THE NEURAL BASIS OF HUMAN PROSOCIAL BEHAVIOR

EDITED BY: Yefeng Chen, Hang Ye, Chao Liu and Qi Li

PUBLISHED IN: Frontiers in Psychology and Frontiers in Neuroscience

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88963-218-3
DOI 10.3389/978-2-88963-218-3

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

THE NEURAL BASIS OF HUMAN PROSOCIAL BEHAVIOR

Topic Editors:
Yefeng Chen, Zhejiang University, China
Hang Ye, Zhejiang University of Finance and Economics, China
Chao Liu, Beijing Normal University, China
Qi Li, Chinese Academy of Sciences, China

Citation: Chen, Y., Ye, H., Liu, C., Li, Q., eds. (2019). The Neural Basis of Human Prosocial Behavior. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88963-218-3

Table of Contents

frontiers in Psychology

EDITORIAL
published: 10 September 2019
doi: 10.3389/fpsyg.2019.02058

Editorial: The Neural Basis of Human Prosocial Behavior

Yefeng Chen[1], *Hang Ye*[2*], *Chao Liu*[3] *and Qi Li*[4]

[1] College of Economics and Interdisciplinary Center for Social Sciences, Zhejiang University, Hangzhou, China, [2] Center for Economic Behavior and Decision-Making, School of Economics, Zhejiang University of Finance and Economics, Hangzhou, China, [3] State Key Laboratory of Cognitive Neuroscience and Learning & IDG, McGovern Institute for Brain Research, Beijing Normal University, Beijing, China, [4] Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

Keywords: neural basis, prosocial behavior, neural mechanisms, neural imaging, brain stimulation

Editorial on the Research Topic

The Neural Basis of Human Prosocial Behavior

With the rise of laboratory and field experimental economics, the famous prisoner's dilemma, public good, dictator, ultimatum, and trust games have become the classical paradigms of studying prosocial behavior (Güth et al., 1982; Berg et al., 1995; Fehr and Gächter, 2002; Camerer, 2003). Due to the increasing use of functional magnetic resonance imaging (fMRI), transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (tDCS) with human subjects playing economic games, the neural basis of prosocial behavior has been uncovered by a large amount of neural imaging and stimulating research (Rilling et al., 2002; Sanfey et al., 2003; de Quervain et al., 2004; Knoch et al., 2006; Krueger et al., 2007). A wide range of brain areas including, but not limited to the prefrontal cortex, orbitofrontal cortex, cingulate cortex, striatum, and amygdale have been revealed highly correlated or causally related with prosocial behaviors.

A number of hypotheses such as empathy, altruism, reciprocity, inequality aversion, or guilt aversion preferences have been considered as motives promoting prosocial behavior. However, the neural bases of these different preferences have seldom been revealed and the mechanisms of how these preferences influence prosocial behavior have rarely been discussed. Moreover, since prosocial behavior may be due to the cooperative work of several brain areas (neural network), it is essential to integrate findings from difference disciplines including psychology, economics, neuroscience, and to nearly all the social and behavioral sciences.

The present Research Topic of Frontiers in Psychology aims to bring a collection of research revealing the neural basis of human prosocial behavior. Totally 14 articles composing this unique Frontiers Research Topic in different types of prosocial behavior.

There are 3 review articles included in this volume. Luo summarize the research on the neural basis of different types of pro-social behaviors and describe a common shared neural circuitry of these pro-social behaviors. This review introduces several widely used approaches to develop new insights into understanding prosocial behaviors by combining the game theory of economics with neuroscience technologies. Zheng et al. summarize models of the emotional influence on fairness-related decision making and the corresponding behavioral and neural evidence. In their view, the future research on fairness-related decision making should focus on inducing incidental social emotion, avoiding irrelevant emotion when regulating, exploring the individual differences in emotional dispositions, and strengthening the ecological validity of the paradigm. Liu et al. review neuroimaging studies on social networks, and probe into the connection between individuals'

Edited and reviewed by:
Walter Adriani,
Istituto Superiore di Sanità (ISS), Italy

****Correspondence:***
Hang Ye
hangye@zju.edu.cn

Specialty section:
This article was submitted to
Decision Neuroscience,
a section of the journal
Frontiers in Psychology

Received: *25 June 2019*
Accepted: *23 August 2019*
Published: *10 September 2019*

Citation:
Chen Y, Ye H, Liu C and Li Q (2019)
Editorial: The Neural Basis of Human
Prosocial Behavior.
Front. Psychol. 10:2058.
doi: 10.3389/fpsyg.2019.02058

social network size and neural mechanisms. They find there are two main methods to measure the social network size. One is Social Network Index and the other is Social Network Questionnaires. These two measurements in view of the hierarchical organization of social networks are carefully examined in this paper. And the authors reveal that the two assessments are dissimilar in effect. This finding sheds new light on the understanding of the subtle distinctions among various social network assessments.

Adopting givesome games and public good dilemma, Liu et al. explored social interaction patterns between the disabled and abled people. This is the only one behavioral study but not neural study in this volume. However, this study is quite interesting using a special sample. They found disabled people were more likely to interact with the disabled people, while the abled people preferred to interact with the abled people; comparing with the abled people, the disabled people had higher cooperation; they also revealed that advantage in the number of the disabled people could reverse their disadvantage in the identity. The results provide related theoretical support for the disabled people's federation and communities when carrying out activities for the disabled people.

All the remaining 10 papers explore the neural basis of different types of prosocial behavior using neuroimaging and brain stimulation approaches such as fMRI, TMS, tDCS, ERP, and so on.

Using the event-related potential (ERP) technique, Liu et al. explored neural mechanisms underlying the processing of evaluating altruistic outcomes when self-interests are sacrificed. Their ERP results showed that when evaluating another person's outcomes in the low-empathy condition, an inversed FRN effect occurred. But this kind of effect did not appear in the high-empathy condition. This study suggest that empathy could modulate the neural responses to altruistic outcomes in which increasing welfare of others could result in a cost of the self.

On the topic of fairness and inequity aversion, Li et al. provided behavioral and electrophysiological data to demonstrate that advantageous inequity aversion may differ as a function of the individual's role in determining allocations. If the individual cannot decide to distribute, this kind of inequity aversion will disappear. In their functional MRI study, Wei et al. investigated how social support affects the responders' fairness considerations and related decision-making processes in the ultimatum game. They demonstrated that the fairness-related decision-making processes are context-dependent and are modulated by social support.

By manipulating prestige-based social status, Blue et al. found that participants who played the role of investors in TG tended to be more affected by higher status Trustee promises than by lower status Trustee promises, despite the equal reinforcement schedule across conditions. Their findings suggest that honesty perception is affected by social status at both a behavioral and neural level, and that subjective socio-economic status may modulate this effect.

In the research on cooperation and punishment, using a linear asymmetric PG, Li et al. demonstrated the effect of the rLPFC on a priori normative beliefs without threats of external punishment through tDCS. Their finding reveals that rLPFC stimulation affects beliefs in the cooperation norm. As the author said, this research is a promising step toward understanding how neurobiological mechanisms are connected to beliefs in cooperation norms. In another study, for the first time, Li et al. compared the different neural processes of fourth-party evaluation on third-party help and punishment. Their ERP results revealed that fourth-party bystanders' FRN amplitudes were modulated by the third-party behaviors.

Regarding the study of deception, Gao et al. investigated the effect of modulating the activity of the DLPFC on deception. They conducted a between-subject design in a signaling framework of deception. Their results demonstrated the important role of DLPFC in modulating self-interested driven deceptive behavior. And they also found that in the sham stimulation treatment, males were more honest than females, while such gender difference disappeared in the right anodal/left cathodal stimulation treatment. Moreover, Tang et al. is the first study to investigate how activity in rTPJ affects deception in fairness related moral hypocrisy. They used a revised version of dictator game to examine the role of self-centered and other-regarding concerns in deception through stimulating rTPJ by tDCS. They found that deception in moral hypocrisy was increased by revealing appearing fair without true fairness to recipients than not. And this effect was decreased by anodal stimulation on rTPJ rather than cathodal and sham stimulation.

Finally, there are 2 paper focus on the moral judgment. In Ying et al.'s functional magnetic resonance imaging study, the participants evaluated the degree of disgust using sentences related to mild moral violations with different types of behavioral agents including the mother and stranger. They doubly dissociated two insular components in the processing of moral transgression events, and found that in the stranger condition, the component located in the posterior region was more activated. While in the case of mother condition, the other component located in the anterior region was more activated. This study provided key evidence for understanding the principle of embodied cognition. In addition, they also demonstrated that high-level moral disgust is built on more basic disgust via a mental construction approach through a process of embodied schemata. Using tDCS which allows cortical excitability to be directly manipulated, Zheng et al. investigated whether modulating the excitability of the bilateral DLPFC (or TPJ) can directly influence participants' moral judgments by affecting their cognitive reasoning or emotional processes. They observed that activating the right DLPFC as well as inhibiting the left DLPFC led to less utilitarian judgments especially in moral-personal conditions, indicating that the right DLPFC plays an crucial role in moral judgments. Their findings provide important information regarding the impact of tDCS on the DLPFC of healthy participants, especially with respect to moral-personal dilemmas.

Overall, we believe that the research presented in this topic can promote a better understanding of neural basis of prosocial behavior.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was supported by the National Nature Science Foundation of China (71773111) and the National Social Science Foundation of China (15ZDB134, 13AZD061).

REFERENCES

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Game. Econ. Behav*. 10, 122–142. doi: 10.1006/game.1995.1027

Camerer, C. F. (2003). Behavioural studies of strategic thinking in games. *Trends Cogn. Sci*. 7, 225–231. doi: 10.1016/S1364-6613(03)00094-9

de Quervain, D. J., Fischbacher, U., Treyer, V., and Schellhammer, M. (2004). The neural basis of altruistic punishment. *Science* 305:1254. doi: 10.1126/science. 1100735

Fehr, E., and Gächter, S. (2002). Altruistic punishment in humans. *Nature* 415, 137–140. doi: 10.1038/415137a

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ*. 3, 367–388. doi: 10.1016/0167-2681(82)90011-7

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., and Fehr, E. (2006).Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832. doi: 10.1126/science.1129156

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., et al. (2007). Neural correlates of trust. *Proc. Natl. Acad. Sci. U.S.A*. 104, 20084–20089. doi: 10.1073/pnas.0710103104

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., and Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron* 35, 395–405. doi: 10.1016/S0896-6273(02)00755-9

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755–1758. doi: 10.1126/science.1082976

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

REVIEW
published: 05 June 2018
doi: 10.3389/fpsyg.2018.00859

The Neural Basis of and a Common Neural Circuitry in Different Types of Pro-social Behavior

*Jun Luo**

Neuro & Behavior EconLab, School of Economics, Center for Economic Behavior and Decision-Making, Zhejiang University of Finance & Economics, Hangzhou, China

Pro-social behaviors are voluntary behaviors that benefit other people or society as a whole, such as charitable donations, cooperation, trust, altruistic punishment, and fairness. These behaviors have been widely described through non self-interest decision-making in behavioral experimental studies and are thought to be increased by social preference motives. Importantly, recent studies using a combination of neuroimaging and brain stimulation, designed to reveal the neural mechanisms of pro-social behaviors, have found that a wide range of brain areas, specifically the prefrontal cortex, anterior insula, anterior cingulate cortex, and amygdala, are correlated or causally related with pro-social behaviors. In this review, we summarize the research on the neural basis of various kinds of pro-social behaviors and describe a common shared neural circuitry of these pro-social behaviors. We introduce several general ways in which experimental economics and neuroscience can be combined to develop important contributions to understanding social decision-making and pro-social behaviors. Future research should attempt to explore the neural circuitry between the frontal lobes and deeper brain areas.

