



AFRL-AFOSR-JP-TR-2022-0043

Revealing Neurocircuitry for Prosociality in Human

Hackjin Kim
Korea University Research and Business Foundation
1-2 Anam-dong 5-ga Seongbuk-gu
Seoul, , 136713
KR

06/01/2022
Final Technical Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
Air Force Office of Scientific Research
Asian Office of Aerospace Research and Development
Unit 45002, APO AP 96338-5002

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE 20220601		2. REPORT TYPE Final		3. DATES COVERED <table style="width: 100%; border: none;"> <tr> <td style="width: 50%; border: none;">START DATE 20150928</td> <td style="width: 50%; border: none;">END DATE 20200327</td> </tr> </table>		START DATE 20150928	END DATE 20200327
START DATE 20150928	END DATE 20200327						
4. TITLE AND SUBTITLE Revealing Neurocircuitry for Prosociality in Human							
5a. CONTRACT NUMBER		5b. GRANT NUMBER FA2386-15-1-4065		5c. PROGRAM ELEMENT NUMBER 61102F			
5d. PROJECT NUMBER		5e. TASK NUMBER		5f. WORK UNIT NUMBER			
6. AUTHOR(S) Hackjin Kim							
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Korea University Research and Business Foundation 1-2 Anam-dong 5-ga Seongbuk-gu Seoul 136713 KR					8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AOARD UNIT 45002 APO AP 96338-5002				10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/AFOSR IOA	11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-AFOSR-JP-TR-2022-0043		
12. DISTRIBUTION/AVAILABILITY STATEMENT A Distribution Unlimited: PB Public Release							
13. SUPPLEMENTARY NOTES							
14. ABSTRACT This report contains the results from three studies: Study #1: Intrinsic functional and structural brain connectivity in humans predicts individual social comparison orientation (under revision) Study #2: Roles of the MPFC and Insula in Impression Management under Social Observation (under revision) Study #3: A Hierarchical Allostatic Regulation Model of Medial Prefrontal Cortex Function for Social Valuation (results published in Frontiers in Decision Neuroscience).							
15. SUBJECT TERMS							
16. SECURITY CLASSIFICATION OF:				17. LIMITATION OF ABSTRACT SAR			
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	18. NUMBER OF PAGES 77				
19a. NAME OF RESPONSIBLE PERSON ALAN LIN					19b. PHONE NUMBER (Include area code) 227-7009		

TABLE OF CONTENTS

Section	Page
List of Figures	ii
List of Tables	ii

Study #1: Intrinsic functional and structural brain connectivity in humans predicts individual social comparison orientation

1.0 SUMMARY	1
2.0 INTRODUCTION	1
3.0 METHODS, ASSUMPTIONS, AND PROCEDURES	3
3.1 Participants	3
3.2 Measuring social comparison orientation	3
3.3 Image acquisition	3
3.4 Structural connectivity analysis within the reward network	4
3.5 Functional connectivity analysis within the reward network	4
3.6 Exploratory voxel-level whole-brain analysis	5
3.7 Mediation analysis with the neural features from the exploratory whole-brain analyses	6
3.8 Inter-network connectivity between the frontoparietal and cingulo-opercular networks	6
3.9 Testing the effectiveness of neural predictors of social comparison orientation	7
4.0 RESULTS AND DISCUSSION	7
4.1 Intrinsic structural and functional connectivity within the reward network	7
4.2 Exploratory voxel-wise whole-brain analysis	8
4.3 Mediation effect	9
4.4 SCO associated with inter-network connectivity.....	10
4.5 Regression models predicting individual difference in social comparison	11
4.6 Discussion	11
5.0 CONCLUSIONS	15
6.0 REFERENCES	16

Study #2: Roles of the MPFC and Insula in Impression Management under Social Observation

1.0 SUMMARY	20
2.0 INTRODUCTION	20
3.0 METHODS, ASSUMPTIONS, AND PROCEDURES	23
3.1 Participants	23
3.2 Stimuli	24
3.3 Impression ratings from independent samples	24
3.4 Experimental procedures	25
3.5 fMRI task	25
3.6 Image acquisition	26
3.7 Data analysis	26
3.7.1 Behavioral analysis	26
3.7.1.1 Subject-level analysis	26
3.7.1.2 Group-level analysis	27
3.7.2 Analysis of Independent Sample Data: Correlational Analysis	27
3.7.3 fMRI data analysis	27
3.7.3.1 Preprocessing	27
3.7.3.2 Definition of ROI	28
3.7.3.3 Statistical threshold	28
3.7.3.4 Subject-level analysis	28
3.7.3.5 Group-level analysis: Three-way ANOVA	29
4.0 RESULTS AND DISCUSSION	29
4.1 Behavioral Results: Three-way ANOVA and descriptive statistics	29
4.2 Results of Independent Sample Data: Correlational Analysis	30
4.3 fMRI Results: Three-way ANOVA	31
4.3 Discussion	32

5.0 CONCLUSIONS	35
6.0 REFERENCES	35

Study #3: A Hierarchical Allostatic Regulation Model of Medial Prefrontal Cortex Function for Social Valuation

1.0 SUMMARY	42
2.0 INTRODUCTION	42
3.0 Anatomically and functionally dissociable subregions in the medial prefrontal cortex	43
3.1 Anatomical boundaries between the medial prefrontal cortex subregions	43
3.2 Role of the ventromedial prefrontal cortex (vmPFC) in social valuation	45
3.3 Role of the dorsomedial prefrontal cortex (dmPFC) in social valuation	47
3.4 Role of the rostromedial prefrontal cortex (rmPFC) in social valuation	48
4.0 Hierarchical allostatic regulation model of mPFC function for social valuation	52
4.1 mPFC subregions encoding gradient of internal-to-external valuation	52
4.2 Role of the thalamic reticular nucleus in switching between thalamo-cortical loops	54
4.3 Factors affecting the optimal balance between stability and plasticity	57
4.4 Self-control, self-efficacy, and metacognition as types of allostatic regulation	59
5.0 CONCLUSIONS	60
6.0 REFERENCES	60

LIST OF FIGURES

Figure 1 Associations between social comparison orientation (SCO) and the reward network	8
Figure 2. Results from the voxel-level, whole-brain resting state fMRI analyses	9
Figure 3. Results of inter-network connectivity analysis	10
Figure 4. A schematic diagram of a typical trial in a trait-evaluation task	26
Figure 5. Behavioral results from the trait-evaluation task	30
Figure 6. Impression rating data from independent sample	31
Figure 7. fMRI Results	32
Figure 8. Schematic diagram of anatomical segregation within the medial prefrontal cortex	45
Figure 9. Functional segregation of mPFC function during ethical consumption under social observation	51
Figure 10. The hierarchical allostatic regulation model of mPFC function for social valuation	53
Figure 11. Role of the thalamic reticular nucleus (TRN) in shifting between the thalamo-cortical loops	56

LIST OF TABES

Table 1. Summary of linear regression models with each brain measure individually and together	59
--	----

Study #1: Intrinsic functional and structural brain connectivity in humans predicts individual social comparison orientation (Authors: Wi Hoon Jung, Hackjin Kim; Currently Under Revision in Frontiers in Psychiatry)

1.0 SUMMARY

Social comparison orientation (SCO), the tendency to compare oneself with others, is universal, varies widely across individuals, and predicts important life and health outcomes. However, the neural mechanism underlying individual differences in SCO is still not well-understood. In the present study, we identified intrinsic neural markers of SCO in healthy young adults ($n = 42$) using a multimodal neuroimaging approach that included diffusion tensor imaging and resting-state functional MRI data. We found that higher SCO was associated with weaker structural and functional connectivity (SC and FC) strengths between the ventral striatum and the medial prefrontal cortex, which are core regions of the brain reward network. Additionally, individual SCO was negatively associated with neural fluctuations in the intraparietal sulcus (IPS), part of the frontoparietal network, and positively with FC between the IPS and anterior insula/amygdala cluster. This finding was further confirmed by the observation of independently-defined, large-scale, inter-network FC between the frontoparietal network and cingulo-opercular network. These results show how the intrinsic functional and structural architecture of the brain is linked to individual differences in SCO, providing novel evidence for an association between individual SCO and the brain's various networks.

2.0 INTRODUCTION

Social comparison—comparing one's own opinions and abilities with those of others—is a constant and ubiquitous experience that occurs throughout life. People use social comparison to evaluate or groom their social reputation and relationships (Festinger 1954; Buunk and Gibbons 2007). However, individuals vary in their tendency to engage in social comparison, and a person's tendency to compare oneself with others is referred to as social comparison orientation (SCO). This can be measured using the Iowa-Netherlands Comparison Orientation Measure (INCOM) (Gibbons and Buunk 1999). Understanding individual differences in SCO is important, because higher SCO is associated with a variety of real-life behaviors and outcomes, including several psychological dimensions such as poorer self-perception and lower self-esteem (Vogel et al. 2015), lower job satisfaction (White et al., 2006), more altruistic/helping behavior to unfamiliar others (Buunk and Dijkstra 2014), and an increased susceptibility to mental illness such as depression (Swallow and Kuiper 1988). Despite its importance, the neural markers that may underlie individual differences in SCO have been only sparsely investigated.

Over the past decade, researchers have observed links between social comparison and reward processing at both the psychological and neurological levels. For example, social comparison is associated with increased activity in the ventral striatum (VS) and medial prefrontal cortex (MPFC) (Fliessbach et al. 2007; Dvash et al. 2010; Denny et al. 2012; Kang et al. 2013a; Kedia et al. 2014a), both of which are

considered core regions of the brain reward system (Delgado et al. 2000; O'Doherty et al. 2003; Haber and Knutson 2010). More recent studies have reported that functional communication between the VS and MPFC is critically involved in social comparison and may underlie individual differences in SCO (Bault et al. 2011; Kang et al. 2013a).

Besides the reward-related neural network, several other neural structures may be involved in general comparison in both social and non-social domains (Kedia et al. 2014a; Swencionis and Fiske 2014). These regions include the intraparietal sulcus (IPS), anterior insula (aINS), and anterior cingulate cortex (ACC). The IPS is known to play a role in encoding numerical quantities and the activity of the region can be modulated by the distance between two magnitudes (i.e., comparison of numbers, size, luminance, or height) (Pinel et al. 2004; Cohen et al. 2005; Dormal et al. 2012). The activity of the aINS and ACC increases when people compare themselves to better-off others, and it is stronger for self-other than for familiar-other comparisons (Kedia et al. 2017; Luo et al. 2018). However, it is not yet clear whether structural and functional features of these areas at rest are linked to individual variability in SCO.

Recently, several studies have used task-independent neural measures to predict specific behaviors. These measures are more likely to be of practical use, because they are likely less tied to a specific context and are therefore relatively stable over time (Kable and Levy 2015; Jung et al. 2018). However, to our knowledge, no neuroimaging studies have yet investigated task-independent neural markers of individual differences in SCO. In the present study, we thus examined neural predictors of individual differences in SCO using a multimodal, task-independent neuroimaging approach, including diffusion tensor imaging (DTI) and resting-state functional magnetic resonance imaging (RS-fMRI), that focused on individual variations in the brain's intrinsic structural and functional network architecture. Because the reward network plays a central role in social comparison, we first aimed to identify potential associations between individuals' SCO scores and the strength of structural connectivity (SC) and resting-state functional connectivity (FC) between the VS and the MPFC, which are the central components of the reward network (i.e., intra-network connectivity within the reward network). For exploratory purposes to test whether SCO is associated with other areas and networks outside the reward network, we next performed the following additional analyses at the whole-brain voxel level using data-driven approach examining RS-fMRI and DTI data. In other words, to identify neural signatures of SCO at the whole-brain voxel level, we searched for brain areas associated with individual SCO in terms of local features (e.g., local connectivity and local fluctuation) in neural activity and in SC. This analysis highlighted the IPS, the neural structure previously linked to aspects of general comparison. To further characterize this association, we searched for neural regions whose FC strength with the IPS is associated with SCO, by using seed-based FC analysis, and found a significant association between the IPS–aINS/amygdala FC strength and SCO. Given that core areas of the frontoparietal network (FPN) and of the cingulo-opercular network (CON) include the IPS and aINS (Seeley et al. 2007; Menon 2015) respectively, we investigated whether individual SCO scores are associated with the inter-network FC between these two large-scale functional networks. Finally, linear regression analyses with identified neural variables showed that each of the identified features, particularly that from the RS-fMRI data, uniquely explains the variance in

SCO. These results provide a set of unique multi-modal intrinsic neural markers associated with individual differences in SCO.

3.0 METHODS, ASSUMPTIONS, AND PROCEDURES

3.1 Participants

A total of 47 participants were recruited from Korea University and the surrounding community. From each participant, we collected high-resolution T1-weighted anatomical MRI, RS-fMRI, DTI, and fMRI during an incentive delay task in the context of social comparison. In this study, we focused on task-independent measures of brain function and structure (i.e., RS-fMRI and DTI data) to examine the link between individual variations in SCO and individual differences in intrinsic functional and structural brain features. Among all, 43 had both DTI and RS-fMRI data available, and four participants were lost due to technical problems. One additional participant was excluded due to excessive head motion during RS-fMRI—that is, > 2.5 mm of translation or 2.5° of rotation and > 0.24 mm mean frame-wise displacement (FD; > 2 standard deviations from the group mean) (Power et al. 2011). Ultimately, the data of 42 participants (27 women, 15 men, age [mean $\pm$ SD]: 22.29 ± 3.04 years, all right-handed, SCO: 3.79 ± 0.58) were used in the final analyses. We confirmed that the final sample size was rational to obtain scientifically meaningful results, based on a power analysis performed before data analysis using G*Power software (Faul et al. 2007). Assuming an effect size of 0.5, an alpha level of 0.05, and a power of 0.90 to ensure correlation with the bivariate normal model, the G*Power analysis resulted in a required sample size of 37. All study procedures were approved by the Institutional Review Board of Korea University, and all participants provided written informed consent.

3.2 Measuring social comparison orientation

The degree of SCO for each participant was assessed using the Iowa-Netherlands Comparison Orientation Measure scale (INCOM) (Gibbons and Buunk 1999; Buunk and Gibbons 2007), which is a widely-used scale to test an individual's SCO. It consists of 11 items, each scored using a 5-point Likert scale (1 = I disagree strongly, 5 = I agree strongly). The INCOM measures an individual's tendency towards social comparison (e.g., "I often compare myself with others with respect to what I have accomplished in life"). All participants filled out the debriefing questionnaires, including the INCOM, scale before completing the scans.

3.3 Image acquisition

All images were scanned using a 3-T scanner (Siemens Magnetom Trio; Erlangen, Germany). High-resolution, T1-weighted anatomical images were acquired using a 3D magnetization-prepared, rapid-acquisition gradient echo (MPRAGE) sequence

(repetition time [TR] = 1900 ms, echo time [TE] = 2.52 ms, flip angle [FA] = 9°, voxel size = $1.0 \times 1.0 \times 1.0$ mm, 192 sagittal slices). Next, functional images were obtained using T2*-weighted, echo-planar imaging (EPI; TR = 2000 ms, TE = 20 ms, FA = 90°, voxel size = $3.0 \times 3.0 \times 3.0$ mm, 42 interleaved axial slices, and 155 volumes). During RS-fMRI, participants were instructed to keep their eyes open and maintain fixation. An eye-tracker mounted on a head coil was used to monitor the participants' eyes and ensure they did not fall asleep during the scan. Finally, DTI data were acquired with a 32-channel head coil using a single-shot, multiband EPI sequence (TR = 3000 ms, TE = 70 ms, FA = 90°, multiband acceleration factor = 3, phase partial Fourier = 6/8, voxel size = $2.0 \times 2.0 \times 2.0$ mm, 75 interleaved axial slices, and 64 diffusion directions with b-values of 1000 s/mm² and eight images with b-values of 0 s/mm²).

3.4 Structural connectivity analysis within the reward network

DTI data were preprocessed using PANDA v1.3.1 (Cui et al. 2013) (<https://www.nitrc.org/projects/panda/>): a pipeline tool for diffusion MRI that uses the processing functions of established packages, including FSL (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>) and the Diffusion Toolkit (<https://www.nitrc.org/projects/trackvis/>). Briefly, a brain mask was made using the b0 images. Diffusion images were registered to the average of the b0 images using an affine transformation to correct for eddy current-induced distortions and simple head-motion. Whole-brain fiber tracking was performed using the fiber assignment by continuous tracking (FACT) algorithm (Mori et al. 1999), with the fractional anisotropy threshold set at 0.20 and the tracking turning angular threshold set at 45°. Afterwards, spline filtering was applied to smooth the streamline tractography. To quantify the degree of connection between the left VS and MPFC, as well as between the right VS and MPFC in the native space, we first identified these three regions-of-interest (ROIs) based on a previous meta-analysis involving the valuation system in the human brain (Figure 1A) (Bartra et al. 2013). Next, these ROIs were transformed from the MNI space to each subject's native space. The number and average length of the fibers connecting each pair of ROIs were then calculated (Figure 1B). To normalize the fiber number, we divided it by the average volume and length of the two connecting regions. This counteracted bias where it was larger; closer brain regions inherently project/receive more fibers. Because the values were non-normally distributed, they were log-transformed before subsequent statistical analysis. We performed partial correlation, with age and sex as covariates, between SCO scores and normalized fiber numbers.

3.5 Functional connectivity analysis within the reward network

In the case of RS-fMRI data, the first five volumes were discarded to avoid instability in the initial data signal. Preprocessing steps, which included slice-acquisition timing, motion correction, nuisance signal regression, and spatial normalization, were performed using SPM12 (www.fil.ion.ucl.ac.uk/spm) and DPARSFA toolbox (Chao-Gan and Yu-Feng 2010) (www.rfmri.org/DPARSF). To

remove the effects of head motion and non-neuronal fluctuations on signals, as in recent studies (Parkes et al. 2018), the following nuisance parameters were included as regressors within the general linear model: Friston 24-motion parameters, five principal components estimated from both the white matter and cerebrospinal fluid regions using a component-based noise correction method (Behzadi et al. 2007), mean global signal, head motion scrubbing regressors one volume before and two volumes after the bad time point (root mean square volume-to-volume displacement > 0.25) (Jenkinson et al. 2002), and two polynomial trends (linear trend and quadratic trend). The spatially normalized residual images were smoothed using a 6-mm Gaussian kernel and band-pass filtering (0.01–0.1 Hz). The Pearson correlation coefficients were then computed between the mean time series of an ROI pair (i.e., MPFC, left VS, and right VS) as the strength of FC. The correlations were then converted to z-values using Fisher r-to-z transformation. We also performed partial correlation (covariates: age, sex, and motion indexed by mean FD) between SCO scores and z-transformed FC values, with the following co-variables.

3.6 Exploratory voxel-level whole-brain analysis

To ensure the study was complete and identify the brain areas or networks, mentioned in the Introduction, that are associated with individual SCO, we performed multiple regression analysis on data-driven structural and functional brain maps generated from DTI and RS-fMRI data together with SCO scores. Specifically, for voxel-level whole-brain DTI analysis, preprocessed DTI images were used to estimate four DTI metrics in the DTIFIT function of FSL: fractional anisotropy (FA), which measures the directionality of water diffusion, axial diffusivity (AD), which measures diffusion parallel to the white matter tract, radial diffusivity (RD), which measures diffusion perpendicular to the tract, and mean diffusivity (MD), which measures the diffusion speed of water molecules. The MD was estimated as the mean of all three eigenvalues ($[\lambda_1 + \lambda_2 + \lambda_3]/3$), AD as the mean of the second and third eigenvalues ($[\lambda_2 + \lambda_3]/2$), and RD as the principal eigenvalue (λ_1). To estimate the voxel-wise values of the DTI metrics of each subject's skeleton, we performed TBSS (Smith et al. 2006). All the subjects' FA images were aligned into the MNI standard space using the non-linear registration tool FNIRT. Next, a mean FA image was created and skeletonized/thinned to produce an image representing the center of all tracts common to the group (threshold = 0.2) (Figure S1). Each subject's aligned data (including FA, AD, RD, and MD) were then projected onto this skeleton. Finally, for the FA, AD, RD, and MD maps, we performed permutation-based statistics (using FSL's Randomise with 5,000 permutations) to determine the areas in which DTI metrics were associated with SCO. Age and sex were included as covariates. Threshold-free cluster enhancement (TFCE) was used to correct for multiple comparisons (corrected $p < 0.05$).

For voxel-wise, whole-brain RS-fMRI analysis, the following local FC maps were generated with a default setting of DPARSFA: (1) ReHo—a measure of localized intraregional connectivity (Zang et al. 2004), (2) DC—a measure of local network connectivity (Buckner et al. 2009; Cole et al. 2010), and (3) fALFF—a measure of the relative contribution of specific local frequency fluctuations in neural

activity to the whole frequency range (Zou et al. 2008). The ReHo maps were created using the Kendall coefficient of concordance of each voxel's time series with those of its 26 neighboring voxels (Smith et al. 2006). The DC maps were obtained by summing of the weights of the significant connections ($r > 0.25$) (Zang et al. 2004) for each voxel. For each of these three RS-fMRI maps, we implemented multiple regression analysis in SPM to identify areas in which the values of each map were associated with SCO, controlling for age, sex, and mean FD as covariates. Using the IPS cluster that had significant fALFF activity in the whole-brain RS-fMRI analysis, we generated and analyzed seed-based FC maps that were seeded using IPS. The maps were regressed against SCO scores to identify regions that were functionally coupled with IPS as a function of individual SCO differences. All results were corrected for multiple comparisons to a significance level of $p < 0.05$ (uncorrected height threshold of $p < 0.001$ combined with FWE-corrected extent threshold of $p < 0.05$).

3.7 Mediation analysis with the neural features from the exploratory whole-brain analyses

To further examine the relationship between SCO and the above IPS findings (i.e., IPS fALFF and IPS-aINS/amygdala seed-based FC), we tested whether the direct effect of the IPS fALFF strength (X) on SCO (Y) could be explained in terms of the indirect influence of IPS-aINS/amygdala FC strength (M) as a mediator. To this end, we used the M3 Mediation Toolbox (<https://github.com/canlab/MediationToolbox>). Age, sex, and mean FD were included as covariates. Bootstrapping with 10,000 resamples was used for statistical inference in each path ($p < 0.05$).

3.8 Inter-network connectivity between the frontoparietal and cingulo-opercular networks

Based on the results of the multiple regression analysis with the above seed-based FC maps, we hypothesized that SCO is associated with functional interactions between two largely independent neural networks: the frontoparietal network (FPN) and the cingulo-opercular network (CON; also often referred to as the salience network), because the IPS and aINS/amygdala clusters reported above are the key components, respectively (Seeley et al. 2007; Menon 2015). To validate this hypothesis, we evaluated the data within a network framework. Specifically, the nodes of each network (25 nodes in the FPN and 14 nodes in the CON) consisted of 6-mm radius spheres centered on the coordinates taken from the corresponding networks in the Power-264 atlas, as defined in terms of the task-based fMRI and resting-state FC techniques (Power et al. 2011). Next, to estimate inter-network FC, we extracted the mean time series from each of the nodes, computed the average connectivity across all node-to-node connections between the two networks using Pearson's correlation, and converted the correlations into z-values using Fisher r-to-z transformation. For exploratory purposes, we also computed the average connectivity across node pairs within the same network, defining this as intra-network FC. We then performed partial correlation (covariates: age, sex, and mean FD) between SCO scores and z-

transformed FC strengths.

3.9 Testing the effectiveness of neural predictors of social comparison orientation

Finally, we examined whether the neural features found in this study (i.e., right VS–MPFC SC, right VS–MPFC FC, fALFF in the IPS, IPS–aINS/amygdala FC, and inter-network FC between FPN and CON) explain independent or overlapping variance in SCO. This was done by performing a multiple linear regression analysis including all of the neural features as independent variables to explain the variance in SCO. Notably, significant features in the multiple regression model explain variance in SCO over and above that explained by all other remaining features (Jung et al., 2018). We used SPSS Statistics version 25 to perform the linear regression analysis on each brain variable alone, as well as on all five identified brain variables together. Before performing this statistical analysis, the effects of age and sex were regressed out of all the neural variables.

4.0 RESULTS AND DISCUSSION

4.1 Intrinsic structural and functional connectivity within the reward network

The SCO scores were negatively associated with both SC ($r = -0.350$, $p = 0.027$) and FC ($r = -0.479$, $p = 0.001$) between the right VS and the MPFC, whereas no correlation was found in the left hemisphere (SC: $r = 0.044$, $p = 0.787$; FC: $r = 0.054$, $p = 0.734$) (**Figure 1**). There were no significant correlations between the strengths of SC and FC ($r = -0.134$, $p = 0.399$ for the left VS–MPFC connection; $r = 0.133$, $p = 0.403$ for the right).

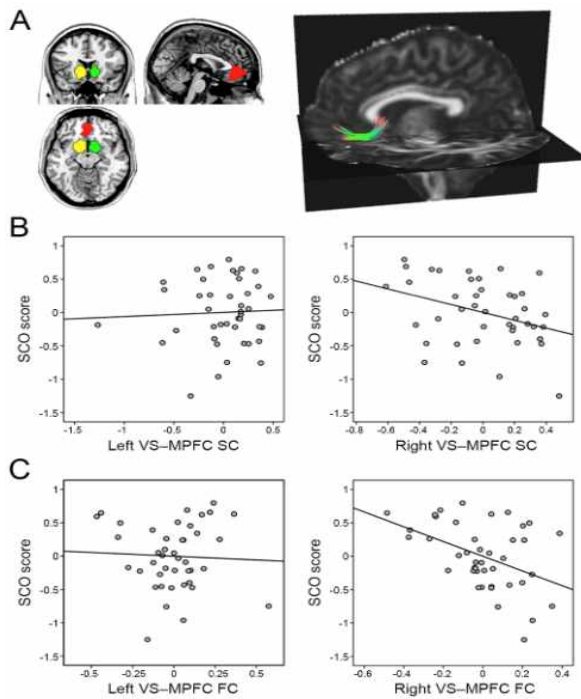


Figure 1 Associations between social comparison orientation (SCO) and the reward network. (A) The right panel shows three regions of interest—the left ventral striatum (VS; yellow), the right VS (green), and the medial prefrontal cortex (MPFC; red). The left panel displays a single subject's tracking map between the right VS and the MPFC, identified using tractography analysis (for illustration purposes only) (B) Partial correlation scatterplot between SCO score and normalized fiber number, displayed as the strength of structural connectivity between the VS and MPFC. (C) Partial correlation scatterplot between SCO score and VS-MPFC functional connectivity strength (z-transformed). For illustration purposes, (B) and (C) were generated using Pearson's correlation analysis between residuals after age and sex were regressed out.

4.2 Exploratory voxel-wise whole-brain analysis

Voxel-level whole-brain RS-fMRI analysis revealed that the fractional amplitude of low-frequency fluctuation (fALFF) value in the right IPS (peak MNI x, y, z coordinates = 60, -42, 42; peak z-value = 4.45) was negatively associated with SCO score (**Figure 2A**). No regions showed any significant correlation with other voxel-level whole-brain RS-fMRI maps including regional homogeneity (ReHo) and degree centrality (DC) maps at an uncorrected significance level of $p < 0.001$ and a family-wise error (FWE)-corrected extent of $p < 0.05$.

Further multiple regression analysis using seed-based FC maps, with the right IPS acting as the seed point, revealed that SCO score was positively associated with FC strength between the right IPS seed and right aINS extending to amygdala (referred to as “aINS/amygdala” cluster here), areas belonging to the FPN and CON respectively (x, y, z coordinates = 30, 3, -18; z-value = 3.95; uncorrected significance level $p < 0.001$; FWE-corrected extent $p < 0.05$; **Figure 2B**).

No regions showed any significant correlation with the whole-brain structural maps created using DTI data (fractional anisotropy [FA], mean diffusivity [MD], axial diffusivity [AD], and radial diffusivity [RD] tract-based spatial statistics [TBSS] maps).

