INDIVIDUALIZED PREDICTION OF THIRD-PARTY PUNISHMENT BEHAVIOR FROM INTRINSIC FUNCTIONAL BRAIN CONNECTIVITY

by

Ko-Tsung H	lsu
A Thesis	
Submitted to	the
Graduate Fact	ulty
of	5
George Mason Ur	niversity
in Partial Fulfillr	nent of
The Requirements for	• the Degree
of	
Master of Scie	ence
Bioinformatics and Comp	Itational Biology
Biomiormanes and Compe	
Committee:	
	Dr. Frank Krueger, Thesis Director
	Dr. William Kannady, Committee
	Member
	Wentber
	Dr. Nadine Kabbani, Committee
	Member
	Dr Josif Vaisman Director
	School of Systems Biology
	Dr. Donna Fox, Associate Dean,
	Office of Student Affairs & Special
	Programs, College of Science
	Dr. Deggy Agouris Deen College of
	DI. Peggy Agouris, Dean, Conege of
	Science
Date:	Spring Semester 2019
	George Mason University
	Fairfax, VA

Individualized Prediction of Third-Party Punishment Behavior from Intrinsic Functional Brain Connectivity

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at George Mason University

by

Ko-Tsung Hsu Bachelor of Science Ming Chuan University, 2015

Director: Frank Krueger, Associate Professor School of Systems Biology

> Spring Semester 2019 George Mason University Fairfax, VA

Copyright 2019 Ko-Tsung Hsu All Rights Reserved

ACKNOWLEDGMENTS

I sincerely acknowledge all the people who participated in this experiment and all those responsible for collecting the data. I am responsible for performing the statistical analyses and writing the thesis. Finally, I want to especially acknowledge Dr. Krueger for directing this thesis.

TABLE OF CONTENTS

List of Tables	v
List of Figures	vi
List of Abbreviations and/or Symbols	vii
Abstract	. viii
1. Theoretical background	1
1.1 Introduction	1
1.2 Economic Games as Instruments to Measure Third-Party Punishment	2
1.3 Neuropsychological Model of Third-Party Punishment	3
1.4 Resting-State Functional Magnetic Resonance Imaging	6
1.5 Prediction-Analytics Framework	10
1.6 Study Goal and Hypotheses	11
2. Methods	12
2.1 Participants	12
2.2 Paradigm	13
2.3 Procedure	15
2.4 Behavioral Data Analyses	16
2.5 MRI Data Acquisition	16
2.6 Neuroimaging Preprocessing	17
2.7 Resting-State Functional Connectivity Analyses	18
2.8 Multivariate Regression Analyses	20
2.9 Network Ranking Procedure for Determining the Best Prediction Model	21
3. Results	22
3.1 Results for Behavioral Analyses	22
3.2 Results for Prediction Analyses	23
4. Discussion	27
5. Appendix	31
6. References	35

LIST OF TABLES

Table	Page
Table 1. P-value Matrix for Ranking the Models Performance	

LIST OF FIGURES

Figure	Page
Figure 1. Illustration of the Third-Party Punishment Game.	3
Figure 2. Neuropsychological Model of Third-Party Punishment	4
Figure 3. Illustration of the Seed Method Identifying Resting-State Functional Netwo	rk. 8
Figure 4. Resting Brain Networks Identified Through Independent Component Analy	sis.9
Figure 5. Setup of Dictator Game, Second-Party, and Third-Party Punishment Game	13
Figure 6. Third-Party Punishment based on Fairness of Offer.	15
Figure 7. Five Resting-State Brain Networks Defined by the Dosenbach's Atlas	19
Figure 8. Behavioral Results for Third-Party and Second-Party Punishment.	23
Figure 9. Results for Prediction Performances for Third-Party Punishment Behavior	24
Figure 10. Resting-State Functional Connectivity of Best Model Predicting Third-Par	ty
Punishment Behavior	26

LIST OF ABBREVIATIONS

Analysis Of Variance	ANOVA
Blood Oxygen Level Dependent	BOLD
Central-Executive Network	CEN
Cingulo-Opercular Network	CON
Dictator Game	DG
Dorsomedial Prefrontal Cortex	dmPFC
Frontoparietal Network	FPN
Functional Connectivity Toolbox	CONN
Functional Connectivity	FC
Functional Magnetic Resonance Imaging	fMRI
Independent Component Analysis	ICA
Leave-One-Subject-Out Cross-Validation	LOSOCV
Medial Prefrontal Cortex	mPFC
Monetary Units	MU
Multivariate Prediction Analysis	MVPA
Occipital Network	OccN
Principal Component Analysis	PCA
Region Of Interest	ROI
Resting-State Functional Magnetic Resonance Imaging	RS-fMRI
Second-Party Punishment	SPP
Sensorimotor Network	SMN
Standardized Mean Squared Error	SMSE
Statistical Parametric Mapping Toolbox	SPM
Support Vector Machine	SVM
Theory Of Mind	ТоМ
Third-Party Punishment	TPP
Ultimatum Game	UG
Ventromedial Prefrontal Cortex	vmPFC

ABSTRACT

INDIVIDUALIZED PREDICTION OF THIRD-PARTY PUNISHMENT BEHAVIOR FROM INTRINSIC FUNCTIONAL BRAIN CONNECTIVITY

Ko-Tsung Hsu, M.S.

George Mason University, 2019

Thesis Director: Dr. Frank Krueger

A robust human society is developed normally on the ground of social cooperation, serving a critical role in human relationships. Importantly, social cooperation is subjected to the establishment of social norms. To maintain human society, third-party punishment (TPP) as a consistently sanctioning behavior facilitates the enforcement of social norms. At the psychological level, TPP is based on blame which is an amalgam of intent and harm to victim and the offender's intention in violating social norms. At the neural level, TPP behavior builds on the interaction of the salience network (determining the degree of norm violation), default-mode network (determining the degree of blame), and central-executive network (determining the degree of punishment). Although task-based functional magnetic resonance imaging (fMRI) has been extensively used to investigate individual differences in the propensity to punish, whether TPP behavior can be predicted through task-free fMRI based on resting-state functional connectivity (RSFC) remains open. The goal of this study was to apply multivariate prediction analysis (MVPA) to RSFC patterns of large-scale brain networks to predict individual difference in TPP behavior measured with a TPP exchange game. The findings demonstrated that RSFC between the default-mode network (DMN) and the central-executive network (CEN) predicted TPP behavior, indicating a signal transmission from blame (DMN) to punishment behavior (CEN). In conclusion, investigating the individual difference in TPP behavior based on RSFC provides us with a new comprehensive understanding of sustaining cooperation and enforcement of social norms in human society.

1. THEORETICAL BACKGROUND

<u>1.1 Introduction</u>

Cooperation is an essential component to function in normal human society (Fehr & Fischbacher, 2004a). Therefore, facilitating widespread cooperation in human society becomes an issue worth discussing in the field of psychology. So far, many studies have engaged in this intriguing field and shed light on principle factors sustaining cooperation in human society (Fehr, Fischbacher, & Gächter, 2002). Social punishment and reciprocity are factors involved in functioning normality of human society; for instance, both factors facilitate cooperation between human beings. Within established social norms, social cooperation can be promoted under legal enforcement of rules (Fehr & Fischbacher, 2004b; Zhong, Chark, Hsu, & Chew, 2016). To sustain social cooperation in human society, punishment based on normative legitimacy of the rules against norm violators is an approach to enforcing social norms. Costly punishment refers to the use of one own's resources to enforce social norms without obtaining any obvious benefits (Fehr & Fischbacher, 2004a). Norm violators (i.e., offenders) are not only punished by the victims themselves (defined as costly second-party punishment, SPP), but also by witnesses (defined as the third-party punishment, TPP) who are not victims of the norm violation (Bendor & Swistak, 2001; Sober, Wilson, & Wilson, 1999). On the one hand, in SPP, second parties who experience unfair treatment or norm violation experience negative emotions (e.g., anger, disgust, aversion, and frustration) may lead to punishment (Pillutla & Murnighan, 1996). On the other hand, third parties who are not affected economically, physically, or psychologically by the norm violators are also willing to punish when they witness a norm violation (Fehr & Fischbacher, 2004b; Henrich, 2006). Therefore, third-party punishment (TPP) enhances the scope of investigation into the motivation of norm violators compared with the SPP, which mostly emphasizes on decision making through emotions (Baumgartner, Götte, Gügler, & Fehr, 2012).

1.2 Economic Games as Instruments to Measure Third-Party Punishment

A number of different experimental paradigms —hypothetical crime scenarios (Glass, Moody, Grafman, & Krueger, 2016) and norm-enforcement economic games (Fehr & Camerer, 2007)— are used to establish neuropsychological factors of TPP. As for hypothetical crime scenarios, participants are given examples of crime narratives and asked to make a punishment decision in accordance to the blameworthiness of these crime scenarios.