Keywords: pro-social behaviors, neural basis, neural circuitry, functional magnetic resonance imaging, transcranial direct current stimulation

OPEN ACCESS

Edited by:
Chao Liu,
Beijing Normal University, China

Reviewed by:
Karen Lisa Bales,
University of California, Davis, United States
Xiaochu Zhang,
University of Science and Technology of China, China
Xiaolan Yang,
Shanghai International Studies University, China

****Correspondence:***
Jun Luo
luojun@zufe.edu.cn

Specialty section:
This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Psychology

Received: *08 February 2018*
Accepted: *14 May 2018*
Published: *05 June 2018*

Citation:
Luo J (2018) The Neural Basis of and a Common Neural Circuitry in Different Types of Pro-social Behavior. Front. Psychol. 9:859. doi: 10.3389/fpsyg.2018.00859

INTRODUCTION

Humans are the most successful species at restraining their self-interest motives, even in interactions with unfamiliar strangers, through the development and enforcement of social norms (Fehr and Fischbacher, 2003; Boyd and Richerson, 2005). This human behavioral feature is thought to be a social adaptation that underlies our evolutionary success (Hrdy, 2009; de Waal, 2010). Pro-social behaviors, in particular, play a crucial role in social life across many cultures (Henrich et al., 2001). They represent a broad category of acts that are defined by significant regions of society as generally beneficial to other people or one's group (Penner et al., 2005). Pro-social behavior involves trade-offs between our own well-being and the well-being of others, including a donation to charity, reciprocal exchange, interpersonal trust, mutual cooperation, costly punishment of norm violations. Pro-social behaviors that are exhibited in game tasks have been found and replicated under controlled environments in many behavioral experiments; players like to share wealth with strangers (Eckel and Grossman, 1996; Fehr and Fischbacher, 2003), punish defectors at a cost (Fehr and Gächter, 2002; Fehr and Fischbacher, 2003; Dawes et al., 2007), invest money in a stranger (Berg et al., 1995; Kosfeld et al., 2005) and reject unfair divisions of a sum of money (Güth et al., 1982; Camerer, 2003).

In this paper, we review studies on the neural activity of going against pure self-interest behaviors. This evidence is based on neuroimaging and brain stimulation approaches that provide a micro-foundation of pro-social behaviors with regard to the underlying neural networks. These studies that involve social preferences are based on neuroscientific methods that include the neural networks and motivational forces involved in charitable donations, rejections to unfair divisions, punishments for non-cooperation behavior at a cost, or decisions to trust in an investment game. The combination of economic game models with modern neuroscientific methods enables researchers to investigate the neural mechanisms of pro-social behaviors and to advance theoretical models of how we make decisions in a social context.

There has been a gradual appearance of studies that reveal the mechanisms of action of social preferences on the brain's reward system, the role that affective factors play in economic decisions, and the neural model of the capacity to infer an actor's mental state during a strategic game. According to these neuroscientific findings, we thus propose an integrated model for a common shared neural circuitry for various kinds of pro-social behaviors, involving the theory of mind network, the reward system, emotion-related brain regions and prefrontal cortical areas. Indeed, this review would be a fruitful starting point for future studies on a model of the neural circuitry involved in pro-social behaviors, by describing the relationship between the behavioral patterns of social preferences and the empirically verified parameters of the brain model. This will bring about an improved model of social decisions and a better understanding of the nature of pro-social behaviors.

EMPATHY/CHARITABLE GIVING

We can empathize with others, that is, understand and share their emotions, feelings, motivations without any exogenous emotional stimulation. This crucial phenomenon of human social interactions occurs in various situations. Prior work from cognitive and behavioral psychology reveals the complex emotion process of empathy, including cognitive appraisal, cognitive perspective taking, and affect sharing (Decety and Jackson, 2004; Lamm et al., 2007; Olsson and Ochsner, 2008; Hein and Singer, 2008).

In accordance with these studies, advances in neuroscience enable us to gain new insights into the neural basis of empathy (Eisenberg, 2000; Hoffman, 2001; Preston and de Waal, 2002; de Vignemont and Singer, 2006; Batson, 2009). First, neuroscientific experiments about empathy indicate that the same neural circuits underlying both affective and cognitive processes are activated when we have a feeling and when others have this feeling. Preston and de Waal (2002) proposed a neuroscientific model of empathy, which specifically states that attended perception to another person in an emotional state automatically activates the participant's representation of that state and that activation of these representations are associated with autonomic and somatic responses.

Moreover, imaging studies have also investigated the brain activity of empathic responses in the field of touch, smell, and pain. Wicker et al. (2003) have performed a functional magnetic resonance imaging (fMRI) study that reveal the same brain regions are activated when observing a facial expression of disgust and when inhaling disgusting odorants. Keysers et al. (2004) have found that there are similar neural mechanisms involved when participants are touched and when they observe someone else being touched by objects. Another study has assessed the brain activity associated with empathy for pain (Singer et al., 2004, 2006). They indicated that activity in the anterior cingulate cortex (ACC) and anterior insula (AI) was observed when participants either felt pain or observed pain in someone else. These brain areas compose the affective pain circuits that represent our responses to pain and our understanding of how others feel pain. Further studies have investigated the temporal dynamics of the neural mechanisms underlying empathy for pain using event-related brain potentials (ERPs). These results showed that the early and late responses to empathy are separately adjusted by the situational reality of the stimuli, and these results support the hypothesis that empathy for pain consists of early emotional sharing and later cognition evaluations (Fan and Han, 2008).

In addition, responses in DMPFC regions while mentalizing with others who have similar and dissimilar thoughts and beliefs have also been shown to predict empathy (Zaki et al., 2009; Majdandžić et al., 2016). Neuroimaging studies have also examined the relations between activation in specific brain areas related to social preferences and self-reported empathy and willingness to help (Tankersley et al., 2007; Mathur et al., 2010; Powers et al., 2015), and found the correlations between the reflexive engagement of neural mechanisms of mentalizing and altruistic behaviors for monetary allocation and time spent helping others (Waytz et al., 2012). In fact, additional research has also demonstrated that the brain activation of brain areas involved in empathy predicts pro-social behaviors toward social exclusion (Masten et al., 2011) and that such activation occurs when participants make decisions to donate money to their family members (Telzer et al., 2011); thus, the neural basis of empathy during in tasks involving charitable donations has received much attention.

A prior attempt on the neural basis of giving showed that the mesolimbic reward system, including ventral tegmental (VTA) and striatal areas were both engaged by receiving money and by anonymous donations to charitable organizations, suggesting that giving has its own reward (Moll et al., 2006). Further study has clarified that there are different neural mechanisms for purely altruistic and warm-glow motives for charitable giving (Harbaugh et al., 2007). To test these two motives, researchers have assessed fMRI while participants played a dictator game in which participants were required to make decisions about whether to give money to a charitable organization. All the participants were randomly assigned to mandatory and voluntary conditions. In the mandatory condition, participants observed money being transferred tax-like to a charitable organization. In the voluntary condition, subjects could make transfers voluntarily to the charity. Similar neural substrates linked to reward processing were elicited while participants received money themselves, when they performed free transfers, and when they observed the charity receiving money. However, this neural activation was higher when charitable giving was voluntary rather than mandatory.

In another study, the motivational mechanisms of charitable giving were identified by multivariate decoding techniques (Tusche et al., 2016). Neural responses in the AI predicted affective empathy for beneficiaries, while temporoparietal junction (TPJ) activity was associated with the degree of cognitive perspective taking, suggesting that these distinct paths of social cognition and psychological mechanisms differentially lead to intraindividual and interindividual heterogeneities in charitable giving. Indeed, there was specific neural evidence of a correlation between individual differences in helpful decisions and the neural activation of AI, ACC, and TPJ (Greening et al., 2014), and neural mechanisms of individual differences in empathy and pro-social behaviors were further revealed by reinforcement learning theory (Lockwood et al., 2016). However, how affective empathy is linked to pro-social behaviors in charitable giving and the neural circuitry underlying empathy in terms of multi-faceted cognitive and emotional process remain poorly understand. Thus, one possible direction is to integrate various constructs of the neural mechanisms of empathy and provide connections between the neural responses to empathy and charitable giving in future studies.

FAIRNESS/INEQUITY AVERSION

People tend to helped those who helped them, and to hurt those who hurt them. Consequences that represent such preferences are called fairness equilibria (Rabin, 1993). This fairness effect has also been recognized in formal theory models of reciprocal fairness (Rabin, 1993) and inequity aversion (Fehr and Schmidt, 1999), both of which assume that there is a trade-off between fairness and individual benefits. To examine decisions about fairness, an ultimatum game (UG) has been proposed (Güth et al., 1982) involving strategic interaction behaviors. As the hypothesis of self-interest motivation, the responder in the UG should accept any non-zero offer from the other party. The proposer can expect this self-interest response, and then will give a smallest non-zero offer to responder.

However, a number of studies have found that offers are commonly around 50% of the sum amount no matter of the total monetary, and lower than 20% of the total offers have more than 50% probability of being rejected (Güth et al., 1982; Roth et al., 1991; Bolton and Zwick, 1995; Henrich et al., 2001). Strong evidence indicates that many subjects reject low offers from proposers in the UG (Henrich et al., 2001; Camerer, 2003). It is thus clear that the actual decisions in the game do not agree with the behaviors of the model predicted to be driven by self-interest motivation, and neuroscience research has begun to provide evidence for the mechanism underlying these decisions in an UG.

An fMRI study first investigated the neural basis of response decisions in an UG (Sanfey et al., 2003). They found that unfair proposals elicited neural activity in brain regions involved in both the processing of cognition (DLPFC) and emotion (bilateral AI); these areas showed greater activation with an unfair offer that was subsequently rejected, whereas a greater response was seen in the DLPFC when an unfair offer was accepted. Further, there was significantly stronger activity in the AI when a participant received an unfair offer from another human compared to the same offer from a computer partner. Finally, the unfair offer was also related to heightened activity in ACC, and may imply the conflict between cognitive and emotion process in the response decision-making for the unfair offer of UG. Thus, receiving unfair offers in an UG was weakly associated with increased activity in these brain areas (see Gabay et al., 2014; Feng et al., 2015 for meta-analyses). Indeed, activation of the AI region involved in emotional arousal and measured as an autonomic index of affective status, indicated that skin conductance responses were stronger for unfair offers and related to the rejection rate of unfair offers in an UG (van't Wout et al., 2006).

Compared to unfair offers in an UG, fair offers led to greater activation in the VMPFC region. Importantly, the choice to reject unfair transfers is associated with improved activity in the AI region (Tabibnia et al., 2008). The key role of the ventromedial prefrontal cortex (VMPFC) in response decisions involving fairness preferences of the UG is also supported by neural evidence (Koenigs and Tranel, 2007) that patients with brain injuries in the VMPFC reject unfair offers in the UG more frequently than healthy participants, implying that the cost of declining non-zero offers is of less concern in the response decisions of the UG when the VMPFC is damaged. An ERP study (Boksem and De Cremer, 2010) showed that medial frontal negativity amplitude was greater for unfair offers than fair offers. Moreover, this effect was shown to be the greatest for responders with high fairness concerns.

To distinguish the functions of different brain areas in response decision-making in an UG, Knoch et al. (2006) used repetitive transcranial magnetic stimulation (rTMS) to inhibit the activation of the right DLPFC (rDLPFC) when responders in an UG faced unfair offers and observed a reduction in responders' willingness to reject unfair offers from proposers, which suggests that participants are more unable to resist the temptation to accept unfair offers from partners. However, participants did not change their judgment for such offers to be unfair after receiving rTMS, which reveals that the rDLPFC is crucial in implementing fairness-related decisions. In terms of transfer decisions from a proposer, another rTMS study indicated that reducing the activity of the right lateral PFC (rLPFC) led to a significant decrease in transfers in the UG, but neither the expected rejection from responders nor the fairness judgments were changed by rTMS (Strang et al., 2014). To modulate the neural excitability (activate or reduce) of specific regions, Ruff et al. (2013) employed transcranial direct current stimulation (tDCS) to demonstrate whether fairness-related decisions in the UG rely causally on neural activation of the rLPFC region. This study revealed that anodal tDCS in rLPFC caused transfers improvement significantly while cathodal tDCS to the rLPFC decreased transfers in the UG compared to sham stimulation. Together, these results provide strong causal evidence for the rLPFC in the implementation of fairness preference.

A pervasive notion in social science is that people have a preference to reduce inequality gaps in wealth distribution (Fehr and Schmidt, 1999). Studies have thus used inequality aversion to represent a fairness motive. To explore the tendency for inequity aversion in distributive decisions, participants performed in a distribution task (similar to UG) while scanning fMRI (Hsu et al., 2008). The experimental results suggested that the putamen encodes efficiency, whereas the insula represents inequity, and the caudate/septal subgenual area responds to a trade-off in efficiency and inequity. Strikingly, the choice about inequitable allocation was related to greater insula activity.

