activations

BSPRY	FER1L4	LRRC49	ONECUT2	SV2C
BTBD17	HS3ST5	LRRN1	PCP4L1	TRPC3
COCH	IPW	MGST2	RBMS1	VAV3
EIF2A	ITM2A	MRPL33	S100A13	WDR6
FABP6	KLHL13	NCOA3	SLC39A13	ZNF268
FANCI	LCA5	NEB	SRGAP1	

Table S3. Genes associated with perspective-taking functional activations

SUPPLEMENTARY REFERENCES

1. T. Singer *et al.*, *The ReSource Project. Background, design, samples and measurements (2nd ed)*. L. Max Planck Institute for Human Cognitive and Brain Sciences, Germany, Ed., MPI special issue in Human Cognitive and Brain Sciences (2016).
2. J. Kabat-Zinn, *Full catastrophe living: using the wisdom of your body and mind to face stress, pain, and illness.*, (Delacorte, New York, NY, 1990).
3. D. Salzberg, *Lovingkindness. The Revolutionary Art of Happiness.*, (Shambala, Boston, MA, 1995).
4. B. E. Kok, T. Singer, Effects of Contemplative Dyads on Engagement and Perceived Social Connectedness Over 9 Months of Mental Training: A Randomized Clinical Trial. *JAMA Psychiatry*, (2016).
5. O. M. Klimecki, S. Leiberg, M. Ricard, T. Singer, Differential pattern of functional brain plasticity after compassion and empathy training. *Soc Cogn Affect Neurosci* **9**, 873-879 (2014).
6. K. D. Neff, C. K. Germer, A pilot study and randomized controlled trial of the mindful self-compassion program. *J Clin Psychol* **69**, 28-44 (2013).
7. T. Holmes, *Parts word: An illustrated guide to your inner life*. (Winged Heart Press, Kalamazoo, 2007).
8. R. C. Schwarz, *Internal familly systems therapy.*, (Guilford, New York, 1997).
9. T. Singer, The neuronal basis and ontogeny of empathy and mind reading: review of literature and implications for future research. *Neurosci Biobehav Rev* **30**, 855-863 (2006).
10. T. Singer, The past, present and future of social neuroscience: a European perspective. *Neuroimage* **61**, 437-449 (2012).
11. S. L. Valk *et al.*, Structural plasticity of the social brain: Differential change after socio-affective and cognitive mental training. *Sci Adv* **3**, e1700489 (2017).
12. P. Kanske, A. Bockler, F. M. Trautwein, T. Singer, 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 (2015).
13. M. Schurz, L. Maliske, P. Kanske, Cross-network interactions in social cognition: A review of findings on task related brain activation and connectivity. *Cortex*, (2020).
14. B. T. Yeo *et al.*, The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol* **106**, 1125-1165 (2011).
15. M. J. Hawrylycz *et al.*, An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391-399 (2012).
16. J. D. Dougherty, E. F. Schmidt, M. Nakajima, N. Heintz, Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. *Nucleic Acids Res* **38**, 4218-4230 (2010).
17. R. A. I. Bethlehem *et al.*, Dispersion of functional gradients across the lifespan. *NeuroImage* (2020).