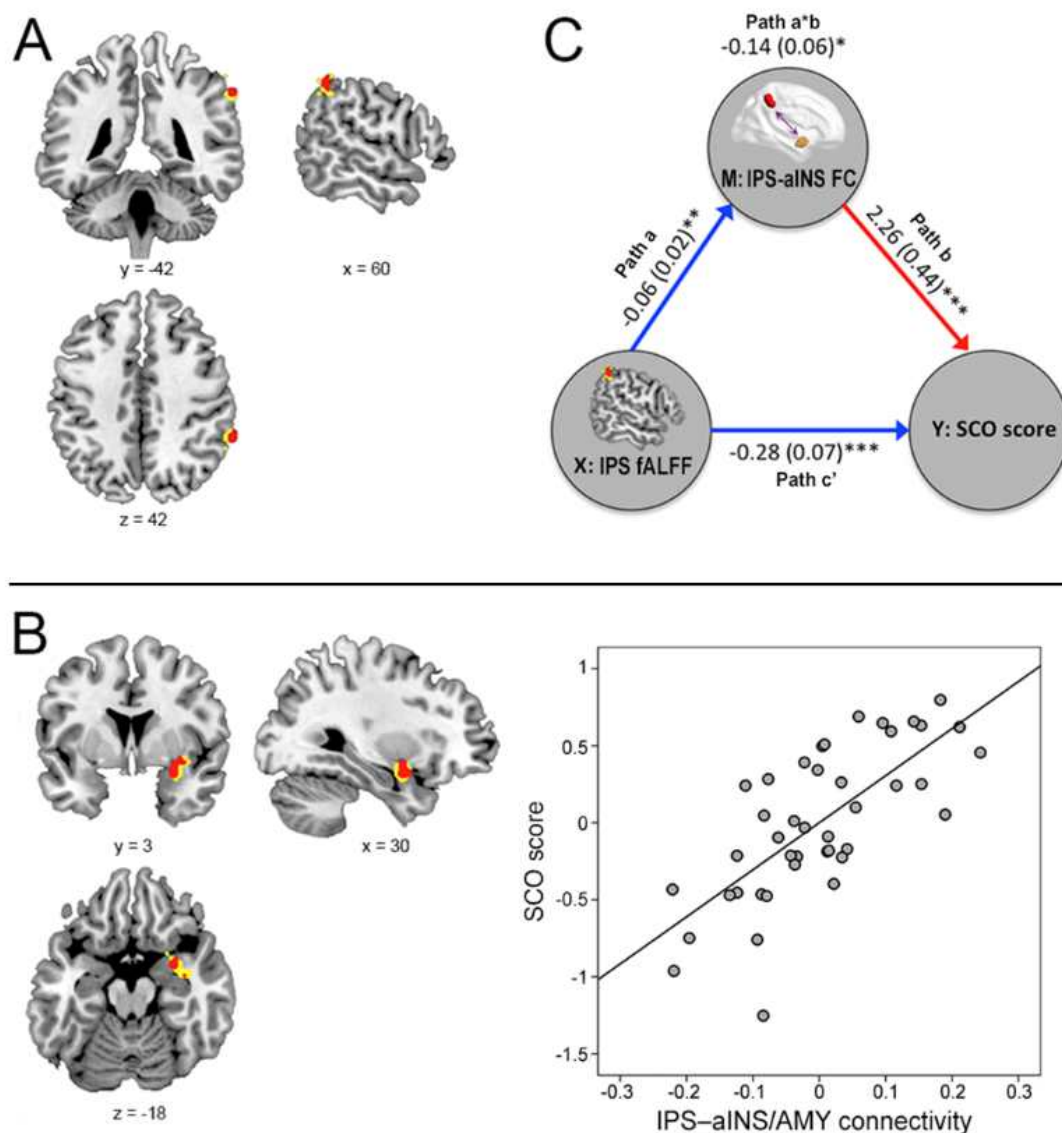


Figure 2. Results from the voxel-level, whole-brain resting state functional magnetic resonance imaging analyses. (A) The fractional amplitude of low-frequency fluctuation (fALFF) value in the right intraparietal sulcus (IPS) was negatively associated with social comparison orientation (SCO) score (height $p < 0.001$ [red] or $p < 0.005$ [yellow]). (B) Functional connectivity strength between the right IPS seed and right anterior insula (aINS)/amygdala cluster was positively associated with SCO score (height $p < 0.001$ [red] or 0.005 [yellow]). For illustration purposes, this scatterplot was generated by performing Pearson correlation analysis between residuals age, sex, and motion were regressed out. (C) The mediation effect of functional connectivity strength in the IPS–aINS/amygdala on the right IPS fALFF and SCO scores. All paths (Paths a, b, and c') and mediation effects (Path a*b) are labeled with path coefficients and their standard errors in parenthesis. Blue and red arrows indicate negative and positive relationships, respectively. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.3 Mediation effect

Figure 2C shows the mediation effect of IPS–aINS/amygdala FC on the relationship between SCO score and fALFF in the IPS. In particular, fALFF was negatively correlated with the FC between IPS and aINS/amygdala (Path a). The same FC was positively correlated with SCO score (Path b). Finally, IPS–aINS/amygdala FC exhibited a negative mediation effect (negative Path a*b) that resulted from an IPS fALFF-associated reduction in IPS–aINS/amygdala FC (negative Path a), and there was a positive relationship between IPS–aINS/amygdala FC and SCO score (positive Path b). This finding indicates that stronger FC between the IPS and the aINS/amygdala mediates the reduced fALFF in the IPS among individuals with high SCO scores.

4.4 SCO associated with inter-network connectivity

In line with our hypothesis, SCO score correlated positively with inter-network connectivity strength between the FPN and CON ($r = 0.393$, $p = 0.013$; **Figure 3**). An exploratory analysis with intra-network connectivity revealed that there were no associations between SCO score and intra-network connectivity ($r = 0.282$, $p = 0.082$ for the FPN; $r = 0.223$, $p = 0.173$ for the CON).

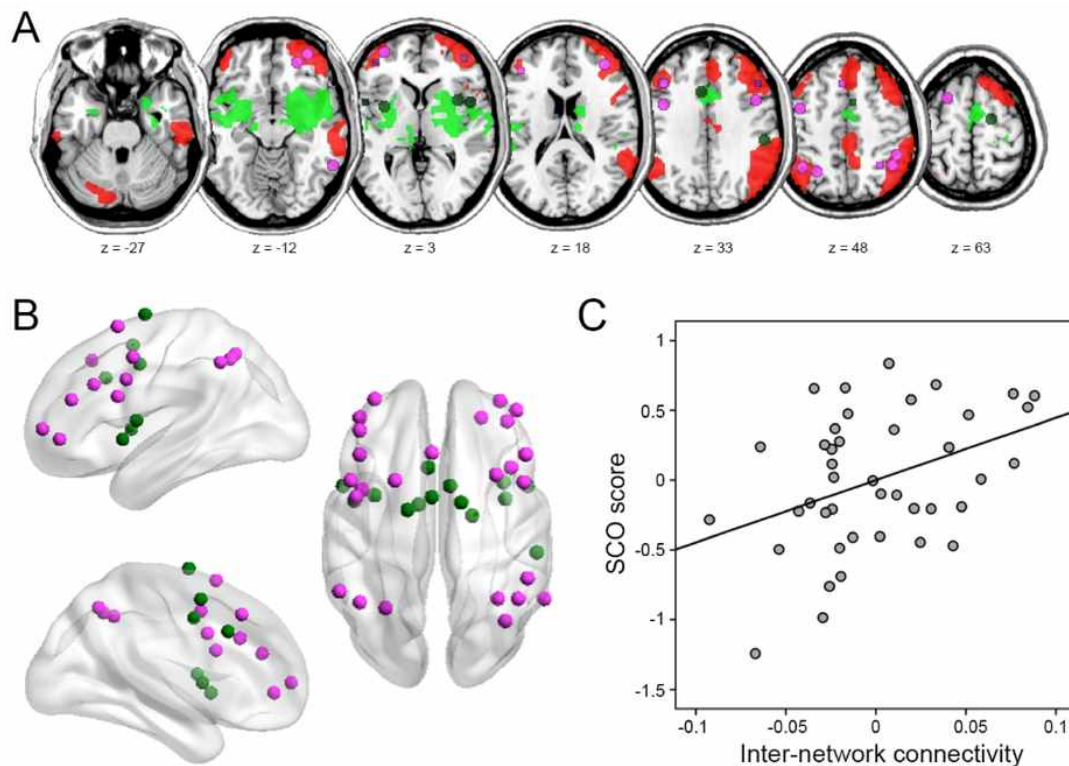


Figure 3. Results of inter-network connectivity analysis. (A) Figure illustrating the location of regions (i.e., nodes) in the frontoparietal network (FPN) (violet) and cingulo-opercular network (CON) (green), identified from the atlas by Power et al. (44). (B) Identified nodes overlaid on within-group, seed-based functional connectivity map of each right IPS and right aINS/amygdala seed, identified from voxel-level, whole-brain analysis (uncorrected height $p < 0.001$ and FWE-corrected extent $p < 0.05$). (C) Partial correlation between SCO score and FPN–CON inter-network connectivity strength. For illustration purposes, this scatterplot was generated using Pearson correlation analysis between residuals after age, sex, and motion

were regressed out.

4.5 Regression models predicting individual difference in social comparison

A linear regression model using all neural variables revealed that the FC within the reward network (right VS–MPFC FC), neural fluctuation (i.e., fALFF) in IPS activity, and the FC between the IPS and aINS/amygdala were significant predictors ($p < 0.05$) of SCO (**Table 1**). In such a combined model, significant measures explain the variance in SCO more than all other measures. We also ran linear regressions with each of the measures individually, allowing us to compare the variance explained by each measure (**Table 1**). The variance (R^2) estimated from these analyses, arranged in ascending order, was as follows: 0.12 for the right VS–MPFC SC alone, 0.14 for the FPN–CON inter-network FC alone, 0.19 for the right VS–MPFC FC alone, 0.45 for the fALFF in the IPS alone, and 0.52 for the IPS–aINS/amygdala FC alone. The variance of the combined model was 0.75.

Table 1. Summary of linear regression models with each brain measure individually¹ and together

	Model with VS–MPFC SC	Model with VS–MPFC FC	Model with fALFF in IPS	Model with IPS–aINS/AMY FC	Model with FPN–CON FC	Model with all brain variables
Constant	-2.38E-5(0.07)	-2.46E-5(0.07)	-2.38E-5(0.06)	-2.24E-5(0.05)	-2.18E-5(0.07)	-2.25E-5 (0.04)
VS–MPFC SC	-0.59(0.25)*	–	–	–	–	-0.12(0.15)
VS–MPFC FC	–	-1.10(0.36)**	–	–	–	-0.51(0.22)*
fALFF in IPS	–	–	-0.43(0.08)***	–	–	-0.26(0.06)***
IPS–aINS/AMY FC	–	–	–	3.05(0.47)***	–	1.83(0.42)***
FPN–CON FC	–	–	–	–	4.18(1.64)*	1.56(1.02)
R^2	0.12	0.19	0.45	0.52	0.14	0.75

¹Linear regressions with each individual measure was performed for comparison purposes, with the amount of the variance (R^2) explained in terms of the model with all brain variables. Data are given as unstandardized coefficients, B (standard errors). Coefficients significantly different from zero are indicated by asterisks: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

R^2 values indicate the amount of variance, explained by the model.

VS–MPFC SC: structural connectivity strength between right ventral striatum and medial prefrontal cortex, as part of the reward network

VS–MPFC FC: functional connectivity strength between right ventral striatum and medial prefrontal cortex

fALFF in IPS: fractional amplitude of low-frequency fluctuation in intraparietal sulcus

IPS–aINS/AMY FC: functional connectivity strength between IPS and anterior insula/amygdala

FPN–CON FC, inter-network functional connectivity between the frontoparietal network and cingulo-opercular network

4.6 Discussion

In the present study, we investigated whether individual differences in social comparison, as measured by SCO, were related to multimodal, context-independent, brain measures, estimated using DTI and RS-fMRI data. In so doing, we identified several intrinsic functional and structural neural markers of SCO. Most importantly, individuals with higher SCO showed weaker SC and FC between the right VS and the MPFC—regions belonging to the reward-related neural network. We also found several exploratory results from the whole-brain voxel level analyses and network analysis. Individuals with higher SCO showed reduced spontaneous neural activity in the IPS—a region belonging to the FPN, increased FC between the IPS and aINS/amygdala, and large-scale inter-network FC between the FPN and CON. Of these measures, right VS–MPFC FC, fALFF in IPS, and IPS–aINS/amygdala FC contributed most to the neural prediction of SCO. The predictive model using all neural markers identified in the present study was highly effective—accounting for a substantial amount of variance in SCO ($R^2 = 0.75$). Taken together, these findings suggest that individual differences in social comparison can be characterized in terms of specific patterns in neural structures as well as intrinsic neural activity, particularly in the neural networks engaged in reward processing and comparative processing of external stimuli.

Our findings of SC and FC between the right VS and MPFC are broadly consistent with previous studies linking the same markers with reward processing. Both the VS and MPFC play a critical role in reward processing, showing elevated activity in response to both primary (e.g., food) and secondary reward stimuli (e.g., money) (Chib et al. 2009; Kim et al. 2011). Importantly, these two regions contribute to the appraisal or representation of the subjective value of either social or non-social rewards (Izuma et al. 2008; Izuma et al. 2010; Bartra et al. 2013; Tamir and Hughes 2018), providing strong evidence for a common neural currency (Saxe and Haushofer 2008; Levy and Glimcher 2012). Relatedly, activity in the VS and MPFC are modulated by the absolute outcome and by the relative payoff differences derived from social comparison (Fliessbach et al. 2007; Dvash et al. 2010; Vostroknutov et al. 2012), and this mechanism can vary depending on cultural membership (Kang et al. 2013a). Several studies have indicated that functional interaction between the VS and MPFC reflects variability in the behavioral changes caused by social comparison (Bault et al. 2011; Kang et al. 2013a). For example, in one study, the VS response to social gains (winning more than a counterpart) during the earlier outcome phase predicted MPFC activity during the subsequent decision phase, and experienced social gains induced behavioral changes in later trials (Bault et al. 2011). In addition, the VS-MPFC FC strength predicted individual variability in the degree to which participants' decisions were affected by relative income (Kang et al. 2013a).

In this study, the association of SCO with SC and FC between VS and vmPFC was significant only in right hemisphere. Though this was not expected, many previous studies have proposed hemispheric specialization of the reward network and of social processing. For instance, a recent functional MRI meta-analysis study shows that hemispheric dominance of striatum activation varies across different types of reward, including food, erotic, and money stimuli (Arsalidou et al., 2020). In addition, a right-lateralized connectivity of the VS to the parietal cortices during resting-state has been reported (Zhang et al., 2017). Studies in different types of social contexts demonstrate

right hemisphere superiority in detecting and processing social stimuli (e.g., voices, faces, and gestures) (Brancucci et al., 2009; Watson et al., 2014) as well as understanding the intentions behind other's actions (Liepelt et al., 2008; Ortigue et al., 2009). Future neuroimaging studies combining both behavioral task on social comparison and resting-state fMRI with a larger sample size will help to verify the observed hemispheric lateralization of FC associated with social comparison.

The present study showed that higher social comparison was associated with weaker VS-MPFC FC during resting-state. Previous studies have reported the involvement of VS-MPFC FC in the manifestation of clinical symptoms, such as addiction and depression (Contreras-Rodriguez et al., 2017; Rupprechter et al., 2020), as well as reward learning and valuation (Bault et al. 2011; Petersen et al., 2018). In the studies on social comparison, an increased activity and FC within the reward network have been observed specifically when people compare themselves to worse-off others (referred to as downward comparison), which is often associated with positive feelings (Kang et al., 2013a; Luo et al., 2018). Thus, it can be speculated that such a hypo-connectivity in the reward network during resting-state observed in the present study may reflect reduced baseline intrinsic reward sensitivity, which may cause people to seek excessive extrinsic social rewards possibly through increased social comparison to others, leading to positive (downward comparison) as well as negative (upward comparison) feelings. In line with this explanation, patients with hyperactivity or increased reward-seeking behavior showed reduced neural responsiveness in the VS, a key part of the reward network (Scheres et al. 2007). Notably, usage-dependent selective synapse elimination (Huttenlocher and Dabholkar 1997) is often observed as an example of day-to-day experience-dependent neural plasticity (Trachtenberg et al. 2002), which may be the mechanism underlying decreases in neural activity and cortical thickness after training (Kang et al. 2013b). Another possible explanation is that the reduced SC and FC between the VS and MPFC may indicate that the number of available alternatives is reduced because subjects engage in the excessive pursuit of a limited number of rewards. One good example of such a state may be approval addiction, which involves the excessive pursuit of approval to gain superior social status to others (i.e., downward comparison). The desire of social approval may be the main cause of social comparison. Though we speculate above on interpretations for our findings, we caution against these interpretations as we did not have any behavioral data to prove these interpretations. Therefore, further study may be necessary to investigate whether weaker resting-state VS-MPFC FC is associated with FC in the same circuit during certain social comparison behavior. Such research would provide a more accurate understanding of the functional implication of VS-MPFC FC in social comparison.

In the present study, individuals with higher SCO exhibited less fALFF in the right IPS, which is part of the FPN. While FC quantifies temporal synchrony between remote brain areas, the fALFF indicates quantifiable magnitudes of spontaneous regional neural activity across the whole brain (Zou et al. 2008). In other words, the fALFF allow us to probe local brain regions where individual differences in resting state activity are correlated with their phenotype (in this case, SCO) across the whole brain at voxel level. The IPS plays a crucial role in visuospatial attention and arithmetic processing (Hubbard et al. 2005), and it activates during cognitive and perceptual

comparison of stimuli that differ in various ways (e.g., number, size, or luminance) (Pinel et al. 2004; Cohen Kadosh et al. 2005; Dormal et al. 2012). Notably, Previous studies have demonstrated that the degree of IPS activity increases with the difficulty of comparison (Cohen Kadosh et al. 2005; Kedia et al. 2014b). IPS activity is also increased during the comparison of social status (Kedia et al. 2014b), as well as during comparison of one's own height against those of acquaintances (Seeley et al. 2007). Interestingly, using the IPS as the seed region, further regression analysis between the IPS seed-based FC maps and the SCO revealed that higher SCO was associated with greater FC between the IPS seed and the aINS/amygdala cluster. Additionally, IPS–aINS/amygdala FC partially mediated the link between fALFF in the IPS and the SCO. Therefore, our findings suggest that the SCO is associated with various features of intrinsic neural activity in the IPS, including the power of local neural activity and the patterns of FC.

Considering that the IPS and aINS/amygdala are the core regions of the FPN and CON, respectively, we examined inter-network FC between the FPN and CON, which were independently identified in a previous study (Power et al. 2011). Thus, we confirmed that higher SCO scores are associated with stronger inter-network FC between the FPN and CON. The aINS/amygdala cluster, which comprises key elements of the CON, together with the ACC, has been strongly implicated in social and non-social emotions, including disgust (Phillips et al. 1997), pain (Wager et al. 2013), unfairness (Sanfey et al. 2003), and empathy (Singer et al. 2009), and interoceptive and emotional awareness (Critchley et al. 2004; Craig 2011; Zaki et al. 2012), as well as in saliency detection (Uddin 2015). Relevant to the present study, the aINS is often engaged during social comparison (Kedia et al. 2017; Luo et al. 2018). In particular, a recent meta-analysis of functional neuroimaging studies emphasized the roles of the aINS and ACC in upward comparison (Luo et al. 2018). Previous studies have reported competitive and cooperative interactions between FPN and CON (Dosenbach et al. 2007; Santangelo 2018). For example, Dosenbach et al. (2007) suggested that these networks communicate with each other, and that each of them carries out dissociable control functions, such as adaptive control in the FPN and stable set-maintenance functions in the CON. Furthermore, the interaction between FPN and CON may be involved in the integration of salient cognitive and affective information to promote goal-directed behavior (Pessoa 2008; Vincent et al. 2008). Thus, we cautiously speculate that the tendency towards higher social comparison can be characterized in terms of increased FC between the CON and FPN, and that this increased FC integrates affective and cognitive/comparative information in the pursuit of self-promotional goals, even during resting periods. Given that the SCO showed negative association with VS–MPFC FC and positive association with IPS–aINS/amygdala FC, we also speculate that individuals with higher social comparison operate more within the external valuation system (IPS–aINS/amygdala FC) and less within the internal valuation system (VS–MPFC FC) than those with lower social comparison during rest. However, there is a lack of additional data supporting this speculation, so future studies should clarify this issue by using functional neuroimaging data obtained simultaneously with behavioral indices of social comparison.

For the exploratory whole-brain analyses, significant relationships of SCO

were found only with the RS-fMRI measures but not with the DTI measures. While DTI measures quantify properties related to the direct anatomical links (i.e., white matter fibers) between voxels, RS-fMRI measures quantify the voxel itself and local or remote connections between voxels, especially in the absence as well as in the presence of direct anatomical links. From this point of view, our results for IPS connectivity may reflect FC derived from indirect anatomical connections (Honey et al., 2009). ROI approach has the advantage of alleviating the multiple comparisons problem by the limiting the number of statistical tests when there are specific hypotheses. Therefore, because of the aforementioned advantage, it may be that the relationship with SC in the present study was found in ROI analysis, but not in voxel-level analysis. In this regard, another possible interpretation is that our DTI measures may be less sensitive in detecting relationships with SCO at the whole-brain level due to more stringent threshold. SC estimation is challenging owing to complex fiber orientations, such as crossing fibers within a voxel. This problem may cause false-positive and false-negative connections, generating spurious and overlooked links of fiber tracts, respectively. In this regard, the current spatial resolution and analytical techniques for DTI data are not sufficient to solve the issue referred to as the “crossing-fiber problem”. Future studies using data with more gradient directions (e.g., high angular resolution diffusion imaging, HARDI; Hosey et al., 2005) and multiple tensor models (e.g., Q-ball; Descoteaux et al., 2007) will clarify the relationship between SCO and SC without the crossing-fiber problem.

The present study had some limitations that should be addressed in future research. Firstly, our interpretations of the findings were necessarily limited by the paucity of information about the directionality of SC and FC. Secondly, it is unclear whether the observed associations reflect the causes or the results of the different levels of social comparison, mainly because the study was cross-sectional in design. The strength of VS–MPFC FC during rest declines with age (Fareri et al. 2015), so future research with longitudinal design should address whether the observed associations change with age. Finally, because the exploratory nature of additional analyses to test whether SCO is associated with certain areas and networks outside the reward networks, hence no further correction for the number of all analyses performed (including mediation analysis, SCO and inter-network connectivity, and regression models) was performed, though each of all these separate analyses was corrected for multiple comparisons.

5.0 CONCLUSIONS

to our knowledge, the present study was the first to demonstrate that task-independent neural markers can explain individual variabilities in social comparison. Using multimodal, task-independent neuroimaging data, including DTI and RS-fMRI data, we identified several brain networks associated with individual differences in SCO, including the reward network (comprising the MPFC and VS), the FPN (containing the IPS), and the CON (containing the aINS/amygdala). These networks have previously been implicated in either social comparison or general comparative information processing. The present study provided novel and important insights regarding the neural mechanisms underlying individual differences in SCO,

suggesting that social comparison is a multidimensional process that engages the networks associated with various motivational, affective, and cognitive components.

6.0 REFERENCES

- Arsalidou M, Vijayarajah S, Sharaev M. 2020. Basal ganglia lateralization in different types of reward. *Brain Imaging Behav.* In press. doi: 10.1007/s11682-019-00215-3.
- Bartra O, McGuire JT, Kable JW. 2013. The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage.* 76:412–427.
- Bault N, Joffily M, Rustichini A, Coricelli G. 2011. Medial prefrontal cortex and striatum mediate the influence of social comparison on the decision process. *Proc Natl Acad Sci USA.* 108:16044–16049.
- Behzadi Y, Restom K, Liao J, Liu TT. 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage.* 37:90–101.
- Brancucci A, Lucci G, Mazzatenta A, Tommasi L. 2009. Asymmetries of the human social brain in the visual, auditory and chemical modalities. *Philos Trans R Soc Lond B Biol Sci.* 364:895–914.
- Buckner RL, Sepulcre J, Talukdar T, Krienen FM, Liu H, Hedden T, Andrews-Hanna JR, Sperling RA, Johnson KA. 2009. Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. *J Neurosci.* 29:1860–1873.
- Buunk AP, Dijkstra P. 2014. Social comparison orientation and perspective taking as related to responses to a victim. *Psychology.* 5:441–450.
- Buunk AP, Gibbons FX. 2007. Social comparison: The end of a theory and the emergence of a field. *Organ Behav Hum Decis Process.* 102:3–21.
- Chao-Gan Y, Yu-Feng Z. 2010. DPARSF: A MATLAB toolbox for "pipeline" data analysis of resting-state fMRI. *Front Syst Neurosci.* 4:13.
- Chib VS, Rangel A, Shimojo S, O'Doherty JP. 2009. Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *J Neurosci.* 29:12315–12320.
- Cohen Kadosh R, Henik A, Rubinsten O, Mohr H, Dori H, van de Ven V, Zorzi M, Hendler T, Goebel R, Linden DE. 2005. Are numbers special? The comparison systems of the human brain investigated by fMRI. *Neuropsychologia.* 43:1238–1248.
- Cole MW, Pathak S, Schneider W. 2010. Identifying the brain's most globally connected regions. *Neuroimage.* 49:3132–3148.
- Contreras-Rodríguez O, Martín-Pérez C, Vilar-López R, Verdejo-García A. 2017. Ventral and Dorsal Striatum Networks in Obesity: Link to Food Craving and Weight Gain. *Biol Psychiatry.* 81:789–796.
- Craig AD (2011) Significance of the insula for the evolution of human awareness of feelings from the body. *Ann N Y Acad Sci.* 1225:72–82.
- Critchley HD, Wiens S, Rotshtein P, Ohman A, Dolan RJ. 2004. Neural systems supporting interoceptive awareness. *Nat Neurosci.* 7:189–195.

- Cui Z, Zhong S, Xu P, He Y, Gong G. 2013. PANDA: A pipeline toolbox for analyzing brain diffusion images. *Front Hum Neurosci.* 7:42.
- Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA. 2000. Tracking the hemodynamic responses to reward and punishment in the striatum. *J Neurophysiol.* 84:3072–3077.
- Denny BT, Kober H, Wager TD, Ochsner KN. 2012. A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *J Cogn Neurosci.* 24:1742–1752.
- Descoteaux M, Angelino E, Fitzgibbons S, Deriche R. 2007. Regularized, fast, and robust analytical Q-ball imaging. *Magn Reson Med.* 58:497–510.
- Dormal V, Andres M, Pesenti M. 2012. Contribution of the right intraparietal sulcus to numerosity and length processing: An fMRI-guided TMS study. *Cortex.* 48:623–629.
- Dosenbach NU, Fair DA, Miezin FM, Cohen AL, Wenger KK, Dosenbach RA, Fox MD, Snyder AZ, Vincent JL, Raichle ME, Schlaggar BL, Petersen SE. 2007. Distinct brain networks for adaptive and stable task control in humans. *Proc Natl Acad Sci USA.* 104:11073–11078.
- Dvash J, Gilam G, Ben-Ze'ev A, Hendler T, Shamay-Tsoory SG. 2010. The envious brain: The neural basis of social comparison. *Hum Brain Mapp.* 31:1741–1750.
- Fareri DS, Gabard-Durnam L, Goff B, Flannery J, Gee DG, Lumian DS, Caldera C, Tottenham N. 2015. Normative development of ventral striatal resting state connectivity in humans. *Neuroimage.* 118:422–437.
- Faul F, Erdfelder E, Lang AG, Buchner A. 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods.* 39:175–191.
- Festinger L. 1954. A theory of social comparison processes. *Hum Relat.* 7:117–140.
- Fliessbach K, Weber B, Trautner P, Dohmen T, Sunde U, Elger CE, Falk A. 2007. Social comparison affects reward-related brain activity in the human ventral striatum. *Science.* 318:1305–1308.
- Gibbons FX, Buunk BP. 1999. Individual differences in social comparison: Development of a scale of social comparison orientation. *J Pers Soc Psychol.* 76:129–142.
- Haber SN, Knutson B. 2010. The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology.* 35:4–6.
- Honey CJ, Sporns O, Cammoun L, et al. 2009. Predicting human resting-state functional connectivity from structural connectivity. *Proc Natl Acad Sci U S A.* 106:2035–2040.
- Hosey T, Williams G, Ansorge R. 2005. Inference of multiple fiber orientations in high angular resolution diffusion imaging. *Magn Reson Med.* 54:1480–1489.
- Hubbard EM, Piazza M, Pinel P, Dehaene S. 2005. Interactions between number and space in parietal cortex. *Nat Rev Neurosci.* 6:435–448.
- Huttenlocher PR, Dabholkar AS. 1997. Regional differences in synaptogenesis in human cerebral cortex. *J Comp Neurol.* 387:167–178.
- Izuma K, Saito DN, Sadato N. 2008. Processing of social and monetary rewards in the human striatum. *Neuron.* 58:284–294.
- Izuma K, Saito DN, Sadato N. 2010. Processing of the incentive for social approval in

- the ventral striatum during charitable donation. *J Cogn Neurosci*. 22:621–631.
- Jenkinson M, Bannister P, Brady M, Smith S. 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*. 17:825–841.
- Jung WH, Lee S, Lerman C, Kable JW. 2018. Amygdala functional and structural connectivity predicts individual risk tolerance. *Neuron*. 98:394–404.e4.
- Kable JW, Levy I. 2015. Neural markers of individual differences in decision-making. *Curr Opin Behav Sci*. 5:100–107.
- Kang DH, Jo HJ, Jung WH, Kim SH, Jung YH, Choi CH, Lee US, An SC, Jang JH, Kwon JS. 2013b. The effect of meditation on brain structure: cortical thickness mapping and diffusion tensor imaging. *Soc Cogn Affect Neurosci*. 8:27–33.
- Kang P, Lee Y, Choi I, Kim H. 2013a. Neural evidence for individual and cultural variability in the social comparison effect. *J Neurosci*. 33:16200–16208.
- Kedia G, Mussweiler T, Adam R, Ischebeck A, Ihssen N, Linden DEJ. 2017. So pretty! The neural correlates of self-other vs familiar-other attractiveness comparisons. *Soc Neurosci*. 14:41–52.
- Kedia G, Mussweiler T, Linden DE. 2014a. Brain mechanisms of social comparison and their influence on the reward system. *Neuroreport*. 25:1255–1265.
- Kedia G, Mussweiler T, Mullins P, Linden DE. 2014b. The neural correlates of beauty comparison. *Soc Cogn Affect Neurosci*. 9:681–688.
- Kim H, Shimojo S, O'Doherty JP. 2011. Overlapping responses for the expectation of juice and money rewards in human ventromedial prefrontal cortex. *Cereb Cortex*. 21:769–776.
- Levy DJ, Glimcher PW. 2012. The root of all value: A neural common currency for choice. *Curr Opin Neurobiol*. 22:1027–1038.
- Liepelt R, Von Cramon DY, Brass M. 2008. How do we infer others' goals from non-stereotypic actions? The outcome of context-sensitive inferential processing in right inferior parietal and posterior temporal cortex. *Neuroimage*. 43:784–792.
- Luo Y, Eickhoff SB, Héту S, Feng C. 2018. Social comparison in the brain: A coordinate-based meta-analysis of functional brain imaging studies on the downward and upward comparisons. *Hum Brain Mapp*. 39:440–458.
- Menon V. 2015. Salience Network. *Brain mapping: An encyclopedic reference vol. 2.*, ed Toga AW (Elsevier Academic Press), pp 597–611.
- Mori S, Crain BJ, Chacko VP, van Zijl PCM. 1999. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annal Neurol*. 45:265–269.
- O'Doherty J, Dayan P, Friston K, Critchley H, Dolan RJ. 2003. Temporal difference models and reward-related learning in the human brain. *Neuron*. 38:329–337.
- Ortigie S, King D, Gazzaniga M, Miller M, Grafton S. 2009. Right hemisphere dominance for understanding the intentions of others: evidence from a split-brain patient. *BMJ Case Rep*. 2009:bcr07.2008.0593.
- Parkes L, Fulcher B, Yücel M, Fornito A. 2018. An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. *Neuroimage*. 171:415–436.
- Pessoa L. 2008. On the relationship between emotion and cognition. *Nat Rev Neurosci*. 9:148–158.