Neural evidence for preference of inequality aversion in distributive decision was also revealed by Tricomi et al. (2010). They have employed fMRI to demonstrate the existence of inequality aversion preferences in the brain. Inequality was created in experiments by recruiting pairs of participants and giving one of them an endowment. The participant who received the endowment showed greater neural reward activation while providing transfers to "other" rather than "self," whereas the participants who did not receive endowment showed a significantly greater activation in reward areas while providing transfers to "self" rather than "other." These results suggest that people are rewarded for reductions in the wealth gap, and the neural mechanisms of reward are strongly related to both advantageous and disadvantageous inequality. Civai et al. (2012) were more concerned with the differential roles of the AI and MPFC in equality versus self-interest in distributive decisions, especially for disadvantageous unequal offers and consequent rejections. The researchers found that the AI region was active during unequal offers, whereas the activity of the MPFC was negatively associated with rejection decisions. When inequity and efficiency were in conflict, participants showed greater activity in a simplified prefrontal network, including the rDLPFC, VMPFC, and the connectivity between them, according to fMRI signals (Baumgartner et al., 2011). Individual differences in inequity aversion were predicted by the blood-oxygen-level dependent (BOLD) signals of the amygdala (AMYG) during a resource-sharing task involving inequitable distributions to one's self and others (Haruno and Frith, 2010).

Taken together, social interactions with inequitable outcomes are linked to neural systems, including the AI, AMYG and prefrontal cortex, that are associated with affective and emotional signaling that alter distribution decisions by modulating fairness perceptions. In addition, inequity may induce a punishment action; thus, the neural networks implicated in inequity aversion could lead to the decision to punish at a cost to the punisher. On the other hand, the preference for inequity aversion may reflect how the neural processes that conform to inequity detection are influenced and further processed through emotional circuitry. Converging evidence indeed suggests that decision tasks related to inequity normally activate brain regions involved in affective processing. Further studies are needed to determine how these signals transform the decision to punish.

COSTLY PUNISHMENT

Across cultures, human always engage in individual costs in readiness to punish violators (Henrich et al., 2001; Fehr and Fischbacher, 2004; Bernhard et al., 2006), who propose an unfair offer during monetary allocation or take a self-interest strategy during a social exchange (Fehr and Gächter, 2002; Egas and Riedl, 2008). Why would humans punish defectors of universally maintained rules while diminishing their personal benefits? The view from evolutionary economics (Boyd et al., 2003; Bowles, 2009) indicates that human behavior in costly punishment has profound evolutionary foundation, and promoting pro-social behaviors, such as reciprocity and cooperation (Nakamaru and Iwasa, 2006; Rand et al., 2010; Rand and Nowak, 2013; Peysakhovich et al., 2014). These suggest that sanction at the cost of personal gain evolved as a spontaneous mechanism rather than as an intended or deliberate pattern; people thus feel satisfaction when punishing norm defectors. It is obvious that costly punishment brings a huge array of discusses about its behavioral mechanism, and start to focus on neural basis of costly punishment in recent years to further explain why we have willingness to costly punish.

A Neuroimaging research (de Quervain et al., 2004) first provided essential insight into the neural networks that shape such costly punishment actions. They designed a context of economic exchange in which investors transferred endowments to agents, but agents did not send back money to investors. This action of non-reciprocity was observed by a third party. Subjects could choose to punish these violators, and symbolic and effective punishments were available. Symbolic punishments did not influence the material benefits of the violator, while effective punishments did decrease the violator's payoff. They used positron emission tomography (PET) to scan the third party's brains while they confronted with the defection and determined the sanction. The neuroimaging results suggested that punishing defections effectively instead of symbolically activated the dorsal striatum (DS) region, which plays an important role in the processing of reward. Furthermore, subjects with higher activity in the DS were ready to pay more costs to punish. These findings proved the hypothesis that humans may achieve satisfaction from the action of punishing violators, even when this punishment causes a monetary loss to themselves.

To further assess this satisfaction through punishing defectors, another neuroimaging study using fMRI scanned the brain reward regions of participants during two-person economic game involving costly punishments (Strobel et al., 2011). They found that, indeed, brain reward areas such as the nucleus accumbens (NAC) and DLPFC were activated by the action of punishment. In addition, this activation was similarly affected by genetic variation of dopamine turnover during both first player and third party punishments. Overall, these results suggest that the interactive network of cognition, affect and motivation form the driving force in costly punishments.

A recent study has also investigated the brain mediation mechanisms during costly punishments based on the BOLD responses in related brain areas (White et al., 2014). Subjects showed greater modulation of BOLD signals based on the level of costly punishment in regions of the reward neural network, for example, the AI cortex and caudate, whereas subjects showed negative modulation of BOLD signals as a level of costly punishment within posterior cingulate cortex (PCC) and VMPFC regions. Converging evidence seems to indicate a transform via the reward circuitry in mediating costly punishment. In addition, the neurobiological determinants have been found an influence in decisions of punishing costly (Crockett et al., 2013). Manipulating the serotonin system of participants during economic exchange game alters the possibility of punishment through modulating the activity of striatum, indicating that serotonin may create the sensitivity threshold for punishment processing.

Some brain stimulation studies provide a causal evidence of prefrontal cortex regions on decisions of costly punishment through changing the activity of prefrontal cortex (van't Wout et al., 2005; Knoch et al., 2006, 2007). Subjects have a lower propensity to punish unfair behavior at a personal cost when rLPFC activity is restrained compared with the sham condition (Knoch et al., 2007). Based on this result, it can be expected that distinctions in the brain functions of the prefrontal cortex could illustrate individual variations in the willingness to punish, that is, the higher the individual baseline level of rLPFC activity, the greater the punishment behavior performed by the individual. To demonstrate whether individual differences in the activity levels of the rLPFC region predict participants' willingness to provide costly punishments to other people, a neuroscience study measured participants' resting-state electroencephalography (EEG) activity (Knoch et al., 2010) before they executed punishments for unfair proposals. A positive relationship was found between resting alpha activity in the rLPFC and the likelihood of a costly punishment. It is well known that the bilateral LPFC was associated with implementing of self-control and cognition processes (Miller and Cohen, 2001; Knoch et al., 2006; Cohen and Lieberman, 2010).

Another brain stimulation study on sanctions (Buckholtz et al., 2015) combined rTMS with fMRI to verify the explicit role of the DLPFC in pro-social behaviors induced by blame and punishment. The participants reduced punishments for violation activities when their brain activity in the DLPFC was inhibited by rTMS, but these participants' blameworthiness ratings were not influenced. The researchers also used fMRI to observe punishment-selective DLPFC region recruitment. These results indicated that these two aspects of decisions are neurobiologically dissociable and confirm a selective causal effect of the DLPFC on punishment behavior. Thus, brain stimulation to related brain regions has a significant effect on norm compliance induced by social punishment threats, whereas stimulation left beliefs of what the norm regulated and subject expectations about social sanctions unaffected.

However, perhaps it is still unclear what the fundamental driving force for neural responses in decisions for costly punishments is. Du and Chang (2015) concluded that three main cognitive and affective functions occur in costly punishment contexts that might have a crucial effect on activating neural regions, such as cost-benefit calculations, inequity aversions and social reference frames. The previous studies show that these three cognitive and affective functions have different neural circuitries underlying the complicated decision process of costly punishments. Furthermore, these neural mechanisms, involving distinct cognitive and affective processes, are likely to interact with one another during the decision to punish at a cost, and such interactions may lead to individual deliberations on the execution of this decision. Therefore, how to differentiate the neural circuitries of these cognitive and affective functions during decision-making in costly punishment is a key issue that needs to be solved.

COOPERATION

Humans often cooperate with each other in society, even with irrelevant strangers and people they will never meet again. This behavioral feature in human is considered as a social adaptation, implying human success in evolutionary progress (Hrdy, 2009; de Waal, 2010). Some behavioral studies focused on what motivation promoted evolution of human cooperation, such as costly punishment, altruistic rewarding and strong reciprocity (Fehr and Gächter, 2002; Boyd et al., 2003; Bowles and Gintis, 2004). Cooperation behavior has also been illustrated extensively in the economic exchange game, for example the prisoner's dilemma game (PDG) (Sally, 1995). In the standard PDG, two players' payoffs depend on interaction of their decisions. The player can get the most payoff if she or he choose to defect and the partner choose to cooperate, while the least to the player happens if she or he choose to cooperate and the partner choose to defect. In addition, mutual cooperation takes a modest amount to each player, whereas mutual defection leads to a lesser payoff to the two players.

Neuroscience methods combined with the paradigms of game theory have examined the neural basis of cooperative behaviors. In two neuroimaging studies (Rilling et al., 2002, 2004), it was revealed that the ventral stratum was activated when playing in mutual cooperation with a partner in a game, as compared to playing with a computer partner. Rilling et al. (2002) first employed fMRI to scan subjects when they played with a paired partner in a repeated PDG to explore the neural substrates of cooperative behavior. They found that the activity of related brain regions, such as NAC, rostral ACC, orbitofrontal cortex (OFC) and the caudate nucleus, involved in reward processing were associated with cooperative behavior, and a crucial role of the striatum in mutual cooperative behavior was demonstrated. Participants' mutual cooperative behavior leads to a higher BOLD signal in the related neural network during a PDG but results in a lower BOLD signal in the same regions if the partner defects. In subsequent research (Rilling et al., 2004), the reward neural network was also activated during cooperation in a sequential PDG, and subjects showed higher anterior paracingulate cortex and posterior STS activity when playing with person rather than with a computer. Cooperation following the defection of a partner would be characterized as an action against one's anticipation of the reciprocity norm and, thus, increase activation of the left AMYG and the bilateral AI (Rilling et al., 2008).

In another experimental paradigm, pairs of subjects were required to perform the same estimation task and received a monetary reward for right answer (Fliessbach et al., 2007). A higher activation of the ventral striatum was linked with the amount of reward earned by the subject, while a lower activation of the same area was linked with the amount of reward paid to the partner. That is, when people are assessed and rewarded by an identical standard, the ventral striatum activity is more closely related to personal relative earnings than payments to the partner. This finding indicates the likelihood that the striatal involvement in rewarding processes seems to vary depend on whether a social exchange was considered to be competition or cooperation. Similarly, when participants were asked to play with a partner competitively or cooperatively during a board game, differential brain regions were activated in two distinct patterns of interaction. The results showed that cooperation caused higher activity in the medial orbitofrontal cortex (MOFC) and anterior frontal cortex (AFC) compared to competition (Decety et al., 2004; Babiloni et al., 2007). However, whether and how these cortical regions are linked to the striatal activity involved in cooperation are currently unknown.

To further elucidate the neural mechanisms of cooperation, King-Casas et al. (2008) recruited subjects suffering from borderline personality disorder (BPD) to play an iterative social interaction game with healthy subjects. Healthy subjects exhibited a linear correlation between AI activity and both the amount of monetary payoff received from the partner and the magnitude of money sent back to their partner. In contrast, subjects with BDP only showed a relationship between AI activity and the amount of money repaid to the partner, not the amount of money received from their partner. These results are evidence that individuals with BPD show impaired AI activity that leads to an inhibition of their ability to benefit from mutual cooperation. Thus, the insula and the ventral striatum track the social interaction decision of the partner of whether to reciprocate cooperation, representing an encoding of the reward processes for the satisfaction gained through mutual cooperation (Sanfey, 2007). In addition, a computational model of social value was provided to predict individual cooperative behavior, which indicated that people receive a signal of social value reward for mutual cooperation (Fareri et al., 2015). This signal of social value was strongly associated with greater activation of the ventral striatum and MPFC, which suggests that this signal predicts cooperative behavior in an iterative social exchange game.

In summary, the implications and motivations behind pro-social behaviors in economic games have been widely discussed, and scans of related brain areas when people play social interaction economic games with a partner could reveal individual differences in cooperative behavior. However, why are people willing to cooperate with the other people in a game? What are motives driving this behavior for all humans? Whereas some economic games have been used mainly to investigate cooperative behavioral consistency (Yamagishi et al., 2013; Peysakhovich et al., 2014), manipulating different economic games with the same subjects could also enable researchers to isolate within-subject motives in order to more accurately examine the nature of cooperative decisions (Brañas-Garza et al., 2014). Studies that have used this methodology have indicated that cooperative behavior is always multi-determined and can be assigned to completely different motives.

Prior work using imaging tools such as fMRI have allowed the identification of the neural networks involved in cooperative behavior. Nonetheless, these tools can provide only limited support to this ambitious purpose as they lack temporal resolution. In addition, they do not permit an on-line, real-life social exchange environment. However, social interaction is an essential part of cooperative behavior. Therefore, how our brains specifically exploit social cues and contexts when considering whether to cooperate remains unclear (Jahng et al., 2017). To account for the complexity of this event, the hyperscanning approach supports a high temporal resolution that allows the capture of simultaneous recordings of brain activity as a possible research direction.