As for norm-enforcement economic games, a modified version of the ultimatum game (UG) has been utilized to explore TPP behavior (Civai, Corradi-Dell'Acqua, Gamer, & Rumiati, 2010). In this game, third parties observe offers of money between proposers and receivers based on an endowment and decide whether to punish the proposers based on the fairness of offer by spending a portion of their endowment (i.e., costly punishment) (Brown, 1991). By definition, this scenario is designated as the TPP, a selective extension of the SPP where the second party retaliating against the proposers as unfair or norm violation occurred (Brown, 1991; Hoffman, 2014; Krueger & Hoffman, 2016) (**Figure 1**).



Figure 1. Illustration of the Third-Party Punishment Game.

Third-party punishment game includes three roles. Proposers share offers with receivers, receivers have to accept the proposers' offers, and third parties witnessing the proposers' offers decide whether to punish the proposers.

1.3 Neuropsychological Model of Third-Party Punishment

A heuristic model of TPP states that the magnitude of TPP behavior is driven by blame, which is based on the proposer's intent and the amount of harm inflicted onto the victim (Krueger & Hoffman, 2016) (**Figure 2A**).



Figure 2. Neuropsychological Model of Third-Party Punishment.

(A) Neuropsychological factors of the third-party punishment including harm (shown in red), intent and blame (shown in blue), and punishment (shown in green). (B) Three neural networks in the brain are responsible for driving punishment behavior, including the salience network (SN) (red), default-mode network (DMN) (blue), and central-executive network (CEN) (green). Abbreviations: AI, anterior insula; dACC, dorsal anterior cingulate cortex; Amyg, amygdala. mPFC, medial prefrontal cortex; dmPFC, dorsomedial prefrontal cortex; vmPFC, ventromedial prefrontal cortex; PCC, posterior cingulate cortex; TPJ, temporoparietal junction. dlPFC, dorsolateral prefrontal cortex; PPC, posterior parietal cortex. The figure is taken from Krueger and Hoffman (Krueger & Hoffman, 2016).

Three large-scale brain networks are associated with TPP behavior (Krueger & Hoffman, 2016) (**Figure 2B**). The salience network (SN) —comprised of anterior insula (AI), dorsal anterior cingulate cortex (dACC), and amygdala (Amyg)— is involved in processing of aversive sensations. The SN begins with detecting the occurrence of a norm violation (dACC) and generating an aversive response (AI). Subsequently, Amyg determines the extent of aversive emotion based on the severity of harm that is caused to the receivers.

Afterward, SN modulates the engagement of a second network, the default mode network (DMN), generally involved in processing autobiographical memory, self-monitoring, and mentalizing (Bressler & Menon, 2010). The medial prefrontal cortex (mPFC) is a core brain region of the DMN network, integrating emotional processes related to harm to the receivers and is also in charge of processing social cognition related to the intent of proposers. The mPFC integrates two pathways into a reasonable assessment of blame (Krueger & Hoffman, 2016). Specifically, the harm-integrating portion is apparently processed through the ventromedial prefrontal cortex (vmPFC), which has functional connectivity with the SN (Krueger & Hoffman, 2016). The intention-integrating portion is processed by the dorsomedial prefrontal cortex (dmPFC) which has functional connectivity with the regions of posterior cingulate cortex (PCC; in charge of self-referential processes), and temporoparietal junction (TPJ; in charge of inferring proposers' intentions, beliefs, or desires in others) (Treadway et al., 2014).

Finally, the central-executive network (CEN) is involved in processing the information of context-dependent assessments for higher-order cognition and decision making (Bressler & Menon, 2010). Punishment decisions result from converting a blame signal through the dorsolateral prefrontal cortex (dlPFC) in the CEN (Bellucci et al., 2017). In addition, before making a punishment decision, the posterior parietal cortex (PPC) constructs a scale of punishment for the dlPFC to select an appropriate punishment decision (Buckholtz et al., 2008).

TPP behavior heavily relies on DMN because impartial witnesses are not directly affected by the social norm violation (i.e., unfair offer) and, therefore, tend to focus on the proposer's intention instead of the outcome related to the receiver (i.e., victim). Hence, DMN is served as a major neural component when making a TPP punishment decision, engaging the CEN (Buckholtz et al., 2008). Simply put, TPP behavior is mostly driven by the signal from DMN rather than SN, implying that punishment decisions are driven by brain regions of DMN related to inferring proposer's intentions.

1.4 Resting-State Functional Magnetic Resonance Imaging

The blood oxygen level dependent (BOLD) signal of functional magnetic resonance imaging (fMRI) is a representative signal that indicates the extent of neural activity (Ogawa, Lee, Kay, & Tank, 1990). The amount of oxygen in the blood is associated with relative neural activity; therefore, changes in the ratio of oxygenated hemoglobin/de-oxygenated hemoglobin can lead to fluctuation of the BOLD signal. Oxygenated hemoglobin has diamagnetic property and does not distort the surrounding magnetic field, whereas deoxygenated hemoglobin has paramagnetic property and distorts the surrounding magnetic field during fMRI (Pauling & Coryell, 1936). Because of the different magnetic susceptibility of oxygenated and de-oxygenated hemoglobin, the BOLD signal either decreases or increases depending on the hemodynamic response in a given brain region.

Brain functions can be quantified via *task-based fMRI* studies that correlate experimental conditions or behavioral performance with neural activation, as indexed by changes in BOLD signals (Ogawa, Lee, Kay, & Tank, 1990). Task-based fMRI, combined with a TPP paradigm, has been previously used in a number of studies (Bellucci et al., 2017;

Buckholtz et al., 2008; David, Hu, Krüger, & Weber, 2017; Ginther et al., 2016; Sun, Tan, Cheng, Chen, & Qu, 2015).

RS-fMRI (*i.e., intrinsic functional connectivity*) provides an appealing alternative to characterize neurodiversity of TPP behavior by scanning individuals while they lie in the scanner with no specific instructions except to close their eyes, relax, and hold still (van den Heuvel & Hulshoff Pol, 2010). RS measures can be acquired quickly (5-10 min), consistently, and reliably, yielding large sample sizes and good compliance in adolescents, enabling developmental studies using a single imaging dataset.

RS-fMRI observed at rest refers to the functional integration of brain areas that is represented by coherent low-frequency (0.01–0.1 Hz) BOLD fluctuations in distant grey matter regions (Biswal, Yetkin, Haughton, & Hyde, 1995). Using a variety of MR scanners (e.g., vendors, field strengths) and analysis techniques (e.g., seed methods, independent component analysis, clustering), RS studies have consistently reported that anatomically separated, but functionally linked brain regions show a high level of ongoing FC during rest. Such networks are often called *RS networks* (Fox et al., 2005).

Several functionally linked sub-networks have been identified (Beckmann, DeLuca, Devlin, & Smith, 2005; Damoiseaux et al., 2006), including SAN, DMN, CEN, sensorimotor network (SMN), and occipital network (OccN). Those networks are *stable across time* and strikingly similar to the networks activated by a broad spectrum of task-based studies (Kelly, Biswal, Craddock, Castellanos, & Milham, 2012).

Two different methods exist to measure RS-fMRI, including model-dependent methods (e.g., seed method) (D. Cordes et al., 2000; Fransson, 2005) and model-free

methods (e.g., independent component analysis; ICA) (Beckmann et al., 2005; Calhoun, Adali, Pearlson, & Pekar, 2001). For the seed method as a model-dependent method, the RS-fMRI time-series of a selected region of interest (ROI) is correlated with the resting-state time-series of other brain regions (Jiang, He, Zang, & Weng, 2004) (**Figure 3**).





(A) To determine the functional connectivity between the seed voxel and randomly selected voxel, the BOLD signals of each voxel were recorded during the resting-state functional scan. (B) The resting-state time-series of the seed voxel is correlated with the resting-state time-series of the randomly selected voxel. If the correlation between two voxels' BOLD signal is relatively high, then these two voxels were recognized as having a high level of functional connectivity and vice versa. (C) Functional connectivity maps can be constructed by correlating all other voxels in the brain with the seed voxel, indicating that survival voxels are highly correlated and having a high level of functional connectivity.

Model-free methods to identify intrinsic FC include principal component analysis (PCA) (Friston, 1998), ICA (Beckmann et al., 2005; Calhoun et al., 2001), hierarchical clustering (Dietmar Cordes, Haughton, Carew, Arfanakis, & Maravilla, 2002; Wang & Li,

2013), and Laplacian clustering (Thirion et al., 2006). Among these approaches, ICA is the most commonly used method (Damoiseaux et al., 2006). For instance, ICA has been used to identify RSFC networks, including SAN (Seeley et al., 2007), DMN (Supekar et al., 2010), and CEN (Sridharan, Levitin, & Menon, 2008) (**Figure 4**).



Figure 4. Resting Brain Networks Identified Through Independent Component Analysis.

The central-executive network responsible for determining the degree of punishment is shown in blue. The salience network responsible for determining the degree of norm violation is shown in yellow. The default-mode network responsible for determining the degree of blame is shown in red. The figure is taken from Menon (Menon, 2011).