- Petersen K, Van Wouwe N, Stark A, et al. 2018. Ventral striatal network connectivity reflects reward learning and behavior in patients with Parkinson's disease. *Hum Brain Mapp.* 39:509–521.
- Phillips ML, Young AW, Senior C, Brammer M, Andrew C, Calder AJ, Bullmore ET, Perrett DI, Rowland D, Williams SC, Gray JA, David AS. 1997. A specific neural substrate for perceiving facial expressions of disgust. *Nature.* 389:495–498.
- Pinel P, Piazza M, Le Bihan D, Dehaene S. 2004. Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron.* 41:983–993.
- Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, Petersen SE. 2011. Functional network organization of the human brain. *Neuron.* 72:665–678.
- Ruppelchter S, Romaniuk L, Series P, et al. 2020. Blunted medial prefrontal cortico-limbic reward-related effective connectivity and depression. *Brain.* in press. doi:10.1093/brain/awaa106
- Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD. 2003. The neural basis of economic decision-making in the Ultimatum Game. *Science.* 300:1755–1758.
- Santangelo V. 2018. Large-scale brain networks supporting divided attention across spatial locations and sensory modalities. *Front Integr Neurosci.* 12:8.
- Saxe R, Haushofer J. 2008. For love or money: A common neural currency for social and monetary reward. *Neuron.* 58:164–165.
- Scheres A, Milham MP, Knutson B, Castellanos FX. 2007. Ventral striatal hypo-responsiveness during reward anticipation in attention-deficit/hyperactivity disorder. *Biol Psychiatry.* 61:720–724.
- Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius MD. 2007. Dissociable intrinsic connectivity networks for salience processing and executive control. *J Neurosci.* 27:2349–2356.
- Singer T, Critchley HD, Preusschoff K. 2009. A common role of insula in feelings, empathy and uncertainty. *Trends Cogn Sci.* 13:334–340.
- Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, Watkins KE, Ciccarelli O, Cader MZ, Matthews PM, Behrens TE. 2006. Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *Neuroimage.* 31:1487–1505.
- Swallow SR, Kuiper NA. 1988. Social comparison and negative self-evaluations: An application to depression. *Clin Psychol Rev.* 8:55–76.
- Swencionis JK, Fiske ST. 2014. How social neuroscience can inform theories of social comparison. *Neuropsychologia.* 56:140–146.
- Tamir DI, Hughes BL. 2018. Social rewards: From basic social building blocks to complex social behavior. *Perspect Psychol Sci.* 13:700–717.
- Trachtenberg JT, Chen BE, Knott GW, Feng G, Sanes JR, Welker E, Svoboda K. 2002. Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. *Nature.* 420:788–794.
- Uddin LQ. 2015. Salience processing and insular cortical function and dysfunction. *Nat Rev Neurosci.* 16:55–61.
- Vincent JL, Kahn I, Snyder AZ, Raichle ME, Buckner RL. 2008. Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *J*

- Neurophysiol. 100:3328–3342.
- Vogel EA, Rose JP, Okdie BM, Eckles K, Franz B. 2015. Who compares and despairs? The effect of social comparison orientation on social media use and its outcomes. *Pers Individ Dif*. 86:249–256.
- Vostroknutov A, Tobler PN, Rustichini A. 2012. Causes of social reward differences encoded in human brain. *J Neurophysiol*. 107:1403–1412.
- Wager TD, Atlas LY, Lindquist MA, Roy M, Woo CW, Kross E. 2013. An fMRI-based neurologic signature of physical pain. *N Engl J Med*. 368:1388–1397.
- Watson R, Latinus M, Charest I, Crabbe F, Belin P. 2014. People-selectivity, audiovisual integration and heteromodality in the superior temporal sulcus. *Cortex*. 50:125–136.
- White JB, Langer EJ, Yariv L, Welch JC. Frequent social comparisons and destructive emotions and behaviors: The dark side of social comparisons. *J Adult Dev*. 2006;13:36-44.
- Zaki J, Davis JI, Ochsner KN. 2012. Overlapping activity in anterior insula during interoception and emotional experience. *Neuroimage*. 62:493–499.
- Zang Y, Jiang T, Lu Y, He Y, Tian L. 2004. Regional homogeneity approach to fMRI data analysis. *Neuroimage*. 22:394–400.
- Zhang S, Hu S, Chao HH, Li CR. 2017. Hemispheric lateralization of resting-state functional connectivity of the ventral striatum: an exploratory study. *Brain Struct Funct*. 222:2573–2583.
- Zou QH, Zhu CZ, Yang Y, Zuo XN, Long XY, Cao QJ, Wang YF, Zang YF. 2008. An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: fractional ALFF. *J Neurosci Methods*. 172:137–141.

Study #2: Roles of the MPFC and Insula in Impression Management under Social Observation (Authors: Leehyun Yoon, Kwangwook Kim, Daehyun Jung, Hackjin Kim; Currently Under Revision in Social, Cognitive, and Affective Neuroscience)

1.0 SUMMARY

People often engage in impression management by presenting themselves and others in a socially desirable manner. However, its specific behavioral manifestation and underlying neural mechanism are unknown. In this functional magnetic resonance imaging study, we investigated the neural mechanism of impression management during self- and friend-evaluation. Only participants assigned to the observation (OBS) group, not the control (CON) group, were told that their responses would be monitored by the experimenters. They answered how well positive and negative trait adjectives described themselves or their friends. The behavioral results showed that the OBS group were more likely to reject negative traits for self-evaluation and to accept positive traits for friend-evaluation. An independent study revealed that demoting negative traits for oneself and promoting positive traits for a friend serve to manage one's impression. The neuroimaging results showed that, in the OBS vs. the CON group, the rostromedial prefrontal cortex (RMPFC) and anterior insula (AI) activity showed stronger correlations with negatively-valenced adjectives during self-evaluation and stronger correlations with positively-valenced adjectives during friend-evaluation. The present study suggests the role of RMPFC and AI in impression management specifically through the socially desirable target-evaluation during observation.

2.0 INTRODUCTION

Impression management plays a critical role in many interpersonal behaviors (Leary and Kowalski, 1990). People present themselves in a socially acceptable manner to build and maintain favorable images of themselves (Leary, 1995; Schlenker, 2012) and display socially desirable behavior (e.g., prosocial behavior) in the presence of others (Soetevent, 2005; Bereczkei et al., 2010; Giskevicius et al., 2010; Kimura et al., 2012; Van Hoorn et al., 2016). However, its neurobiological mechanism is unclear. The present study aims to identify impression management neurocircuitry by investigating how subjects' evaluation of themselves and others changes to be more socially desirable under observation.

Although few studies have directly investigated neural mechanisms of impression management during target-evaluation, several have reported the neural mechanisms associated with some key factors closely related to impression management, e.g., social observation (Leary and Kowalski, 1990), socially desirable behavior (Leary et al., 2011), and self-conscious emotions (Doherty and Schlenker, 1991; Leary, 2004). Notably, RMPFC function has been consistently associated with these key factors.

Firstly, being socially observed, which often facilitates impression management, has elicited RMPFC activity. For example, RMPFC activity increased during judgment about self and social appropriateness under social observation

(Izuma et al., 2010), when receiving public success or failure in a cognitive estimation task (Müller-Pinzler et al., 2015), and when playing an economic game under social observation (Van Hoorn et al., 2016). Furthermore, trait social anxiety, or a dispositional sensitivity to anticipated social evaluation, was positively associated with RMPFC activity during social observation (Müller-Pinzler et al., 2015). Importantly, increased RMPFC activity during these situations was accompanied by a subjective experience of self-conscious emotions (Somerville et al., 2013). People also showed increased prosocial consumer decisions and stronger RMPFC activity encoding the value of prosocial decisions under social observation (Jung et al., 2018).

Secondly, RMPFC activity has been strongly associated with norm-compliant or socially desirable behaviors i.e., the typical way impression management is expressed, especially in the absence of specific instructions about an audience's traits or values (Leary et al., 2011). Specifically, an increase in RMPFC activity was associated with more generous donations (Hare et al., 2010; Mathur et al., 2010; Tusche et al., 2016), more frequent daily helping behavior (Rameson et al., 2012), sending a more prosocial email to a victim of social exclusion (Masten et al., 2011), and an increase in purchasing ethical products (Jung et al., 2018). RMPFC activity has been linked to compliant behavior after social influence e.g., after receiving a persuasive message, RMPFC activity predicted message-compliant behavior, such as increased sunscreen use (Falk et al., 2010), reduced smoking behavior (Cooper et al., 2015; Pegors et al., 2017), and increased engagement in sharing the anti-smoking message with peers (Pei et al., 2019).

Thirdly, RMPFC activity has been associated with self-conscious emotions, which are often elicited by concerns about how one is perceived by others (Leary, 2004), and whether one's behavior is aligned with social standards (Leary and Kowalski, 1990; Tracy and Robins, 2004; Tangney et al., 2007). Previous neuroimaging studies have linked RMPFC activity to the subjective experience of embarrassment (Takahashi et al., 2004; Burnett et al., 2009; Bas-Hoogendam et al., 2017), shame (Michl et al., 2012), guilt (Shin et al., 2000; Takahashi et al., 2004; Zahn et al., 2008; Burnett et al., 2009; Basile et al., 2011; Wagner et al., 2011; Fourie et al., 2014; Gilead et al., 2016), and pride (Zahn et al., 2008). Brain lesion studies and structural neuroimaging data also support the role of the RMPFC in self-conscious emotions. Specifically, patients with brain lesions in the RMPFC failed to exhibit self-conscious emotions (Sturm et al., 2006; Sturm et al., 2008; Krajbich et al., 2009; Moll et al., 2011) and engaged in socially inappropriate self-disclosure (Beer et al., 2006). Furthermore, reduced RMPFC volume was associated with reduced embarrassment, as measured by physiological response and facial expression while watching oneself singing in front of others (Sturm et al., 2012). Together with the RMPFC, the anterior insula (AI) has been involved in behaviors and psychological processes related to impression management. An increase in AI-RMPFC functional coupling predicted prosocial behavior under social observation (Jung et al., 2018). Moreover, lack of embarrassment following an inappropriate social behavior has been reported among patients with behavior-variant frontotemporal dementia (Snowden et al., 2001; Sturm et al., 2006; Sturm et al., 2008; Moll et al., 2011), which is associated with neurodegeneration in the AI and

RMPFC (Rosen et al., 2002; Seeley, 2010; Ibañez and Manes, 2012). In addition, people with social anxiety or social phobia exhibit more activity in these two regions when receiving a social evaluation (Heitmann et al., 2014), and when reading stories of embarrassing social transgressions (Blair et al., 2010).

Based on these findings, we predicted that neurocircuitry involving the RMPFC and AI would be implicated in impression management or pursuing a favorable self-image, particularly under social observation. We defined the RMPFC as an MPFC subregion vertically located between $z = -10$ and $z = 20$ in Montreal Neurological Institute (MNI) coordinates (Dixon et al., 2017) and caudally extended to the genu of the corpus callosum. While the functional division of MPFC has not been clearly defined, recent endeavors to define distinctive roles of MPFC subregions have suggested that the RMPFC has a unique function in self-monitoring based on social standards (Dixon et al., 2017), while adjacent regions, like the dorsomedial prefrontal cortex and ventromedial prefrontal cortex, are implicated in inference of others' mental states or traits and affective processing or integrated situational knowledge, respectively (De La Vega et al., 2016; Lieberman et al., 2019). Our region of interest (ROI) extends to the genu of the corpus callosum, thus covering clusters from many studies involving self-conscious emotions (Somerville et al., 2013; Bas-Hoogendam et al., 2017) and social observation (Izuma et al., 2010).

In our experimental design, participants were randomly allocated to an observation (OBS) or control (CON) group and performed a trait-evaluation task where they rated how well each trait adjective described themselves or their close friends. Based on previous findings that people strategically use self- (Sedikides et al., 2007; Sedikides et al., 2015) as well as other-evaluation (Bolino et al., 2008; Xiaohua Chen et al., 2009) to promote a favorable self-image, we predicted that the trait-evaluation of self and others of OBS groups would be biased toward the goal of impression management compared to CON group. At the neural level, we expected that RMPFC and AI activity would be consistent with behavioral evidence for impression management during the task. In addition, to confirm that the observation effect of target-evaluation reflects the enhanced motivation for impression management, we asked the independent sample to rate their impression of a person who described themselves or their friends with the adjectives we used for the main fMRI task. Then, we examined whether the impression ratings are associated with the valence level of adjectives in the conditions where we found the observation effect.

3.0 METHODS, ASSUMPTIONS, AND PROCEDURES

3.1 Participants

Forty-five participants (mean age = 23.7 years, 21 males) were recruited; 23 were assigned to the CON group (mean age = 23.1 years, 12 males) and 22 to the OBS group (mean age = 24.3 years, 9 males). Two participants in the OBS group were excluded from the functional magnetic resonance imaging (fMRI) data analyses due to excessive head movement (>3 mm), leaving 20 participants in the OBS group

(mean age = 24.2 years, 8 males). We performed a power analysis with the G*Power software to ensure the sample size was sufficient to obtain scientifically meaningful results. Assuming a medium effect size of 0.06 and an alpha level of 0.05, to achieve a power of 0.90 for a mixed ANOVA with 2 groups and 4 repeated measurements, the G*Power analysis yielded a required sample size of 30, that is, 15 participants per group. In addition, we referred two between-subject-design fMRI studies (Heitmann et al., 2014; Peterburs et al., 2016), which examined the closely related topic, specifically the neural mechanism of heightened sensitivity to social evaluation. These two studies recruited 16 or 18 participants for each group. All participants were right-handed and screened for a history of psychiatric or neurological disease and MRI testing eligibility. Experimental procedures were approved by the Institutional Review Board of Korea University. All potential health-related risks were comprehensively explained to participants. All participants were screened using a self-report questionnaire. They all provided informed consent and compensated KRW 30,000 for their participation.

3.2 Stimuli

For the trait-evaluation task, 150 adjectives were selected from a list of personality trait adjectives originally used in a previous study (Anderson, 1968), which had been translated into Korean (Sul *et al.*, 2012). Valence rating of adjectives was obtained from the previous study (Sul *et al.*, 2012), where the participants ($N = 80$) were asked to rate the social desirability of a person with each trait using a 7-point Likert scale (1 = Highly undesirable, 7 = Highly desirable). Among the 150 adjectives, 60 were negative (mean rating ≤ 3), and 60 were positive (mean rating ≥ 5). For the main analysis, we analyzed the participants' decisions for positive and negative adjectives separately in order to distinguish between target-promoting and target-demoting evaluations. The remaining 30 adjectives had a moderate valence level ($3 < \text{mean rating} < 5$) and served as filler trials. Descriptive statistics of valence ratings for negative, positive, and filler conditions and example adjectives are presented in **Supplementary Table 1**.

3.3 Impression Ratings from Independent Samples

To estimate the values of impression management for self- and friend-evaluation with each adjective, by recruiting two independent samples for self-evaluation ($N = 15$, mean age = 24.1 years, 7 males) and friend-evaluation ($N = 16$, mean age = 24.4 years, 8 males), we measured the impression of a hypothetical person who evaluates themselves and a friend with 150 trait adjectives. Participants were instructed to rate (1 = very negative, 7 = very positive) their impression of a person who described themselves with a particular adjective (e.g., "I am competent"), and a person who described their friend with a particular adjective (e.g., "My friend is kind"). We ensured that impression ratings of self-evaluation and friend-evaluation were obtained from two separate groups of participants to avoid potential rating contamination. While our primary purpose of collecting impression ratings for each adjective was to confirm whether the valence level of SN, SP, FN, FP conditions has

a role of modulating the impression, we additionally used impression ratings to provide data about the effect of impression value of adjectives on target-evaluation of our fMRI sample (see **Supplementary Text 1** for analytical procedure and the result).

3.4 Experimental Procedures

Upon arrival, participants were escorted to the MRI control room, received detailed experiment instructions, and randomly assigned to two groups (OBS or CON). OBS participants were told that due to an unexpected technical problem, two researchers would monitor the experiment and manually record their responses during the task (**Figure 1**), whereas CON participants were not informed of any monitoring. In our task, the target of friend was specified as one's fifth close friend. We avoided choosing a friend that was too close to self as high closeness can elicit overlapping representation between self and others (Aron *et al.*, 1992; Mashek *et al.*, 2003; Aron *et al.*, 2004). We also avoided choosing an unfamiliar person to ensure participants possessed sufficient person knowledge necessary for completing the task. To ensure that participants will do the friend-evaluation with a specific friend in mind and let participants vividly remind of the friend, we assessed a questionnaire (**Supplementary Text 2**) that requires the brief profile of the fifth close friend immediately before scanning.

3.5 fMRI Task

The task contained 75 self-evaluation and 75 friend-evaluation trials. In each trial, after the fixation cross (2~4 s), a target for evaluation was presented for 1 s (i.e., "I consider myself to be" for the self-evaluation condition, and "I consider my friend to be" for the friend-evaluation condition). Then, a trait adjective and a 4-point rating scale were presented for 5 s (**Figure 1**). Participants evaluated how appropriately each trait adjective described themselves or their fifth closest friend by pressing one of four buttons corresponding to a value on the scale (1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree). The two conditions (self and friend) were presented in a pseudo-randomly mixed order. Pairings between trait adjectives and targets were counterbalanced across participants. MATLAB toolbox Cogent 2000 (www.vislab.ucl.ac.uk/cogent.php) was used to run the task and present the stimuli. Participants were debriefed upon task completion.

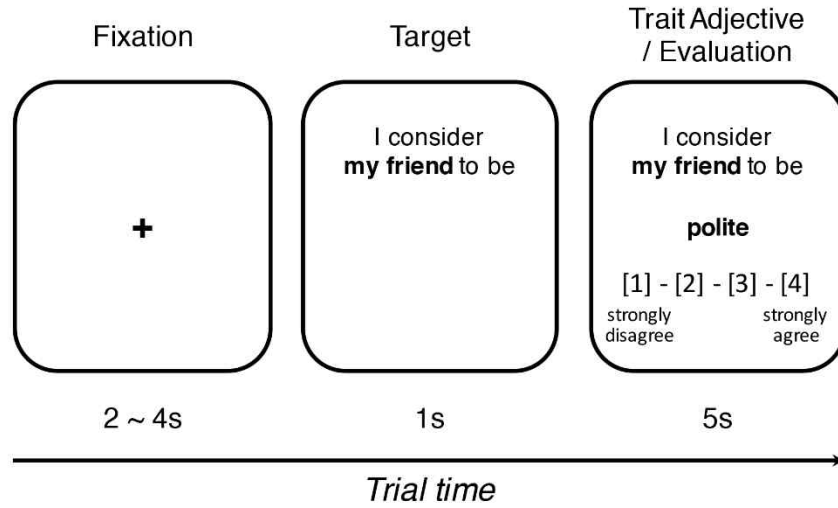


Figure 4. A schematic diagram of a typical trial in a trait-evaluation task. Participants were asked to evaluate how each trait adjective describes themselves or their fifth closest friend. ‘My friend’ was replaced with ‘myself’ in the self-evaluation trials. The OBS Group was told that two researchers would be monitoring the task, whereas the CON Group was not informed of any possibility of being monitored.

3.6 Image Acquisition

Neuroimaging data were acquired using a Siemens MAGNETOM Trio, a 3T Tim system with a 12-channel head matrix coil. We obtained T1-weighted anatomical images (repetition time [TR] = 1900 ms; echo time [TE] = 2.52 ms; flip angle [FA] = 9°; 1 × 1 × 1 mm in-plane resolution; 256 × 256 matrix size). T2*-weighted functional images were obtained using gradient-echo echo-planar pulse sequences (TR = 2000 ms; TE = 30 ms; FA = 90°; field-of-view (FOV) = 220 mm; 78 × 78 matrix; 36 slices; voxel size = 2.8 × 2.8 × 3.0 mm). Experimental stimuli were presented on an MR-compatible LCD monitor mounted on a head coil (refresh rate: 85 Hz; display resolution: 800 × 600 pixels). Participants used a response grip to respond during the task. Task duration was approximately 15 minutes.

3.7 Data Analysis

3.7.1 Behavioral Analysis

3.7.1.1 Subject-level Analysis

$$y = \beta_{SN} x_{SN} + \beta_0 \quad \text{Equation (1)}$$

$$y = \beta_{SP} x_{SP} + \beta_0 \quad \text{Equation (2)}$$

$$y = \beta_{FN} x_{FN} + \beta_0 \quad \text{Equation (3)}$$

$$y = \beta_{FP} x_{FP} + \beta_0 \quad \text{Equation (4)}$$

For the subject-level analysis, we ran four linear regression models (Equations (1) – (4)) to estimate the extent to which the valence level of positive and negative adjectives presented for self- and friend-evaluation (i.e., self-negative [SN], self-positive [SP], friend-negative [FN], or friend-positive [FP]) influenced participants' decisions. In each model, y denotes the participants' decision about how much they agree or disagree with the applicability of the given adjective to the target. Meanwhile, x denotes the mean-centered valence level of the presented adjective, and β_0 represents the intercept. The estimate of β (i.e., $\hat{\beta}$) indicates the degree to which the participant's decision was modulated by an adjective's valence level. To ensure that the high beta estimates of all four conditions uniformly represent desirable target-evaluation, valence level (i.e., x) was reversed-coded in Equation (1) and Equation (3). After reverse-coding, high $\hat{\beta}_{SN}$ and $\hat{\beta}_{FN}$ indicate that participants were more likely to disagree as a function of the presented negative adjective's valence level during self- and friend-evaluation, respectively. Likewise, high $\hat{\beta}_{SP}$ and $\hat{\beta}_{FP}$ indicate that participants were more likely to agree as a function of the presented positive adjective's valence level during self- and friend-evaluation, respectively.

3.7.1.2 Group-level Analysis: Three-way Analysis of Variance (ANOVA)

For the group-level analysis, four parameter estimates ($\hat{\beta}_{SN}, \hat{\beta}_{SP}, \hat{\beta}_{FN}, \hat{\beta}_{FP}$) from each participant were entered into a mixed ANOVA with group (OBS or CON) as the between-subject factor, and target (self or friend) and valence (negative or positive) as the within-subject factors. Post-hoc analyses were conducted to confirm the observation effects in specific conditions. IBM SPSS 23.0 statistical software was used for this analysis.

3.7.2 Analysis of Independent Sample Data: Correlational Analysis

Using data collected from the independent sample, we calculated the mean impression ratings of each trait adjective after excluding the outliers (i.e., greater than three standard deviations from the mean). Four correlational analyses were performed for each condition (i.e., SN, SP, FN, and FP conditions) to examine whether impression ratings are modulated by the valence level of adjectives that the hypothetical speaker used for target-evaluation.

3.7.3 fMRI Data Analysis

3.7.3.1 Preprocessing

Statistical Parametric Mapping (SPM) 8 was used for all neuroimaging data analyses, unless otherwise stated. For data pre-processing, the slice timing differences were corrected for calibrating the interleaved image acquisition sequence and the realignment was conducted to correct for head motion with the reference of the mean brain image. The resulting images were then spatially normalized with respect to the MNI echo planar imaging template, and each voxel was resampled at $2 \times 2 \times 2$ mm

resolution. Lastly, the resampled data were smoothed using a 6-mm full-width-at-half-maximum (FWHM) Gaussian filter to further enhance the signal-to-noise ratio.

3.7.3.2 Definition of ROI

Based on our prediction of the role of the RMPFC and AI in impression management, we conducted the analysis restricting the search area to combined anatomical masks of the two regions. The RMPFC mask combined the bilateral medial orbitofrontal regions, the rectus regions, the superior medial frontal regions, and the anterior cingulate cortex from the anatomical automatic labeling atlas (Tzourio-Mazoyer *et al.*, 2002), within the boundaries mentioned in the introduction (i.e., $-10 < z < 20$ and anterior to the genu of the corpus callosum). MNI coordinates for the RMPFC mask boundary were $-17 < x < 17$, $36 < y < 72$, and $-10 < z < 20$. The AI mask combined the ventral and dorsal anterior insula from a previous study reporting functional subdivisions of the insula (Deen *et al.*, 2010). The AI mask boundaries were $-46 < x < -27$ (left), $27 < x < 46$ (right), $-12 < y < 24$, and $-20 < z < 8$. Anatomical masks were generated and combined using MRIcron (<http://people.cas.sc.edu/rorden/mricron/index.html>) and MarsBaR (Brett *et al.*, 2002).

3.7.3.3 Statistical Threshold

To correct for multiple comparisons, we first ran a small volume correction (SVC) with the a priori ROI, followed by an exploratory analysis with the whole-brain gray matter as the search volume. The statistical threshold was determined using Analysis of Functional NeuroImages (AFNI) software (Cox, 1996). Intrinsic smoothness was estimated using the spatial autocorrelation function (ACF) option in AFNI's 3dFWHMx, and the resulting ACF estimates were used to run Monte Carlo simulations via AFNI's 3dClustSim. The desired cluster size for surviving multiple comparison at $\alpha < 0.05$ over the ROI or whole-brain gray matter was obtained with an initial uncorrected p-value of 0.001. The desired cluster size for the ROI analysis was 15.5, and 67.3 for the whole-brain gray matter.

3.7.3.4 Subject-level Analysis

General linear modeling (GLM) was used to estimate the neural responses to the experimental conditions of interest. Each individual's GLM included the regressors of hemodynamic responses time-locked to four events of interest: presentation of a (1) negative and (2) positive trait adjective during self-evaluation, and presentation of a (3) negative and (4) positive trait adjective during friend-evaluation. Each regressor was parametrically modulated by the trait adjective valence level, and valence level of negative adjectives was reverse-coded such that a higher value indicated more negative valence. Additionally, the onset times of filler trials (i.e., moderately-valenced adjectives) were entered as a regressor of no interest. All button-press responses were added to the model to reduce noise caused by button-pressing. Six motion regressors were also included as regressors of no interest.