TRUST/TRUSTWORTHINESS

It is well known that trust penetrates into many aspects of our life, including working relations, friendships, and family relations. Interpersonal trust is also a core element for deeply understanding economic among people and the loss of trust between exchange partners seriously hinders market exchange. Thus, there are many reasons for researchers to concern the decision of trust. To investigate the decision of trust in economic interaction, Berg et al. (1995) firstly constructed a trust game, in which a player (the investor) has to decide how many amounts of endowment to invest with the other player (the trustee), and then the trustee can choose whether to give back and how much money return to the investor. As the model hypothesis of rational and self-interested people, the trustee will never return money to the investor. The investor can expect this rational decision from the trustee, and should never invest any amounts of money with the trustee.

Despite the predictions of game theory, in fact, most of the investors are still quite willing to transfer considerable amounts of money to a partner, and the trustees often repay some amount of money to the investor. Extensive studies have also discussed the potential factors that induce both trusting and trustworthiness behaviors among people that are not consistent with the hypothesis of the Homo economicus (Cook and Cooper, 2003; Bohnet and Zeckhauser, 2004; Cox, 2004; Ashraf et al., 2006; Schechter, 2007). In laboratory experiments, the subjects robustly showed behavior of trust although with completely strangers, or even when reputation is absent (McCabe and Smith, 2000; King-Casas et al., 2005).

Based on the results of behavioral studies, neuroscientists have attempted to provide the neural basis of trusting behavior. Krueger et al. (2007) employed hyper fMRI to scan pairs of subjects while they were playing against each other in a trust game. According to the within-brain and between brains analyses, several lines of evidence from functional brain activity indicated that the differential activation of related neural systems involves two trust strategies. First, the paracingulate cortex region is linked with the building of a relation of trust by inferring the partner's intentions to predict the subsequent decision. Second, the more recently evolved brain areas could be distinctly involved in interactions with more primitive neural systems developing conditional and unconditional trust relations. Conditional trust decisions significantly activated the VTA region, associated with the estimation of expected rewards, while unconditional trust decisions activated the septal region, associated with social interaction behaviors.

Interestingly, the evidence from neuroendocrinology shows that humans can secrete two hormones in opposite ways that are linked to establish a subtle balance in adjustable trust behaviors. A hormone that promotes social trust is oxytocin (OT). There is evidence that the brain differentiates between interpersonal trust and risk-seeking, derived from a study where the synthetic neuropeptide OT was injected intranasally to subjects while playing a trust game (Kosfeld et al., 2005). The hypothesis was that the betrayal aversion reducing effect of OT might result in decreased activity in the AMYG, suggesting that OT reduces AMYG activity. AMYG function has been demonstrated to be involved in evaluating the trustworthiness of faces (Winston et al., 2002; Adolphs et al., 2005) and ambiguous incidents (Hsu et al., 2005), which both have relevance with decisions in a trust game. It should be noted that these effects of OT might not extend to all people, because a neuropathological study found that OT can inversely inhibit trust behavior in individuals with BPD (Bartz et al., 2010; Bos et al., 2010). These results demonstrate the necessity of taking into account personal heterogeneity while reporting the effects of hormones on individual behaviors (Bartz et al., 2011).

Similarly, in a testosterone administration and placebo-controlled experiment, Bos et al. (2012) used fMRI to provide insights into the neural mechanisms involved in the effect of testosterone on trusting behavior. They found that testosterone improved social vigilance to untrustworthy faces by affecting neuropeptide systems in the central AMYG region, enhancing the communication between the AMYG and brainstem areas. However, testosterone can also change the functional connectivity between the OFC and AMG while judging unfamiliar faces, which then induces an improvement in social vigilance by decreasing top–down control over the AMYG. Although speculative, a neurobiological interpretation based on these results is that testosterone leads to the continuous reduction, in an uncertain social interaction, of the connectivity between the OFC and AMG via a prefrontal-dopaminergic mechanism, which results in more vigilant AMYG responses to signals of untrustworthiness.

Other studies have specifically examined the neural basis of trustworthiness of trustees, and the mechanisms involved in decision factors such as risks, benefits, and reputation have been examined (Baumgartner et al., 2009; Knoch et al., 2009; van den Bos et al., 2009; Aimone et al., 2014). Specifically, researchers have paid attention to the correlation between altruism and trustworthiness in neuroscience studies. A clinical lesion example indicated that patients with injuries to the VMPFC region offer less in a dictator game and show less trustworthy behaviors in a trust game, suggesting that the VMPFC plays an indispensable role in both altruism and trustworthiness decisions (Krajbich et al., 2009; Moretto et al., 2013). There was evidence that the trustee playing the game showed activation in the VMPFC, the posterior cingulated cortex (PCC), the lateral OFC, and the right AMYG, and that the VMPFC response was linked with altruistic behavior (Li et al., 2009). Previous studies in neural cognition have also illustrated that the VMPFC is crucial for evaluating social information and that impairments in the VMPFC caused serious disruptions of emotion and resulted in impaired to decision making, behavior regulation and planning (Damasio, 1994; Anderson et al., 2006).

In addition to the VMPFC region, the roles that other brain areas play when realizing a partner's trustworthiness have also been tested in other studies. In one study, the caudate nucleus activity predicted whether the trustee showed trustworthiness for a partner in a trust game (King-Casas et al., 2005). In a second study, researchers used the same paradigm of a trust game to investigate specializations of the cingulate cortex in encoding trustworthy decisions in a social domain (Tomlin et al., 2006). A further study arranged investors in a trust game to sequentially face three trustees and provided profiles that made them seem morally positive, neutral or negative in order to instill a prior belief about trustworthiness (Delgado et al., 2005). The researchers found that the caudate nucleus activity in the investors was involved in the decision of whether the trustees were weakened when the investors depended on the trustees' information about moral character. Irrespective of the exact neural mechanisms, fMRI results indeed indicate that the AMYG is also associated with facial detections of trustworthiness (Winston et al., 2002; Engell et al., 2007; Todorov et al., 2008a,b; Said et al., 2009), and patients with damage to the bilateral AMYG show more facial evaluations of trustworthiness compared to healthy participants (Adolphs et al., 1998). The ACC and insula are also responsible to the processing of interpersonal trust and social threat (Rushworth et al., 2007).

Prior neuroimaging studies on trustworthiness have led to a well-founded discussion about the correlation between trustworthiness and empathy (Adolphs, 2002; Winston et al., 2002; Engell et al., 2007; Said et al., 2009). In particular, two meta-analyses studies have summarized the differential neural circuits linked to the process of identifying trustworthy and untrustworthy faces (Bzdok et al., 2011; Mende-Siedlecki et al., 2013). Notably, faces considered to be trustworthy primarily involve activity in reward-related brain areas, while faces considered to be untrustworthy primarily engage activation of the ventral AMYG, which quickly responds to a potential threat. A recent ERP study showed that faces perceived as trustworthy are implicitly evaluated even during an unrelated task, as when subjects must memorize characteristics, as seen by the modulation of neural activity linked with visual working memory that processes faces (Meconi et al., 2014).

However, these neuroimaging studies have failed to provide a direct causal effect between the activity in related brain areas and behavioral decisions. In contrast, recent tDCS studies, by affecting brain activity non-invasively, have established causal links between brain activity and trust or trustworthiness decisions. Colzato et al. (2015) used tDCS over the VMPFC region while participants played a trust game and did not found a correlation between VMPFC activity and trust behavior. Other tDCS studies showed that the modulation of activity in several brain areas, such as the OFC, DLPFC and VMPFC may change the subjects' trustworthy behavior (Nihonsugi et al., 2015; Wang et al., 2016; Zheng et al., 2016).

On the whole, these results provide insight into understanding how related brain areas work together when subjects exhibit reciprocal trusting by showing how these neural substrates are distinctly derived from reciprocated trust, betrayal aversion, risk preferences and perspective-taking motives. However, it is still not clear which mechanisms connect these neural substrates that underlie the different motivations in trust behavior, or which neural structures factors determine individual levels of trust behavior. Additionally, identifying the effect of social factors, such as social status, social information and peer influence, on trust behavior based on neural results is necessary in further studies.

In summary, we describe previous experimental studies that used neuroscience methods to explore the role of related brain areas in various pro-social behaviors (see **Table 1**).

NEURAL CIRCUITRIES OF PRO-SOCIAL BEHAVIORS

Based on previous neuroscience studies of different types of pro-social behaviors, it can be seen that there are some similar properties in the neural basis of these behaviors, which all activate related brain areas, including the theory of mind network, reward system, and prefrontal cortex. This implies there may be specific connections between the functions of these areas that lead to people often making decisions according to their other-regarding preferences. Our knowledge of this connection can help us better understand the neural mechanisms of pro-social behaviors. Thus, it is important to find the shared, common neural substrates that link these different types of pro-social behaviors. Here, an integrated model is proposed that depicts the neural circuits by which decision making is firstly primed in the theory of mind network while receiving input from other's information, then the social cognition signal activates reward system, and this action then activates the brain areas associated with emotion to reinforce the reward experience, and, finally, the decision is reflected upon

TABLE 1 | Summary of the study for the neural basis of pro-social behaviors.

Study	Technology	Experimental Task	Pro-social behaviors	Brain areas	Experimental design	Sample size
Rilling et al., 2002	fMRI	Prisoner's Dilemma Game	Cooperation	Ventral striatum	Between (human versus computer) and within (iterated game) subjects.	36
Sanfey et al., 2003	fMRI	Ultimatum game	Fairness	AI and DLPFC	30 rounds in all, 10 playing the game with a human, 10 with a computer, and a further 10 control rounds.	19
de Quervain et al., 2004	PET	Third-party punishment game	Costly punishment	Dorsal striatum	Participants experienced four different conditions.	14
Rilling et al., 2004	fMRI	UG and PDG	Fairness	aPCC and posterior STS.	Between (human versus computer) and within (iterated game) subjects.	19
Decety et al., 2004	fMRI	Computer game	Cooperation	OFC and MPFC	Between (alone, cooperation, or against) subjects.	12
Moll et al., 2006	fMRI	Charitable donation	Altruistic behavior	VTA and STR	Different payoff types were designed: (*i*) pure monetary reward, (*ii*) non-costly donation, and (*iii*) costly donation.	19
Knoch et al., 2006	rTMS	Ultimatum game	Fairness	DLPFC	Applied rTMS to the right or to the left DLPFC and a control group.	52
Knoch et al., 2010	EEG	Ultimatum game	Costly punishment	rPFC	Responder played with 12 different proposers.	20
Spitzer et al., 2007	fMRI	Dictator game with the sanction threat	Costly punishment	OFC and DLPFC	Control and punishment conditions	45
Krueger et al., 2007	hyperfMRI	Trust game	Trust	pACC and VTA	Sequential decisions for monetary payoffs (low, medium, or high).	44
Emonds et al., 2011	fMRI	PDG and coordination game	Reciprocity	DLPFC and STS	Between (proself or prosocial) and within (two games) subjects.	28
Bos et al., 2012	fMRI	Rate facial pictures	Trustworthiness	OFC and AMYG	A randomized, counterbalanced, placebo-controlled, testosterone administration paradigm.	16
Ruff et al., 2013	tDCS	Ultimatum game	Fairness	rLPFC	Randomly assigned to one of three groups: anodal, sham, or cathodal.	64
Aimone et al., 2014	fMRI	Trust game	Betrayal aversion	AI	Both within- and between-subject	30
Strang et al., 2014	TMS	Dictator game	Strategic fairness	DLPFC	Randomly assigned to one of three groups: anodal, sham, or cathodal.	17
Zheng et al., 2016	tDCS	Trust game	Trustworthiness	VMPFC	Randomly assigned to one of three groups: anodal, sham, or cathodal.	60

in the prefrontal cortex to execute a pro-social behavior (see **Figure 1**).