RS activity predicts both task-evoked activity and behavior (van den Heuvel & Hulshoff Pol, 2010). Support for a neural basis of RS-fMRI comes from electrophysiological studies showing a strong association between spontaneous BOLD

fluctuations and fluctuations in neuronal spiking (Sadaghiani et al., 2010). Further, concern about subjects' resting cognitions (Shirer, Ryali, Rykhlevskaia, Menon, & Greicius, 2012), mood (Harrison et al., 2008), instruction (maintain eyes open or closed (Van Dijk et al., 2010), and other factors (substance withdrawal (Cole et al., 2010), drowsiness (Sämann et al., 2011), and sleep (Horovitz et al., 2009) have been addressed by the moderate-to-high test-retest reliability of intrinsic activity and intrinsic FC. Therefore, FC patterns are a *powerful predictor* of a person's age (Dosenbach et al., 2010), memory (Yamashita, Kawato, & Imamizu, 2015), and even personality traits (Yamashita et al., 2015).

1.5 Prediction-Analytics Framework

In a *prediction-analytics framework*, a model relating measures (whole-brain RSFC patterns) to phenotypic variables is learned from a training dataset via multivariate prediction analysis (MVPA) (i.e., multivariate classification and regression methods) (Kelly et al., 2012). The *predictive model* is then applied to an independent test dataset to predict phenotypes. The resulting predictions are compared to the true phenotypes to estimate how well the model generalizes to the test dataset. Meaningful information can be extracted from the learned model by using feature selection methods to reduce the input variables to only those essential for prediction. MVPA has been applied to fMRI and RSFC data. Complementary to inferential statistics that measure the likelihood of such relationships arising by chance, MVPA evaluates the significance of an entire RSFC pattern using a single statistic without correcting for multiple comparisons and measures how well a model generalizes to independent data (*out-of-sample generalization*)

(Richiardi, Achard, Bunke, & Ville, 2013). MVPA has identified RSFC neuromarkers for disorders such as Alzheimer's disease (Dai et al., 2012), schizophrenia (Shen, Wang, Liu, & Hu, 2010), and autism (Anderson et al., 2011) and also for age (Dosenbach et al., 2010), personality traits (Hahn et al., 2015), and prosocial behavior. However, this framework has not been applied to predict individual variation in TPP behavior.

1.6 Study Goal and Hypotheses

Whether RS-fMRI can be utilized to predict individual difference in TPP behavior based on RSFC is still an open question. Therefore, the goal of this study was to apply MVPA to RSFC patterns to predict individual difference in TPP behavior measured with a TPP exchange game. On the behavioral level, we hypothesized that TPP behavior increases with more unfair offers and decreases in comparision to SPP, since TPP is based on blame, which is more driven by the intent than outcome in relation to an unfair offer. On the neural level, we hypothesized that individual differences in average TPP behavior can be predicted by the RSFC between DMN and CEN, because changes in blame (DMN) determine the adjustment of punishment (CEN).

2. METHODS

2.1 Participants

Fifty-two healthy volunteers (28 females, 24 males, *Mean* (M) = 23.5, *Standard Deviation* (SD) = 3.20) participated for financial compensation. All participants were right-handed, had normal or corrected-to-normal vision, and were enrolled either at the University of Mannheim or at the University of Heidelberg in Germany. Participants were excluded from this study, if they met one of the following conditions: having any metal parts in the body like prostheses, nails, valvular transplant, pacemaker or implants, record of operation on head or heart, artificial lenses, therapeutic patches (e.g., nicotine, hormones), non-removable piercing, large area tattoos or permanent make-up – grounded on a preliminary brief psychological survey. Further, only German speaking participants, aged between 19 and 31 years, with no history of epilepsy or agoraphobia as well as no pregnant women were allowed to participate in this study.

The study, approved by the local ethics committee, was conducted in accordance with the ethical guidelines and principles of the Declaration of Helsinki. After being informed about the procedure of the study, written consent was obtained from participants. They were told that their participation was voluntary and that they could dropout at any time point during the study. Participants received 25€ as financial compensation in addition to the monetary units (MU) earned from the economic games that were converted into fiat money at the end of the study.

2.2 Paradigm

In this study, each participant was required to complete three economic games, including the dictator game (DG) (Kahneman, Knetsch, & Thaler, 1986), the TPP game (Fehr & Fischbacher, 2004b), and the SPP game (Forsythe, Horowitz, Savin, & Sefton, 1994) (**Figure 5**).



Figure 5. Setup of Dictator Game, Second-Party, and Third-Party Punishment Game.

(A) In the dictator game, proposers share an offer that must be accepted by receivers. (B) In the third-party punishment game, third parties —observing proposers as witnesses sharing an offer that must be accepted by receivers— decide whether to punish proposers based on unfairness of the offer. (C) In the second party punishment game, proposers share an offer with responders and responders decide whether to punish proposers based on the unfairness of the offer.

The DG game was conducted as a control game, where participants played the role of proposers (i.e., dictator) and were asked to share their MUs (maximum: 12 MUs) with receivers who had to accept their offers (**Figure 5A**). Afterward, participants completed two costly punishment games in a counter-balanced manner to examine TPP and SPP.

For the TPP game, participants were assigned to the role of a third party, who witnessed a series of seven random offers by proposers ranging from fair to unfair (6:6; 7:5; 8:4; 9:3; 10:2; 11:1; 12:0) to receivers (as describes in DG game) (**Figure 5B**). At the beginning of the game, the proposer and third party were assigned equally with 6 MUs. Proposers received additional 12 MUs that they could share with receivers. As impartial third parties, participants could either keep all MUs or punish proposers based on the witnessed offers. The MUs used for punishment by the third parties were tripled in value to reduce the proposers' payoff.

For a fair offer, the prosper, for example, keeps 6 MUs and sends 6 MUs to the receiver. The third party would not punish the proposer, therefore, keeping the original 6 MUs (**Figure 6A**). For an unfair offer, the proposer, for example, keeps 12 MUs (having now a total of 18 MUs) and sends 0 MUs to the receiver (**Figure 6B**). The third party would punish the proposer using the maximum of 6 MUs (having now a total of 0 MUs). The MUs used for punishment will be tripled in value to reduce proposer's payoff; therefore, the maximum deduction on the proposer would be 18 MUs (having now a total of 0 MUs). Consequently, the MUs owned by each player would be 0 MU.



Figure 6. Third-Party Punishment based on Fairness of Offer.

(A) Fair offer: Third party decides not to punish the proposer based on the observation of a fair offer (6:6) shared with the receiver.(B) Unfair offer: Third party decides to punish the proposer based on the observation of an unfair offer (12:0) shared with the receiver.

Finally, participants in the SPP game acted as responders and were asked to decide whether to punish the proposers based on seven randomly received offers ranging from fair to unfair (6:6; 7:5; 8:4; 9:3; 10:2; 11:1; 12:0) (**Figure 5C**). If responders decided to punish proposers, each MUs that they spent were then tripled in value and used to reduce the proposers' payoff.

2.3 Procedure

For the behavioral part of the study, participants were asked to complete the economic games (DG, TPP, SPP) as well as surveys with demographic questions (e.g., gender, education) using the Qualtrics online platform (<u>https://www.qualtrics.com</u>).

For the neuroimaging part of the study, they completed a structural MRI and an RSfMRI scan, each lasting about 10 minutes. During the resting scan, participants were instructed to keep their eyes on the cross-hair displayed on a screen in front of them and not fall asleep or to think systematically about a specific topic.

2.4 Behavioral Data Analyses

The behavioral data analyses were carried out using SPSS 24.0 (IBM Corp, 2016) with alpha set to p < 0.05 (two-tailed). First, the means and standard errors for the average punishment as well as for each of the seven offers were calculated. Second, to identify the effects of variations in fairness of offers and type of punishment, a repeated measure 7 x 2 analysis of variance (ANOVA) on punishment behavior was calculated with Offer (6:6, 7:5, 8:4, 9:3, 10:2, 11:1, 12:0) and Type (TPP vs. SPP) as within-subjects factors. Third, a one-way ANOVA on TPP behavior with Offer (6:6, 7:5, 8:4, 9:3, 10:2, 11:1, 12:0) as a within-subjects factor was performed to determine the slope in punishment from most fair offer (i.e., 6:6) to most unfair offer (i.e., 12:0). Finally, a paired-samples t-test was computed to compare the average punishment behaviors between TPP and SPP.