3.7.3.5 Group-level Analysis: Three-way ANOVA

We hypothesized that RMPFC and AI activity would reflect the observation effect shown in the behavioral results. To test this and search for other brain regions showing the three-way interaction effect, we used a $2 \times 2 \times 2$ mixed factorial design in the second-level analysis of SPM, with group (OBS or CON) as the between-subject factor, and target (self or friend) and valence (negative or positive) as the two within-subject factors. Parametric modulation (PM) maps of each individual were entered into the six cells in the following order: (1) SN condition of CON participants, (2) SN condition of OBS participants, (3) SP condition of CON participants, (4) SP condition of OBS participants, (5) FN condition of CON participants, (6) FN condition of OBS participants, (7) FP condition of CON participants, and (8) FP condition of OBS participants.

To examine the three-way interaction effect parallel to the behavioral results, we planned the following contrasts: [OBS/Self/Neg, CON/Self/Pos, CON/Friend/Neg, OBS/Friend/Pos > CON/Self/Neg, OBS/Self/Pos, OBS/Friend/Neg, CON/Friend/Pos]. Using MarsBaR, we extracted the beta estimates from the regions showing a significant interaction effect and carried out post-hoc analyses to confirm the specific pattern of the three-way interaction. Exploratory analyses were performed to search brain regions showing significant three 2 way interactions (group * valence, group * target, valence * target) and three main effects (group, valence, target). Additionally, we explored significant regions correlated with valence level during each condition (i.e., observation, control, negative, positive, self, friend).

4.0 RESULTS AND DISCUSSION

4.1. Behavioral Results: Three-way ANOVA and descriptive statistics

The behavioral results revealed a significant three-way interaction ($F(1,41) = 5.24, p = 0.027$) (**Figure 2**). Post-hoc analysis indicated that OBS participants were more likely than CON participants to disagree with adjectives as the level of negativity increased during self-evaluation ($F(1,41) = 8.823, p = 0.005$), and were more likely to agree with adjectives as the level of positivity increased during friend-evaluation ($F(1,41) = 16.051, p < 0.001$). There was also a significant main effect of group ($F(1,41) = 14.66, p < 0.001$), indicating that OBS participants exhibited a greater tendency to demote a negative evaluation or promote a positive evaluation. A significant main effect of valence ($F(1,41) = 31.018, p < 0.001$) indicated a greater tendency to demote a negative evaluation than promote a positive evaluation. As our main behavioral analysis investigated the modulation of valence level on ratings within SN, SP, SN, FP trials, for illustrative purpose, we also presented the mean ratings for three divided valence level of each condition (i.e., high, medium, low valence level).

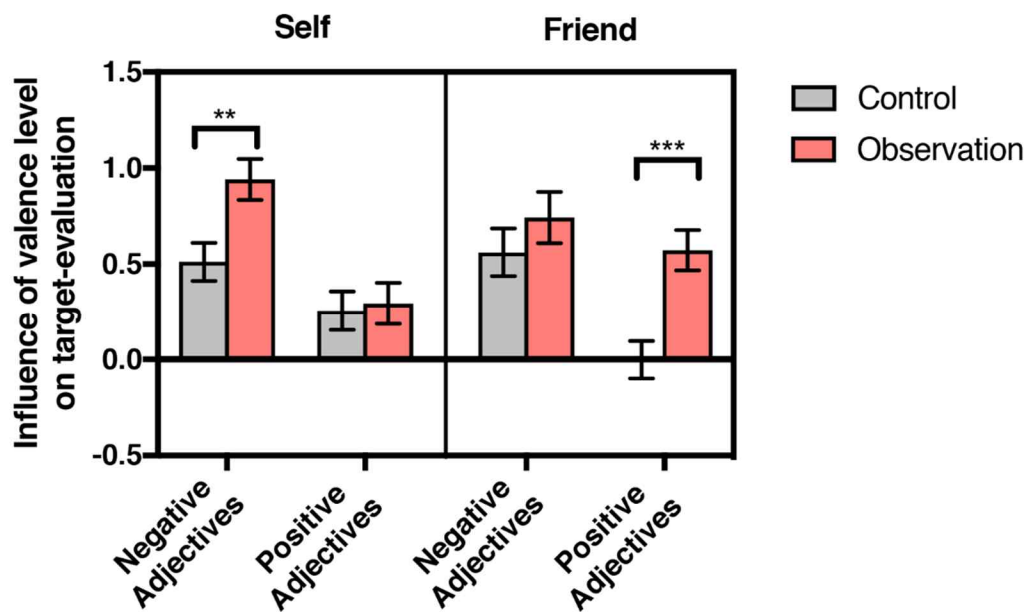


Figure 5. Behavioral results from the trait-evaluation task. Bar graphs showing the degrees to which the valence level of trait adjectives affected self- and friend-evaluation. A significant three-way (i.e., Group $\times$ Target $\times$ Valence) interaction effect indicates that the observation group (vs. control group) displayed less negative self-evaluation and more positive friend-evaluation (** $p < 0.01$, *** $p < 0.001$).

4.2 Results of Independent Sample Data: Correlational Analysis

The correlational analysis of independent sample data was conducted to confirm the observation effects found in the behavioral results, which reflected a strategic effort towards impression management. The mean impression of the hypothetical speaker significantly decreased as a function of the valence level of negative adjectives the speaker used for self-evaluation ($r = 0.69$, $p < 0.0001$; **Figure 3A**), whereas mean impression was not associated with the valence level of positive adjectives the speaker used for self-evaluation ($r = 0.13$, $p = 0.32$; **Figure 3A**). In addition, mean impression significantly decreased as a function of the valence level of negative adjectives the speaker used for friend-evaluation ($r = 0.6$, $p < 0.0001$; **Figure 3B**) and significantly increased as a function of the valence level of positive adjectives the speaker used for friend-evaluation ($r = 0.64$, $p < 0.0001$; **Figure 3B**). This suggests that demoting the negative traits of both self and a friend, and promoting the positive traits of a friend, could serve the goal of impression management.

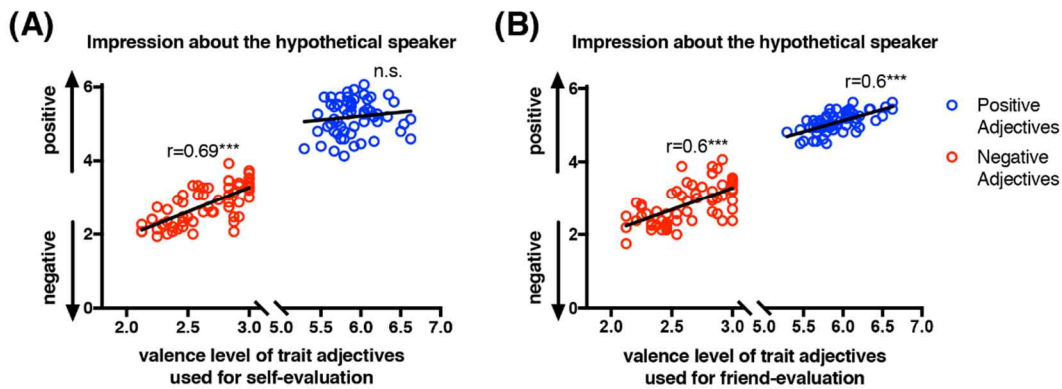


Figure 6. Impression rating data from independent sample. (A) Impression ratings during self-evaluation: Participants from the independent sample formed a more negative impression of a hypothetical speaker who used more negative trait adjectives for self-evaluation, whereas they did not necessarily form a more positive impression of a hypothetical speaker who used more positive trait adjectives for self-evaluation ($***p < 0.001$). (B) Impression ratings during friend-evaluation: Participants from the independent sample formed a more negative impression of a hypothetical speaker who used more negative trait-adjectives for friend-evaluation and formed a more positive impression of a hypothetical speaker who used more positive trait-adjectives for friend-evaluation ($***p < 0.001$).

4.3 fMRI Results: Three-way ANOVA

Similar to the behavioral data analysis, we performed a voxel-wise three-way ANOVA with group (observation or control), target (self or friend), and valence (negative or positive). To examine the three-way interaction effects with the PM maps of valence level, the following planned contrasts were used: $[-1, 1, 1, -1, 1, -1, -1, 1]$. This analysis showed significant clusters in the RMPFC (peak coordinate = -12, 48, 6; cluster size = 17; SVC family-wise error rate (FWE) $p < 0.05$) (**Figure 4A**) and AI (peak coordinate = 30, 14, -14; cluster size = 26; SVC FWE $p < 0.05$) (**Figure 4B**). Post-hoc analyses revealed that RMPFC and AI activity in the OBS vs. the CON group was more strongly correlated with the valence level of negative adjectives during self-evaluation (RMPFC: $F(1, 41) = 9.089, p = 0.004$; AI: $F(1, 41) = 8.295, p = 0.006$) and the valence level of positive adjectives during friend-evaluation (RMPFC: $F(1, 41) = 4.499, p = 0.04$; AI: $F(1, 41) = 5.727, p = 0.021$), which was consistent with the behavioral results. In addition, the CON group showed greater AI activity than the OBS group when tracking the valence level of negative adjectives during friend-evaluation ($F(41, 1) = 4.689, p = 0.036$). Exploratory whole-brain analyses did not reveal any other significant clusters.

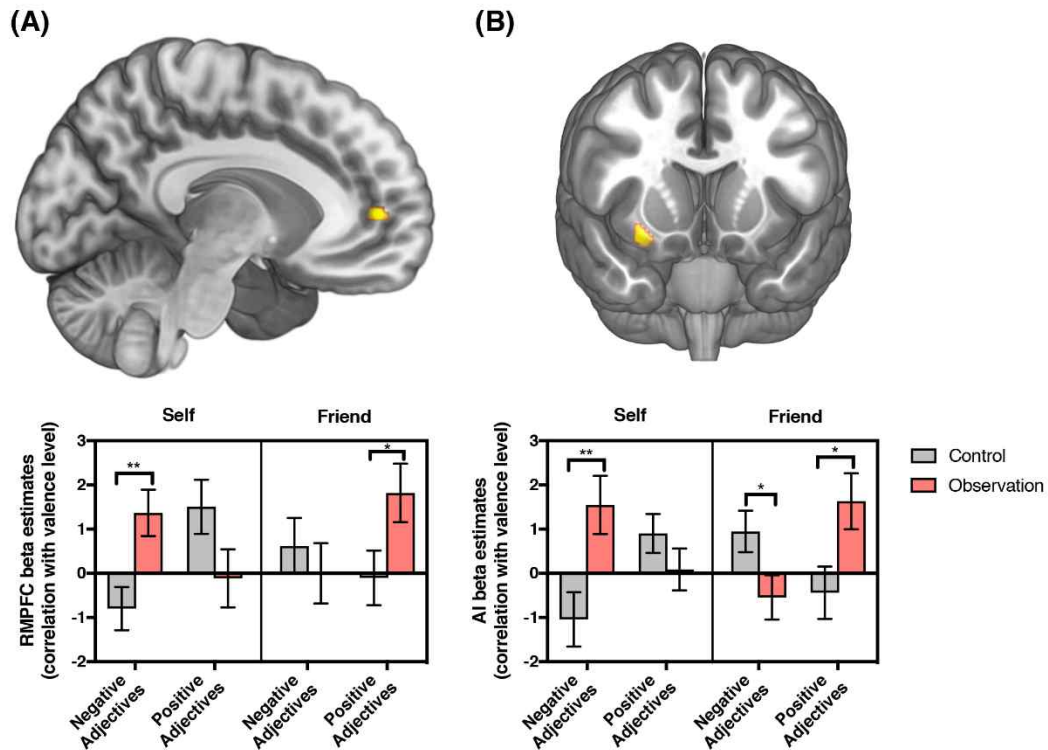


Figure 4. fMRI Results. (A) The RMPFC cluster (peak coordinates: -12, 48, 6; upper panel) showing a significant three-way interaction effect in the voxel-wise three-way ANOVA, and a bar graph (bottom panel) showing condition-specific parameter estimates in the cluster (* $p < 0.05$, ** $p < 0.01$). (B) The insula cluster (peak coordinates: 30, 14, -14; upper panel) showing a significant three-way interaction effect in the voxel-wise three-way ANOVA, and a bar graph (bottom panel) showing condition-specific parameter estimates in the cluster (* $p < 0.05$, ** $p < 0.01$).

The exploratory analyses of 2 way interactions and main effects showed no significant region. We found significant MPFC activity correlated with valence level in observation, self, friend, negative, and positive condition.

4.4 Discussion

This study aimed to investigate the behavioral manifestation and neural mechanism of impression management during self- and friend-evaluation under social observation. The behavioral results showed that people evaluated themselves and a friend in a more socially desirable way under social observation. Specifically, OBS participants demoted negative evaluation of self and promoted positive evaluation of their friend, compared to the CON group. An independent behavioral study confirmed that this behavioral strategy serves the goal of transmitting a favorable self-image to observers by showing that people form a better impression of a speaker who evaluates oneself with less negative adjectives and one's friend with more positive adjectives. The neuroimaging analysis revealed that the OBS group (vs. CON group) show more

RMPFC and AI activity when tracking the valence level of negative adjectives during self-evaluation and the valence level of positive adjectives during friend-evaluation. Combined with the behavioral findings, these neuroimaging data suggest the RMPFC and AI contribute to impression management, playing critical roles in computing the value of preventing an unfavorable impression, and promoting a favorable self-image on a trial-by-trial basis, the function of which is facilitated under social observation.

Consistent with abundant evidence that social observation promotes socially desirable responding (Soetevent, 2005; Bereczkei *et al.*, 2010; Griskevicius *et al.*, 2010; Kimura *et al.*, 2012; Van Hoorn *et al.*, 2016), our data confirmed that social observation suppressed negative self-evaluation and facilitated positive friend-evaluation. Importantly, both behaviors may serve the common goal of impression management, as our independent behavioral study demonstrated that less negative self-evaluation and more positive friend-evaluation convey a more favorable impression to observers.

Self-evaluation: Social observation reduced negative self-evaluation but did not facilitate positive self-evaluation. Consistent with this, the independent behavioral study also revealed that people prefer those who evaluate themselves less negatively, but not necessarily those who evaluate themselves more positively. Together, this suggests that avoiding negative self-evaluation, but not increasing positive self-evaluation, is suitable for the goal of transmitting a favorable self-image to observers, which is in line with previous findings that a modesty bias is often observed during self-evaluation (Sedikides *et al.*, 2015) and that self-enhancing self-presentation is often perceived as dislikable (Powers and Zuroff, 1988; Wosinska *et al.*, 1996; Van Tongeren *et al.*, 2014).

Friend-evaluation: Social observation increased positive friend-evaluation, consistent with the result from the independent sample that people prefer those who evaluate their friends more positively. These findings are also in line with previous findings that people with other-oriented personality (e.g., warm, trustworthy, and humble), e.g., those who speak positively about others, are well-liked by others (Wortman and Wood, 2011). For the negative traits, however, our behavioral results showing that observation did not reduce negative friend-evaluation seem inconsistent with the result from the independent behavioral study that people who speak unwell of others were perceived less favorably. A possible account of such a discrepancy could be that (1) avoiding a negative friend-evaluation (but not promoting a positive friend-evaluation) is fully internalized such that it can be entertained even in private situations (i.e., CON group), or that (2) participants may have chosen a friend who does not possess severe enough negative traits for the trait-evaluation task, which could lead to a possible ceiling effect in avoiding negative friend-evaluation, as observed in the present study.

In the present study, OBS participants vs. CON participants showed increased RMPFC activity tracking the trial-by-trial fluctuations in the valence level of trait adjectives during negative self-evaluation and positive friend-evaluation. The observed role of RMPFC activity in impression management is consistent with its activity during social observation (Izuma *et al.*, 2010; Somerville *et al.*, 2013; Müller-Pinzler *et al.*, 2015; Van Hoorn *et al.*, 2016; Jung *et al.*, 2018), during the experience of self-conscious emotions (Shin *et al.*, 2000; Takahashi *et al.*, 2004; Zahn *et al.*, 2008; Burnett *et al.*, 2009; Basile *et al.*, 2011; Wagner *et al.*, 2011; Michl *et al.*, 2012;

Somerville *et al.*, 2013; Fourie *et al.*, 2014; Gilead *et al.*, 2016), and when engaging in socially desirable behavior (Hare *et al.*, 2010; Mathur *et al.*, 2010; Masten *et al.*, 2011; Rameson *et al.*, 2012; Tusche *et al.*, 2016). Moreover, our result is consistent with recent studies that have demonstrated the role of this region in dynamically monitoring self-value in the eyes of others e.g., RMPFC activity tracked the discrepancy between others' and one's own view of oneself (Korn *et al.*, 2012). In a similar vein, RMPFC function and its development have shown to play a role in updating self-worth by encoding prediction errors between experienced and expected social evaluation (Will *et al.*, 2017; Yoon *et al.*, 2018).

Importantly, even in the absence of social observation, RMPFC activity has been strongly implicated in trait-judgment (Van Overwalle, 2009; Schurz *et al.*, 2014), especially about self and close others (Kelley *et al.*, 2002; Ochsner *et al.*, 2005; Vanderwal *et al.*, 2008; Murray *et al.*, 2012; Araujo *et al.*, 2015). Although several studies have linked RMPFC activity during trait-judgment to similarity (Mitchell *et al.*, 2005; Mitchell *et al.*, 2006), closeness (Krienen *et al.*, 2010), or knowledgeability (Welborn and Lieberman, 2015), the specific role of the RMPFC in trait-judgment remains disputed. By showing that social observation facilitates RMPFC function critical for impression management, our study provides a novel view that the increased RMPFC activity during trait judgment of self and close others reported in previous studies reflects its allostatic regulatory function largely related to impression management or the motivation for socially desirable behavior. Participants in a typical trait-judgment task may be concerned how others (e.g., an MRI operator or an experimenter) would think about their judgments about self/close others. Alternatively, participants may be concerned about whether their trait-judgments are aligned with social expectations even without being conscious of an external observer. Despite the difficulty of measuring private concerns about social desirability, future study should more systematically elucidate the effect of motivation for social desirability on RMPFC activity during trait-judgment tasks. Specifically, this can be achieved by adding a condition with an instruction that guarantees complete confidentiality of participants' responses (i.e., a private instruction condition) to examine parametric differences in RMPFC activity across three conditions (i.e., private instruction, no instruction, and observation instruction).

In addition to the RMPFC, the AI was also engaged in impression management during self- and friend-evaluation. AI has been theorized as a central hub of human emotion processing (Damasio, 1999; Craig and Craig, 2009), playing a key role in integrating bodily signals and linking them to higher-order cognition — critical for adjusting behaviors adaptively depending on emotional and social contexts (Ibañez and Manes, 2012). In addition, structural impairment of these two regions is often associated with reduced self-conscious emotions and socially inappropriate behavior (Rosen *et al.*, 2002; Seeley, 2010; Ibañez and Manes, 2012).

It has been theorized that the RMPFC and AI may comprise key neurocircuitry for visceromotor prediction or allostatic regulatory control (Stephan *et al.*, 2016; Kleckner *et al.*, 2017). These two regions are known to communicate directly with the midbrain homeostatic control system in the brainstem (Fischer *et al.*, 2016), possibly via von Economo neurons known to support rapid brain-body communications (Allman *et al.*, 2010, 2011). According to this view, the basic function of the RMPFC

and AI would be to predict potential future homeostatic imbalance and apply preventive physiological responses or behavioral actions. Impression management in our study may therefore be the behavioral manifestation of allostatic regulatory processes, subserved by the RMPFC and AI to prevent foreseen homeostatic disturbances frequently accompanied by social rejection or social threat (Dickerson *et al.*, 2009; Kemeny, 2009; Slavich *et al.*, 2010; Murphy *et al.*, 2013). To illustrate, during friend-evaluation with positive adjectives, knowledge leading to positive trait rejection will clash with expected social rewards that typically follow praise for others. This conflict could be detected by the AI, communicated to the RMPFC, and elicit socially desirable friend-promoting actions, thus increasing chances of survival in a social world. We believe that allostatic regulatory control parsimoniously explains RMPFC and/or AI activity observed in studies regarding social observation, socially desirable behavior, and self-conscious emotions, which all share a common psychological process — concern for potential social evaluation.

It should be noted that the neural correlates of impression management we found by measuring neural activity encoding valence level could be seen as indirect evidence. The alternative way of analysis could be to measure neural activity encoding participants' actual decisions (see result from this analysis in **Supplementary Table 4**). However, this parameter could reflect not only the motivation of impression management but also actual beliefs about the self and a friend. Therefore, we believed that the best way to identify the neural correlates of impression management in our study is to measure neural activity encoding valence level, which is shown to have a significant effect on managing and forming impression at the behavioral level. Moreover, given that our sample size is smaller than the recent recommendation of fMRI studies (Geuter *et al.*, 2018), more studies with the large sample size are necessary to replicate and confirm the roles of MPFC and AI on impression management.

5.0 CONCLUSIONS

Social observation motivates people to demote negative self-evaluation and promote positive friend-evaluation, serving the common goal of communicating positive self-image. Social observation also facilitates RMPFC and AI activity when tracking the valence level of negative adjectives during self-evaluation and positive adjectives during friend-evaluation, which is consistent with previous literature on RMPFC and AI involvement in social observation, socially desirable behavior, and self-conscious emotions. Revealing such a mechanism of impression management would greatly expand our understanding of complex human social behaviors, which may be rooted in a simpler core principle — a self-serving motive.

6.0 REFERENCES

- Allman, J. M., Tetreault, N. A., Hakeem, A. Y., Manaye, K. F., Semendeferi, K., Erwin, J. M., Park, S., Goubert, V., & Hof, P. R. (2010) The von Economo neurons in frontoinsular and anterior cingulate cortex in great apes and humans. *Brain Structure and Function*, 214, 495-517.

- Allman, J. M., Tetreault, N. A., Hakeem, A. Y., Manaye, K. F., Semendeferi, K., Erwin, J. M., Park, S., Goubert, V., & Hof, P. R. (2011) The von Economo neurons in the frontoinsula and anterior cingulate cortex. *Annals of the New York Academy of Sciences*, 1225, 59-71.
- Anderson, N. H. (1968) Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, 9, 272.
- Araujo, H. F., Kaplan, J., Damasio, H., & Damasio, A. (2015) Neural correlates of different self domains. *Brain and Behavior*, 5, e00409.
- Aron, A., Aron, E. N., & Smollan, D. (1992) Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, 63, 596.
- Aron, A., Aron, E. N., Tudor, M., & Nelson, G. (2004) Close relationships as including other in the self. In: H. T. Reis, C. E. Rusbult (eds). *Close relationships: Key readings*. Philadelphia PA: Taylor & Francis. p. 365-379.
- Bas-Hoogendam, J. M., van Steenbergen, H., Kreuk, T., van der Wee, N. J., & Westenberg, P. M. (2017) How embarrassing! The behavioral and neural correlates of processing social norm violations. *PLOS ONE*, 12, e0176326.
- Basile, B., Mancini, F., Macaluso, E., Caltagirone, C., Frackowiak, R. S., & Bozzali, M. (2011) Deontological and altruistic guilt: Evidence for distinct neurobiological substrates. *Human Brain Mapping*, 32, 229-239.
- Beer, J. S., John, O. P., Scabini, D., & Knight, R. T. (2006) Orbitofrontal cortex and social behavior: Integrating self-monitoring and emotion-cognition interactions. *Journal of Cognitive Neuroscience*, 18, 871-879.
- Bereczkei, T., Birkas, B., & Kerekes, Z. (2010) Altruism towards strangers in need: Costly signaling in an industrial society. *Evolution and Human Behavior*, 31, 95-103.
- Blair, K. S., Geraci, M., Hollon, N., Otero, M., DeVido, J., Majestic, C., Jacobs, M., Blair, R., & Pine, D. S. (2010) Social norm processing in adult social phobia: Atypically increased ventromedial frontal cortex responsiveness to unintentional (embarrassing) transgressions. *American Journal of Psychiatry*, 167, 1526-1532.
- Bolino, M. C., Kacmar, K. M., Turnley, W. H., & Gilstrap, J. B. (2008) A multi-level review of impression management motives and behaviors. *Journal of Management*, 34, 1080-1109.
- Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002) Region of interest analysis using an SPM toolbox. 8th International Conference on Functional Mapping of the Human Brain, Sendai.
- Burnett, S., Bird, G., Moll, J., Frith, C., & Blakemore, S. J. (2009) Development during adolescence of the neural processing of social emotion. *Journal of Cognitive Neuroscience*, 21, 1736-1750.
- Cooper, N., Thompson, S., Brook O'Donnell, M., & Falk, E. B. (2015) Brain activity in self-and value-related regions in response to online antismoking messages predicts behavior change. *Journal of Media Psychology: Theories, Methods, and Applications*, 27, 93.
- Cox, R. W. (1996) AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29, 162-173.