Pro-social behaviors surely requires theory of mind during cognition signal input, as it is the neural network underlying our ability to attribute other's mental states, providing us with a prediction of their intentions and actions. Theory of mind is likely to provide our other-regarding, which allows the ability to share others' thoughts and feelings and, therefore, motivates pro-social behaviors (de Waal, 2008). A number of studies in cognitive neuroscience have discovered the neural network for considering another person's thoughts, comprising the precuneus, bilateral TPJ and right superior temporal sulcus (RSTS) (Rilling et al., 2004; Saxe, 2006; Young et al., 2007; van Overwalle and Baetens, 2009; Ye et al., 2015). In particular, the TPJ shows increased activity when participants read about a person's beliefs in non-moral (Saxe and Powell, 2006) and moral (Young et al., 2007) contexts. Accordingly, it has also been suggested that the ACC might play an important role in representing the mental states of others (Gallagher et al., 2002; Gallagher and Frith, 2003; Rilling et al., 2004). This brain area is involved not only when mentalizing about the thoughts, intentions or beliefs of others but also when people are attending to their own states. Frith and Frith (2003) suggest that this area subserves the formation of decoupled representations of beliefs about the world. In addition, it has been shown that activity in this area is strongly associated with the level of an individual's pro-social behavior (Tankersley et al., 2007). These results address long-standing discussions about the sources of social decision making by indicating that pro-social behaviors might first derive from the theory of mind network, with its the proclivity toward social-cognitive thoughts of other's mental states.

Pro-social behaviors are always accompanied by the activation of the reward system, and these behaviors were marked and intensified after other-regarding thoughts of mental states. The reward system is a neural network responsible for incentive salience (i.e., craving, motivation, or desiring for a reward), related learning (mainly positive reinforcement for actions), and the activation of emotions, especially ones that involve pleasure as a central constituent (e.g., happiness, joy, and euphoria). Importantly, correlations between different kinds of pro-social behaviors and the reward system have been shown in many previous neuroscience studies. Moll et al. (2006) found that the reward system, such as VTA and striatum areas were both activated when participants give donations to the charity. A significant activation in the reward system in response to inequity aversion has also been found (Tricomi et al., 2010). Neuroimaging results suggest that punishment defection in an economic game activates the participants' brain reward regions, such as the NACs and thalamus (de Quervain et al., 2004; Strobel et al., 2011). It was also demonstrated that the activities of related brain regions, such as NACs and the striatum area involved in reward processing were strongly correlated with cooperation behavior (Rilling et al., 2002, 2004; Sanfey, 2007). In addition, trust behaviors are primarily linked to reward-related brain areas when participants identify trustworthy faces (Meconi et al., 2014). These findings provide strong evidence of the rewarding process of different types of pro-social behaviors, which in return demonstrates the crucial role the reward system plays in the shared neural substrates of pro-social behaviors.

Pro-social behaviors inevitably activate emotion-related regions in the brain when these decisions are rewarded, and these reward experiences must be stored in memory. The emotion-related system is a group of neural structures that are primarily involved in many of our feelings and motivations, including fear and disgust. The AI is thought to process convergent information to create a relevant context for the emotions involved in a sensory experience. Certain structures in this system are involved in memory processing as well. The AMYG is responsible for determining where memories are stored in the brain and which the memories are stored. It is believed that this determination is based on how great an emotional response the action invokes. The hippocampus sends out memories to related brain regions for long-term storage. In sum, this system supports various functions, such as interpreting emotional signals, regulating hormones, storing memories and processing motivation. Therefore, it is easy to infer that these neural structures are related with pro-social behaviors. A recent study has also found that the BOLD signals from the AI predicted affective empathy and helpful decisions (Greening et al., 2014; Tusche et al., 2016), and the BOLD signals from the AMYG region can be used to assess individual differences in fairness (Haruno and Frith, 2010). In fact, neural responses in the AI region associated with emotional arousal through the measurement of an autonomic index in affective status indicated that there was a strong correlation between AI activity and inequity aversion (van't Wout et al., 2006; Civai et al., 2012), and healthy participants exhibited a strong relationship between trusting behaviors and activity in this neural system including AI, AMYG, and PCC regions (Adolphs et al., 2005; Krueger et al., 2007; King-Casas et al., 2008; Watanabe et al., 2014). Bos et al. (2012) explored the neural mechanisms regarding the cause effect of testosterone on trusting behaviors via neuropeptide systems in the central AMYG region. In addition, it was shown that participants showed a significant modulation of neural responses within the PCC as a function of costly punishments and trusting behaviors (Li et al., 2009; White et al., 2014). Rilling et al. (2008) has also found that cooperative behaviors increase AMYG and AI activity, and subjects showed a higher anterior paracingulate cortex activity when playing cooperation strategies (Rilling et al., 2004). In conclusion, there is enough evidence to indicate that these neural structures, triggered by the need for arousing emotions and long-term memory play an important role in the production and reinforcement of pro-social behaviors.

Pro-social behavior truly needs to be controlled and planned as a whole while balancing various motivations, and the prefrontal cortex region is considered to be the control center of pro-social behavior and is the key part of the common shared neural substrates of different types of pro-social behaviors. The prefrontal cortex region is known to be associated with planning and modulating pro-social behaviors (Yang and Raine, 2009). In terms of psychology, the most typical functions carried out through the prefrontal cortex are executive functions (Shimamura, 2000). Executive functions involve the abilities to make decisions among different conflicting considerations

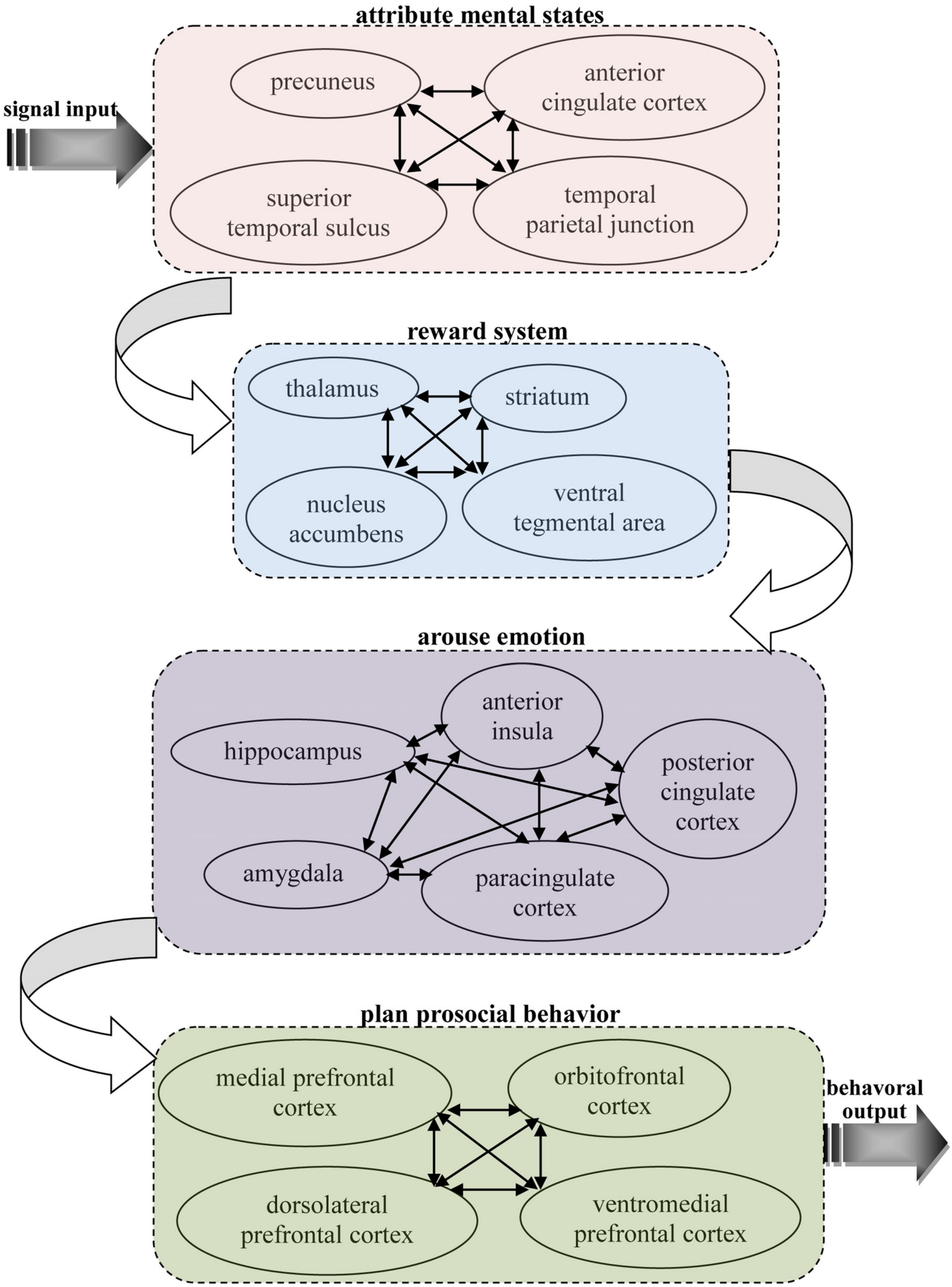

FIGURE 1 | A possible common neural circuitry associated with different types of pro-social behavior in the human brain.

(Goldberg, 2002); determine good or bad, self-interest or other-regarding, and same or different; create expectations according to events; create predictions of consequences and future outcomes of current actions; and enable social "control" (the capability to inhibit desires that, if not inhibited, could cause socially unacceptable consequences). The effect of this neural network on carrying out pro-social behaviors has been demonstrated in many neuroscience studies. For example, several studies have shown that individual differences in signals from the MPFC are correlated with empathy and altruistic behaviors (Zaki et al., 2009; Wagner et al., 2011; Powers et al., 2015). The indispensable roles of the VMPFC in the decision making involved in fairness, altruism and trustworthiness are also supported by previous neural evidence (Koenigs and Tranel, 2007; Krajbich et al., 2009; Moretto et al., 2013). A fMRI study (Spitzer et al., 2007) investigated the relationship between costly punishments and activity in the lateral OFC and rDLPFC and the causal effect of activity in the LPFC region on decisions of costly punishment by altering the activation of this region (van't Wout et al., 2005; Knoch et al., 2006, 2007). Interestingly, other brain stimulation studies have indicated that reducing the activation of the rLPFC leads to a significant change in fairness-related behaviors, but neither expected punishment from others nor fairness norms were altered (Ruff et al., 2013; Strang et al., 2014). These findings reveal that activity in the prefrontal cortex is a crucial biological prerequisite for an important aspect of evolutionary and social human behavior. These findings suggest that the prefrontal cortex is responsible for behavioral control although it is dissociated from the neural structures that enable people to anticipate social norms and attribute other's mental states. The structural connectivity and situation-dependent functions of prefrontal cortex (Duncan, 2010) make it possible to integrate and coordinate activation of the neural networks related to pro-social behavior during action control.

DISCUSSION

This review introduces several widely used methods that combined the game theory of economics with neuroscience technologies to develop new insights into understanding pro-social behaviors. These findings contribute important progress for measuring the neural mechanisms involved in pro-social behaviors and yield the guarantee of identifying and accurately characterizing both the mechanisms and the factors that influence interactions and engagements in pro-social behavior.

The economic game approach has several some advantages over typical paradigms of decision-making, not only for use within real, sequential, social interactions that make it possible to study complex exchange contexts such as fairness, trust, cooperation, and norm compliance. On the other hand, neuroscience methods of assessing pro-social behavior have obtained several prominent achievements in examining how the utility of parameters of behavior are represented in neural systems (Knutson et al., 2005; Sugrue et al., 2005; Padoa-Schioppa and Assad, 2006). Thus, neuroscience studies of pro-social behaviors could explore the neural circuitries of parameters that game models both expect (such as strategic payoffs) and do not expect (such as social preference). In addition, behavioral and neural data created through this method can confirm significance in offer special restraints, based on neural networks, for any theory that attempts to build precise models of pro-social behavior.

However, there are important challenges to address for any novel approach. These challenges are involved in disciplines that manipulate diverse analysis perspectives and have distinct theoretical assumptions. In particular, there are crucial differences in study methodologies, such as the use of deception, which is strictly prohibited in economics but widely used in neuroscience and psychology. Furthermore, it is important to be prudent when understanding brain activity measured by neuroimaging methods. For instance, the correlation of a brain area with either reward encoding or emotion processing in prior studies does not necessarily mean that this brain region's activity during at social interaction game can automatically be considered, respectively, as involved in reward or punishments (Sanfey, 2007). Therefore, one should be cautious in this field and support these conclusions by collecting evidence from other methodologies or, at a minimum, illustrating that behavioral results are consistent with the identified neural activities, for example, high levels of activation of the rewarding neural network being associated with an individual's preference for social decisions (de Quervain et al., 2004).