2.5 MRI Data Acquisition

Brain images were collected using a Siemens Trio-3T MRI scanner (Siemens Medical Systems, Erlangen, Germany) with a 32-channel head coil. For each participant, a high-resolution anatomical scan of the entire brain was collected using a T1-weighted 3D magnetization prepared rapid acquisition with gradient echo (MP-RAGE) sequence: time of repetition (TR), 2300 ms; TE, 3.03 ms; flip angle, 9°; number of slices, 192; field of view (FOV), 256 mm; matrix size, 256 x 256; voxel size, 1 x 1 x 1 mm. The blood oxygen

level-dependent (BOLD) signal for functional images was measured using a T2-weighted gradient EPI with the following parameters: TR, 2000 ms; TE, 30ms; flip angle, 80° ; thickness, 3mm; number of slices, 36; FOV, 192 mm; matrix size, 64 x 64 mm; voxel size, 3 x 3 x 3 mm. The first five scans of the EPI were discarded to minimize T1 effects.

2.6 Neuroimaging Preprocessing

The MRI data analysis was performed using the Statistical Parametric Mapping Toolbox (SPM 12, Wellcome Trust Centre for Neuroimaging, London, UK; http://www.fil.ion.ucl.ac.uk/spm/software/spm12/) running under Matlab R2018a (Mathworks Inc., Natick, MA, USA). Pre-processing of the EPI images included the following steps. To allow signal equilibrium, the first ten volumes of functional images were discarded. Afterward, the functional images were bias-corrected for field inhomogeneity. Subsequently, the slice-timing method was performed on functional images, using interpolation to correct the time point of each functional image. Next, all functional images were aligned to the mean image for movement correction to reduce the artifact during the imaging. Functional images were then co-registered to their anatomical images, and both anatomical images and functional images were spatially normalized to MNI template through deformation fields derived from anatomical segmentation (resampling voxel size was $2 \times 2 \times 2$ mm³). Lastly, functional images were smoothed with isotropic Gaussian kernel of 4 x 4 x 4 mm³ full width at half maximum (FWHM) to increase the ratio of signal-to-noise.

Artifact of functional images was detected and excluded using the ART toolbox (https://www.nitrc.org/projects/artifact_detect/) based on the following criteria. First, head displacement in x, y, or z-direction greater than 2 mm from the previous frames was rejected. Second, rotational displacement greater than 0.02 radians from the previous frame was rejected. Third, compared with the intensity of the mean image, global mean intensity in the functional images greater than 3 standard deviations was rejected. Subsequently, those outliers were included as nuisance regressors within the first-level general linear model. Ultimately, high-frequency noise and linear drift artifacts were removed by implementing the band-pass filter (0.01-0.1 Hz).

2.7 Resting-State Functional Connectivity Analyses

The RSFC was determined based on the Dosenbach's atlas, consisting of 142 (ROIs, nodes) (Dosenbach et al., 2010) (**Figure 7**). The distance among pairs of ROIs was at least greater than 10-mm. The ROIs (n=142) were subdivided into five pre-defined RSFC networks: cingulo-opercular network (CON) (overlapping with SN), sensorimotor network (SMN), default-mode network (DMN), frontoparietal network (FPN) (overlapping with CEN), and occipital network (OccN).



Figure 7. Five Resting-State Brain Networks Defined by the Dosenbach's Atlas.

Region of interest (n=142) taken from the Dosenbach's Atlas can be subdivided into five RSFC networks: sensorimotor network (yellow), occipital network (indigo), cingulo-opercular network (red), default mode network (green), and frontoparietal network (cyan). Abbreviation: RSFC, resting-state functional connectivity.

For each participant, the BOLD signal of all voxels residing in every ROI was averaged. Between each pair of ROIs, the network edge was defined as functional connectivity and FC between edges was calculated by the Pearson correlation coefficient using the Functional Connectivity (CONN) toolbox (https://www.nitrc.org/projects/conn). The Pearson's correlation coefficients were transformed into Fisher's z values to represent the degree of FC. For each participant, a correlation matrix with 10,082 unique connections was created, which were subsequently used in the multivariate regression analyses (i.e., prediction-analytics framework).

2.8 Multivariate Regression Analyses

For the multivariate regression analyses to predict variations in participants' average TPP behavior, a support vector machine (SVM) algorithm was implemented for training and testing the prediction models, using the sci-kit-learn toolbox (https://scikit-learn.org/stable/) in Python (https://www.python.org/). A total of 15 predictive models were computed, including five intra-network models (i.e., intra-network RSFC: CON-CON, SMN-SMN, DMN-DMN, FPN-FPN, and OccN-OccN) and ten inter-network models (i.e., inter-network RSFC: CON-SMN, CON-DMN, CON-FPN, CON-OccN, SMN-DMN, SMN-FPN, SMN-OccN, DMN-FPN, DMN-OccN, FPN-OccN).

A leave-one-subject-out cross-validation (LOSOCV) approach was implemented for each prediction model. For every iteration, the SVM algorithm was used to train the data on a training set, where one of the participants was left out. Because of the highdimensionality of the features (ROI-to-ROI FCs coefficients, n=10,082), a feature selection approach was applied, where five percent of the strongest correlations between the features and the targets (i.e., TPP behavior) were kept as the most relevant features. The RSFC matrix of the left-out participant served as a testing set for the prediction the TPP behavior. This iteration was repeated *n* times (n = total number of participants). Therefore, *n* models were generated, yielding each time a behavioral prediction for each participant.

The standardized mean squared error (SMSE) was used to assess the performance of a model, i.e., the error of the algorithm's performance divided by the targets' variance. A permutation (n=1000) approach was implemented to determine the significance of each model. For each permutation, every cross-validated model was run with randomly permuted targets. The number of SMSE scores lower than true targets was calculated (n_{perm}); and the p-value was computed by dividing this number by the total number of permutations (i.e., p = $(1 + n_{perm})/(1 + 1,000)$).

2.9 Network Ranking Procedure for Determining the Best Prediction Model

To compare the regression model performances between the 15 pre-defined networks (i.e., five intra-networks and ten inter-between networks), a 15*15 p-value matrix was constructed by applying paired-samples t-tests that compared the distributions between the original and permutation-based prediction errors for each network. A significant p-value indicated whether the prediction error for one network is significantly greater than that for another one. From the obtained p values can be inferred how significantly a given network predicts TPP behavior in comparison to all other networks. Therefore, a sorting of p-values allows a ranking of networks, in which a network that predicts the target behavior better than the maximum number of other network pairs will be ranked the highest.

3. RESULTS

3.1 Results for Behavioral Analyses

The means and standard errors of mean for the average punishments across offers as well as for each offer for each costly punishment type (TPP and SPP) are displayed in Figure 8. The two-way ANOVA revealed a main effect for the factors Offer (F(6,306)=4.71, p<0.035) and Type (F(1,51)=30.44, p<0.001), but not significant interaction effect for Offer x Type (F(6,306)=0.65, p=0.648). For the factor Offer, punishment increased linearly from the fairest offer (i.e., 6:6) to the most unfair offer (i.e. 12:0) (F(1,51)=48.89, p<0.001). For the factor Type, participant punished on average less in the TPP game compared to the SPP game (t(51)=-2.17, p<0.035).





Costly punishment behavior (mean \pm standard error mean) increased linearly from the fairest offer to the most unfair offer. Participant punished on average for third-party punishment (TPP) compared to second-party punishment (SPP) (*p<0.05).

3.2 Results for Prediction Analyses

Prediction Performance. A prediction framework (i.e. machine-learning algorithm) was applied to predict participants' TPP behavior based on RSFC of 15 networks (i.e., five intra-networks and ten inter-networks) using the Dosenbach's atlas (Dosenbach et al., 2010) (**Figure 9**).



Figure 9. Results for Prediction Performances for Third-Party Punishment Behavior.

The significance of performance (i.e., standard mean square error, SMSE, red dot line) for the 15 cross-validated prediction models (five intra-networks and ten inter-networks) based on the permutation approach are shown. *p < 0.05; **p < 0.01; ***p < 0.001. CON, cingulo-opercular network; SMN, sensorimotor network; DMN, default mode network; FPN, frontoparietal network; OccN, occipital network.

The performances of the 12 cross-validated network models were significantly better than chance for predicting the average TPP behavior: CON (SMSE=0.96, p<0.002), CON-DMN (SMSE=0.89, p<0.001), CON-FPN (SMSE=0.93 p<0.05), CON-OccN (SMSE=0.96, p<0.01), CON-SMN (SMSE=0.93, p<0.002), DMN (SMSE=0.88, p<0.005), DMN-FPN (SMSE=0.83, p<0.001), DMN-OccN (SMSE=0.97, p<0.05), DMN-SMN (SMSE=0.91, p<0.001), FPN (SMSE=0.86, p<0.001), FPN-OccN (SMSE=0.97, p<0.001), SMN (SMSE=0.95, p<0.01).

Network Ranking Procedure. Table 1 shows the 15*15 p-value matrix that resulted from comparing the distributions of the cross-validated and permutation-based prediction errors between the 15 networks (i.e., five intra-networks and ten inter-between networks).