- Craig, A. D., & Craig, A. (2009) How do you feel—now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, 10, 59-70.
- Damasio, A. (1999) The feeling of what happens: Body and emotion in the making of consciousness. Fort Worth, TX: Harcourt College Publishers.
- De La Vega, A., Chang, L. J., Banich, M. T., Wager, T. D., & Yarkoni, T. (2016) Large-scale meta-analysis of human medial frontal cortex reveals tripartite functional organization. *Journal of Neuroscience*, 36, 6553-6562.
- Deen, B., Pitskel, N. B., & Pelphrey, K. A. (2010) Three systems of insular functional connectivity identified with cluster analysis. *Cerebral Cortex*, 21, 1498-1506.
- Dickerson, S. S., Gable, S. L., Irwin, M. R., Aziz, N., & Kemeny, M. E. (2009) Social-evaluative threat and proinflammatory cytokine regulation: An experimental laboratory investigation. *Psychological Science*, 20, 1237-1244.
- Dixon, M. L., Thiruchselvam, R., Todd, R., & Christoff, K. (2017) Emotion and the prefrontal cortex: An integrative review. *Psychological Bulletin*, 143, 1033.
- Doherty, K., & Schlenker, B. R. (1991) Self-consciousness and strategic self-presentation. *Journal of Personality*, 59, 1-18.
- Falk, E. B., Berkman, E. T., Mann, T., Harrison, B., & Lieberman, M. D. (2010) Predicting persuasion-induced behavior change from the brain. *Journal of Neuroscience*, 30, 8421-8424.
- Fischer, D. B., Boes, A. D., Demertzi, A., Evrard, H. C., Laureys, S., Edlow, B. L., Liu, H., Saper, C. B., Pascual-Leone, A., & Fox, M. D. (2016) A human brain network derived from coma-causing brainstem lesions. *Neurology*, 87, 2427-2434.
- Fourie, M. M., Thomas, K. G., Amodio, D. M., Warton, C. M., & Meintjes, E. M. (2014) Neural correlates of experienced moral emotion: An fMRI investigation of emotion in response to prejudice feedback. *Social Neuroscience*, 9, 203-218.
- Gilead, M., Katzir, M., Eyal, T., & Liberman, N. (2016) Neural correlates of processing “self-conscious” vs. “basic” emotions. *Neuropsychologia*, 81, 207-218.
- Griskevicius, V., Tybur, J. M., & Van den Bergh, B. (2010) Going green to be seen: Status, reputation, and conspicuous conservation. *Journal of Personality and Social Psychology*, 98, 392.
- Hare, T. A., Camerer, C. F., Knoepfle, D. T., O'Doherty, J. P., & Rangel, A. (2010) Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *Journal of Neuroscience*, 30, 583-590.
- Heitmann, C. Y., Peterburs, J., Mothes-Lasch, M., Hallfarth, M. C., Böhme, S., Miltner, W. H., & Straube, T. (2014) Neural correlates of anticipation and processing of performance feedback in social anxiety. *Human Brain Mapping*, 35, 6023-6031.
- Ibañez, A., & Manes, F. (2012) Contextual social cognition and the behavioral variant of frontotemporal dementia. *Neurology*, 78, 1354-1362.
- Izuma, K., Saito, D. N., & Sadato, N. (2010) The roles of the medial prefrontal cortex and striatum in reputation processing. *Social Neuroscience*, 5, 133-147.
- Jung, D., Sul, S., Lee, M., & Kim, H. (2018) Social observation increases functional

- segregation between MPFC subregions predicting prosocial consumer decisions. *Scientific Reports*, 8, 3368.
- Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002) Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience*, 14, 785-794.
- Kemeny, M. E. (2009) Psychobiological responses to social threat: Evolution of a psychological model in psychoneuroimmunology. *Brain, Behavior, and Immunity*, 23, 1-9.
- Kimura, A., Mukawa, N., Yamamoto, M., Masuda, T., Yuasa, M., Goto, S.-I., Oka, T., & Wada, Y. (2012) The influence of reputational concerns on purchase intention of fair-trade foods among young Japanese adults. *Food Quality and Preference*, 26, 204-210.
- Kleckner, I. R., Zhang, J., Touroutoglou, A., Chanes, L., Xia, C., Simmons, W. K., Quigley, K. S., Dickerson, B. C., & Barrett, L. F. (2017) Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nature Human Behaviour*, 1, 0069.
- Korn, C. W., Prehn, K., Park, S. Q., Walter, H., & Heekeren, H. R. (2012) Positively biased processing of self-relevant social feedback. *Journal of Neuroscience*, 32, 16832-16844.
- Krajchich, I., Adolphs, R., Tranel, D., Denburg, N. L., & Camerer, C. F. (2009) Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *Journal of Neuroscience*, 29, 2188-2192.
- Krienen, F. M., Tu, P.C., & Buckner, R. L. (2010) Clan mentality: Evidence that the medial prefrontal cortex responds to close others. *Journal of Neuroscience*, 30, 13906-13915.
- Leary, M. R. (1995) *Self-presentation: Impression management and interpersonal behavior*. Brown & Madison, WI: Benchmark Publishers.
- Leary, M. R. (2004) Digging deeper: The fundamental nature of “self-conscious” emotions. *Psychological Inquiry*, 15, 129-131.
- Leary, M. R., Allen, A. B., & Terry, M. L. (2011) Managing social images in naturalistic versus laboratory settings: Implications for understanding and studying self-presentation. *European Journal of Social Psychology*, 41, 411-421.
- Leary, M. R., & Kowalski, R. M. (1990) Impression management: A literature review and two-component model. *Psychological Bulletin*, 107, 34.
- Lieberman, M. D., Straccia, M. A., Meyer, M. L., Du, M., & Tan, K. M. (2019) Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): Causal, multivariate, and reverse inference evidence. *Neuroscience & Biobehavioral Reviews*, 99, 311-328.
- Mashek, D. J., Aron, A., & Boncimino, M. (2003) Confusions of self with close others. *Personality and Social Psychology Bulletin*, 29, 382-392.
- Masten, C. L., Morelli, S. A., & Eisenberger, N. I. (2011) An fMRI investigation of empathy for ‘social pain’ and subsequent prosocial behavior. *NeuroImage*, 55, 381-388.
- Mathur, V. A., Harada, T., Lipke, T., & Chiao, J. Y. (2010) Neural basis of extraordinary empathy and altruistic motivation. *NeuroImage*, 51, 1468-1475.
- Michl, P., Meindl, T., Meister, F., Born, C., Engel, R. R., Reiser, M., & Hennig-Fast,

- K. (2012) Neurobiological underpinnings of shame and guilt: A pilot fMRI study. *Social Cognitive and Affective Neuroscience*, 9, 150-157.
- Mitchell, J. P., Banaji, M. R., & MacRae, C. N. (2005) The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience*, 17, 1306-1315.
- Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006) Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, 50, 655-663.
- Moll, J., Zahn, R., de Oliveira-Souza, R., Bramati, I. E., Krueger, F., Tura, B., Cavanagh, A. L., & Grafman, J. (2011) Impairment of prosocial sentiments is associated with frontopolar and septal damage in frontotemporal dementia. *NeuroImage*, 54, 1735-1742.
- Müller-Pinzler, L., Gazzola, V., Keysers, C., Sommer, J., Jansen, A., Frässle, S., Einhäuser, W., Paulus, F. M., & Krach, S. (2015) Neural pathways of embarrassment and their modulation by social anxiety. *NeuroImage*, 119, 252-261.
- Murphy, M. L., Slavich, G. M., Rohleder, N., & Miller, G. E. (2013) Targeted rejection triggers differential pro-and anti-inflammatory gene expression in adolescents as a function of social status. *Clinical Psychological Science*, 1, 30-40.
- Murray, R. J., Schaer, M., & Debbane, M. (2012) Degrees of separation: A quantitative neuroimaging meta-analysis investigating self-specificity and shared neural activation between self- and other-reflection. *Neuroscience and Biobehavioral Reviews*, 36, 1043-1059.
- Ochsner, K. N., Beer, J. S., Robertson, E. R., Cooper, J. C., Gabrieli, J. D., Kihlstrom, J. F., & D'Esposito, M. (2005) The neural correlates of direct and reflected self-knowledge. *NeuroImage*, 28, 797-814.
- Pegors, T. K., Tompson, S., O'Donnell, M. B., & Falk, E. B. (2017) Predicting behavior change from persuasive messages using neural representational similarity and social network analyses. *NeuroImage*, 157, 118-128.
- Pei, R., Schmälzle, R., Kranzler, E. C., O'Donnell, M. B., & Falk, E. B. (2019) Adolescents' neural response to tobacco prevention messages and sharing engagement. *American Journal of Preventive Medicine*, 56, S40-S48.
- Powers, T. A., & Zuroff, D. C. (1988) Interpersonal consequences of overt self-criticism: A comparison with neutral and self-enhancing presentations of self. *Journal of Personality and Social Psychology*, 54, 1054.
- Rameson, L. T., Morelli, S. A., & Lieberman, M. D. (2012) The neural correlates of empathy: Experience, automaticity, and prosocial behavior. *Journal of Cognitive Neuroscience*, 24, 235-245.
- Rosen, H. J., Gorno-Tempini, M. L., Goldman, W., Perry, R., Schuff, N., Weiner, M., Feiwell, R., Kramer, J., & Miller, B. L. (2002) Patterns of brain atrophy in frontotemporal dementia and semantic dementia. *Neurology*, 58, 198-208.
- Schlenker, B. R. (2012) Self-presentation. In: M. R. Leary & J. P. Tangney (eds). *Handbook of self and identity*. New York, NY: The Guilford Press.
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014) Fractionating theory of mind: A meta-analysis of functional brain imaging studies.

- Neuroscience & Biobehavioral Reviews, 42, 9-34.
- Sedikides, C., Gaertner, L., & Cai, H. (2015) On the panculturality of self-enhancement and self-protection motivation: The case for the universality of self-esteem. *Advances in Motivation Science*, 2, 185-241.
- Sedikides, C., Gregg, A. P., & Hart, C. M. (2007) The importance of being modest. In: C. Sedikides & S. Spencer (eds). *The self: Frontiers in social psychology* (pp. 163–184). New York, NY, Psychology Press.
- Seeley, W. W. (2010) Anterior insula degeneration in frontotemporal dementia. *Brain Structure and Function*, 214, 465-475.
- Shin, L. M., Dougherty, D. D., Orr, S. P., Pitman, R. K., Lasko, M., Macklin, M. L., Alpert, N. M., Fischman, A. J., & Rauch, S. L. (2000) Activation of anterior paralimbic structures during guilt-related script-driven imagery. *Biological Psychiatry*, 48, 43-50.
- Slavich, G. M., Way, B. M., Eisenberger, N. I., & Taylor, S. E. (2010) Neural sensitivity to social rejection is associated with inflammatory responses to social stress. *Proceedings of the National Academy of Sciences*, 107, 14817-14822.
- Snowden, J., Bathgate, D., Varma, A., Blackshaw, A., Gibbons, Z., & Neary, D. (2001) Distinct behavioural profiles in frontotemporal dementia and semantic dementia. *Journal of Neurology, Neurosurgery & Psychiatry*, 70, 323-332.
- Soetevent, A. R. (2005) Anonymity in giving in a natural context—a field experiment in 30 churches. *Journal of Public Economics*, 89, 2301-2323.
- Somerville, L. H., Jones, R. M., Ruberry, E. J., Dyke, J. P., Glover, G., & Casey, B. J. (2013) The medial prefrontal cortex and the emergence of self-conscious emotion in adolescence. *Psychological Science*, 24, 1554-1562.
- Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A., Paliwal, S., Gard, T., Tittgemeyer, M., Fleming, S. M., Haker, H., & Seth, A. K. (2016) Allostatic self-efficacy: A metacognitive theory of dyshomeostasis-induced fatigue and depression. *Frontiers in Human Neuroscience*, 10, 550.
- Sturm, V. E., Ascher, E. A., Miller, B. L., & Levenson, R. W. (2008) Diminished self-conscious emotional responding in frontotemporal lobar degeneration patients. *Emotion*, 8, 861.
- Sturm, V. E., Rosen, H. J., Allison, S., Miller, B. L., & Levenson, R. W. (2006) Self-conscious emotion deficits in frontotemporal lobar degeneration. *Brain*, 129, 2508-2516.
- Sturm, V. E., Sollberger, M., Seeley, W. W., Rankin, K. P., Ascher, E. A., Rosen, H. J., Miller, B. L., & Levenson, R. W. (2012) Role of right pregenual anterior cingulate cortex in self-conscious emotional reactivity. *Social Cognitive and Affective Neuroscience*, 8, 468-474.
- Sul, S., Choi, I., & Kang, P. (2012) Cultural modulation of self-referential brain activity for personality traits and social identities. *Social Neuroscience*, 7, 280-291.
- Takahashi, H., Yahata, N., Koeda, M., Matsuda, T., Asai, K., & Okubo, Y. (2004) Brain activation associated with evaluative processes of guilt and embarrassment: an fMRI study. *NeuroImage*, 23, 967-974.
- Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007) Moral emotions and moral behavior. *Annual Review of Psychology*, 58, 345-372.

- Tracy, J. L., & Robins, R. W. (2004) Putting the self into self-conscious emotions: A theoretical model. *Psychological Inquiry*, 15, 103-125.
- Tusche, A., Böckler, A., Kanske, P., Trautwein, F.-M., & Singer, T. (2016) Decoding the charitable brain: empathy, perspective taking, and attention shifts differentially predict altruistic giving. *Journal of Neuroscience*, 36, 4719-4732.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15, 273-289.
- Van Hoorn, J., Van Dijk, E., Güroğlu, B., & Crone, E. A. (2016) Neural correlates of prosocial peer influence on public goods game donations during adolescence. *Social Cognitive and Affective Neuroscience*, 11, 923-933.
- Van Overwalle, F. (2009) Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, 30, 829-858.
- Van Tongeren, D. R., Davis, D. E., & Hook, J. N. (2014) Social benefits of humility: Initiating and maintaining romantic relationships. *The Journal of Positive Psychology*, 9, 313-321.
- Vanderwal, T., Hunyadi, E., Grupe, D. W., Connors, C. M., & Schultz, R. T. (2008) Self, mother and abstract other: An fMRI study of reflective social processing. *NeuroImage*, 41, 1437-1446.
- Wagner, U., N'diaye, K., Ethofer, T., & Vuilleumier, P. (2011) Guilt-specific processing in the prefrontal cortex. *Cerebral Cortex*, 21, 2461-2470.
- Welborn, B. L., & Lieberman, M. D. (2015) Person-specific theory of mind in medial pFC. *Journal of Cognitive Neuroscience*, 27, 1-12.
- Will, G.-J., Rutledge, R. B., Moutoussis, M., & Dolan, R. J. (2017) Neural and computational processes underlying dynamic changes in self-esteem. *Elife*, 6, e28098.
- Wortman, J., & Wood, D. (2011) The personality traits of liked people. *Journal of Research in Personality*, 45, 519-528.
- Wosinska, W., Dabul, A. J., Whetstone-Dion, R., & Cialdini, R. B. (1996) Self-presentational responses to success in the organization: The costs and benefits of modesty. *Basic and Applied Social Psychology*, 18, 229-242.
- Xiaohua Chen, S., Bond, M. H., Chan, B., Tang, D., & Buchtel, E. E. (2009) Behavioral manifestations of modesty. *Journal of Cross-Cultural Psychology*, 40, 603-626.
- Yoon, L., Somerville, L. H., & Kim, H. (2018) Development of MPFC function mediates shifts in self-protective behavior provoked by social feedback. *Nature Communications*, 9, 3086.
- Zahn, R., Moll, J., Paiva, M., Garrido, G., Krueger, F., Huey, E. D., & Grafman, J. (2008) The neural basis of human social values: evidence from functional MRI. *Cerebral Cortex*, 19, 276-283.

Study #3: A Hierarchical Allostatic Regulation Model of Medial Prefrontal Cortex Function for Social Valuation (Authors: Hackjin Kim; Published in Frontiers in Decision Neuroscience, March 2020)

1.0 SUMMARY

The medial prefrontal cortex (mPFC) has long been recognized as the key component of the neurocircuitry involved in various social as well as non-social behaviors; however, little is known regarding the organizing principle of distinctive subregions in the mPFC that integrates a wide range of mPFC functions. The present study proposes a hierarchical model of mPFC functionality, where three functionally dissociable subregions, namely, the ventromedial prefrontal cortex (vmPFC), rostromedial prefrontal cortex (rmPFC), and dorsomedial prefrontal cortex (dmPFC), are differentially involved in computing values of decision-making. According to this model, the mPFC subregions interact with each other in such a way that more dorsal regions utilize additional external sensory information from environment to predict and prevent conflicts occurring in more ventral regions tuned to internal bodily signals, thereby exerting the hierarchically organized allostatic regulatory control over homeostatic reflexes. This model also emphasizes the role of the thalamic reticular nucleus (TRN) in arbitrating the transitions between different thalamo-cortical loops, detecting conflicts between competing options for decision-making, and in shifting flexibly between decision modes. The hierarchical architecture of the mPFC working in conjunction with the TRN may play a key role in adjusting the internal (bodily) needs to suit the constraints of external (environmental) variables better; thus, effectively addressing the stability-plasticity dilemma.

2.0 INTRODUCTION

Imagine you are an international student who came to study in a country that is culturally very different from that you grew up. How would you maintain stable codes of conduct while updating other codes in a novel social situation? The problem of acquiring new knowledge without disrupting the existing knowledge, the so-called stability-plasticity dilemma, is one of the major obstacles encountered by any adaptive agent (Grossberg, 2013). The human brain is known to be one of the most successful systems for dealing with the stability-plasticity dilemma. Multitudes of researchers in a wide range of academic disciplines, from computer scientists to neuroscientists, have extensively investigated the way the human brain resolves this dilemma. Our knowledge on the exact mechanisms for such a capacity, particularly under social contexts, remains limited.

This article suggests that such a stability-plasticity dilemma can be effectively addressed by the brain's capacity to predict and prevent homeostatic imbalance, which is called allostasis (McEwen & Stellar, 1993; Schulkin, 2003; Sterling & Eyer, 1988), to maintain a state of homeostasis. For example, the brain constantly seeks the optimal regulation of bodily homeostasis by adding increasing amounts of external sensory inputs (e.g., visual, auditory, tactile stimuli) in order to predict and prevent anticipated homeostatic imbalance as early and accurately as

possible (External valuation). As a result, the most parsimonious pattern of external inputs that led to a successful prediction and prevention of homeostatic imbalance becomes associated with a specific coordinated pattern of somatic or visceral reflexes or both, which is engaged again in a reflex-like fashion whenever the same or a similar input pattern is presented (Internal valuation). Importantly, when an internal valuation fails to achieve the state of homeostasis, then external valuation will be engaged again to update the internal valuation. This internal-external valuation cycle may be at the heart of the allostatic regulation and also likely to reflect how the brain deals with the stability-plasticity dilemma.

To illustrate how such an allostatic regulation can be linked to social valuation, this article first reviews recent findings of the roles of the medial prefrontal cortex (mPFC) in decision-making in social situations. Next, the hierarchical allostatic regulation model of the mPFC function for computing values of social decision-making is proposed, with an emphasis on the three functionally and anatomically dissociable subregions of the mPFC: the ventromedial prefrontal cortex (vmPFC), the rostromedial prefrontal cortex (rmPFC), and the dorsomedial prefrontal cortex (dmPFC). In this model, the mPFC subregions are organized such that more ventral and more dorsal regions are involved in internal and external valuation, respectively, and the intermediate areas are functionally and spatially graded concerning such dimensions. More ventral regions are involved in intuitive value computation to meet internal needs, prioritizing stability, whereas more dorsal regions are involved in deliberative value computation to utilize external information, prioritizing plasticity. External valuation in more dorsal regions is engaged to resolve a conflict that occurs when mutually competing units are simultaneously activated in more ventral regions. Following repeated engagements, such external valuation can serve to update internalized values encoded in the ventral regions. Finally, it will be discussed how this model can efficiently address the stability-plasticity dilemma and why such functionality is critical for adaptive behavior in constantly changing social situations.

This review includes literatures from human as well as nonhuman research including monkeys and rats. Despite some evidence for remarkable cross-species homology among rats, macaques, and humans in the anatomy of the medial prefrontal cortex (Vogt et al., 2013), attention should be paid to the interpretation of the cross-species comparisons reported in the present study regarding functional differences among different subregions of the medial prefrontal cortex.

3.0 Anatomically and functionally dissociable subregions in the medial prefrontal cortex

3.1 Anatomical boundaries between the medial prefrontal cortex subregions

According to the influential anatomical studies (Joyce & Barbas, 2018; Mackey & Petrides, 2014; Palomero-Gallagher et al., 2019; Vogt, 2005), human mPFC can be broadly divided into three functionally and anatomically dissociable subregions: (1) the vmPFC [roughly corresponds to the medial aspect of Brodmann area (BA) 11, BA 12, BA 14, and BA 25] (2) dmPFC [BA 9, BA 24 (the pregenual anterior

cingulate cortex), and BA 32 (the anterior midcingulate cortex)], and (3) rmPFC [BA 10, BA 24 (the pregenual anterior cingulate cortex), and BA 32 (the pregenual anterior cingulate cortex)]. For a practical purpose, a recent neuroimaging study utilized the MNI coordinate system such that the vmPFC and rmPFC are divided by the z-plane of -10 (MNI coordinate system), and the dmPFC and the rmPFC are divided by the z-plane of $+22$ (Lieberman, Straccia, Meyer, Du, & Tan, 2019) (Figure 1).

It is well-known that these subregions have unique patterns of anatomical (Haber & Knutson, 2010; Haruno & Kawato, 2006; Yin & Knowlton, 2006) as well as functional (Bzdok et al., 2013; de la Vega, Chang, Banich, Wager, & Yarkoni, 2016) connections with other neural structures. More specifically, recent meta-analyses of functional connectivity revealed that the vmPFC is functionally connected with the nucleus accumbens, amygdala, and thalamus; the rmPFC with the nucleus accumbens, hippocampus, posterior cingulate cortex, and retrosplenial cortex; and the dmPFC with the inferior frontal gyrus, temporo-parietal junction, and middle temporal gyrus (Bzdok et al., 2013; de la Vega et al., 2016). It should be noted, however, that these meta-analyses may not reflect the full connectivity because they are based on fMRI studies with limited spatial and temporal resolution.

The mPFC subregions do not seem to be mutually independent, but functionally inter-connected with each other. For example, it has been proposed that reinforcement learning occurs through multiple mutually interacting cortico-thalamo-striatal loops, propagating information mostly from a ventral to a dorsal direction (Yin & Knowlton, 2006). Despite the recent suggestion that the more dorsal mPFC handles more abstract and complex information than the more ventral mPFC (Denny, Kober, Wager, & Ochsner, 2012; Suzuki et al., 2012), the specific roles of distinctive mPFC subregions and the exact nature of the interaction between them in the service of social behavior are currently unknown. In the next section, I will review the empirical and theoretical works on the functional properties of each subregion in more detail before proposing an integrative model of the mPFC function.

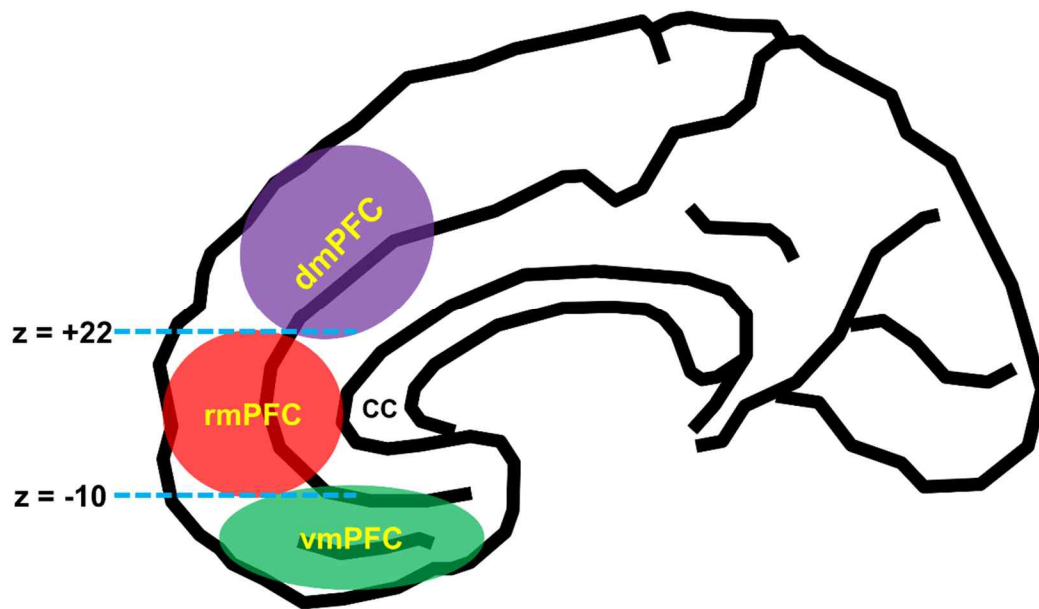


Figure 8. Schematic diagram of anatomical segregation within the medial prefrontal cortex. mPFC can be broadly divided into three functionally and anatomically dissociable subregions: the dorsomedial prefrontal cortex (dmPFC), the rostromedial prefrontal cortex (rmPFC), and the ventromedial prefrontal cortex (vmPFC). These regions roughly correspond to the medial aspect of the Brodmann area (BA) 9 & BA 32 (the anterior midcingulate cortex), BA 10 & BA 32 (the perigenual anterior cingulate cortex), and BA 11 & BA 12, respectively. The dmPFC and rmPFC are divided by the z-plane of +22, and the rmPFC and vmPFC are divided by the z-plane of -10 (Lieberman et al., 2019). Abbreviations: cc, corpus callosum.

3.2 Role of the ventromedial prefrontal cortex (vmPFC) in social valuation

Many animal and human studies have described various aspects of the function of the vmPFC, including the regulation of emotions (Delgado et al., 2016; LaLumiere, Niehoff, & Kalivas, 2010; Milad et al., 2009; Quirk & Beer, 2006), valuation for decision-making (Bartra, McGuire, & Kable, 2013; Kable & Glimcher, 2007; H. Kim, Adolphs, O'Doherty, & Shimojo, 2007), and goal-directed actions (O'Doherty, 2011). The vmPFC is also thought to be a key cortical component of the central autonomic network (Beissner, Meissner, Bar, & Napadow, 2013). Consistent with this, there is additional evidence that the vmPFC may be involved in encoding internally driven valuation. For example, the vmPFC is the main target of the dopaminergic projection from the midbrain modulated by either food intake (de Araujo, Ferreira, Tellez, Ren, & Yeckel, 2012) or direct vagus nerve stimulation (Han et al., 2018). The vmPFC activity is also modulated by the experimentally-induced dopamine level (Jocham, Klein, & Ullsperger, 2011), visceral signals like hunger and satiety (Howard, Gottfried, Tobler, & Kahnt, 2015; Roy, Shohamy, & Wager, 2012), and outcome devaluation (de Wit, Corlett, Aitken, Dickinson, & Fletcher, 2009; Valentin, Dickinson, & O'Doherty, 2007). Also, vmPFC activity covaries with heart rate variability (Ziegler, Dahnke, Yeragani, & Bar, 2009), and

vmPFC lesions impair expression of normal physiological responses during decision-making (Bechara, Tranel, Damasio, & Damasio, 1996).

The vmPFC is known to have dense anatomical connections with both the nucleus accumbens (Haber, Kim, Mailly, & Calzavara, 2006) and the amygdala (Carmichael & Price, 1995; Ghashghaei, Hilgetag, & Barbas, 2007). The former is mainly involved in learning reward-seeking behavior by reinforcing actions to obtain rewards (Demos, Heatherton, & Kelley, 2012; Ikemoto & Panksepp, 1999; Lawrence, Hinton, Parkinson, & Lawrence, 2012), whereas the latter is primarily involved in learning defensive behaviors to avoid potentially dangerous or unpleasant stimuli (De Martino, Camerer, & Adolphs, 2010; Mason, Capitanio, Machado, Mendoza, & Amaral, 2006; Schwartz, Wright, Shin, Kagan, & Rauch, 2003). These connections may allow the vmPFC to generate rapid avoidance or approach responses, making it an ideal system for a cost-benefit analysis to achieve homeostatic balance in a given situation (Basten, Biele, Heekeren, & Fiebach, 2010; H. Kim, Shimojo, & O'Doherty, 2006; Schneirla, 1959).

In the field of social neuroscience, the vmPFC has been strongly implicated in processing “first-person” information (Bzdok et al., 2013; Denny et al., 2012; H. Kim et al., 2007; Lebreton, Jorge, Michel, Thirion, & Pessiglione, 2009; I. Levy, Lazzaro, Rutledge, & Glimcher, 2011), although several studies have also shown that the vmPFC activity can be commonly involved in decisions for both self and others (Janowski, Camerer, & Rangel, 2013; Nicolle et al., 2012). Such an inconsistency can be reconciled by assuming the role of the vmPFC function in intuitive, internalized valuation for other-regarding decisions. Supporting this idea, the vmPFC was shown to be involved in decision-making for others, when people apply self-simulation to estimate a stranger’s preferences (Janowski et al., 2013; Kang, Lee, Sul, & Kim, 2013) and when people are fully familiarized with others’ preferences through practice (Nicolle et al., 2012). These findings, therefore, indicate that the vmPFC could be involved in computing the value of choices for others, only when such valuation process is internally driven via familiarization of others’ preferences or through egocentric simulations.

Similarly, in the specific context of prosociality, the vmPFC seems to encode decision values for highly internalized forms of altruistic behaviors (i.e., internalized prosocial valuation) as in harm-aversion in social dilemma and moral emotions (Buckholtz & Marois, 2012; Crockett, 2013; Hare, Camerer, Knoepfle, & Rangel, 2010; Moll et al., 2006; Shenhav & Greene, 2010; Sul et al., 2015; Tricomi, Rangel, Camerer, & O'Doherty, 2010; Zaki & Mitchell, 2011). For example, a more recent study showed that selfish people used the vmPFC only when calculating the value of the choices for themselves but not those for strangers, unlike altruistic people who used the vmPFC for both self and other (Sul et al., 2015). Besides, more prosocial people showed higher vmPFC activity during prosocial choice, whether they are observed by others or not, and higher vmPFC activity was associated with faster response time for prosocial choices (Jung, Sul, Lee, & Kim, 2018). Taken together, these findings suggest that prosocial valuation encoded by vmPFC may be intuitively engaged and immune to social context.

According to recent theories on morality and altruism (Haidt, 2007), the ultimate desire for survival and reproduction can be extended to creating an altruistic

instrumental desire to sacrifice oneself for others. That is, people can learn the belief that the act of helping others is an effective way to draw a favorable impression from others, and such a belief can be internalized to create a new instrumental desire. Such an instrumental desire for altruism may be internalized in the vmPFC, which may then facilitate prosocial behavior automatically and intuitively, more or less independently of social context (Jung et al., 2018; Rand, Greene, & Nowak, 2012; Sul et al., 2015). This idea is also consistent with the findings that the vmPFC is associated with seeking social status (Hughes & Beer, 2012, 2013; Milad et al., 2009). For example, an altruistic decision may result from an incentive to avoid the possibility of losing reputation due to selfish behavior. In this sense, the vmPFC activity associated with prosocial behavior may indicate the degree to which one's valuation for social reward is internalized, and, therefore, is resistant to contextual changes.

3.3 Role of the dorsomedial prefrontal cortex (dmPFC) in social valuation

The dorsomedial prefrontal cortex (dmPFC) has been implicated in numerous aspects of psychological functions (Ebitz & Hayden, 2016), such as detecting and resolving conflicts among competitive responses (Shenhav, Cohen, & Botvinick, 2016), searching for a new value beyond the current familiar state (Kolling et al., 2016), and computing decision values based on external sensory signals from the environment, unlike the vmPFC involved in internal valuation (Bouret & Richmond, 2010; Howard et al., 2015; Nakao, Ohira, & Northoff, 2012). Consistent with the functional dissociation between dmPFC and vmPFC in humans, recent studies on rats have also shown a functionally competitive relationship between dorsal and ventral subregion of the mPFC (Coutureau & Killcross, 2003). For example, the prelimbic cortex, a more dorsal part of the mPFC in rats comparable to BA 32 in primates (Vogt et al., 2013), is responsible for voluntary and goal-directed initial responses, whereas the infralimbic cortex, a more ventral part of mPFC in rats comparable to BA 25 in primates, is responsible for developing habit-like behaviors, which are formed progressively through overtraining (Killcross & Coutureau, 2003). In addition, the BA 32 along with its neighboring cortical regions including the rmPFC or the dmPFC have heavy anatomical connections with the hippocampus and the related rhinal cortex (Barbas, 2015), and lesioning these cortices resulted in significant impairment in the mnemonic retrieval of context within which external sensory stimuli are experienced (Chapados & Petrides, 2015). Based on these findings, it can be speculated that, when two or more competing responses come into conflict, the dmPFC is engaged to search for a new and more appropriate response to resolve the conflict by directing attention to external sensory information from the environment or information available in memory (Cabeza, Dolcos, Graham, & Nyberg, 2002; Horst & Laubach, 2009), which may have little to do with fulfilling the immediate internal needs of the body.