Although neuroimaging data cannot provide causal inferences, it is likely close to causality by predicting decision making during a treatment based on brain region activity during another treatment. For example, individual heterogeneity in the activity of the caudate nucleus when punishment is free predicts the level of willingness to punish when the punishing is costly (de Quervain et al., 2004). Similarly, individual differences in the activation of striatal regions when donations are mandatory are associated with participants' willingness to give money when this is a voluntary behavior (Harbaugh et al., 2007). These results further provide evidence of the rewarding process of pro-social behavior, which in return supports the hypothesis of shared neural circuitries of social reward and other primary and secondary reward (Montague and Berns, 2002).

Future studies in neuroscience should exploit the wide range of available tools, by using multiple measurement methods simultaneously (such as fMRI and rTMS, tDCS, or hormone measurement) along with the valuable behavioral parameters and theoretical predictions from complex game models. It is well known that non-invasive brain stimulation can establish causal correlations between related brain activities and individual behaviors by altering these neural processes and subsequent individual behaviorally expressed preferences. These methods have been used to not only support a biological basis for a mathematical characterizations in pro-social behaviors that are based on neural systems but also provide predictions of brain activity in social interactions and economic exchanges and how these behaviors can be transformed when manipulated by rTMS, tDCS, and other tools.

Nevertheless, brain stimulation technology is not the only approach to establish a causal inference between identified neural circuits and individuals' pro-social behaviors. Several pharmacological studies show great potential in this domain. For example, testosterone can increase the fairness transfers of a proposer (Eisenegger et al., 2010) and the probability that a responder rejects unfair offers (Burnham, 2007) in an UG; the neurohormone oxytocin improves behaviors of trust but not trustworthiness (Kosfeld et al., 2005); the depletion of the neurotransmitter serotonin increases the number of rejection decisions for unfair offers in an UG (Crockett et al., 2008; Crockett, 2009); and benzodiazepine can decrease the number of decisions to reject (Gospic et al., 2011). These studies of pharmacological interventions combined with fMRI allow observations of how neural systems causally influence behavioral changes (Baumgartner et al., 2008; Gospic et al., 2011).

In addition, several neuroscience studies have began to consider how the neural systems involved in pro-social behaviors are influenced by various factors. One factor is "social image," that is, how does knowing that other people are watching you influence your decisions and brain activity? This topic has been a concern of economists (Andreoni and Bernheim, 2009) and is important because social image could be changed by different organizational norms in information and institutions. An fMRI study indicated that the bilateral striatum was more highly activated when participants' charitable donations were observed than in the control treatment (Izuma et al., 2008), which supports the hypothesis that a social image rooted in charitable giving is rewarding. Consistent with a wide range of inequity aversion, one study was concerned with whether an awareness of high-status people suffering a failure would create a positive reward. Activity in the ventral striatum was found in response to these hypothetical contexts, and the BOLD signal predicted self-rated decisions (Takahashi et al., 2009). Emotions can also have an impact on pro-social behaviors. A neuroimaging study exploring this topic was based on real crime cases with "mitigating circumstances" (Yamada et al., 2012) and found that the activity in the insula, an identified neural correlate of empathy, was related to the level of sentence reduction.

However, our current understanding of neural mechanisms of pro-social behavior is still limited. This understanding will be improved if we obtain additional interpretations of the genetic and neurophysiological mechanisms of information processing in neural reward systems. Previous studies have revealed that social reward generally activate the ventral or DS, and there is a substantial overlap between the activity in these regions and the activity observed in studies about anticipated monetary reward or reinforcement learning (Fehr and Camerer, 2007; Fehr, 2009). This overlap is in accordance with the hypothesis that social preferences are similar to preferences for physical reward in terms of brain activity, which supports the theory about which decisions reflect a tradeoff between one's own benefits and the benefits of others.

More importantly, our brain has to compare social welfare and individual benefits and solve a conflict between these aspects when we exhibit pro-social behaviors. Previous studies have demonstrated that the prefrontal cortical regions that evolved lately (in evolutionary perspective) play crucial roles in this process of conflict resolution. For example, the VMPFC region is more activated when subjects can punish defectors at a personal cost than when punishment is free (de Quervain et al., 2004). Both the VMPFC and dorsal ACC regions show high activation levels when participants give charitable donations involving a cost (Moll et al., 2006). The ACC is known that have an important effect on conflict monitoring (Botvinick et al., 2001), so the activation of this area aligns with the presence of a conflict between pro-social motivations and self-interest incentives. In addition, the value of the response of the VMPFC is influenced by other responses of the posterior superior temporal cortex (PSTC) that have been shown to be crucial in overwhelming egocentricity prejudice, suggesting that the activity in both the VMPFC and the PSTC are important constituents of the neural network of pro-social behaviors.

In addition, the crucial role of DLFPC region in the processing of pro-social behavior has also been demonstrated (Sanfey et al., 2003). This study investigated the neural networks involved in the response decisions of an UG in which a rejection of unfair transfers indicates a balance between a self-interest motive and a fairness motive. Indeed, a function of the DLPFC may be enabling an individual to make choices for their long-term benefit for a good reputation in social interactions rather than the short-term benefit of the individual (van den Bos et al., 2009). The effect of the DLPFC on overwhelming short-term self-interest has also been investigated, and results show that the behavior of compliance with norms under the threat of punishment is positively related with the level of activity in the DLPFC (Spitzer et al., 2007).

More importantly, neuroscientists have indicated that we cannot consider brain regions as separate mini-brains but rather as widely interconnected regions. Notably, the frontal lobes are more linked to other brain regions than any other parts of the brain (Goldberg, 2002). In addition, the frontal lobes perform the most complex and developed functions of all parts of the brain, namely, the executive functions. This region is involved in complex, purposeful, and intentional decision making. However, previous studies also show that other brain areas, especially activation of brain–stem structures and the prefrontal cortex, are commonly associated with pro-social behaviors, such as the AI and the rostral ACC, which are activated when subjects empathize with other people experiencing pain (Singer et al., 2004); the ventral striatum, whose activity is strongly correlated with the amount of money given to charity (Harbaugh et al., 2007); the amygdale, whose BOLD signals can predict individual differences in aversion to inequity (Haruno and Frith, 2010); and both the DLPFC and caudate nucleus, whose activities are high when individuals receive unfair transfers from partners (Harbaugh et al., 2007). We also propose a common shared neural network of pro-social behaviors involving the theory of mind network, emotion-related regions, the reward system and the prefrontal cortex. In this neural circuitry, the control function of the prefrontal cortex plays a key role in coordinating human rationales, emotion, perception, cognition, motivation and reinforcement learning.

Among all species, humans are unique in terms of the ways in which they govern social life by executing pro-social behaviors. In addition to having the longer period during which brain development has been shaped by living environment, human beings change their environment, which shapes brains to an unprecedented extent among other species (Wexler, 2006). We thus suggest that the evolutionary consequence of promoting pro-social behaviors is the development of the prefrontal cortex, which plays a crucial role in the neural circuitry of social preferences. The prefrontal cortex connected with other brain regions when executing social cognition functions, but the region-to-region interaction mechanisms between the prefrontal cortex and deeper brain areas that represent pro-social behaviors are still not clear.

Finally, the overall target of this effort is the identification of a complete and general model of the neural network of decisions involved in pro-social preferences. There have been previous attempts from both "cognitive neuroscience in social decisions" and "neuroeconomics" to interpret the social brain and the related moral emotions (Adolphs, 2001, 2003; Greene et al., 2001; McCabe et al., 2001; Moll et al., 2002; Rilling et al., 2002; Sanfey et al., 2003; Singer et al., 2004). In addition, a paradigm with high sensitivity may indicate how affective and cognitive responses in the brain diverge or converge throughout decision making. Future studies combining fMRI with another method with a higher temporal resolution, such as EEG or functional near-infrared spectroscopy (fNIRS), may describe novel data on the region-to-region interactions between neural activities associated with cost-benefit calculations, social preferences, and the processing of information across self and others. Interestingly, finding a non-human primate model for pro-social behaviors can complement studies in humans by demonstrating specific neuronal circuitries in the core neural processes using single-unit recordings and pharmacological interventions in particular populations of neurons. Another important aspect of concern in the neuroscientific study of pro-social behavior is the social context in which pro-social behavior occurs. Understanding how the social context transforms the neural processes involved in cost-benefit calculations, social preferences, and the process of information between self and others will lead to better interpretations of the complex events behind human pro-social behaviors.

AUTHOR CONTRIBUTIONS

JL drew the table, wrote and revised the manuscript, and finally approved the version to be published.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 71703145) and MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Project No. 17YJCZH120).

REFERENCES

Adolphs, R. (2001). The neurobiology of social cognition. *Curr. Opin. Neurobiol* 11, 231–239. doi: 10.1016/S0959-4388(00)00202-6

Adolphs, R. (2002). Trust in the brain. *Nat. Neurosci.* 5, 192–193. doi: 10.1038/nn0302-192

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nat. Rev. Neurosci.* 4, 165–178. doi: 10.1038/nrn1056

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., and Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature* 433, 68–72. doi: 10.1038/nature03086

Adolphs, R., Tranel, D., and Damasio, A. R. (1998). The human amygdala in social judgment. *Nature* 393, 470–474. doi: 10.1038/30982

Aimone, J. A., Houser, D., and Weber, B. (2014). Neural signatures of betrayal aversion: an fMRI study of trust. *Proc. R. Soc. B.* 281:20132127. doi: 10.1098/rspb.2013.2127

Anderson, S. W., Barrash, J., Bechara, A., and Tranel, D. (2006). Impairments of emotion and real-world complex behavior following childhood-or adult-onset damage to ventromedial prefrontal cortex. *J. Int. Neuropsych. Soc.* 12, 224–235. doi: 10.1017/S1355617706060346

Andreoni, J., and Bernheim, B. D. (2009). Social image and the 50–50 norm: a theoretical and experimental analysis of audience effects. *Econometrica* 77, 1607–1636. doi: 10.3982/ECTA7384

Ashraf, N., Bohnet, I., and Piankov, N. (2006). Decomposing trust and trustworthiness. *Exp. Econ.* 9, 193–208. doi: 10.1007/s10683-006-9122-4

Babiloni, F., Astolfi, L., Cincotti, F., Mattia, D., Tocci, A., Tarantino, A., et al. (2007). "Engineering in medicine and biology society, EMBS 2007," *Proceedings of the 29th Annual International Conference of the IEEE*, Paris, 4953–4956.

Bartz, J., Simeon, D., Hamilton, H., Kim, S., Crystal, S., Braun, A., et al. (2010). Oxytocin can hinder trust and cooperation in borderline personality disorder. *Soc. Cogn. Affect. Neur.* 6, 556–563. doi: 10.1093/scan/nsq085

Bartz, J. A., Zaki, J., Bolger, N., and Ochsner, K. N. (2011). Social effects of oxytocin in humans: context and person matter. *Trends. Cogn. Sci.* 15, 301–309. doi: 10.1016/j.tics.2011.05.002

Batson, C. D. (2009). "These things called empathy: eight related but distinct phenomena," in *The Social Neuroscience of Empathy*, ed. J. Decety, and W. Ickes (Cambridge, MA: MIT Press).

Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K., and Fehr, E. (2009). The neural circuitry of a broken promise. *Neuron* 64, 756–770. doi: 10.1016/j.neuron.2009.11.017

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., and Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron* 58, 639–650. doi: 10.1016/j.neuron.2008.04.009

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., and Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nat. Neurosci.* 14, 1468–1474. doi: 10.1038/nn.2933

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Game. Econ. Behav.* 10, 122–142. doi: 10.1006/game.1995.1027

Bernhard, H., Fischbacher, U., and Fehr, E. (2006). Parochial altruism in humans. *Nature* 442, 912–915. doi: 10.1038/nature04981

Bohnet, I., and Zeckhauser, R. (2004). Trust, risk and betrayal. *J. Econ. Behav. Organ.* 55, 467–484. doi: 10.1016/j.jebo.2003.11.004

Boksem, M. A., and De Cremer, D. (2010). Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Soc. Neurosci.* 5, 118–128. doi: 10.1080/17470910903202666

Bolton, G. E., and Zwick, R. (1995). Anonymity versus punishment in ultimatum bargaining. *Game. Econ. Behav.* 10, 95–121. doi: 10.1006/game.1995.1026

Bos, P. A., Hermans, E. J., Ramsey, N. F., and Van Honk, J. (2012). The neural mechanisms by which testosterone acts on interpersonal trust. *Neuroimage.* 61, 730–737. doi: 10.1016/j.neuroimage.2012.04.002

Bos, P. A., Terburg, D., and Van Honk, J. (2010). Testosterone decreases trust in socially naive humans. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9991–9995. doi: 10.1073/pnas.0911700107

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., and Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol. Rev.* 108, 624–652. doi: 10.1037/0033-295X.108.3.624

Bowles, S. (2009). *Microeconomics: Behavior, Institutions, and Evolution*. Princeton, NJ: Princeton University Press.