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. DMN-FPN	-	0	0	0.36	0	0	0	0.382	0	0	0.422	0	0.389	0.192	0
2. CON-FPN	0.01	_	0.137	0.002	0.036	0	0	0.001	0.033	0	0.007	0.025	0.024	0	0.024
3. OccN-SMN	0.049	0	-	0.018	0.182	0	0	0.021	0.166	0	0.043	0.098	0.085	0.004	0.13
4. DMN-OccN	0	0	0	_	0	0	0	0.488	0	0	0.508	0	0.467	0.237	0
5. CON-OccN	0.136	0	0	0.059	-	0	0	0.068	0.387	0	0.105	0.249	0.178	0.024	0.302
6. FPN-FPN	0.021	0.303	0.11	0.012	0.044	-	0	0.014	0.035	0	0.016	0.028	0.03	0.006	0.035
7. CON-CON	0.003	0.174	0.037	0.001	0.01	0.379	_	0.001	0.006	0	0.003	0.005	0.004	0	0.01
8. CON-DMN	0	0	0	0	0	0	0	-	0	0	0.511	0	0.466	0.243	0
9. FPN-OccN	0.2	0	0	0.119	0	0	0	0.133	_	0	0.163	0.362	0.233	0.049	0.39
10. FPN-SMN	0.001	0.058	0.016	0.004	0.007	0.187	0.233	0.001	0.008	_	0.002	0.005	0.002	0	0.003
11. OccN-OccN	0	0	0	0	0	0	0	0	0	0	_	0	0.457	0.263	0
12. DMN-DMN	0.356	0	0	0.233	0	0	0	0.242	0	0	0.298	_	0.315	0.109	0.546
13. DMN-SMN	0	0	0	0	0	0	0	0	0	0	0	0	_	0	0
14. SMN-SMN	0	0	0	0	0	0	0	0	0	0	0	0	0.629	_	0
15. CON-SMN	0.336	0	0	0.208	0	0	0	0.222	0	0	0.27	0	0.279	0.105	_

Table 1. P-Value Matrix for Ranking Performance of Models.

CON, cingulo-opercular network; SMN, sensorimotor network; DMN, default mode network; FPN, frontoparietal network; OccN, occipital network.

After applying the network ranking procedure, allowing to determine the network that predicts the target behavior better than the maximum number of other network pairs, the DMN-FPN inter-network ranked the highest, indicating that the RSFC between these two networks predicted the average TPP behavior the best (**Figure 10**).





4. DISCUSSION

The goal of the study was to investigate the prediction of individual differences in average TPP behavior based on RSFC —combining a TPP game and a prediction-analytics framework. On the one hand, the behavioral results revealed that the amount of punishment —increasing with the degree of unfairness— was less in TPP than in SPP. On the other hand, the neuroimaging results indicated that RSFC between DMN and FPN (overlapping with CEN) predicted best the individual differences in average TPP behavior.

The first hypothesis — TPP behavior increasing with the observation of more unfair offers in the TPP game— was confirmed. The proposed neuropsychological model indicated that punishment behavior hinges on blame, which is based on the unfair outcome to the receiver and the intent of the proposer (Krueger & Hoffman, 2016). Our findings are supported by previous investigations, indicating a linear association between the degree of unfairness and amount of punishment (Jordan, Mcauliffe, & Rand, 2015; Stallen et al., 2018; Wu et al., 2014; Zhong et al., 2016). Furthermore, findings from other economic exchange games (e.g., SPP game) confirm this behavioral pattern, implying consistency of this phenomenon among different game paradigms (Fehr & Fischbacher, 2004b; Gummerum & Chu, 2014; McCall, Steinbeis, Ricard, & Singer, 2014; Stallen et al., 2018).

Further, our findings also demonstrated that the average costly punishment behavior was significantly lower in TPP than in SPP, which also has been sown in previous studies (Buckholtz & Marois, 2012; Zhou, Jiao, & Zhang, 2017). For SPP behavior, the second party is directly affected by the unfair offers made by the proposer; and, therefore, the punishment behavior is more driven by harm (i.e., unfair outcome). For TPP behavior, in contrast, the third party is not directly affected by the unfair offers, putting the focus on the intention of the proposer, which probably reduces costly punishment in TPP compared to SPP.

The second hypothesis —individual differences in average TPP behavior can be predicted by the interaction of RSFC between DMN and CEN— was confirmed. The results showed that RSFC between DMN and CEN (i.e., FPN) predicted best individual differences in average TPP behavior. TPP behavior engages three interacting large-scale networks. The SN (associated with aversive self-related emotional experiences that guide behavior) detects social norm violations (dACC), responds by generating an aversive experience (AI), and provides an emotional signal coding for the severity of harm (Amyg). The DMN (associated with social cognition, mentalizing, and theory of mind, ToM (Premack & Woodruff, 1978) is anchored in mPFC. This network assesses blame by integrating harm to the victim via the vmPFC's inter-network connectivity with SN (allowing the experience of feelings congruent with another's emotional situation, affective ToM) with the intent of the perpetrator evaluated via the dmPFC's intra-network connectivity with regions associated with self-referential processing (PCC) and inferring intentions in others (TPJ; understanding others' mental states and feelings, cognitive ToM). The CEN (anchored in the dIPFC and associated with higher-order cognition and decision making) converts the blame signal from the DMN into an actual decision, selecting a punishment (when focusing on the perpetrator) or compensation (when focusing on the victim) that fits the norm violation.

Previous research employing task-based fMRI has shown that both DMN (associated with inferring someone's intention) and CEN (associated with the determination of the appropriate punishment amount) are engaged in TPP (Heekeren, Wartenburger, Schmidt, Schwintowski, & Villringer, 2003; Schleim, Spranger, Erk, & Walter, 2011). Presumably, prior to converting blame into punishment, the mPFC anchored in DMN needs to integrate intention-integrating signals from TPJ and harm-integrating signals from vmPFC into an assessment of blame (Krueger & Hoffman, 2016). TPP is more driven from inferring the intention of the proposers, leading to a lower punishment behavior in comparison to SPP. According to previous task-based fMRI research, a higher engagement of the social inference system anchored in DMN is required (Buckholtz & Marois, 2012; Krueger, Hoffman, Walter, & Grafman, 2014) to assess the intentions of others (Harris, Todorov, & Fiske, 2005).

After integrating these two signals, the CEN converts blame assessment into an appropriate punishment decision (Bellucci et al., 2017), where the PCC constructs a scale of punishment for the dIPFC to select an appropriate punishment decision (Buckholtz et al., 2008). Previous evidence has shown that that the magnitude of TPP punishment correlates proportionally to the activation in the right dIPFC, a core brain region of CEN engaged in determining the appropriate punishment (Buckholtz et al., 2008). In sum, RSFC between DMN and FPN (i.e., CEN), significantly involved in signal conversion from blame to punishment, accounts for individual differences in TPP.

Although those findings provide us with a comprehensive understanding of the neural architecture underlying TPP behavior; however, some limitations exist that have to

be addressed in future research. First, future studies investigating TPP behavior based on RSFC using task-free fMRI need to be investigated in larger study samples. As a result, larger study samples will allow multivariate regression models to accomplish better accuracy and lower error as well as lower variance. Second, the prediction of costly punishment was based on measurements at a single time-point at the day of the experiment. Future studies should investigate whether RSFC can also predict TPP behavior at different time-points in the future —characterizing temporally stable consistency of the underlying RSFC.

In summary, the behavioral results demonstrated that the amount of TPP increased as the offer became increasingly unfair and the amount of punishment in the TPP was less than SPP. Furthermore, RSFC between DMN and CEN predicted individual differences in TPP behavior. The process of signal conversion from blame into punishment engages the interaction between DMN (i.e., inferring the intentions and goals of others) and CEN (i.e., determining the appropriate punishment). In conclusion, investigating individual differences in costly punishment behavior allow comprehensive understandings for sustaining cooperation and fulfilling enforcement of social norms in human society. Furthermore, in our research, behavioral analysis and prediction-analytics framework were conducted on the normal participants. Consequently, with these observed results and established prediction-analytics frameworks, investigating individual difference with social disorder disease become possible.

5. APPENDIX

Table S1. Characteristics of 142 ROIs defined by Dosenbach's Atlas.

Characteristics (i.e., coordinates, hemisphere, label, and network) for each region of interest (ROI) of the Dosenbach' Atlas are shown.