The dmPFC has been often shown to respond to negative outcomes such as pain (Rainville, Duncan, Price, Carrier, & Bushnell, 1997), monetary loss (Liu, Hairston, Schrier, & Fan, 2011), as well as social rejection (Eisenberger, Lieberman, & Williams, 2003). Some recent theoretical works also suggested a more general

function of the dmPFC, that is, to integrate multiple sources of information from a wide range of brain network to guide our thoughts and actions (Shackman et al., 2011), or to maintain the representation of expected rewards and to allocate available physiological resources to meet or exceed task demands (Touroutoglou, Andreano, Dickerson, & Barrett, 2020). One can speculate that experiencing negative outcome may trigger neural processes of re-allocating attention to the environment in order to search for a new potentially better alternative, whereas experiencing positive outcome may elicit a simpler strategy of maintaining previously chosen behavior that have led to the successful consequence. Consistent with the evolutionarily advantageous decision heuristic of win-stay lose-shift (Nowak & Sigmund, 1993), this view suggests that positive and negative outcomes are naturally associated with internal and external valuation process, which are mainly subserved by the vmPFC and the dmPFC, respectively.

In social neuroscience, contrary to the role of the vmPFC in processing “first-person” information, the dmPFC appears to be more involved in processing “third-person” information, which includes mentalization or perspective-taking (Amodio & Frith, 2006; Behrens, Hunt, & Rushworth, 2009; Frith & Frith, 2006; Hampton, Bossaerts, & O'Doherty, 2008; Kang et al., 2013; Mitchell, Macrae, & Banaji, 2006), valuation of decisions for others (Hutcherson, Bushong, & Rangel, 2015; Jung, Sul, & Kim, 2013; Sul et al., 2015; Suzuki et al., 2012), evaluation of outcomes given to others (Apps & Ramnani, 2014; Chang, Gariépy, & Platt, 2013; Lockwood, Apps, Roiser, & Viding, 2015), and prosocial behavior (Waytz, Zaki, & Mitchell, 2012). Despite these other-centered functions, the dmPFC activity is not always associated with prosocial behavior. For example, the dmPFC activity encoding value of decision for others was more prominent among selfish compared to prosocial people (Sul et al., 2015), and the value-related dmPFC activity was stronger for self-centered than other-oriented decisions under social observation (Jung et al., 2018). These inconsistencies about the role of the dmPFC in prosocial decisions should be examined more carefully by considering differences in experimental contexts across studies. Given that the dmPFC is also associated with strategic decisions that maximize profits (Behrens et al., 2009; Hampton et al., 2008; Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004; Seo, Cai, Donahue, & Lee, 2014), it can be inferred that the dmPFC activity can lead to prosocial behavior only when such deliberate decisions regarding others are strategically beneficial to decision-makers.

Taken together, these findings suggest that the dmPFC activity may predict prosocial behavior only when the context intuitively triggers selfish behavior, however, prosocial behavior can be strategically more beneficial. Conversely, the same region may be engaged even when the context automatically triggers prosocial motivation, however, economic value maximization can be strategically more beneficial. Therefore, the dmPFC can be engaged whenever a conflict occurs among two or more responses, and consideration of additional (external) information is necessary for value-maximization, regardless of whether its activity leads to a prosocial outcome or not.

3.4 Role of the rostromedial prefrontal cortex (rmPFC) in social valuation

The rmPFC, which lies between the vmPFC and the dmPFC, has unique and privileged anatomical features because of its widespread anatomical connections with many cortical and subcortical structures including the brainstem, the insula, and most of the other mPFC subregions (Dixon, Thiruchselvam, Todd, & Christoff, 2017). This region has been implicated in various functions such as default-mode processing (Andrews-Hanna, Reidler, Sepulcre, Poulin, & Buckner, 2010; Uddin, Kelly, Biswal, Castellanos, & Milham, 2009), far-sighted decisions, where one needs to choose between immediate smaller and delayed more substantial rewards (Kable & Glimcher, 2007), and, most notably, cognitive branching, that is, pursuing a long-term mental plan by tracking the values of ongoing and alternative behavioral strategies and switching to the better option (Koechlin & Hyafil, 2007; Mansouri, Koechlin, Rosa, & Buckley, 2017).

In social neuroscience, the rmPFC has been best known for its prioritized role in self-referential processing (Kelley et al., 2002; Moran, Macrae, Heatherton, Wyland, & Kelley, 2006; Northoff et al., 2006), although it also has been shown to encode decision values for both self and others (Hutcherson et al., 2015; Sul et al., 2015). For example, in a typical self-referential task where participants view a list of trait-related words and report whether they are self- or other-descriptive, increased activity is found in the rmPFC during conditions of self vs. other (Kelley et al., 2002). Different groups of researchers have interpreted such a self-referential activity in the rmPFC as perceived similarity (Mitchell et al., 2006), personal significance (K. Kim & Johnson, 2015; Krienen, Tu, & Buckner, 2010), and social valuation (D'Argembeau, 2013). An alternate, possibly more plausible, reason for the rmPFC activity during a self-referential task might be that it reflects heightened motivation for seeking self-enhancement, including both self-promotion (approach) and self-protection (avoidance), which is similar to its suggested role in reputation management (Amodio & Frith, 2006; Izuma, 2012). According to this account, the rmPFC activity increases during self- vs. other-referential task because one feels a greater need to engage motivation for self-enhancement. This alternative view can be supported by several recent findings listed below.

First, the rmPFC activity is often associated with different types of self-conscious emotions that can occur depending on whether one's behavior is appropriate to social standards or not (Edelmann, 1987; Keltner & Buswell, 1997; Leary & Kowalski, 1990; Tangney, Stuewig, & Mashek, 2007; Tracy & Robins, 2004). For example, in many studies using emotion-evoking scenarios, rmPFC activity has been linked to subjective experience of various self-conscious emotions such as embarrassment (Bas-Hoogendam, van Steenbergen, Kreuk, van der Wee, & Westenberg, 2017; Burnett, Bird, Moll, Frith, & Blakemore, 2009; Takahashi et al., 2004), shame (Michl et al., 2012), guilt (Basile et al., 2011; Burnett et al., 2009; Fourie, Thomas, Amodio, Warton, & Meintjes, 2014; Gilead, Katzir, Eyal, & Liberman, 2016; Shin et al., 2000; Takahashi et al., 2004; Wagner, N'diaye, Ethofer, & Vuilleumier, 2011; Zahn et al., 2008), and pride (Gilead et al., 2016; Zahn et al., 2008).

Second, the structural and functional integrity of the rmPFC is linked to individual differences in motivation for self-enhancement. For example, people with high rejection sensitivity showed increased rmPFC activity when anticipating social

evaluation (Powers, Somerville, Kelley, & Heatherton, 2013), and those with high trait social anxiety are characterized with heightened rmPFC activity during social observation (Müller-Pinzler et al., 2015). Also, patients with lesions in areas including rmPFC failed to exhibit self-conscious emotion (Krajchich, Adolphs, Tranel, Denburg, & Camerer, 2009; Moll et al., 2011; Sturm, Ascher, Miller, & Levenson, 2008; Sturm, Rosen, Allison, Miller, & Levenson, 2006) and expressed socially inappropriate self-disclosing behavior (Beer, John, Scabini, & Knight, 2006), and those with smaller volume of rmPFC showed reduced physiological and behavioral indices of self-conscious emotional responses when watching a video clip of themselves singing (Sturm et al., 2012). Recently, it was also demonstrated that the developmental maturity of the rmPFC may be critical for a more sophisticated and socially appropriate expression of self-protective motivation in response to negative evaluation from others (Yoon, Somerville, & Kim, 2018).

Third, social observation, one of the most potent situational factors boosting self-enhancement motivation, can modulate the rmPFC activity, often leading to increase in a socially desirable behavior. For example, social observation increased rmPFC activity during judgment about self and social appropriateness (Izuma, Saito, & Sadato, 2010), public success or failure on a cognitive task (Müller-Pinzler et al., 2015), and an economic game (Van Hoorn, Van Dijk, Güroğlu, & Crone, 2016), often being accompanied by subjectively experienced self-conscious emotion (Somerville et al., 2013). Consistent with the “costly signaling theory,” which views altruistic or prosocial behavior as a signal of willingness and ability to help others (Fehr & Fischbacher, 2003; Hardy & Van Vugt, 2006; Nowak & Sigmund, 1998; Zahavi, 1975), social observation by others or even subtle surveillance cues can be powerful enough to increase prosocial behavior (Bereczkei, Birkas, & Kerekes, 2010; Griskevicius, Tybur, & Van den Bergh, 2010; Izuma, 2012; Kimura et al., 2012; Soetevent, 2005). Supporting the behavioral evidence of costly signaling theory, social observation also increased the rmPFC activity encoding the value of prosocial decisions, and such a context-dependent prosocial valuation of the rmPFC was clearly distinguishable from those of the vmPFC and the dmPFC (Figure 2) (Jung et al., 2018). It should be noted, however, that context-dependent functionality of rmPFC for social valuation does not necessarily require social observation. For example, increased rmPFC activity was associated with strategic prosocial behavior when participants had been explicitly instructed to make a donation with money endowed by the experimenter, which may have posed a substantial social pressure, similar to a social observation (Cutler & Campbell-Meiklejohn, 2019; Fukuda et al., 2019; Tusche, Bockler, Kanske, Trautwein, & Singer, 2016).

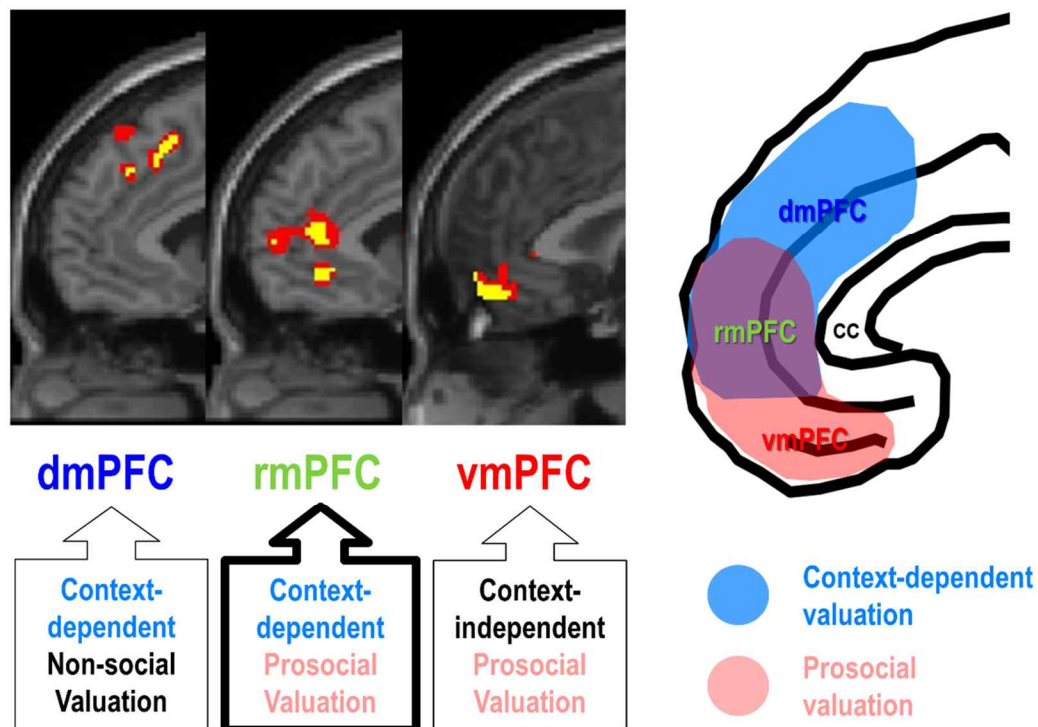


Figure 9. Functional segregation of mPFC function during ethical consumption under social observation. The rmPFC and the dmPFC encode subject-specific values of purchasing social products (prosocial decision) and non-social products (self-centered decision), respectively, under social observation (context-dependent), whereas the vmPFC encodes subject-specific values of purchasing social products regardless of social observation (context-independent) (Adapted from Jung et al., 2018).

Lastly, the rmPFC integrates social information to update the subjective estimation of self-efficacy or social status under competitive or cooperative social contexts. For example, several recent studies have shown that the rmPFC activity tracks trial-by-trial fluctuation of expected social dominance during competitive perceptual decision tasks (Ligneul, Obeso, Ruff, & Dreher, 2016), dynamic updates of self-efficacy, estimated based on self and other's performances in perceptual decision task (Wittmann et al., 2016), prediction error signals between expected and observed social feedback from others (Korn, Prehn, Park, Walter, & Heekeren, 2012; Will, Rutledge, Moutoussis, & Dolan, 2017), updating knowledge of one's own social hierarchy (Kumaran, Banino, Blundell, Hassabis, & Dayan, 2016), and value of self-protective behavior in response to negative social feedback (Yoon et al., 2018). According to a recent hypothesis, self-efficacy can be metacognitive beliefs about the brain's capacity to successfully regulate bodily states and the rmPFC plays a crucial role in such "allostatic self-efficacy" (Stephan et al., 2016; Wager et al., 2009). Taken together, these studies indicate that at the core of self-enhancement motivation lies the allostatic regulatory function of the rmPFC, that is, shaping the internal drive for (bodily) homeostasis so that it better fits into the constraints of

external (environmental) contextual variables (Smith, Thayer, Khalsa, & Lane, 2017).

In summary, under situations where one's impression or reputation is at stake, the rmPFC may arbitrate between intuitive motivation for self-enhancement (i.e., internal valuation) and careful consideration of contextual information (i.e., external valuation). Such an rmPFC arbitration function may be critical for the neural mechanism of allostatic regulatory control that serves to meet internal bodily needs in a socially relevant manner.

In the next section, I will introduce the hierarchical allostatic regulation model of mPFC function and show how hierarchically organized subregions of mPFC can interact with each other in such a way that more dorsal regions utilize additional external information from environment to predict and prevent conflicts occurring in more ventral regions tuned to internal bodily signals. In the end, I will show how such a hierarchical allostatic regulatory mPFC function can effectively address the stability-plasticity dilemma in a constantly changing social environment.

4.0 Hierarchical allostatic regulation model of mPFC function for social valuation

4.1 mPFC subregions encoding gradient of internal-to-external valuation

In the hierarchical model of social valuation, the mPFC comprises three functionally dissociable and hierarchically organized subregions: vmPFC, rmPFC, and dmPFC. These regions are differentially involved in computing values of decision along the ventral-to-dorsal spatial gradient of increasing external sensory inputs (e.g., via the temporal cortex and the parietal cortex) and decreasing internal inputs (e.g., via the brainstem, the hypothalamus, the amygdala, and the nucleus accumbens) (Dixon et al., 2017). Such an mPFC functional gradient is also consistent with the direction of evolutionary progress revealed by a recent analysis of the sulcal organization pattern across primate species (Amiez et al., 2019).

It should be noted, however, that the boundary between internal and external valuation is only a relative one. A level can be either internal or external depending on whether it is compared with its upper or lower level, respectively. For example, the vmPFC can be an internal valuation system when compared with the rmPFC but can also be an external valuation system when compared with the amygdala and the nucleus accumbens. In addition, the internal valuation of the vmPFC should be distinguished from instinctive or reflex-like responses that may be controlled by other structures, such as the amygdala, nucleus accumbens, and the spinal cord (located at a level further below the vmPFC). Therefore, the vmPFC can also be involved in valuation that requires abstract and sophisticated representations of a task structure if such representations became internalized through repetition (Hampton et al., 2008) as well as in goal-directed decisions when goals are primarily determined by internal signals carrying a homeostatic bodily state (Valentin et al., 2007).

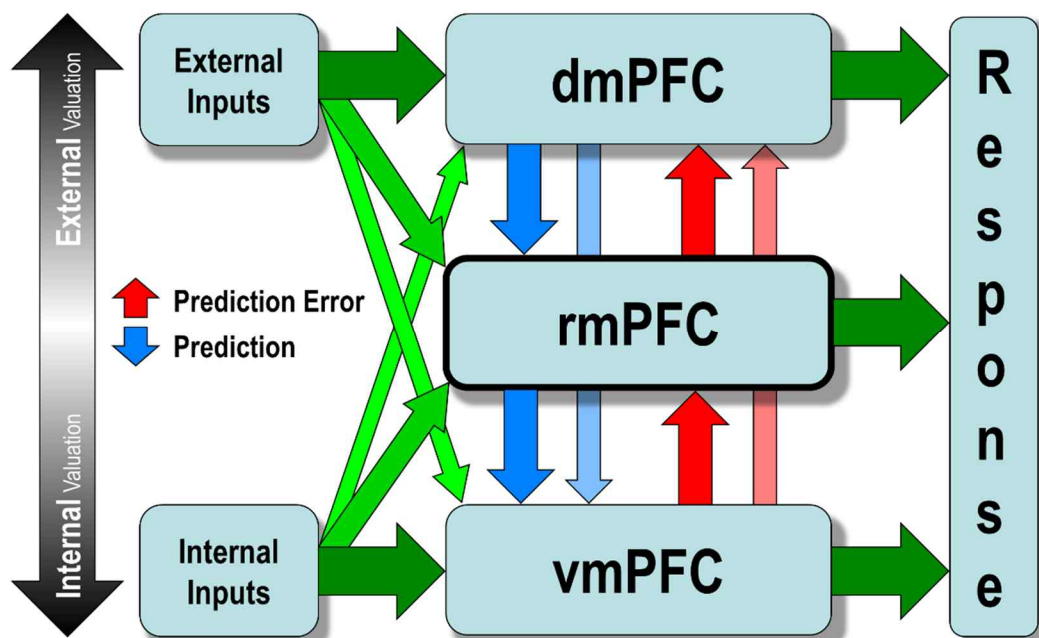


Figure 10. The hierarchical allostatic regulation model of mPFC function for social valuation. The mPFC comprises of three functionally dissociable hierarchically organized subregions: the vmPFC, rmPFC, and dmPFC, which are differentially involved in computing values of decision along the ventral-to-dorsal spatial gradient of increasing external sensory inputs (e.g., via the temporal cortex and the parietal cortex) and decreasing internal inputs (e.g., via the brainstem, the hypothalamus, the amygdala, and the nucleus accumbens). The vmPFC computes the internal valuation that generates interoceptive prediction signals and elicits a familiar intuitive response to prevent foreseen bodily imbalance. When two or more mutually incompatible values are simultaneously activated at the level of vmPFC, a conflict (prediction error) may occur triggering the upper levels (i.e., either rmPFC or dmPFC), which would then disengage internal valuation and increase the sensitivity to incoming sensory signals from the external environment to resolve the conflict. Such a process, called external valuation, sends prediction signals to update the preexisting value encoded at the lower level and continues until it finds a new value that resolves the conflict. The newly updated value will be strengthened and internalized through repetition so that it is activated more quickly and easily in similar situations later without causing a conflict.

Briefly speaking, the model detects prediction error in the internal valuation at the lower level and triggers the external valuation at the upper level to update the preexisting values at the lower level. Both external and internal input can trigger internalized values encoded by the vmPFC, which then activate interoceptive prediction signals that trigger a familiar, intuitive, and habit-like response to prevent anticipated bodily imbalance. Such a process is called internal valuation. However, when two or more mutually incompatible values are simultaneously activated at the level of the vmPFC, a conflict (i.e., prediction error) occurs, which then disengages the internal valuation and engages the upper levels (i.e., either rmPFC or dmPFC). The upper levels would then engage in resolving the conflict at the lower level by increasing sensitivity to incoming sensory signals from the external environment, taking over the decision control temporarily by searching for a new and more sophisticated stimulus-response mapping. Such a process, called external valuation,

sends prediction signals to update the pre-existing mapping at the lower level, and continues until it finds a new mapping that resolves the conflict. The new mapping will be strengthened and internalized through repetition so that it is activated quicker and more comfortably in similar future situations without causing a conflict (Figure 3).

This model naturally predicts that easy decisions should activate the vmPFC due to a weak conflict between options, whereas difficult decisions should elicit the dmPFC due to a strong conflict between options with similar value. Supporting this prediction, larger and smaller differences in value between two alternative options in a choice task were associated with greater activity in the vmPFC and the dmPFC, respectively (Hackel, Zaki, & Van Bavel, 2017; Piva et al., 2019). According to the model, in the case of larger value difference, the vmPFC alone can handle the value computation for decision, whereas, in the case of smaller value difference, the dmPFC needs to be engaged to integrate additional information from the environment to resolve the conflicts in the vmPFC.

To illustrate better how the model works, especially in the social domain, prosocial behavior can be a good example. If the value for a prosocial decision does not conflict with the value for a self-interested decision, then the prosocial behavior can be triggered rapidly via the intuitive internal valuation by the vmPFC. However, in cases where the two values conflict with each other in the vmPFC, additional information must be considered to resolve the conflict and to choose a more appropriate value in each context or to create a new value as a more appropriate alternative. By incorporating increasingly complex external information, the process of creating more efficient and sophisticated behavioral rules can produce various abstract social values, which can later be internalized through repetition. For example, an infant may learn to attract the attention of his/her caregiver to keep the caregiver closer. Such behaviors are often reinforced by the successful avoidance of hunger and insecurity. These primary forms of social reward can later serve as powerful internalized motivation for making friends and pursuing social status as a child grows up. Possibly, various secondary reinforcers, such as money and social rewards, are learned because they serve the common goal of preventing anticipated homeostatic imbalance (e.g., hunger or pain), and these newly acquired values can be internalized in the vmPFC to enable rapid comparisons between various types of rewards, serving as the common neural currency (Chib, Rangel, Shimojo, & O'Doherty, 2009; Izuma, Saito, & Sadato, 2008; H. Kim, Shimojo, & O'Doherty, 2011; D. J. Levy & Glimcher, 2012; Lin, Adolphs, & Rangel, 2012).

4.2 Role of the thalamic reticular nucleus in switching between thalamo-cortical loops

As mentioned above, a successful adaptation in a continually changing environment would require a careful assessment of the efficiency of the currently engaged decision system as well as a flexible transition between intuitive (internal) and deliberative (external) decision systems. What makes our brain achieve such an elegantly complex adaptive function? One potential neural candidate that is crucial for such functions may be the thalamic reticular nucleus (TRN). The TRN consists of

a layer of inhibitory neurons surrounding the thalamic nuclei, and can be divided into several sectors connected to different thalamic nuclei and their associated cortical regions, and is believed to serve as a nexus that moderates the interaction between separate sectors of thalamo-cortical loops and controls the transition between distinctive attentional modes (Crick, 1984; Guillery, Feig, & Lozsadi, 1998; Pinault, 2004). The TRN neurons exert an inhibitory control not only on the thalamo-cortical neurons but also on local inhibitory interneurons, which may lead to the disinhibition of the thalamo-cortical projection neurons (Steriade, Deschenes, Domich, & Mulle, 1985). The connection between the TRN neurons and the local inhibitory cells in the thalamus is believed to be subservient to the processes for focusing attention to relevant signals by suppressing other competing sub-networks processing non-relevant signals (Steriade, 1999), making the TRN neurons capable of detecting changes in the environment and modulating excitability of their target thalamic relay neurons (Yu, Xu, He, & He, 2009).

The TRN neurons can be segregated into at least two functionally and anatomically distinctive groups: the anterior (or limbic) TRN sector and the posterior (or sensory) TRN sector, which have distinct connectivity patterns (Zikopoulos & Barbas, 2012) and are shown to control different behavioral states (Halassa et al., 2014). The former has dense connections with the limbic regions, including the vmPFC and the amygdala, whereas the latter is connected more strongly with the sensory cortices and corresponding thalamic nuclei (Zikopoulos & Barbas, 2006, 2007, 2012), possibly being involved in focusing attention to internal signals and external stimuli, respectively (Halassa et al., 2014; Zikopoulos & Barbas, 2012). How could the TRN contribute to the transitions between internal and external processing? First, the TRN may be capable of hijacking on-going behaviors and mediates a rapid transition from the external (dorsal) to the internal (ventral) thalamocortical sector. More specifically, the internal sector can be engaged by default to address internal signals carrying information on anticipated homeostatic imbalance, and any sudden transition from an external sector to an internal sector appears to be driven by inputs from other internal sectors that are located even closer to the root of the hierarchical structure. Supporting this idea, the anterior sector of the TRN can be quickly engaged by the inputs from the amygdala and the vmPFC (Zikopoulos & Barbas, 2012). Such input signals can allow the TRN to interrupt and control on-going exploratory behaviors or external sensory processing and to initiate a rapid switch to the internal thalamocortical sector, which can then trigger stereotypical instinctive or reflexive behavioral responses that had been engaged repeatedly to deal with previous similar situations. Second, the TRN is also capable of switching between thalamocortical sectors in the opposite direction, that is, from internal to external sectors. Such shifts can be achieved by the densely distributed inhibitory neurons in the TRN, which are responsible for lateral inhibition between different thalamocortical loops via direct mutual inhibition or inhibitory projection to the thalamic projection cells (Crabtree, 2018; John, Zikopoulos, Bullock, & Barbas, 2016). Therefore, these neurons are ideally suited for detecting conflicts between simultaneously activated competing units in the internal sector of the TRN, which can lead to disinhibition of other non-occupied units in the external sector of the TRN network. To better illustrate these

complex dynamics of the TRN network, a simplified diagram of the TRN networks is shown in Figure 4.

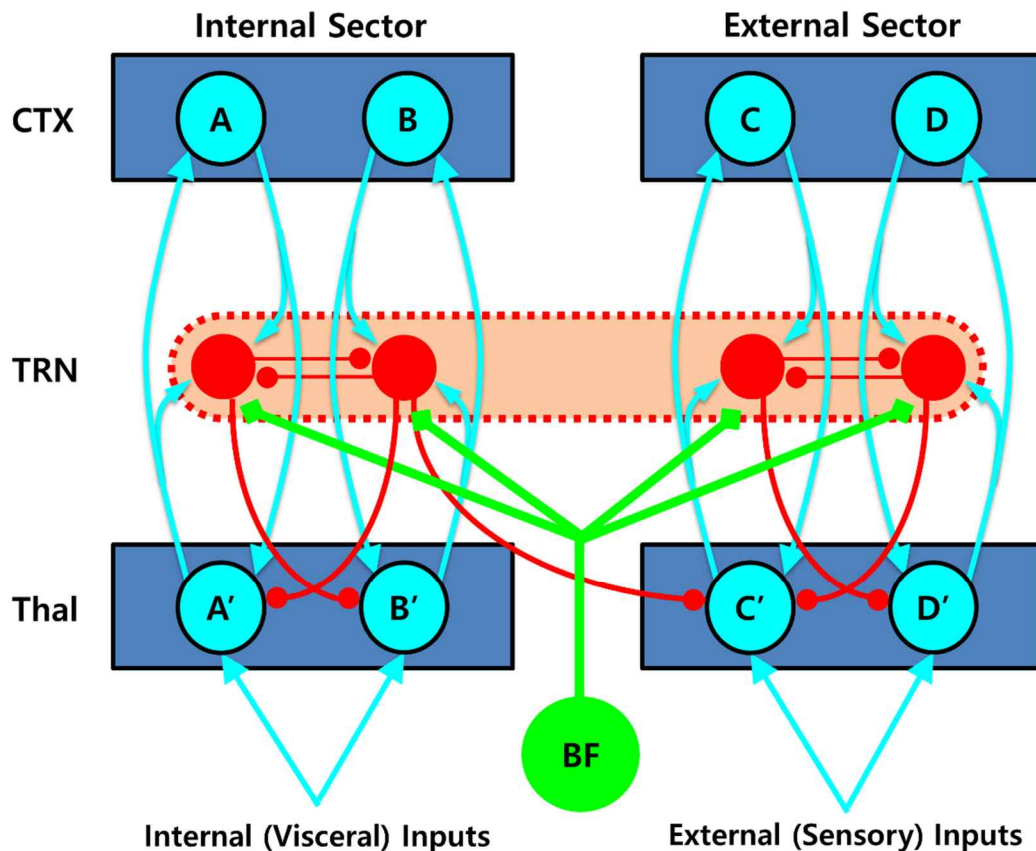


Figure 11. Role of the thalamic reticular nucleus (TRN) in shifting between the thalamo-cortical loops. Mutually inhibiting interneurons densely distributed in the TRN are perfectly suited for detecting conflicts between non-compatible units being engaged simultaneously, which can lead to disinhibition of other non-occupied units in the external sector, resulting in an attentional shift between different thalamo-cortical loops. In this diagram, simultaneous activation of two mutually competitive units (i.e., A and B) in the internal sector would lead to a sector-wide disinhibition of the thalamic projection neurons (i.e., C' and D') in the external sector, which would then gate the external sector, initiating more elaborated processing of additional external sensory information. Such inhibitory connections are likely to be asymmetrical, that is, favoring the direction from internal to external sectors and therefore prioritizing internal over external valuation. Note that not all the necessary connections are shown for visualization purpose. Abbreviations: CTX, cortex; TRN, thalamic reticular nucleus; Thal: thalamus; BF: basal forebrain.