Bowles, S., and Gintis, H. (2004). The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theor. Popul. Biol.* 65, 17–28. doi: 10.1016/j.tpb.2003.07.001

Boyd, R., Gintis, H., Bowles, S., and Richerson, P. (2003). The evolution of altruistic punishment. *Proc. Natl. Acad. Sci. U.S.A.* 100, 3531–3535. doi: 10.1073/pnas.0630443100

Boyd, R., and Richerson, P. J. (2005). *The Origin and Evolution of Cultures*. Oxford: Oxford University Press.

Brañas-Garza, P., Espín, A. M., Exadaktylos, F., and Herrmann, B. (2014). Fair and unfair punishers coexist in the Ultimatum Game. *Sci. Rep.* 4:6025. doi: 10.1038/srep06025

Buckholtz, J. W., Martin, J. W., Treadway, M. T., Jan, K., Zald, D. H., Jones, O., et al. (2015). From blame to punishment: disrupting prefrontal cortex activity reveals norm enforcement mechanisms. *Neuron* 87, 1369–1380. doi: 10.1016/j.neuron.2015.08.023

Burnham, T. C. (2007). High-testosterone men reject low ultimatum game offers. *Proc. R. Soc. B.* 274, 2327–2330. doi: 10.1098/rspb.2007.0546

Bzdok, D., Langner, R., Caspers, S., Kurth, F., Habel, U., Zilles, K., et al. (2011). ALE meta-analysis on facial judgments of trustworthiness and attractiveness. *Brain. Struct. Funct.* 215, 209–223. doi: 10.1007/s00429-010-0287-4

Camerer, C. F. (2003). Behavioural studies of strategic thinking in games. *Trends Cogn. Sci.* 7, 225–231. doi: 10.1016/S1364-6613(03)00094-9

Civai, C., Crescentini, C., Rustichini, A., and Rumiati, R. I. (2012). Equality versus self-interest in the brain: differential roles of anterior insula and medial prefrontal cortex. *Neuroimage* 62, 102–112. doi: 10.1016/j.neuroimage.2012.04.037

Cohen, J. R., and Lieberman, M. D. (2010). "The common neural basis of exerting self-control in multiple domains," in *Self Control in Society, Mind, and Brain*, Vol. 9, eds R. Hassin, K. Ochsner, and Y. Trope, 141–162.

Colzato, L. S., Sellaro, R., Van den Wildenberg, W. P. M., and Hommel, B. (2015). tDCS of medial prefrontal cortex does not enhance interpersonal trust. *J. Psychophysiol.* 29, 131–134. doi: 10.1027/0269-8803/a000144

Cook, K. S., and Cooper, R. M. (2003). "Experimental studies of cooperation, trust, and social exchange," in *Trust and Reciprocity: Interdisciplinary Lessons for Experimental Research* (New York, NY: Russell Sage), 209–244.

Cox, J. C. (2004). How to identify trust and reciprocity. *Game. Econ. Behav.* 46, 260–281. doi: 10.1016/S0899-8256(03)00119-2

Crockett, M. J. (2009). The neurochemistry of fairness. *Ann. Ny. Acad. Sci.* 1167, 76–86. doi: 10.1111/j.1749-6632.2009.04506.x

Crockett, M. J., Apergis-Schoute, A. M., Herrmann, B., Lieberman, M. D., Müller, U., Robbins, T. W., et al. (2013). Serotonin modulates striatal responses to fairness and retaliation in humans. *J. Neurosci.* 33, 3505–3513. doi: 10.1523/JNEUROSCI.2761-12.2013

Crockett, M. J., Clark, L., Tabibnia, G., Lieberman, M. D., and Robbins, T. W. (2008). Serotonin modulates behavioral reactions to unfairness. *Science* 320, 1739–1739. doi: 10.1126/science.1155577

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY: Quill.

Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R., and Smirnov, O. (2007). Egalitarian motives in humans. *Nature* 446, 794–796. doi: 10.1038/nature05651

de Quervain, D. J., Fischbacher, U., Treyer, V., and Schellhammer, M. (2004). The neural basis of altruistic punishment. *Science* 305:1254. doi: 10.1126/science.1100735

de Vignemont, F., and Singer, T. (2006). The empathic brain: how, when and why? *Trends. Cogn. Sci.* 10, 435–441.

de Waal, F. B. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annu. Rev. Psychol.* 59, 279–300. doi: 10.1146/annurev.psych.59.103006.093625

de Waal, F. B. (2010). *The Age of Empathy: Nature's Lessons for a Kinder Society*. New York, NY: Broadway Books.

Decety, J., and Jackson, P. L. (2004). The functional architecture of human empathy. *Behav. Cogn. Neurosci. Rev.* 3, 71–100. doi: 10.1177/1534582304267187

Decety, J., Jackson, P. L., Sommerville, J. A., Chaminade, T., and Meltzoff, A. N. (2004). The neural bases of cooperation and competition: an fMRI investigation. *Neuroimage* 23, 744–751. doi: 10.1016/j.neuroimage.2004.05.025

Delgado, M. R., Frank, R. H., and Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat. Neurosci.* 8, 1611–1618. doi: 10.1038/nn1575

Du, E., and Chang, S. W. (2015). Neural components of altruistic punishment. *Front. Neurosci.* 9:26. doi: 10.3389/fnins.2015.00026

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn. Sci.* 14, 172–179. doi: 10.1016/j.tics.2010.01.004

Eckel, C. C., and Grossman, P. J. (1996). Altruism in anonymous dictator games. *Game. Econ. Behav.* 16, 181–191. doi: 10.1006/game.1996.0081

Egas, M., and Riedl, A. (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proc. R. Soc. B.* 275, 871–878. doi: 10.1098/rspb.2007.1558

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annu. Rev. Psychol.* 51, 665–697. doi: 10.1146/annurev.psych.51.1.665

Eisenegger, C., Naef, M., Snozzi, R., Heinrichs, M., and Fehr, E. (2010). Prejudice and truth about the effect of testosterone on human bargaining behaviour. *Nature* 463, 356–359. doi: 10.1038/nature08711

Emonds, G., Declerck, C. H., Boone, C., Vandervliet, E. J., and Parizel, P. M. (2011). Comparing the neural basis of decision making in social dilemmas of people with different social value orientations, a fMRI study. *J. Neurosci. Psychol. Econ.* 4, 11–24. doi: 10.1037/a0020151

Engell, A. D., Haxby, J. V., and Todorov, A. (2007). Implicit trustworthiness decisions: automatic coding of face properties in the human amygdala. *J. Cognitive. Neurosci.* 19, 1508–1519. doi: 10.1162/jocn.2007.19.9.1508

Fan, Y., and Han, S. (2008). Temporal dynamic of neural mechanisms involved in empathy for pain: an event-related brain potential study. *Neuropsychologia* 46, 160–173. doi: 10.1016/j.neuropsychologia.2007.07.023

Fareri, D. S., Chang, L. J., and Delgado, M. R. (2015). Computational substrates of social value in interpersonal collaboration. *J. Neurosci.* 35, 8170–8180. doi: 10.1523/JNEUROSCI.4775-14.2015

Fehr, E. (2009). On the economics and biology of trust. *J. Eur. Econ. Assoc.* 7, 235–266. doi: 10.1162/JEEA.2009.7.2-3.235

Fehr, E., and Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends. Cogn. Sci.* 11, 419–427. doi: 10.1016/j.tics.2007.09.002

Fehr, E., and Fischbacher, U. (2003). The nature of human altruism. *Nature* 425, 785–791. doi: 10.1038/nature02043

Fehr, E., and Fischbacher, U. (2004). Third-party punishment and social norms. *Evol. Hum. Behav.* 25, 63–87. doi: 10.1016/S1090-5138(04)00005-4

Fehr, E., and Gächter, S. (2002). Altruistic punishment in humans. *Nature* 415, 137–140. doi: 10.1038/415137a

Fehr, E., and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Q. J. Econ.* 114, 817–868. doi: 10.1098/rspb.2015.0392

Feng, C., Luo, Y. J., and Krueger, F. (2015). Neural signatures of fairness-related normative decision making in the ultimatum game: a coordinate-based meta-analysis. *Hum. Brain. Mapp.* 36, 591–602. doi: 10.1002/hbm.22649

Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C. E., et al. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. *Science* 318, 1305–1308. doi: 10.1126/science.1145876

Frith, U., and Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philos. T. R. Soc. B.* 358, 459–473. doi: 10.1098/rstb.2002.1218

Gabay, A. S., Radua, J., Kempton, M. J., and Mehta, M. A. (2014). The Ultimatum Game and the brain: a meta-analysis of neuroimaging studies. *Neurosc. Biobehav. R.* 47, 549–558. doi: 10.1016/j.neubiorev.2014.10.014

Gallagher, H. L., and Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends. Cogn. Sci.* 7, 77–83. doi: 10.1016/S1364-6613(02)00025-6

Gallagher, H. L., Jack, A. I., Roepstorff, A., and Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *Neuroimage.* 16, 814–821. doi: 10.1006/nimg.2002.1117

Goldberg, E. (2002). *The Executive Brain: Frontal Lobes and the Civilized Mind*. New York, NY: Oxford University Press.

Gospic, K., Mohlin, E., Fransson, P., Petrovic, P., Johannesson, M., and Ingvar, M. (2011). Limbic justice—amygdala involvement in immediate rejection in the ultimatum game. *PLoS Biol.* 9:e1001054. doi: 10.1371/journal.pbio.1001054

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., and Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science* 293, 2105–2108. doi: 10.1126/science.1062872

Greening, S., Norton, L., Virani, K., Ty, A., Mitchell, D., and Finger, E. (2014). Individual differences in the anterior insula are associated with the likelihood of financially helping versus harming others. *Cogn. Affect. Behav. Neurosci.* 14, 266–277. doi: 10.3758/s13415-013-0213-3

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* 3, 367–388. doi: 10.1016/0167-2681(82)90011-7

Harbaugh, W. T., Mayr, U., and Burghart, D. R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* 316, 1622–1625. doi: 10.1126/science.1140738

Haruno, M., and Frith, C. D. (2010). Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nat. Neurosci.* 13, 160–161. doi: 10.1038/nn.2468

Hein, G., and Singer, T. (2008). I feel how you feel but not always: the empathic brain and its modulation. *Curr. Opin. Neurobiol.* 18, 153–158. doi: 10.1016/j.conb.2008.07.012

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., et al. (2001). In search of homo economicus: behavioral experiments in 15 small-scale societies. *Am. Econ. Rev.* 91, 73–78. doi: 10.1257/aer.91.2.73

Hoffman, M. L. (2001). *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge, MA: Cambridge University Press.

Hrdy, S. B. (2009). *The Woman That Never Evolved: With a New Preface and Bibliographical Updates*. Cambridge, MA: Harvard University Press.