	MNI-corrdinates			hemisphere			
number	х	У	Z	left/right	ROI label (abbr.)	ROI label	network
1	6	64	3	right	VMPFC	ventromedial prefrontal cortex	default
2	29	57	18	right	aPFC	anterior prefrontal cortex	fronto-parietal
3	-29	57	10	left	aPFC	anterior prefrontal cortex	fronto-parietal
4	0	51	32	-	mPFC	medial prefrontal cortex	default
5	-25	51	27	left	aPFC	anterior prefrontal cortex	default
6	9	51	16	right	VMPFC	ventromedial prefrontal cortex	default
7	-6	50	-1	left	VMPFC	ventromedial prefrontal cortex	default
8	27	49	26	right	aPFC	anterior prefrontal cortex	cingulo-opercular
9	42	48	-3	right	vent-aPFC	ventral anterior prefrontal cortex	fronto-parietal
10	-43	47	2	left	vent-PFC	ventral prefrontal cortex	fronto-parietal
11	-11	45	17	left	VMPFC	ventromedial prefrontal cortex	default
12	39	42	16	right	VLPFC	ventral lateral prefrontal cortex	fronto-parietal
13	8	42	-5	right	VMPFC	ventromedial prefrontal cortex	default
14	9	39	20	right	ACC	anterior cingulate cortex	default
15	46	39	-15	right	VLPFC	ventral lateral prefrontal cortex	default
16	40	36	29	right	DLPFC	dorsolateral prefrontal cortex	fronto-parietal
17	23	33	47	right	sup-front	superior frontal gyrus	default
18	34	32	7	right	VPFC	ventral prefrontal cortex	cingulo-opercular
19	-2	30	27	left	ACC	anterior cingulate cortex	cingulo-opercular
20	-16	29	54	left	sup-front	superior frontal gyrus	default
21	-1	28	40	left	ACC	anterior cingulate cortex	fronto-parietal
22	46	28	31	right	DLPFC	dorsolateral prefrontal cortex	fronto-parietal
23	-52	28	17	left	VPFC	ventral prefrontal cortex	fronto-parietal
24	-44	27	33	left	DLPFC	dorsolateral prefrontal cortex	fronto-parietal
25	51	23	8	right	vFC	ventral frontal cortex	cingulo-opercular
26	38	21	-1	right	AI	anterior insula	cingulo-opercular
27	9	20	34	right	dACC	dorsal anterior cingulate cortex	cingulo-opercular
28	-36	18	2	left	AI	anterior insula	cingulo-opercular
29	40	17	40	right	dFC	dorsal prefrontal cortex	fronto-parietal
30	-6	17	34	left	basal-ganglia	basal ganglia	cingulo-opercular
31	0	15	45	-	mPFC	medial prefrontal cortex	cingulo-opercular

sensorimotor	frontal lobe	frontal	right	14	11	58	32
cingulo-opercular	ventral frontal cortex	vFC	left	14	10	-46	33
fronto-parietal	dorsal prefrontal cortex	dFC	right	34	8	44	34
sensorimotor	dorsal prefrontal cortex	dFC	right	34	8	60	35
fronto-parietal	dorsal prefrontal cortex	dFC	left	36	7	-42	36
sensorimotor	ventral frontal cortex	vFC	left	23	7	-55	37
cingulo-opercular	basal ganglia	basal-ganglia	left	7	6	-20	38
cingulo-opercular	basal ganglia	basal-ganglia	right	7	6	14	39
cingulo-opercular	ventral frontal cortex	vFC	left	1	6	-48	40
sensorimotor	pre-supplementary motor area	pre-SMA	right	51	5	10	41
sensorimotor	ventral frontal cortex	vFC	right	12	1	43	42
sensorimotor	supplementary motor area	SMA	-	52	-1	0	43
cingulo-opercular	middle insula	mid-insula	right	-3	-2	37	44
sensorimotor	frontal lobe	frontal	right	32	-3	53	45
sensorimotor	precentral gyrus	PreCG	right	17	-3	58	46
cingulo-opercular	thalamus	thalamus	left	13	-3	-12	47
sensorimotor	middle insula	mid-insula	left	11	-3	-42	48
sensorimotor	precentral gyrus	PreCG	left	49	-6	-44	49
sensorimotor	parietal lobe	parietal	left	54	-8	-26	50
sensorimotor	precentral gyrus	PreCG	right	24	-8	46	51
sensorimotor	precentral gyrus	PreCG	left	23	-9	-54	52
sensorimotor	precentral gyrus	PreCG	right	38	-11	44	53
sensorimotor	parietal lobe	parietal	left	36	-12	-47	54
sensorimotor	middle insula	mid-insula	right	16	-12	33	55
sensorimotor	middle insula	mid-insula	left	15	-12	-36	56
cingulo-opercular	thalamus	thalamus	left	6	-12	-12	57
cingulo-opercular	thalamus	thalamus	right	6	-12	11	58
cingulo-opercular	middle insula	mid-insula	right	2	-12	32	59
sensorimotor	temporal lobe	temporal	right	8	-13	59	60
cingulo-opercular	middle insula	mid-insula	left	1	-14	-30	61
sensorimotor	parietal lobe	parietal	left	59	-15	-38	62
default	inferior temporal gyrus	IT	right	-13	-15	52	63
sensorimotor	parietal lobe	parietal	left	50	-18	-47	64
sensorimotor	parietal lobe	parietal	right	45	-20	46	65
sensorimotor	parietal lobe	parietal	left	38	-22	-55	66
sensorimotor	precentral gyrus	PreCG	left	22	-22	-54	67
sensorimotor	temporal lobe	temporal	left	9	-22	-54	68
sensorimotor	parietal lobe	parietal	right	55	-23	41	69
sensorimotor	posterior insula	post-insula	right	17	-24	42	70
cingulo-opercular	basal ganglia	basal-ganglia	right	2	-24	11	71

default	inferior temporal gyrus	IT	left	-15	-25	-59	72
default	precuneus	PC	right	31	-26	1	73
sensorimotor	parietal lobe	parietal	right	62	-27	18	74
sensorimotor	parietal lobe	parietal	left	60	-27	-38	75
cingulo-opercular	posterior insula	post-insula	left	9	-28	-30	76
sensorimotor	parietal lobe	parietal	left	64	-30	-24	77
cingulo-opercular	temporal lobe	temporal	right	5	-30	51	78
sensorimotor	posterior parietal cortex	post-parietal	left	48	-31	-41	79
cingulo-opercular	precuneus	PC	left	-4	-31	-4	80
cingulo-opercular	fusiform gyrus	fusiform	right	-18	-31	54	81
sensorimotor	temporal lobe	temporal	left	16	-37	-41	82
sensorimotor	temporal lobe	temporal	left	13	-37	-53	83
default	fusiform gyrus	fusiform	right	-15	-37	28	84
default	precuneus cortex	PreC	left	45	-38	-3	85
sensorimotor	superior parietal lobule	SPL	right	65	-39	34	86
cingulo-opercular	precuneus cortex	PreC	right	50	-40	8	87
fronto-parietal	inferior parietal lobe	IPL	left	42	-40	-41	88
cingulo-opercular	parietal lobe	parietal	right	20	-41	58	89
default	precuneus	PC	left	3	-41	-8	90
default	inferior temporal gyrus	IT	left	-2	-41	-61	91
default	occipital lobe	Occ	left	-11	-42	-28	92
default	precuneus	PC	left	25	-43	-5	93
default	precuneus cortex	PreC	right	25	-43	9	94
cingulo-opercular	temporal lobe	temporal	right	8	-43	43	95
fronto-parietal	inferior parietal lobe	IPL	right	43	-44	54	96
cingulo-opercular	parietal lobe	parietal	left	30	-44	-55	97
fronto-parietal	posterior parietal cortex	post-parietal	left	48	-46	-35	98
cingulo-opercular	superior temporal gyrus	ST	right	21	-46	42	99
fronto-parietal	inferior parietal lobe	IPL	left	49	-47	-48	100
cingulo-opercular	angular gyrus	AG	left	29	-47	-41	101
cingulo-opercular	temporal lobe	temporal	left	11	-47	-59	102
fronto-parietal	inferior parietal lobe	IPL	left	39	-50	-53	103
default	precuneus cortex	PreC	right	33	-50	5	104
occipital	occipital lobe	Occ	left	1	-50	-18	105
fronto-parietal	inferior parietal lobe	IPL	right	47	-52	44	106
default	precuneus	PC	left	17	-52	-5	107
default	precuneus	PC	right	17	-55	10	108
default	precuneus cortex	PreC	left	29	-56	-6	109
fronto-parietal	intra-parietal sulcus	IPS	left	46	-58	-32	110
default	precuneus	PC	left	17	-58	-11	111

112	32	-59	41	right	IPS	intra-parietal sulcus	fronto-parietal
113	51	-59	34	right	AG	angular gyrus	default
114	-34	-60	-5	left	Occ	occipital lobe	occipital
115	36	-60	-8	right	Occ	occipital lobe	occipital
116	46	-62	5	right	temporal	temporal lobe	occipital
117	-48	-63	35	left	AG	angular gyrus	default
118	-52	-63	15	left	TPJ	temporoparietal junction	cingulo-opercular
119	-44	-63	-7	left	Occ	occipital lobe	occipital
120	19	-66	-1	right	Occ	occipital lobe	occipital
121	11	-68	42	right	PreC	precuneus cortex	default
122	17	-68	20	right	Occ	occipital lobe	occipital
123	-36	-69	40	left	IPS	intra-parietal sulcus	default
124	39	-71	13	right	Occ	occipital lobe	occipital
125	-9	-72	41	left	Occ	occipital lobe	default
126	45	-72	29	right	Occ	occipital lobe	default
127	29	-73	29	right	Occ	occipital lobe	occipital
128	-2	-75	32	left	Occ	occipital lobe	default
129	-29	-75	28	left	Occ	occipital lobe	occipital
130	-16	-76	33	left	Occ	occipital lobe	occipital
131	-42	-76	26	left	Occ	occipital lobe	default
132	9	-76	14	right	Occ	occipital lobe	occipital
133	15	-77	32	right	Occ	occipital lobe	occipital
134	20	-78	-2	right	Occ	occipital lobe	occipital
135	-5	-80	9	left	postOcc	posterior occipital lobe	occipital
136	29	-81	14	right	postOcc	posterior occipital lobe	occipital
137	33	-81	-2	right	postOcc	posterior occipital lobe	occipital
138	-37	-83	-2	left	postOcc	posterior occipital lobe	occipital
139	-29	-88	8	left	postOcc	posterior occipital lobe	occipital
140	13	-91	2	right	postOcc	posterior occipital lobe	occipital
141	27	-91	2	right	postOcc	posterior occipital lobe	occipital
142	-4	-94	12	left	postOcc	posterior occipital lobe	occipital