For example, simultaneous activation of two mutually competitive units (i.e., A and B) in the internal sector could lead to a sector-wide disinhibition of the thalamic projection neurons (i.e., C' and D') in the external sector. Such a disinhibition would then gate the external sector, initiating more elaborated processing of additional external sensory information. Considering that intra-TRN connections are mainly formed by gap junctions (Hou, Smith, & Zhang, 2016), the inter-sector inhibition could be best achieved by multiple divergent inhibitory projections from the TRN cells to the thalamic projection cells (Crabtree, 2018).

Supporting this idea, such an ‘open-loop’ TRN network, wherein a thalamic projection neuron is inhibited by the TRN neuron excited not by itself but by another thalamic projection neuron, has been identified in anatomical studies (Kimura, Imbe, Donishi, & Tamai, 2007; Pinault & Deschenes, 1998), and demonstrated to play a major role in signal propagation across distinct thalamocortical loops in a recent computation modeling study (Brown, Taheri, Kenyon, Berger-Wolf, & Llano, 2020). It is noteworthy that such inter-sector inhibitory connections are assumed to be largely asymmetrical, that is, favoring the direction from internal to external sectors and therefore prioritizing internal over external valuation, consistent with the proposal that information propagates preferentially from a ventromedial to a dorsolateral direction across the thalamo-cortico-striatal loops (Yin & Knowlton, 2006). Although this is a speculative hypothesis that needs more concrete additional evidence, such an anatomical structure of the TRN network may be suitable for the mechanism of switching from the ventral (or internal) to the dorsal (external) thalamocortical sector that is functionally equivalent to the transition from an intuitive to a deliberative decision mode.

4.3 Factors affecting the optimal balance between stability and plasticity

In the model mentioned above, the external valuation process in the upper level continues until the conflict is resolved at the lower level. Thus, it is reasonable to question how much conflict in the lower level is large enough to trigger the external valuation process, and to what degree the conflict needs to be resolved before disengaging the external valuation process. Answers to these questions are essential to understand how our brains deal with the challenging problem of the stability-plasticity dilemma. That is, any adaptive agent is expected to effectively address the stability-plasticity dilemma (Grossberg, 2013) to avoid either catastrophic forgetting (i.e., extreme plasticity) (Ratcliff, 1990) or the entrenchment effect (i.e., extreme stability) (Zevin & Seidenberg, 2004). For the optimal balance between stability and plasticity, therefore, an agent needs to carefully determine the optimal level of tolerance for mismatches between incoming sensory inputs and previously learned categorical representation. However, an agent often fails to maintain the balance due to the lack of information about the ultimate goal of an organism, which has to do with how precisely a prediction or model generated at the moment matches forthcoming internal and external states (Friston, 2010).

Pursuing the optimal balance between stability and plasticity can be related to the concept of allostasis, which allows an organism to produce system-wide behavioral and physiological adjustments to environmental challenges through prediction in advance of a need (McEwen & Stellar, 1993; Schulkin, 2003; Sterling & Eyer, 1988). Several recent theories suggest that the mPFC along with the insula form a neurocircuitry thought to convey allostatic predictions that modulate the set points of homeostatic reflexes (Barrett & Simmons, 2015; Critchley & Harrison, 2013; Kleckner et al., 2017; Stephan et al., 2016; Sterling, 2014), possibly via rapid direct or indirect communications with the brainstem (Allman et al., 2010; Fischer et al., 2016). Perhaps, the role of the mPFC in allostasis can be best understood by examining this region as one of the key cortical substrates for heart-rate variability

(HRV) (Thayer, Ahs, Fredrikson, Sollers, & Wager, 2012). HRV is known to reflect the heart's ability to detect and respond adaptively to unpredictable environmental changes and is regarded as a possible indicator of allostatic capacity to integrate behavioral strategies and energy stores in response to environmental demands (Grossman & Taylor, 2007).

Consistent with the hierarchical allostatic regulation model of mPFC function proposed in the present study, different mPFC subregions may be involved in distinctive aspects of HRV-related functions. For example, the vmPFC activity is known to covary with HRV, and the rmPFC is suggested to play a regulatory role over the autonomic response initiated by the vmPFC (Ziegler et al., 2009). This idea is further supported by a recent meta-analysis showing that the vmPFC and the rmPFC are involved in sympathetic and parasympathetic processes, respectively (Beissner et al., 2013). In line with this, a recent finding showed that pharmacological inactivation of the vmPFC and the rmPFC led to decreased and increased autonomic and behavioral responses to negative emotional stimuli in non-human primates, respectively (Wallis, Cardinal, Alexander, Roberts, & Clarke, 2017). These findings suggest that, while the vmPFC quickly elicits familiar, internalized responses to cope with anticipated physical consequences of external stimuli, the rmPFC appears to seek a more holistic solution for harmonization between the fast-autonomic response and the constraints of the external environment. Perhaps, another key contributor to the optimal balance between stability and plasticity may be the neuromodulatory afferent signals targeting the allostatic neurocircuitry mentioned earlier as well as the TRN network. For example, there is evidence that the insula, the mPFC, and the TRN are the major targets of the cholinergic signals from the basal forebrain (Haber & Calzavara, 2009). In general, these neuromodulatory signals can potentiate presynaptic glutamatergic and GABAergic neurotransmission (Alkondon, Pereira, Barbosa, & Albuquerque, 1997; Freund, Jungschafer, Collins, & Wehner, 1988; Radcliffe, Fisher, Gray, & Dani, 1999). These signals can facilitate the competition between simultaneously activated mutually incompatible units, leading to an enhanced signal-to-noise ratio in the target area (Everitt & Robbins, 1997; Sarter & Bruno, 1997). Therefore, such cholinergic neuromodulatory inputs to the TRN inhibitory network could increase the sensitivity to the conflict in the internal sector as well as the likelihood of disinhibiting otherwise suppressed units in the external sector, which could then lead to enhanced precision in the environmental sensory information (Feldman & Friston, 2010). Through this process, cholinergic signals can lower the degree of tolerance for a mismatch between actual and predicted bodily states, which can then lead to frequent and prolonged engagement of the external valuation process that will continue to search for a new categorical representation to resolve the mismatch. In this sense, the role of cholinergic signals in modulating competition in the TRN is analogous to the concept of vigilance parameter that determines the allowable degree of mismatch between any input pattern and any stored patterns, resulting in either crude (i.e., low vigilance) or fine (i.e., high vigilance) categorization of incoming stimuli (Grossberg, 2013).

In addition to cholinergic signals, some other neuromodulatory signals, such as dopamine and noradrenaline, have been shown to serve similar functions; that is,

signaling the degree of sensitivity to the discrepancy between prediction (or belief) and actual sensory information (Friston et al., 2014). For example, the dopamine neurons in the midbrain signal the discrepancy between predicted and actual rewards (Schultz, 1998), and the noradrenergic neurons in the locus coeruleus can interrupt the activity of on-going functional networks and facilitate their reorganization to promote rapid behavioral adaptation (Bouret & Sara, 2005). These different neuromodulatory signals may serve a common goal of adjusting the balance between stability and plasticity, by reporting an integrative sum of internal milieu or interoceptive prediction errors to the brain (Fadel & Burk, 2010). To this end, a higher sensitivity to interoceptive prediction errors can lead to a higher vigilance, resulting in finer mappings between internal needs and external environment, which may indicate a more adaptive capacity for allostatic regulation. As an example in the domain of social neuroscience, people with a polymorphism in the dopamine D4 receptor gene (DRD4), which is associated with a higher sensitivity to environmental reward, are more likely to display behaviors that are more culturally dominant and socially desirable, compared to those with other types of DRD4 polymorphisms (Kitayama et al., 2014). In summary, different types of neuromodulatory signals may share a common goal of allowing the brain to continuously check and modulate the precision of interoceptive prediction and to build an accurate and sophisticated internal model of visceral states (Friston et al., 2014), which can be advantageous for pursuing adaptive behavior in a constantly changing social environment.

4.4 Self-control, self-efficacy, and metacognition as types of allostatic regulation

Recent theories suggest that interoceptive prediction errors reporting homeostatic or allostatic imbalance are essential for valuation of decisions (Gu & FitzGerald, 2014) as well as cognitive and goal-directed control over habitual actions (Pezzulo, Rigoli, & Friston, 2015). These theories can be further refined by considering the allostatic function of the rmPFC in solving the stability-plasticity dilemma. For example, people with high baseline HRV showed greater self-control during a food choice task and higher rmPFC activity when the participants had to overcome their taste preferences to choose the healthier option (Maier & Hare, 2017). Similarly, the rmPFC has also been shown to be involved in meta-decision, that is, choosing between distinctive decision systems. For example, the rmPFC is known to play a vital role in arbitrating the transition from the intuitive (model-free) to the analytical (model-based) decision systems, by tracking prediction error signals arising from the performance of the model-free system (Lee, Shimojo, & O'Doherty, 2014). Based on the model introduced above, such arbitration can be best explained by the function of the rmPFC in detecting and resolving conflict in internal valuation by referring to additional external information. When engaged, the rmPFC may quickly resolve the conflict in the vmPFC, or it may trigger an even higher level of external valuation computed by the dmPFC. The rmPFC may be particularly suitable for such a function of arbitration between internal and external valuation, because of its privileged anatomical feature of integrating balanced inputs from both internal and external sources of incoming information.

The rmPFC has also been implicated in metacognition. For example, the rmPFC has been shown to track changes in the subjective sense of decision confidence (Bang & Fleming, 2018), and lesions to the anterior sectors of the prefrontal cortex, including rmPFC, led to an impairment of perceptual metacognitive accuracy (Fleming, Ryu, Golfinos, & Blackmon, 2014). Similarly, the rmPFC function in context-sensitive reputation management (Jung et al., 2018) may reflect metacognitive monitoring of the appropriateness of intuitive and internalized valuation for self-enhancement under the constraints of social contexts, seeking the optimal decision via referring to interoceptive prediction error signals and engaging external valuation whenever necessary. From this point of view, the role of the rmPFC in seeking the balance between internal and external valuation, which may be critical for successful allostatic regulation, may also be the core neural foundation shared among various forms of adaptive functions such as self-efficacy, self-regulation, and metacognition (Zimmerman & Moylan, 2009).

5.0 CONCLUSIONS

The present study proposes that the mPFC subregions are hierarchically organized and differentially involved in computing values of decision, forming the ventral-to-dorsal spatial gradient of increasing external sensory inputs and decreasing internal inputs and exerting the hierarchical allostatic regulatory control over homeostatic reflexes. This hierarchical allostatic regulation model of mPFC function also emphasizes the role of the TRN in arbitrating the transitions between functionally dissociable thalamo-cortical loops. Because of its unique anatomical architecture with a robust inhibitory network, the TRN is capable of rapid and powerful orchestration over multiple thalamo-cortical loops and is critical for detecting and resolving conflicts between available options for decision-making and in shifting flexibly between decision modes. Importantly, neuromodulatory afferents to the TRN, signaling overall misfit of the interoceptive prediction model, can modulate the degree of sensitivity to the mismatch between predicted and observed bodily states. Such modulatory signals are critical for the mPFC function in allostatic regulation, which seeks an optimal balance between stability and plasticity, which can serve to maximize the probability of survival. In conclusion, the present model of the mPFC function can provide a useful theoretical framework, whereby previous findings once scattered around the mPFC area can be incorporated to generate and test novel hypotheses. Furthermore, future studies focusing on the hierarchical nature of the mPFC function can further expand our knowledge in various clinical symptoms, such as addiction, anxiety, and depression, which may be caused by a failure of the key neural mechanism for dealing with the stability-plasticity dilemma.

6.0 REFERENCES

- Alkondon, M., Pereira, E. F., Barbosa, C. T., & Albuquerque, E. X. (1997). Neuronal nicotinic acetylcholine receptor activation modulates gamma-aminobutyric acid release from CA1 neurons of rat hippocampal slices. *J Pharmacol Exp Ther*, 283(3), 1396-1411.

- Allman, J. M., Tetreault, N. A., Hakeem, A. Y., Manaye, K. F., Semendeferi, K., Erwin, J. M., . . . Hof, P. R. (2010). The von Economo neurons in fronto-insular and anterior cingulate cortex in great apes and humans. *Brain Struct Funct*, 214(5-6), 495-517. doi: 10.1007/s00429-010-0254-0
- Amiez, C., Sallet, J., Hopkins, W. D., Meguerditchian, A., Hadj-Bouziane, F., Hamed, S. B., . . . Petrides, M. (2019). Sulcal organization in the medial frontal cortex reveals insights into primate brain evolution. *Nat Commun*, 10, 3437.
- Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci*, 7(4), 268-277. doi: 10.1038/nrn1884
- Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron*, 65(4), 550-562. doi: 10.1016/j.neuron.2010.02.005
- Apps, M. A., & Ramnani, N. (2014). The anterior cingulate gyrus signals the net value of others' rewards. *J Neurosci*, 34(18), 6190-6200. doi: 10.1523/JNEUROSCI.2701-13.2014
- Bang, D., & Fleming, S. M. (2018). Distinct encoding of decision confidence in human medial prefrontal cortex. *Proc Natl Acad Sci U S A*, 115(23), 6082-6087. doi: 10.1073/pnas.1800795115
- Barbas, H. (2015). General cortical and special prefrontal connections: principles from structure to function. *Annu Rev Neurosci*, 38, 269-289. doi: 10.1146/annurev-neuro-071714-033936
- Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, 16(7), 419-429. doi: 10.1038/nrn3950
- Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage*, 76, 412-427. doi: 10.1016/j.neuroimage.2013.02.063
- Bas-Hoogendam, J. M., van Steenbergen, H., Kreuk, T., van der Wee, N. J. A., & Westenberg, P. M. (2017). How embarrassing! The behavioral and neural correlates of processing social norm violations. *PloS one*, 12(4), e0176326.
- Basile, B., Mancini, F., Macaluso, E., Caltagirone, C., Frackowiak, R. S. J., & Bozzali, M. (2011). Deontological and altruistic guilt: evidence for distinct neurobiological substrates. *Human Brain Mapping*, 32(2), 229-239.
- Basten, U., Biele, G., Heekeren, H. R., & Fiebach, C. J. (2010). How the brain integrates costs and benefits during decision making. *Proc Natl Acad Sci U S A*, 107(50), 21767-21772. doi: 10.1073/pnas.0908104107
- Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cereb Cortex*, 6(2), 215-225. doi: 10.1093/cercor/6.2.215
- Beer, J. S., John, O. P., Scabini, D., & Knight, R. T. (2006). Orbitofrontal cortex and social behavior: integrating self-monitoring and emotion-cognition interactions. *Journal of cognitive neuroscience*, 18(6), 871-879.
- Behrens, T. E., Hunt, L. T., & Rushworth, M. F. (2009). The computation of social behavior. *Science*, 324(5931), 1160-1164. doi: 10.1126/science.1169694
- Beissner, F., Meissner, K., Bar, K. J., & Napadow, V. (2013). The autonomic brain: an activation likelihood estimation meta-analysis for central processing of

- autonomic function. *J Neurosci*, 33(25), 10503-10511. doi: 10.1523/JNEUROSCI.1103-13.2013
- Bereczkei, T., Birkas, B., & Kerekes, Z. (2010). Altruism towards strangers in need: costly signaling in an industrial society. *Evolution and Human Behavior*, 31(2), 95-103. doi: 10.1016/j.evolhumbehav.2009.07.004
- Bouret, S., & Richmond, B. J. (2010). Ventromedial and orbital prefrontal neurons differentially encode internally and externally driven motivational values in monkeys. *J Neurosci*, 30(25), 8591-8601. doi: 10.1523/JNEUROSCI.0049-10.2010
- Bouret, S., & Sara, S. J. (2005). Network reset: a simplified overarching theory of locus coeruleus noradrenaline function. *Trends Neurosci*, 28(11), 574-582. doi: 10.1016/j.tins.2005.09.002
- Brown, J. W., Taheri, A., Kenyon, R. V., Berger-Wolf, T., & Llano, D. A. (2020). Signal Propagation Via Open-Loop Intrathalamic Architectures: A Computational Model. *eNeuro*. doi: 10.1523/ENEURO.0441-19.2020
- Buckholtz, J. W., & Marois, R. (2012). The roots of modern justice: cognitive and neural foundations of social norms and their enforcement. *Nat Neurosci*, 15(5), 655-661. doi: 10.1038/nn.3087
- Burnett, S., Bird, G., Moll, J., Frith, C. D., & Blakemore, S. (2009). Development during adolescence of the neural processing of social emotion. *Journal of cognitive neuroscience*, 21(9), 1736-1750.
- Bzdok, D., Langner, R., Schilbach, L., Engemann, D. A., Laird, A. R., Fox, P. T., & Eickhoff, S. B. (2013). Segregation of the human medial prefrontal cortex in social cognition. *Front Hum Neurosci*, 7, 232. doi: 10.3389/fnhum.2013.00232
- Cabeza, R., Dolcos, F., Graham, R., & Nyberg, L. (2002). Similarities and differences in the neural correlates of episodic memory retrieval and working memory. *Neuroimage*, 16(2), 317-330. doi: 10.1006/nimg.2002.1063
- Carmichael, S. T., & Price, J. L. (1995). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *J Comp Neurol*, 363(4), 615-641. doi: 10.1002/cne.903630408
- Chang, S. W., Gariépy, J. F., & Platt, M. L. (2013). Neuronal reference frames for social decisions in primate frontal cortex. *Nat Neurosci*, 16(2), 243-250. doi: 10.1038/nn.3287
- Chapados, C., & Petrides, M. (2015). Ventrolateral and dorsomedial frontal cortex lesions impair mnemonic context retrieval. *Proc Biol Sci*, 282(1801), 20142555. doi: 10.1098/rspb.2014.2555
- Chib, V. S., Rangel, A., Shimojo, S., & O'Doherty, J. P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *J Neurosci*, 29(39), 12315-12320. doi: 10.1523/JNEUROSCI.2575-09.2009
- Coutureau, E., & Killcross, S. (2003). Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. *Behav Brain Res*, 146(1-2), 167-174.
- Crabtree, J. W. (2018). Functional Diversity of Thalamic Reticular Subnetworks. *Front Syst Neurosci*, 12, 41. doi: 10.3389/fnsys.2018.00041

- Crick, F. (1984). Function of the thalamic reticular complex: the searchlight hypothesis. *Proc Natl Acad Sci U S A*, 81(14), 4586-4590.
- Critchley, H. D., & Harrison, N. A. (2013). Visceral Influences on Brain and Behavior. *Neuron*, 77(4), 624-638. doi: 10.1016/j.neuron.2013.02.008
- Crockett, M. J. (2013). Models of morality. *Trends Cogn Sci*, 17(8), 363-366. doi: 10.1016/j.tics.2013.06.005
- Cutler, J., & Campbell-Meiklejohn, D. (2019). A comparative fMRI meta-analysis of altruistic and strategic decisions to give. *Neuroimage*, 184, 227-241. doi: 10.1016/j.neuroimage.2018.09.009
- D'Argembeau, A. (2013). On the role of the ventromedial prefrontal cortex in self-processing: the valuation hypothesis. *Front Hum Neurosci*, 7, 372. doi: 10.3389/fnhum.2013.00372
- de Araujo, I. E., Ferreira, J. G., Tellez, L. A., Ren, X., & Yeckel, C. W. (2012). The gut-brain dopamine axis: a regulatory system for caloric intake. *Physiol Behav*, 106(3), 394-399. doi: 10.1016/j.physbeh.2012.02.026
- de la Vega, A., Chang, L. J., Banich, M. T., Wager, T. D., & Yarkoni, T. (2016). Large-Scale Meta-Analysis of Human Medial Frontal Cortex Reveals Tripartite Functional Organization. *J Neurosci*, 36(24), 6553-6562. doi: 10.1523/JNEUROSCI.4402-15.2016
- De Martino, B., Camerer, C. F., & Adolphs, R. (2010). Amygdala damage eliminates monetary loss aversion. *Proceedings of the National Academy of Sciences of the United States of America*, 107(8), 3788-3792. doi: 10.1073/pnas.0910230107
- de Wit, S., Corlett, P. R., Aitken, M. R., Dickinson, A., & Fletcher, P. C. (2009). Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. *J Neurosci*, 29(36), 11330-11338. doi: 10.1523/JNEUROSCI.1639-09.2009
- Delgado, M. R., Beer, J. S., Fellows, L. K., Huettel, S. A., Platt, M. L., Quirk, G. J., & Schiller, D. (2016). Viewpoints: Dialogues on the functional role of the ventromedial prefrontal cortex. *Nat Neurosci*, 19(12), 1545-1552. doi: 10.1038/nn.4438
- Demos, K. E., Heatherton, T. F., & Kelley, W. M. (2012). Individual Differences in Nucleus Accumbens Activity to Food and Sexual Images Predict Weight Gain and Sexual Behavior. *Journal of Neuroscience*, 32(16), 5549-5552. doi: 10.1523/Jneurosci.5958-11.2012
- Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *J Cogn Neurosci*, 24(8), 1742-1752. doi: 10.1162/jocn_a_00233
- Dixon, M. L., Thiruchselvam, R., Todd, R., & Christoff, K. (2017). Emotion and the prefrontal cortex: An integrative review. *Psychol Bull*, 143(10), 1033-1081. doi: 10.1037/bul0000096
- Ebitz, R. B., & Hayden, B. Y. (2016). Dorsal anterior cingulate: a Rorschach test for cognitive neuroscience. *Nature Neuroscience*, 19(10), 1278-1279. doi: 10.1038/nn.4387

- Edelmann, R. J. (1987). *The psychology of embarrassment*. Chichester, England: Wiley.
- Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science*, 302(5643), 290-292. doi: 10.1126/science.1089134
- Everitt, B. J., & Robbins, T. W. (1997). Central cholinergic systems and cognition. *Annu Rev Psychol*, 48, 649-684. doi: 10.1146/annurev.psych.48.1.649
- Fadel, J., & Burk, J. A. (2010). Orexin/hypocretin modulation of the basal forebrain cholinergic system: Role in attention. *Brain Res*, 1314, 112-123. doi: 10.1016/j.brainres.2009.08.046
- Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425(6960), 785-791. doi: 10.1038/nature02043
- Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Front Hum Neurosci*, 4, 215. doi: 10.3389/fnhum.2010.00215
- Fischer, D. B., Boes, A. D., Demertzi, A., Evrard, H. C., Laureys, S., Edlow, B. L., . . . Geerling, J. C. (2016). A human brain network derived from coma-causing brainstem lesions. *Neurology*, 87(23), 2427-2434. doi: 10.1212/WNL.0000000000003404
- Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain*, 137(Pt 10), 2811-2822. doi: 10.1093/brain/awu221
- Fourie, M. M., Thomas, K. G. F., Amodio, D. M., Warton, C. M. R., & Meintjes, E. M. (2014). Neural correlates of experienced moral emotion: an fMRI investigation of emotion in response to prejudice feedback. *Social neuroscience*, 9(2), 203-218.
- Freund, R. K., Jungschaffer, D. A., Collins, A. C., & Wehner, J. M. (1988). Evidence for modulation of GABAergic neurotransmission by nicotine. *Brain Res*, 453(1-2), 215-220. doi: 10.1016/0006-8993(88)90160-6
- Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127-138. doi: 10.1038/nrn2787
- Friston, K. J., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2014). The anatomy of choice: dopamine and decision-making. *Philos Trans R Soc Lond B Biol Sci*, 369(1655). doi: 10.1098/rstb.2013.0481
- Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, 50(4), 531-534. doi: 10.1016/j.neuron.2006.05.001
- Fukuda, H., Ma, N., Suzuki, S., Harasawa, N., Ueno, K., Gardner, J. L., . . . Nakahara, H. (2019). Computing Social Value Conversion in the Human Brain. *J Neurosci*, 39(26), 5153-5172. doi: 10.1523/JNEUROSCI.3117-18.2019
- Ghashghaei, H. T., Hilgetag, C. C., & Barbas, H. (2007). Sequence of information processing for emotions based on the anatomic dialogue between prefrontal cortex and amygdala. *Neuroimage*, 34(3), 905-923. doi: 10.1016/j.neuroimage.2006.09.046
- Gilead, M., Katzir, M., Eyal, T., & Liberman, N. (2016). Neural correlates of processing “self-conscious” vs. “basic” emotions. *Neuropsychologia*, 81, 207-218.