6. REFERENCES

- Anderson, J. S., Nielsen, J. A., Froehlich, A. L., DuBray, M. B., Druzgal, T. J., Cariello, A. N., ... Lainhart, J. E. (2011). Functional connectivity magnetic resonance imaging classification of autism. *Brain: A Journal of Neurology*, 134(Pt 12), 3742–3754. https://doi.org/10.1093/brain/awr263
- Baumgartner, T., Götte, L., Gügler, R., & Fehr, E. (2012). The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. *Human Brain Mapping*, 33(6), 1452–1469. https://doi.org/10.1002/hbm.21298
- Beckmann, C. F., DeLuca, M., Devlin, J. T., & Smith, S. M. (2005). Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 360(1457), 1001–1013. https://doi.org/10.1098/rstb.2005.1634
- Bellucci, G., Chernyak, S., Hoffman, M., Deshpande, G., Dal Monte, O., Knutson, K. M., ... Krueger, F. (2017). Effective connectivity of brain regions underlying thirdparty punishment: Functional MRI and Granger causality evidence. *Social Neuroscience*, *12*(2), 124–134. https://doi.org/10.1080/17470919.2016.1153518
- Bendor, J., & Swistak, P. (2001). The Evolution of Norms. American Journal of Sociology, 106(6), 1493–1545. https://doi.org/10.1086/321298
- Biswal, B., Yetkin, F. Z., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, *34*(4), 537–541.
- Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. *Trends in Cognitive Sciences*, 14(6), 277–290. https://doi.org/10.1016/j.tics.2010.04.004

Brown, D. (1991). Human Universals. McGraw Hill.

- Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., & Marois, R. (2008). The Neural Correlates of Third-Party Punishment. *Neuron*, 60(5), 930–940. https://doi.org/10.1016/j.neuron.2008.10.016
- Buckholtz, J. W., & Marois, R. (2012). The roots of modern justice: cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience*, 15, 655–661. https://doi.org/10.1038/nn.3087
- Calhoun, V. D., Adali, T., Pearlson, G. D., & Pekar, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping*, 14(3), 140–151. https://doi.org/10.1002/hbm.1048
- Civai, C., Corradi-Dell'Acqua, C., Gamer, M., & Rumiati, R. I. (2010). Are irrational reactions to unfairness truly emotionally-driven? Dissociated behavioural and emotional responses in the Ultimatum Game task. *Cognition*, 114(1), 89–95. https://doi.org/10.1016/j.cognition.2009.09.001
- Cole, D. M., Beckmann, C. F., Long, C. J., Matthews, P. M., Durcan, M. J., & Beaver, J. D. (2010). Nicotine replacement in abstinent smokers improves cognitive withdrawal symptoms with modulation of resting brain network dynamics. *NeuroImage*, 52(2), 590–599. https://doi.org/10.1016/j.neuroimage.2010.04.251
- Cordes, D., Haughton, V. M., Arfanakis, K., Wendt, G. J., Turski, P. A., Moritz, C. H., ... Meyerand, M. E. (2000). Mapping functionally related regions of brain with functional connectivity MR imaging. *AJNR. American Journal of Neuroradiology*, 21(9), 1636–1644.
- Cordes, Dietmar, Haughton, V., Carew, J. D., Arfanakis, K., & Maravilla, K. (2002). Hierarchical clustering to measure connectivity in fMRI resting-state data. *Magnetic Resonance Imaging*, 20(4), 305–317.
- Dai, Z., Yan, C., Wang, Z., Wang, J., Xia, M., Li, K., & He, Y. (2012). Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3). *NeuroImage*, 59(3), 2187–2195. https://doi.org/10.1016/j.neuroimage.2011.10.003

- Damoiseaux, J. S., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., & Beckmann, C. F. (2006). Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences of the United States of America*, 103(37), 13848–13853. https://doi.org/10.1073/pnas.0601417103
- David, B., Hu, Y., Krüger, F., & Weber, B. (2017). Other-regarding attention focus modulates third-party altruistic choice: An fMRI study. *Scientific Reports*, 7(1), 43024. https://doi.org/10.1038/srep43024
- Dosenbach, N. U. F., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., ... Schlaggar, B. L. (2010). Prediction of Individual Brain Maturity Using fMRI. *Science (New York, N.Y.)*, 329(5997), 1358–1361. https://doi.org/10.1126/science.1194144
- Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences*, 11(10), 419–427. https://doi.org/10.1016/j.tics.2007.09.002
- Fehr, E., & Fischbacher, U. (2004a). *Social norms and human cooperation*. 8(4), 185–190. https://doi.org/10.1016/j.tics.2004.02.007
- Fehr, E., & Fischbacher, U. (2004b). Third-party punishment and social norms. *Evolution and Human Behavior*, 25(2), 63–87. https://doi.org/10.1016/S1090-5138(04)00005-4
- Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, 13(1), 1–25. https://doi.org/10.1007/s12110-002-1012-7
- Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in Simple Bargaining Experiments. 6(3), 347–369. https://doi.org/10.1006/game.1994.1021
- Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the*

United States of America, *102*(27), 9673–9678. https://doi.org/10.1073/pnas.0504136102

- Fransson, P. (2005). Spontaneous low-frequency BOLD signal fluctuations: an fMRI investigation of the resting-state default mode of brain function hypothesis. *Human Brain Mapping*, 26(1), 15–29. https://doi.org/10.1002/hbm.20113
- Friston, K. J. (1998). The disconnection hypothesis. *Schizophrenia Research*, *30*(2), 115–125.
- Ginther, M. R., Bonnie, R. J., Hoffman, M. B., Shen, F. X., Simons, K. W., Jones, O. D., & Marois, R. (2016). Parsing the Behavioral and Brain Mechanisms of Third-Party Punishment. 36(36), 9420–9434. https://doi.org/10.1523/jneurosci.4499-15.2016
- Glass, L., Moody, L., Grafman, J., & Krueger, F. (2016). Neural signatures of third-party punishment: evidence from penetrating traumatic brain injury. *Social Cognitive* and Affective Neuroscience, 11(2), 253–262. https://doi.org/10.1093/scan/nsv105
- Gummerum, M., & Chu, M. T. (2014). Outcomes and intentions in children's, adolescents', and adults' second- and third-party punishment behavior. *Cognition*, *133*(1), 97–103. https://doi.org/10.1016/j.cognition.2014.06.001
- Hahn, T., Notebaert, K., Anderl, C., Reicherts, P., Wieser, M., Kopf, J., ... Windmann, S. (2015). Reliance on functional resting-state network for stable task control predicts behavioral tendency for cooperation. *NeuroImage*, *118*, 231–236. https://doi.org/10.1016/j.neuroimage.2015.05.093
- Harris, L. T., Todorov, A., & Fiske, S. T. (2005). Attributions on the brain: neuroimaging dispositional inferences, beyond theory of mind. *NeuroImage*, 28(4), 763–769. https://doi.org/10.1016/j.neuroimage.2005.05.021
- Harrison, B. J., Pujol, J., Ortiz, H., Fornito, A., Pantelis, C., & Yücel, M. (2008). Modulation of Brain Resting-State Networks by Sad Mood Induction. *PLOS ONE*, 3(3), e1794. https://doi.org/10.1371/journal.pone.0001794