- Griskevicius, V., Tybur, J. M., & Van den Bergh, B. (2010). Going green to be seen: status, reputation, and conspicuous conservation. *J Pers Soc Psychol*, 98(3), 392-404. doi: 10.1037/a0017346
- Grossberg, S. (2013). Adaptive Resonance Theory: how a brain learns to consciously attend, learn, and recognize a changing world. *Neural Netw*, 37, 1-47. doi: 10.1016/j.neunet.2012.09.017
- Grossman, P., & Taylor, E. W. (2007). Toward understanding respiratory sinus arrhythmia: relations to cardiac vagal tone, evolution and biobehavioral functions. *Biol Psychol*, 74(2), 263-285. doi: 10.1016/j.biopsycho.2005.11.014
- Gu, X., & FitzGerald, T. H. (2014). Interoceptive inference: homeostasis and decision-making. *Trends Cogn Sci*, 18(6), 269-270. doi: 10.1016/j.tics.2014.02.001
- Guillery, R. W., Feig, S. L., & Lozsadi, D. A. (1998). Paying attention to the thalamic reticular nucleus. *Trends Neurosci*, 21(1), 28-32.
- Haber, S. N., & Calzavara, R. (2009). The cortico-basal ganglia integrative network: the role of the thalamus. *Brain Res Bull*, 78(2-3), 69-74. doi: 10.1016/j.brainresbull.2008.09.013
- Haber, S. N., Kim, K. S., Mailly, P., & Calzavara, R. (2006). Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, providing a substrate for incentive-based learning. *J Neurosci*, 26(32), 8368-8376. doi: 10.1523/JNEUROSCI.0271-06.2006
- Haber, S. N., & Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology*, 35(1), 4-26. doi: 10.1038/npp.2009.129
- Hackel, L. M., Zaki, J., & Van Bavel, J. J. (2017). Social identity shapes social valuation: evidence from prosocial behavior and vicarious reward. *Soc Cogn Affect Neurosci*, 12(8), 1219-1228. doi: 10.1093/scan/nsx045
- Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316(5827), 998-1002. doi: 10.1126/science.1137651
- Halassa, M. M., Chen, Z., Wimmer, R. D., Brunetti, P. M., Zhao, S., Zikopoulos, B., . . . Wilson, M. A. (2014). State-dependent architecture of thalamic reticular subnetworks. *Cell*, 158(4), 808-821. doi: 10.1016/j.cell.2014.06.025
- Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc Natl Acad Sci U S A*, 105(18), 6741-6746. doi: 10.1073/pnas.0711099105
- Han, W., Tellez, L. A., Perkins, M. H., Perez, I. O., Qu, T., Ferreira, J., . . . de Araujo, I. E. (2018). A Neural Circuit for Gut-Induced Reward. *Cell*, 175(3), 887-888. doi: 10.1016/j.cell.2018.10.018
- Hardy, C. L., & Van Vugt, M. (2006). Nice guys finish first: the competitive altruism hypothesis. *Pers Soc Psychol Bull*, 32(10), 1402-1413. doi: 10.1177/0146167206291006
- Hare, T. A., Camerer, C. F., Knoepfle, D. T., & Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *J Neurosci*, 30(2), 583-590. doi: 10.1523/JNEUROSCI.4089-09.2010

- Haruno, M., & Kawato, M. (2006). Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning. *Neural Netw*, 19(8), 1242-1254. doi: 10.1016/j.neunet.2006.06.007
- Horst, N. K., & Laubach, M. (2009). The role of rat dorsomedial prefrontal cortex in spatial working memory. *Neuroscience*, 164(2), 444-456. doi: 10.1016/j.neuroscience.2009.08.004
- Hou, G., Smith, A. G., & Zhang, Z. W. (2016). Lack of Intrinsic GABAergic Connections in the Thalamic Reticular Nucleus of the Mouse. *J Neurosci*, 36(27), 7246-7252. doi: 10.1523/JNEUROSCI.0607-16.2016
- Howard, J. D., Gottfried, J. A., Tobler, P. N., & Kahnt, T. (2015). Identity-specific coding of future rewards in the human orbitofrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 112(16), 5195-5200. doi: 10.1073/pnas.1503550112
- Hughes, B. L., & Beer, J. S. (2012). Medial orbitofrontal cortex is associated with shifting decision thresholds in self-serving cognition. *Neuroimage*, 61(4), 889-898. doi: 10.1016/j.neuroimage.2012.03.011
- Hughes, B. L., & Beer, J. S. (2013). Protecting the Self: The Effect of Social-evaluative Threat on Neural Representations of Self. *Journal of Cognitive Neuroscience*, 25(4), 613-622.
- Hutcherson, C. A., Bushong, B., & Rangel, A. (2015). A Neurocomputational Model of Altruistic Choice and Its Implications. *Neuron*, 87(2), 451-462. doi: 10.1016/j.neuron.2015.06.031
- Ikemoto, S., & Panksepp, J. (1999). The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. *Brain Research Reviews*, 31(1), 6-41. doi: Doi 10.1016/S0165-0173(99)00023-5
- Izuma, K. (2012). The social neuroscience of reputation. *Neurosci Res*, 72(4), 283-288. doi: 10.1016/j.neures.2012.01.003
- Izuma, K., Saito, D. N., & Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron*, 58(2), 284-294. doi: 10.1016/j.neuron.2008.03.020
- Izuma, K., Saito, D. N., & Sadato, N. (2010). The roles of the medial prefrontal cortex and striatum in reputation processing. *Social Neuroscience*, 5(2), 133-147.
- Janowski, V., Camerer, C., & Rangel, A. (2013). Empathic choice involves vmPFC value signals that are modulated by social processing implemented in IPL. *Soc Cogn Affect Neurosci*, 8(2), 201-208. doi: 10.1093/scan/nsr086
- Jocham, G., Klein, T. A., & Ullsperger, M. (2011). Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. *J Neurosci*, 31(5), 1606-1613. doi: 10.1523/JNEUROSCI.3904-10.2011
- John, Y. J., Zikopoulos, B., Bullock, D., & Barbas, H. (2016). The Emotional Gatekeeper: A Computational Model of Attentional Selection and Suppression through the Pathway from the Amygdala to the Inhibitory Thalamic Reticular

- Nucleus. *PLoS Comput Biol*, 12(2), e1004722. doi: 10.1371/journal.pcbi.1004722
- Joyce, M. K. P., & Barbas, H. (2018). Cortical Connections Position Primate Area 25 as a Keystone for Interoception, Emotion, and Memory. *J Neurosci*, 38(7), 1677-1698. doi: 10.1523/JNEUROSCI.2363-17.2017
- Jung, D., Sul, S., & Kim, H. (2013). Dissociable neural processes underlying risky decisions for self versus other. *Front Neurosci*, 7, 15. doi: 10.3389/fnins.2013.00015
- Jung, D., Sul, S., Lee, M., & Kim, H. (2018). Social Observation Increases Functional Segregation between MPFC Subregions Predicting Prosocial Consumer Decisions. *Sci Rep*, 8(1), 3368. doi: 10.1038/s41598-018-21449-z
- Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature neuroscience*, 10(12), 1625-1633.
- Kang, P., Lee, J., Sul, S., & Kim, H. (2013). Dorsomedial prefrontal cortex activity predicts the accuracy in estimating others' preferences. *Front Hum Neurosci*, 7, 686. doi: 10.3389/fnhum.2013.00686
- Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *J Cogn Neurosci*, 14(5), 785-794. doi: 10.1162/08989290260138672
- Keltner, D., & Buswell, B. N. (1997). Embarrassment: Its distinct form and appeasement functions. *Psychological Bulletin*, 122, 250-270.
- Killcross, S., & Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb Cortex*, 13(4), 400-408.
- Kim, H., Adolphs, R., O'Doherty, J. P., & Shimojo, S. (2007). Temporal isolation of neural processes underlying face preference decisions. *Proceedings of the National Academy of Sciences*, 104(46), 18253-18258.
- Kim, H., Shimojo, S., & O'Doherty, J. P. (2006). Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biol*, 4(8), e233. doi: 10.1371/journal.pbio.0040233
- Kim, H., Shimojo, S., & O'Doherty, J. P. (2011). Overlapping responses for the expectation of juice and money rewards in human ventromedial prefrontal cortex. *Cereb Cortex*, 21(4), 769-776. doi: 10.1093/cercor/bhq145
- Kim, K., & Johnson, M. K. (2015). Activity in ventromedial prefrontal cortex during self-related processing: positive subjective value or personal significance? *Soc Cogn Affect Neurosci*, 10(4), 494-500. doi: 10.1093/scan/nsu078
- Kimura, A., Imbe, H., Donishi, T., & Tamai, Y. (2007). Axonal projections of single auditory neurons in the thalamic reticular nucleus: implications for tonotopy-related gating function and cross-modal modulation. *Eur J Neurosci*, 26(12), 3524-3535. doi: 10.1111/j.1460-9568.2007.05925.x
- Kimura, A., Mukawa, N., Yamamoto, M., Masuda, T., Yuasa, M., Goto, S., . . . Wada, Y. (2012). The influence of reputational concerns on purchase intention of fair-trade foods among young Japanese adults. *Food Quality and Preference*, 26(2), 204-210. doi: 10.1016/j.foodqual.2012.05.002
- Kitayama, S., King, A., Yoon, C., Tompson, S., Huff, S., & Liberzon, I. (2014). The dopamine D4 receptor gene (DRD4) moderates cultural difference in

- independent versus interdependent social orientation. *Psychol Sci*, 25(6), 1169-1177. doi: 10.1177/0956797614528338
- Kleckner, I. R., Zhang, J., Touroutoglou, A., Chanes, L., Xia, C., Simmons, W. K., . . . Barrett, L. F. (2017). Evidence for a Large-Scale Brain System Supporting Allostasis and Interoception in Humans. *Nat Hum Behav*, 1. doi: 10.1038/s41562-017-0069
- Koechlin, E., & Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science*, 318(5850), 594-598. doi: 10.1126/science.1142995
- Kolling, N., Wittmann, M. K., Behrens, T. E. J., Boorman, E. D., Mars, R. B., & Rushworth, M. F. S. (2016). Value, search, persistence and model updating in anterior cingulate cortex. *Nature Neuroscience*, 19(10), 1280-1285. doi: 10.1038/nn.4382
- Korn, C. W., Prehn, K., Park, S. Q., Walter, H., & Heekeren, H. R. (2012). Positively biased processing of self-relevant social feedback. *J Neurosci*, 32(47), 16832-16844. doi: 10.1523/JNEUROSCI.3016-12.2012
- Krajbich, I., Adolphs, R., Tranel, D., Denburg, N. L., & Camerer, C. F. (2009). Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *Journal of Neuroscience*, 29(7), 2188-2192.
- Krienen, F. M., Tu, P. C., & Buckner, R. L. (2010). Clan mentality: evidence that the medial prefrontal cortex responds to close others. *J Neurosci*, 30(41), 13906-13915. doi: 10.1523/JNEUROSCI.2180-10.2010
- Kumaran, D., Banino, A., Blundell, C., Hassabis, D., & Dayan, P. (2016). Computations Underlying Social Hierarchy Learning: Distinct Neural Mechanisms for Updating and Representing Self-Relevant Information. *Neuron*, 92(5), 1135-1147. doi: 10.1016/j.neuron.2016.10.052
- LaLumiere, R. T., Niehoff, K. E., & Kalivas, P. W. (2010). The infralimbic cortex regulates the consolidation of extinction after cocaine self-administration. *Learning & Memory*, 17(4), 168-175. doi: 10.1101/lm.1576810
- Lawrence, N. S., Hinton, E. C., Parkinson, J. A., & Lawrence, A. D. (2012). Nucleus accumbens response to food cues predicts subsequent snack consumption in women and increased body mass index in those with reduced self-control. *Neuroimage*, 63(1), 415-422. doi: 10.1016/j.neuroimage.2012.06.070
- Leary, M. R., & Kowalski, R. M. (1990). Impression Management: A Literature Review and Two-Component Model. *Psychological Bulletin*, 107(1), 34-47.
- Lebreton, M., Jorge, S., Michel, V., Thirion, B., & Pessiglione, M. (2009). An automatic valuation system in the human brain: evidence from functional neuroimaging. *Neuron*, 64(3), 431-439. doi: 10.1016/j.neuron.2009.09.040
- Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural Computations Underlying Arbitration between Model-Based and Model-free Learning. *Neuron*, 81(3), 687-699. doi: 10.1016/j.neuron.2013.11.028
- Levy, D. J., & Glimcher, P. W. (2012). The root of all value: a neural common currency for choice. *Curr Opin Neurobiol*, 22(6), 1027-1038. doi: 10.1016/j.conb.2012.06.001
- Levy, I., Lazzaro, S. C., Rutledge, R. B., & Glimcher, P. W. (2011). Choice from non-choice: predicting consumer preferences from blood oxygenation level-

- dependent signals obtained during passive viewing. *J Neurosci*, 31(1), 118-125. doi: 10.1523/JNEUROSCI.3214-10.2011
- Lieberman, M. D., Straccia, M. A., Meyer, M. L., Du, M., & Tan, K. M. (2019). Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): Causal, multivariate, and reverse inference evidence. *Neurosci Biobehav Rev*, 99, 311-328. doi: 10.1016/j.neubiorev.2018.12.021
- Ligneul, R., Obeso, I., Ruff, C. C., & Dreher, J. C. (2016). Dynamical Representation of Dominance Relationships in the Human Rostromedial Prefrontal Cortex. *Curr Biol*, 26(23), 3107-3115. doi: 10.1016/j.cub.2016.09.015
- Lin, A., Adolphs, R., & Rangel, A. (2012). Social and monetary reward learning engage overlapping neural substrates. *Soc Cogn Affect Neurosci*, 7(3), 274-281. doi: 10.1093/scan/nsr006
- Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neurosci Biobehav Rev*, 35(5), 1219-1236. doi: 10.1016/j.neubiorev.2010.12.012
- Lockwood, P. L., Apps, M. A., Roiser, J. P., & Viding, E. (2015). Encoding of Vicarious Reward Prediction in Anterior Cingulate Cortex and Relationship with Trait Empathy. *J Neurosci*, 35(40), 13720-13727. doi: 10.1523/JNEUROSCI.1703-15.2015
- Müller-Pinzler, L., Gazzola, Valeria, Keysers, Christian, Sommer, J., Jansen, Andreas, Frässle, Stefan, . . . Krach, Sören. (2015). Neural pathways of embarrassment and their modulation by social anxiety. *NeuroImage*, 119, 252-261.
- Mackey, S., & Petrides, M. (2014). Architecture and morphology of the human ventromedial prefrontal cortex. *Eur J Neurosci*, 40(5), 2777-2796. doi: 10.1111/ejn.12654
- Maier, S. U., & Hare, T. A. (2017). Higher Heart-Rate Variability Is Associated with Ventromedial Prefrontal Cortex Activity and Increased Resistance to Temptation in Dietary Self-Control Challenges. *J Neurosci*, 37(2), 446-455. doi: 10.1523/JNEUROSCI.2815-16.2016
- Mansouri, F. A., Koechlin, E., Rosa, M. G. P., & Buckley, M. J. (2017). Managing competing goals - a key role for the frontopolar cortex. *Nat Rev Neurosci*, 18(11), 645-657. doi: 10.1038/nrn.2017.111
- Mason, W. A., Capitanio, J. P., Machado, C. J., Mendoza, S. P., & Amaral, D. G. (2006). Amygdectomy and responsiveness to novelty in rhesus monkeys (*Macaca mulatta*): Generality and individual consistency of effects. *Emotion*, 6(1), 73-81. doi: 10.1037/1528-3542.6.1.73
- McEwen, B. S., & Stellar, E. (1993). Stress and the individual. Mechanisms leading to disease. *Arch Intern Med*, 153(18), 2093-2101.
- Michl, P., Meindl, T., Meister, F., Born, C., Engel, R. R., Reiser, M., & Hennig-Fast, K. (2012). Neurobiological underpinnings of shame and guilt: a pilot fMRI study. *Social cognitive and affective neuroscience*, 9(2), 150-157.
- Milad, M. R., Pitman, R. K., Ellis, C. B., Gold, A. L., Shin, L. M., Lasko, N. B., . . . Rauch, S. L. (2009). Neurobiological Basis of Failure to Recall Extinction Memory in Posttraumatic Stress Disorder. *Biological Psychiatry*, 66(12), 1075-1082. doi: 10.1016/j.biopsych.2009.06.026

- Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, 50(4), 655-663. doi: 10.1016/j.neuron.2006.03.040
- Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., & Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc Natl Acad Sci U S A*, 103(42), 15623-15628. doi: 10.1073/pnas.0604475103
- Moll, J., Zahn, R., de Oliveira-Souza, R., Bramati, I. E., Krueger, F., Tura, B., . . . Grafman, J. (2011). Impairment of prosocial sentiments is associated with frontopolar and septal damage in frontotemporal dementia. *Neuroimage*, 54(2), 1735-1742.
- Moran, J. M., Macrae, C. N., Heatherton, T. F., Wyland, C. L., & Kelley, W. M. (2006). Neuroanatomical evidence for distinct cognitive and affective components of self. *J Cogn Neurosci*, 18(9), 1586-1594. doi: 10.1162/jocn.2006.18.9.1586
- Nakao, T., Ohira, H., & Northoff, G. (2012). Distinction between Externally vs. Internally Guided Decision-Making: Operational Differences, Meta-Analytical Comparisons and Their Theoretical Implications. *Front Neurosci*, 6, 31. doi: 10.3389/fnins.2012.00031
- Nicolle, A., Klein-Flugge, M. C., Hunt, L. T., Vlaev, I., Dolan, R. J., & Behrens, T. E. (2012). An agent independent axis for executed and modeled choice in medial prefrontal cortex. *Neuron*, 75(6), 1114-1121. doi: 10.1016/j.neuron.2012.07.023
- Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *Neuroimage*, 31(1), 440-457.
- Nowak, M. A., & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature*, 364(6432), 56-58. doi: 10.1038/364056a0
- Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature*, 393(6685), 573-577. doi: 10.1038/31225
- O'Doherty, J. P. (2011). Contributions of the ventromedial prefrontal cortex to goal-directed action selection. *Ann N Y Acad Sci*, 1239, 118-129. doi: 10.1111/j.1749-6632.2011.06290.x
- Palomero-Gallagher, N., Hoffstaedter, F., Mohlberg, H., Eickhoff, S. B., Amunts, K., & Zilles, K. (2019). Human Pregenuel Anterior Cingulate Cortex: Structural, Functional, and Connectional Heterogeneity. *Cereb Cortex*, 29(6), 2552-2574. doi: 10.1093/cercor/bhy124
- Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. *Prog Neurobiol*, 134, 17-35. doi: 10.1016/j.pneurobio.2015.09.001
- Pinault, D. (2004). The thalamic reticular nucleus: structure, function and concept. *Brain Res Brain Res Rev*, 46(1), 1-31. doi: 10.1016/j.brainresrev.2004.04.008
- Pinault, D., & Deschenes, M. (1998). Anatomical evidence for a mechanism of lateral inhibition in the rat thalamus. *Eur J Neurosci*, 10(11), 3462-3469. doi: 10.1046/j.1460-9568.1998.00362.x

- Piva, M., Velnoskey, K., Jia, R., Nair, A., Levy, I., & Chang, S. W. (2019). The dorsomedial prefrontal cortex computes task-invariant relative subjective value for self and other. *Elife*, 8. doi: 10.7554/eLife.44939
- Powers, K. E., Somerville, L. H., Kelley, W. M., & Heatherton, T. F. (2013). Rejection sensitivity polarizes striatal–medial prefrontal activity when anticipating social feedback. *Journal of cognitive neuroscience*, 25(11), 1887-1895.
- Quirk, G. J., & Beer, J. S. (2006). Prefrontal involvement in the regulation of emotion: convergence of rat and human studies. *Curr Opin Neurobiol*, 16(6), 723-727. doi: 10.1016/j.conb.2006.07.004
- Radcliffe, K. A., Fisher, J. L., Gray, R., & Dani, J. A. (1999). Nicotinic modulation of glutamate and GABA synaptic transmission of hippocampal neurons. *Ann N Y Acad Sci*, 868, 591-610. doi: 10.1111/j.1749-6632.1999.tb11332.x
- Rainville, P., Duncan, G. H., Price, D. D., Carrier, B., & Bushnell, M. C. (1997). Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science*, 277(5328), 968-971. doi: 10.1126/science.277.5328.968
- Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, 489(7416), 427-430. doi: 10.1038/nature11467
- Ratcliff, R. (1990). Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. *Psychol Rev*, 97(2), 285-308. doi: 10.1037/0033-295x.97.2.285
- Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *Neuroimage*, 22(4), 1694-1703. doi: 10.1016/j.neuroimage.2004.04.015
- Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends Cogn Sci*, 16(3), 147-156. doi: 10.1016/j.tics.2012.01.005
- Sarter, M., & Bruno, J. P. (1997). Cognitive functions of cortical acetylcholine: toward a unifying hypothesis. *Brain Res Brain Res Rev*, 23(1-2), 28-46.
- Schneirla, T. C. (1959). An Evolutionary and Developmental Theory of Biphasic Processes Underlying Approach and Withdrawal. *Nebraska Symposium on Motivation*, 7, 1-42.
- Schulkin, J. (2003). *Rethinking Homeostasis*. Cambridge: MIT Press.
- Schultz, W. (1998). Predictive reward signal of dopamine neurons. *J Neurophysiol*, 80(1), 1-27. doi: 10.1152/jn.1998.80.1.1
- Schwartz, C. E., Wright, C. I., Shin, L. M., Kagan, J., & Rauch, S. L. (2003). Inhibited and uninhibited infants "grown up": Adult amygdalar response to novelty. *Science*, 300(5627), 1952-1953. doi: DOI 10.1126/science.1083703
- Seo, H., Cai, X., Donahue, C. H., & Lee, D. (2014). Neural correlates of strategic reasoning during competitive games. *Science*, 346(6207), 340-343. doi: 10.1126/science.1256254
- Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., & Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat Rev Neurosci*, 12(3), 154-167. doi: 10.1038/nrn2994

- Shenhav, A., Cohen, J. D., & Botvinick, M. M. (2016). Dorsal anterior cingulate cortex and the value of control. *Nature Neuroscience*, 19(10), 1286-1291.
- Shenhav, A., & Greene, J. D. (2010). Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude. *Neuron*, 67(4), 667-677. doi: 10.1016/j.neuron.2010.07.020
- Shin, L. M., Dougherty, D. D., Orr, S. P., Pitman, R. K., Lasko, M., Macklin, M. L., . . . Rauch, S. L. (2000). Activation of anterior paralimbic structures during guilt-related script-driven imagery. *Biological psychiatry*, 48(1), 43-50.
- Smith, R., Thayer, J. F., Khalsa, S. S., & Lane, R. D. (2017). The hierarchical basis of neurovisceral integration. *Neurosci Biobehav Rev*, 75, 274-296. doi: 10.1016/j.neubiorev.2017.02.003
- Soetevent, A. R. (2005). Anonymity in giving in a natural context—a field experiment in 30 churches. *Journal of Public Economics*, 89(11-12), 2301-2323. doi: 10.1016/j.jpubeco.2004.11.002
- Somerville, L. H., Jones, R. M., Ruberry, E. J., Dyke, J. P., Glover, G., & Casey, B. J. (2013). The medial prefrontal cortex and the emergence of self-conscious emotion in adolescence. *Psychological science*, 24(8), 1554-1562.
- Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A., Paliwal, S., Gard, T., . . . Petzschner, F. H. (2016). Allostatic Self-efficacy: A Metacognitive Theory of Dyshomeostasis-Induced Fatigue and Depression. *Front Hum Neurosci*, 10, 550. doi: 10.3389/fnhum.2016.00550
- Steriade, M. (1999). Coherent oscillations and short-term plasticity in corticothalamic networks. *Trends Neurosci*, 22(8), 337-345.
- Steriade, M., Deschenes, M., Domich, L., & Mulle, C. (1985). Abolition of spindle oscillations in thalamic neurons disconnected from nucleus reticularis thalami. *J Neurophysiol*, 54(6), 1473-1497.
- Sterling, P. (2014). Homeostasis vs allostasis: implications for brain function and mental disorders. *JAMA Psychiatry*, 71(10), 1192-1193. doi: 10.1001/jamapsychiatry.2014.1043
- Sterling, P., & Eyer, J. (1988). Allostasis: a new paradigm to explain arousal pathology. In S. Fisher & J. Reason (Eds.), *Handbook of Life Stress, Cognition, and Health* (pp. 629–649). New York: John Wiley and Sons.
- Sturm, V. E., Ascher, E. A., Miller, B. L., & Levenson, R. W. (2008). Diminished self-conscious emotional responding in frontotemporal lobar degeneration patients. *Emotion*, 8(6), 861.
- Sturm, V. E., Rosen, H. J., Allison, S., Miller, B. L., & Levenson, R. W. (2006). Self-conscious emotion deficits in frontotemporal lobar degeneration. *Brain*, 129(9), 2508-2516.
- Sturm, V. E., Sollberger, M., Seeley, W. W., Rankin, K. P., Ascher, E. A., Rosen, H. J., . . . Levenson, R. W. (2012). Role of right pregenual anterior cingulate cortex in self-conscious emotional reactivity. *Social cognitive and affective neuroscience*, 8(4), 468-474.
- Sul, S., Tobler, P. N., Hein, G., Leiberg, S., Jung, D., Fehr, E., & Kim, H. (2015). Spatial gradient in value representation along the medial prefrontal cortex reflects individual differences in prosociality. *Proceedings of the National*

- Academy of Sciences of the United States of America, 112(25), 7851-7856. doi: 10.1073/pnas.1423895112
- Suzuki, S., Harasawa, N., Ueno, K., Gardner, J. L., Ichinohe, N., Haruno, M., . . . Nakahara, H. (2012). Learning to simulate others' decisions. *Neuron*, 74(6), 1125-1137. doi: 10.1016/j.neuron.2012.04.030
- Takahashi, H., Yahata, N., Koeda, M., Matsuda, T., Asai, K., & Okubo, Y. (2004). Brain activation associated with evaluative processes of guilt and embarrassment: an fMRI study. *Neuroimage*, 23(3), 967-974.
- Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annu Rev Psychol*, 58, 345-372. doi: 10.1146/annurev.psych.56.091103.070145
- Thayer, J. F., Ahs, F., Fredrikson, M., Sollers, J. J., & Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health. *Neuroscience and Biobehavioral Reviews*, 36(2), 747-756. doi: 10.1016/j.neubiorev.2011.11.009
- Touroutoglou, A., Andreano, J., Dickerson, B. C., & Barrett, L. F. (2020). The tenacious brain: How the anterior mid-cingulate contributes to achieving goals. *Cortex*, 123, 12-29. doi: 10.1016/j.cortex.2019.09.011
- Tracy, J. L., & Robins, R. W. (2004). TARGET ARTICLE: "Putting the Self Into Self-Conscious Emotions: A Theoretical Model". *Psychological Inquiry*, 15(2), 103-125. doi: 10.1207/s15327965pli1502_01
- Tricomi, E., Rangel, A., Camerer, C. F., & O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature*, 463(7284), 1089-1091. doi: 10.1038/nature08785
- Tusche, A., Bockler, A., Kanske, P., Trautwein, F. M., & Singer, T. (2016). Decoding the Charitable Brain: Empathy, Perspective Taking, and Attention Shifts Differentially Predict Altruistic Giving. *J Neurosci*, 36(17), 4719-4732. doi: 10.1523/JNEUROSCI.3392-15.2016
- Uddin, L. Q., Kelly, A. M., Biswal, B. B., Castellanos, F. X., & Milham, M. P. (2009). Functional connectivity of default mode network components: correlation, anticorrelation, and causality. *Hum Brain Mapp*, 30(2), 625-637. doi: 10.1002/hbm.20531
- Valentin, V. V., Dickinson, A., & O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. *J Neurosci*, 27(15), 4019-4026. doi: 10.1523/JNEUROSCI.0564-07.2007
- Van Hoorn, J., Van Dijk, E., Güroğlu, B., & Crone, E. A. (2016). Neural correlates of prosocial peer influence on public goods game donations during adolescence. *Social cognitive and affective neuroscience*, 11(6), 923-933.
- Vogt, B. A. (2005). Pain and emotion interactions in subregions of the cingulate gyrus. *Nat Rev Neurosci*, 6(7), 533-544. doi: 10.1038/nrn1704
- Vogt, B. A., Hof, P. R., Zilles, K., Vogt, L. J., Herold, C., & Palomero-Gallagher, N. (2013). Cingulate area 32 homologues in mouse, rat, macaque and human: cytoarchitecture and receptor architecture. *J Comp Neurol*, 521(18), 4189-4204. doi: 10.1002/cne.23409
- Wager, T. D., Waugh, C. E., Lindquist, M., Noll, D. C., Fredrickson, B. L., & Taylor, S. F. (2009). Brain mediators of cardiovascular responses to social

- threat: part I: Reciprocal dorsal and ventral sub-regions of the medial prefrontal cortex and heart-rate reactivity. *Neuroimage*, 47(3), 821-835. doi: 10.1016/j.neuroimage.2009.05.043
- Wagner, Ullrich, N'diaye, Karim, Ethofer, Thomas, & Vuilleumier, Patrik. (2011). Guilt-specific processing in the prefrontal cortex. *Cerebral cortex*, 21(11), 2461-2470.
- Wallis, C. U., Cardinal, R. N., Alexander, L., Roberts, A. C., & Clarke, H. F. (2017). Opposing roles of primate areas 25 and 32 and their putative rodent homologs in the regulation of negative emotion. *Proc Natl Acad Sci U S A*, 114(20), E4075-E4084. doi: 10.1073/pnas.1620115114
- Waytz, A., Zaki, J., & Mitchell, J. P. (2012). Response of dorsomedial prefrontal cortex predicts altruistic behavior. *J Neurosci*, 32(22), 7646-7650. doi: 10.1523/JNEUROSCI.6193-11.2012
- Will, G. J., Rutledge, R. B., Moutoussis, M., & Dolan, R. J. (2017). Neural and computational processes underlying dynamic changes in self-esteem. *Elife*, 6. doi: 10.7554/eLife.28098
- Wittmann, M. K., Kolling, N., Faber, N. S., Scholl, J., Nelissen, N., & Rushworth, M. F. (2016). Self-Other Mergence in the Frontal Cortex during Cooperation and Competition. *Neuron*, 91(2), 482-493. doi: 10.1016/j.neuron.2016.06.022
- Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nat Rev Neurosci*, 7(6), 464-476. doi: 10.1038/nrn1919
- Yoon, L., Somerville, L. H., & Kim, H. (2018). Development of MPFC function mediates shifts in self-protective behavior provoked by social feedback. *Nat Commun*, 9(1), 3086. doi: 10.1038/s41467-018-05553-2
- Yu, X. J., Xu, X. X., He, S., & He, J. (2009). Change detection by thalamic reticular neurons. *Nat Neurosci*, 12(9), 1165-1170. doi: 10.1038/nn.2373
- Zahavi, A. (1975). Mate selection-a selection for a handicap. *J Theor Biol*, 53(1), 205-214.
- Zahn, R., Moll, J., Paiva, M., Garrido, G., Krueger, F., Huey, E. D., & Grafman, J. (2008). The neural basis of human social values: evidence from functional MRI. *Cerebral cortex*, 19(2), 276-283.
- Zaki, J., & Mitchell, J. P. (2011). Equitable decision making is associated with neural markers of intrinsic value. *Proc Natl Acad Sci U S A*, 108(49), 19761-19766. doi: 10.1073/pnas.1112324108
- Zevin, J. D., & Seidenberg, M. S. (2004). Age-of-acquisition effects in reading aloud: tests of cumulative frequency and frequency trajectory. *Mem Cognit*, 32(1), 31-38.
- Ziegler, G., Dahnke, R., Yeragani, V. K., & Bar, K. J. (2009). The relation of ventromedial prefrontal cortex activity and heart rate fluctuations at rest. *Eur J Neurosci*, 30(11), 2205-2210. doi: 10.1111/j.1460-9568.2009.07008.x
- Zikopoulos, B., & Barbas, H. (2006). Prefrontal projections to the thalamic reticular nucleus form a unique circuit for attentional mechanisms. *J Neurosci*, 26(28), 7348-7361. doi: 10.1523/JNEUROSCI.5511-05.2006
- Zikopoulos, B., & Barbas, H. (2007). Circuits formultisensory integration and attentional modulation through the prefrontal cortex and the thalamic reticular nucleus in primates. *Rev Neurosci*, 18(6), 417-438.

- Zikopoulos, B., & Barbas, H. (2012). Pathways for emotions and attention converge on the thalamic reticular nucleus in primates. *J Neurosci*, 32(15), 5338-5350. doi: 10.1523/JNEUROSCI.4793-11.2012
- Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *The educational psychology series. Handbook of metacognition in education* (pp. 299-315). New York, NY, US: Routledge/Taylor & Francis Group.