- Heekeren, H. R., Wartenburger, I., Schmidt, H., Schwintowski, H.-P., & Villringer, A. (2003). An fMRI study of simple ethical decision-making. *Neuroreport*, *14*(9), 1215–1219. https://doi.org/10.1097/01.wnr.0000081878.45938.a7
- Henrich, J. (2006). SOCIAL SCIENCE: Enhanced: Cooperation, Punishment, and the Evolution of Human Institutions. *Science*, *312*(5770), 60–61. https://doi.org/10.1126/science.1126398
- Hoffman, M. (2014). *The Punisher's Brain: The Evolution of Judge and Jury*. Cambridge University Press.
- Horovitz, S. G., Braun, A. R., Carr, W. S., Picchioni, D., Balkin, T. J., Fukunaga, M., & Duyn, J. H. (2009). Decoupling of the brain's default mode network during deep sleep. *Proceedings of the National Academy of Sciences*, 106(27), 11376–11381. https://doi.org/10.1073/pnas.0901435106
- Jiang, T., He, Y., Zang, Y., & Weng, X. (2004). Modulation of functional connectivity during the resting state and the motor task. *Human Brain Mapping*, 22(1), 63–71. https://doi.org/10.1002/hbm.20012
- Jordan, J., Mcauliffe, K., & Rand, D. (2015). *The effects of endowment size and strategy method on third party punishment*. https://doi.org/10.1007/s10683-015-9466-8
- Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a Constraint on Profit Seeking: Entitlements in the Market. *The American Economic Review*, 76(4), 728–741.
- Kelly, C., Biswal, B. B., Craddock, R. C., Castellanos, F. X., & Milham, M. P. (2012). Characterizing variation in the functional connectome: promise and pitfalls. *Trends in Cognitive Sciences*, 16(3), 181–188. https://doi.org/10.1016/j.tics.2012.02.001
- Krueger, F., & Hoffman, M. (2016). The Emerging Neuroscience of Third-Party Punishment. *Trends in Neurosciences*, 39(8), 499–501. https://doi.org/10.1016/j.tins.2016.06.004

- Krueger, F., Hoffman, M., Walter, H., & Grafman, J. (2014). An fMRI investigation of the effects of belief in free will on third-party punishment. *Social Cognitive and Affective Neuroscience*, 9(8), 1143–1149. https://doi.org/10.1093/scan/nst092
- McCall, C., Steinbeis, N., Ricard, M., & Singer, T. (2014). Compassion meditators show less anger, less punishment, and more compensation of victims in response to fairness violations. *Frontiers in Behavioral Neuroscience*, 8. https://doi.org/10.3389/fnbeh.2014.00424
- Menon, V. (2011). Large-scale brain networks and psychopathology: a unifying triple network model. *Trends in Cognitive Sciences*, 15(10), 483–506. https://doi.org/10.1016/j.tics.2011.08.003
- Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 87(24), 9868– 9872.
- Pauling, L., & Coryell, C. D. (1936). The Magnetic Properties and Structure of Hemoglobin, Oxyhemoglobin and Carbonmonoxyhemoglobin. *Proceedings of the National Academy of Sciences of the United States of America*, 22(4), 210–216.
- Pillutla, M. M., & Murnighan, J. K. (1996). Unfairness, Anger, and Spite: Emotional Rejections of Ultimatum Offers. *Organizational Behavior and Human Decision Processes*, 68(3), 208–224. https://doi.org/10.1006/obhd.1996.0100
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? Behavioral and Brain Sciences, 1(4), 515–526. https://doi.org/10.1017/S0140525X00076512
- Richiardi, J., Achard, S., Bunke, H., & Ville, D. V. D. (2013). Machine Learning with Brain Graphs: Predictive Modeling Approaches for Functional Imaging in Systems Neuroscience. *IEEE Signal Processing Magazine*, 30(3), 58–70. https://doi.org/10.1109/MSP.2012.2233865

Sadaghiani, S., Scheeringa, R., Lehongre, K., Morillon, B., Giraud, A.-L., & Kleinschmidt, A. (2010). Intrinsic connectivity networks, alpha oscillations, and tonic alertness: a simultaneous electroencephalography/functional magnetic resonance imaging study. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 30(30), 10243–10250. https://doi.org/10.1523/JNEUROSCI.1004-10.2010

- Sämann, P. G., Wehrle, R., Hoehn, D., Spoormaker, V. I., Peters, H., Tully, C., ... Czisch, M. (2011). Development of the brain's default mode network from wakefulness to slow wave sleep. *Cerebral Cortex (New York, N.Y.: 1991)*, 21(9), 2082–2093. https://doi.org/10.1093/cercor/bhq295
- Schleim, S., Spranger, T. M., Erk, S., & Walter, H. (2011). From moral to legal judgment: the influence of normative context in lawyers and other academics. *Social Cognitive and Affective Neuroscience*, 6(1), 48–57. https://doi.org/10.1093/scan/nsq010
- Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., ... Greicius, M. D. (2007). Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 27(9), 2349–2356. https://doi.org/10.1523/JNEUROSCI.5587-06.2007
- Shen, H., Wang, L., Liu, Y., & Hu, D. (2010). Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI. *NeuroImage*, 49(4), 3110–3121. https://doi.org/10.1016/j.neuroimage.2009.11.011
- Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V., & Greicius, M. D. (2012). Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral Cortex (New York, N.Y.: 1991)*, 22(1), 158–165. https://doi.org/10.1093/cercor/bhr099
- Sober, P. E., Wilson, P. D. S., & Wilson, D. S. (1999). Unto Others: The Evolution and Psychology of Unselfish Behavior (New Ed edition). Cambridge, Mass.: Harvard University Press.

- Sridharan, D., Levitin, D. J., & Menon, V. (2008). A critical role for the right frontoinsular cortex in switching between central-executive and default-mode networks. *Proceedings of the National Academy of Sciences of the United States of America*, 105(34), 12569–12574. https://doi.org/10.1073/pnas.0800005105
- Stallen, M., Rossi, F., Heijne, A., Smidts, A., Dreu, C. K. W. D., & Sanfey, A. G. (2018). Neurobiological Mechanisms of Responding to Injustice. *Journal of Neuroscience*, 38(12), 2944–2954. https://doi.org/10.1523/JNEUROSCI.1242-17.2018
- Sun, L., Tan, P., Cheng, Y., Chen, J., & Qu, C. (2015). The effect of altruistic tendency on fairness in third-party punishment. 6. https://doi.org/10.3389/fpsyg.2015.00820
- Supekar, K., Uddin, L. Q., Prater, K., Amin, H., Greicius, M. D., & Menon, V. (2010). Development of functional and structural connectivity within the default mode network in young children. *NeuroImage*, 52(1), 290–301. https://doi.org/10.1016/j.neuroimage.2010.04.009
- Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J.-B., Lebihan, D., & Dehaene, S. (2006). Inverse retinotopy: inferring the visual content of images from brain activation patterns. *NeuroImage*, 33(4), 1104–1116. https://doi.org/10.1016/j.neuroimage.2006.06.062
- Treadway, M. T., Buckholtz, J. W., Martin, J. W., Jan, K., Asplund, C. L., Ginther, M. R., ... Marois, R. (2014). Corticolimbic gating of emotion-driven punishment. *Nature Neuroscience*, 17(9), 1270–1275. https://doi.org/10.1038/nn.3781
- van den Heuvel, M. P., & Hulshoff Pol, H. E. (2010). Exploring the brain network: a review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology*, 20(8), 519–534. https://doi.org/10.1016/j.euroneuro.2010.03.008
- Van Dijk, K. R. A., Hedden, T., Venkataraman, A., Evans, K. C., Lazar, S. W., & Buckner, R. L. (2010). Intrinsic functional connectivity as a tool for human

connectomics: theory, properties, and optimization. *Journal of Neurophysiology*, *103*(1), 297–321. https://doi.org/10.1152/jn.00783.2009

- Wang, Y., & Li, T.-Q. (2013). Analysis of Whole-Brain Resting-State fMRI Data Using Hierarchical Clustering Approach. *PLOS ONE*, 8(10), e76315. https://doi.org/10.1371/journal.pone.0076315
- Wu, Y., Yu, H., Shen, B., Yu, R., Zhou, Z., Zhang, G., ... Zhou, X. (2014). Neural basis of increased costly norm enforcement under adversity. *Social Cognitive and Affective Neuroscience*, 9(12), 1862–1871. https://doi.org/10.1093/scan/nst187
- Yamashita, M., Kawato, M., & Imamizu, H. (2015). Predicting learning plateau of working memory from whole-brain intrinsic network connectivity patterns. *Scientific Reports*, 5, 7622. https://doi.org/10.1038/srep07622
- Zhong, S., Chark, R., Hsu, M., & Chew, S. H. (2016). Computational substrates of social norm enforcement by unaffected third parties. *NeuroImage*, 129, 95–104. https://doi.org/10.1016/j.neuroimage.2016.01.040
- Zhou, Y., Jiao, P., & Zhang, Q. (2017). Second-party and third-party punishment in a public goods experiment. *Applied Economics Letters*, 24(1), 54–57. https://doi.org/10.1080/13504851.2016.1161709

BIOGRAPHY

Ko-Tsung Hsu graduated from Chang Jung Senior High School, Tainan, Taiwan, in 2011. He received his Bachelor of Science in Biotechnology from Ming Chuan University, Taoyuan, Taiwan, in 2015. He was employed as a grassroots technician for two months in Dong Guan after the graduation from Ming Chuan University. He has studied since 2017 for a Master of Science in Bioinformatics and Computational Biology at George Mason University.