Hsu, M., Anen, C., and Quartz, S. R. (2008). The right and the good: distributive justice and neural encoding of equity and efficiency. *Science* 320, 1092–1095. doi: 10.1126/science.1153651

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., and Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310, 1680–1683. doi: 10.1126/science.1115327

Izuma, K., Saito, D. N., and Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron* 58, 284–294. doi: 10.1016/j.neuron.2008.03.020

Jahng, J., Kralik, J. D., Hwang, D. U., and Jeong, J. (2017). Neural dynamics of two players when using nonverbal cues to gauge intentions to cooperate during the Prisoner's Dilemma Game. *Neuroimage.* 157, 263–274. doi: 10.1016/j.neuroimage.2017.06.024

Keysers, C., Wicker, B., Gazzola, V., Anton, J. L., Fogassi, L., and Gallese, V. (2004). A touching sight: SII/PV activation during the observation and experience of touch. *Neuron* 42, 335–346. doi: 10.1016/S0896-6273(04)00156-4

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., and Montague, P. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science* 321, 806–810. doi: 10.1126/science.1156902

King-Casas, B., Tomlin, D., Anen, C., Camerer, C., Quartz, S. R., and Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science* 308, 78–83. doi: 10.1126/science.1108062

Knoch, D., Gianotti, L. R., Baumgartner, T., and Fehr, E. (2010). A neural marker of costly punishment behavior. *Psychol. Sci.* 21, 337–342. doi: 10.1177/0956797609360750

Knoch, D., Nitsche, M. A., Fischbacher, U., Eisenegger, C., Pascual-Leone, A., and Fehr, E. (2007). Studying the neurobiology of social interaction with transcranial direct current stimulation—the example of punishing unfairness. *Cereb. Cortex.* 18, 1987–1990. doi: 10.1093/cercor/bhm237

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., and Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832. doi: 10.1126/science.1129156

Knoch, D., Schneider, F., Schunk, D., Hohmann, M., and Fehr, E. (2009). Disrupting the prefrontal cortex diminishes the human ability to build a good reputation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20895–20899. doi: 10.1073/pnas.0911619106

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., and Glover, G. (2005). Distributed neural representation of expected value. *J. Neurosci.* 25, 4806–4812. doi: 10.1523/JNEUROSCI.0642-05.2005

Koenigs, M., and Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: evidence from the Ultimatum Game. *J. Neurosci.* 27, 951–956. doi: 10.1523/JNEUROSCI.4606-06.2007

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., and Fehr, E. (2005). Oxytocin increases trust in humans. *Nature* 435, 673–676. doi: 10.1038/nature03701

Krajbich, I., Adolphs, R., Tranel, D., Denburg, N. L., and Camerer, C. F. (2009). Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *J. Neurosci.* 29, 2188–2192. doi: 10.1523/JNEUROSCI.5086-08.2009

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., et al. (2007). Neural correlates of trust. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20084–20089. doi: 10.1073/pnas.0710103104

Lamm, C., Batson, C. D., and Decety, J. (2007). The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *J. Cognitive. Neurosci.* 19, 42–58. doi: 10.1162/jocn.2007.19.1.42

Li, J., Xiao, E., Houser, D., and Montague, P. R. (2009). Neural responses to sanction threats in two-party economic exchange. *Proc. Natl. Acad. Sci. U.S.A.* 106, 16835–16840. doi: 10.1073/pnas.0908855106

Lockwood, P. L., Apps, M. A., Valton, V., Viding, E., and Roiser, J. P. (2016). Neurocomputational mechanisms of prosocial learning and links to empathy. *Proc. Natl. Acad. Sci. U.S.A.* 113, 9763–9768. doi: 10.1073/pnas.1603198113

Majdandžić, J., Amashaufer, S., Hummer, A., Windischberger, C., and Lamm, C. (2016). The selfless mind: How prefrontal involvement in mentalizing with similar and dissimilar others shapes empathy and prosocial behavior. *Cognition* 157, 24–38. doi: 10.1016/j.cognition.2016.08.003

Masten, C. L., Morelli, S. A., and Eisenberger, N. I. (2011). An fMRI investigation of empathy for 'social pain'and subsequent prosocial behavior. *Neuroimage* 55, 381–388. doi: 10.1016/j.neuroimage.2010.11.060

Mathur, V. A., Harada, T., Lipke, T., and Chiao, J. Y. (2010). Neural basis of extraordinary empathy and altruistic motivation. *Neuroimage* 51, 1468–1475. doi: 10.1016/j.neuroimage.2010.03.025

McCabe, K., Houser, D., Ryan, L., Smith, V., and Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11832–11835. doi: 10.1073/pnas.211415698

McCabe, K. A., and Smith, V. L. (2000). A comparison of naive and sophisticated subject behavior with game theoretic predictions. *Proc. Natl. Acad. Sci. U.S.A.* 97, 3777–3781. doi: 10.1073/pnas.97.7.3777

Meconi, F., Luria, R., and Sessa, P. (2014). Individual differences in anxiety predict neural measures of visual working memory for untrustworthy faces. *Soc. Cogn. Affect. Neur.* 9, 1872–1879. doi: 10.1093/scan/nst189

Mende-Siedlecki, P., Verosky, S. C., Turk-Browne, N. B., and Todorov, A. (2013). Robust selectivity for faces in the human amygdala in the absence of expressions. *J. Cognitive. Neurosci.* 25, 2086–2106. doi: 10.1162/jocn_a_00469

Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167

Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourão-Miranda, J., Andreiuolo, P. A., et al. (2002). The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *J. Neurosci.* 22, 2730–2736. doi: 10.1523/JNEUROSCI.22-07-02730.2002

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., and Grafman, J. (2006). Human fronto–mesolimbic networks guide decisions about charitable donation. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15623–15628. doi: 10.1073/pnas.0604475103

Montague, P. R., and Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron* 36, 265–284. doi: 10.1016/S0896-6273(02)00974-1

Moretto, G., Sellitto, M., and di Pellegrino, G. (2013). Investment and repayment in a trust game after ventromedial prefrontal damage. *Front. Hum. Neurosci.* 7:593. doi: 10.3389/fnhum.2013.00593

Nakamaru, M., and Iwasa, Y. (2006). The coevolution of altruism and punishment: role of the selfish punisher. *J. Theor. Biol.* 240, 475–488. doi: 10.1016/j.jtbi.2005.10.011

Nihonsugi, T., Ihara, A., and Haruno, M. (2015). Selective increase of intention-based economic decisions by noninvasive brain stimulation to the dorsolateral prefrontal cortex. *J. Neurosci.* 35, 3412–3419. doi: 10.1523/JNEUROSCI.3885-14.2015

Olsson, A., and Ochsner, K. N. (2008). The role of social cognition in emotion. *Trends Cogn. Sci.* 12, 65–71. doi: 10.1016/j.tics.2007.11.010

Padoa-Schioppa, C., and Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature* 441, 223–226. doi: 10.1038/nature04676

Penner, L. A., Dovidio, J. F., Piliavin, J. A., and Schroeder, D. A. (2005). Prosocial behavior: multilevel perspectives. *Annu. Rev. Psychol.* 56, 365–392. doi: 10.1146/annurev.psych.56.091103.070141

Peysakhovich, A., Nowak, M. A., and Rand, D. G. (2014). Humans display a 'cooperative phenotype' that is domain general and temporally stable. *Nat. Commun.* 5:4939. doi: 10.1038/ncomms5939

Powers, K. E., Chavez, R. S., and Heatherton, T. F. (2015). Individual differences in response of dorsomedial prefrontal cortex predict daily social behavior. *Soc. Cogn. Affect. Neur.* 11, 121–126. doi: 10.1093/scan/nsv096

Preston, S. D., and de Waal, F. B. (2002). Empathy: its ultimate and proximate bases. *Behav. Brain. Sci.* 25, 1–20.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *Am. Econ. Rev.* 83, 1281–1302.

Rand, D. G., Armao, J. J. I. V., Nakamaru, M., and Ohtsuki, H. (2010). Anti-social punishment can prevent the co-evolution of punishment and cooperation. *J. Theor. Biol.* 265, 624–632. doi: 10.1016/j.jtbi.2010.06.010

Rand, D. G., and Nowak, M. A. (2013). Human cooperation. *Trends. Cogn. Sci.* 17, 413–425. doi: 10.1016/j.tics.2013.06.003

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., and Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron* 35, 395–405. doi: 10.1016/S0896-6273(02)00755-9

Rilling, J. K., King-Casas, B., and Sanfey, A. G. (2008). The neurobiology of social decision-making. *Curr. Opin. Neurobiol.* 18, 159–165. doi: 10.1016/j.conb.2008.06.003

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *Neuroimage* 22, 1694–1703. doi: 10.1016/j.neuroimage.2004.04.015

Roth, A., Prasnikar, V., Okuno-Fujiwara, M., and Zamir, S. (1991). Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: an experimental study. *Am. Econ. Rev.* 5, 1068–1095.

Ruff, C. C., Ugazio, G., and Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science* 342, 482–484. doi: 10.1126/science.1241399

Rushworth, M. F., Behrens, T. E., Rudebeck, P. H., and Walton, M. E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. *Trend. Cogn. Sci.* 11, 168–176. doi: 10.1016/j.tics.2007.01.004

Said, C. P., Baron, S. G., and Todorov, A. (2009). Nonlinear amygdala response to face trustworthiness: contributions of high and low spatial frequency information. *J. Cognitive. Neurosci.* 21, 519–528. doi: 10.1162/jocn.2009.21041

Sally, D. (1995). Conversation and cooperation in social dilemmas: a meta-analysis of experiments from 1958 to 1992. *Ration. Soc.* 7, 58–92. doi: 10.1177/1043463195007001004

Sanfey, A. G. (2007). Social decision-making: insights from game theory and neuroscience. *Science* 318, 598–602. doi: 10.1126/science.1142996

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755–1758. doi: 10.1126/science.1082976

Saxe, R. (2006). Why and how to study theory of mind with fMRI. *Brain. Res.* 1079, 57–65. doi: 10.1016/j.brainres.2006.01.001

Saxe, R., and Powell, L. J. (2006). It's the thought that counts: specific brain regions for one component of theory of mind. *Psychol. Sci.* 17, 692–699. doi: 10.1111/j.1467-9280.2006.01768.x

Schechter, L. (2007). Traditional trust measurement and the risk confound: an experiment in rural Paraguay. *J. Econ. Behav. Organ.* 62, 272–292. doi: 10.1016/j.jebo.2005.03.006

Shimamura, A. P. (2000). The role of the prefrontal cortex in dynamic filtering. *Psychobiology* 28, 207–218.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 1157–1162. doi: 10.1126/science.1093535

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature* 439, 466–469. doi: 10.1038/nature04271

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., and Fehr, E. (2007). The neural signature of social norm compliance. *Neuron* 56, 185–196. doi: 10.1016/j.neuron.2007.09.011

Strang, S., Gross, J., Schuhmann, T., Riedl, A., Weber, B., and Sack, A. T. (2014). Be nice if you have to—the neurobiological roots of strategic fairness. *Soc. Cogn. Affect. Neur.* 10, 790–796. doi: 10.1093/scan/nsu114

Strobel, A., Zimmermann, J., Schmitz, A., Reuter, M., Lis, S., Windmann, S., et al. (2011). Beyond revenge: neural and genetic bases of altruistic punishment. *Neuroimage* 54, 671–680. doi: 10.1016/j.neuroimage.2010.07.051

Sugrue, L. P., Corrado, G. S., and Newsome, W. T. (2005). Choosing the greater of two goods: neural currencies for valuation and decision making. *Nat. Rev. Neurosci.* 6, 363–375. doi: 10.1038/nrn1666

Tabibnia, G., Satpute, A. B., and Lieberman, M. D. (2008). The sunny side of fairness: preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychol. Sci.* 19, 339–347. doi: 10.1111/j.1467-9280.2008.02091.x

Takahashi, H., Kato, M., Matsuura, M., Mobbs, D., Suhara, T., and Okubo, Y. (2009). When your gain is my pain and your pain is my gain: neural correlates of envy and schadenfreude. *Science* 323, 937–939. doi: 10.1126/science.1165604

Tankersley, D., Stowe, C. J., and Huettel, S. A. (2007). Altruism is associated with an increased neural response to agency. *Nat. Neurosci.* 10, 150–151. doi: 10.1038/nn1833

Telzer, E. H., Masten, C. L., Berkman, E. T., Lieberman, M. D., and Fuligni, A. J. (2011). Neural regions associated with self control and mentalizing are recruited during prosocial behaviors towards the family. *Neuroimage* 58, 242–249. doi: 10.1016/j.neuroimage.2011.06.013

Todorov, A., Baron, S. G., and Oosterhof, N. N. (2008a). Evaluating face trustworthiness: a model based approach. *Soc. Cogn. Affect. Neur.* 3, 119–127. doi: 10.1093/scan/nsn009

Todorov, A., Said, C. P., Engell, A. D., and Oosterhof, N. N. (2008b). Understanding evaluation of faces on social dimensions. *Trend. Cogn. Sci.* 12, 455–460. doi: 10.1016/j.tics.2008.10.001

Tomlin, D., Kayali, M. A., King-Casas, B., Anen, C., Camerer, C. F., Quartz, S. R., et al. (2006). Agent-specific responses in the cingulate cortex during economic exchanges. *Science* 312, 1047–1050. doi: 10.1126/science.1125596

Tricomi, E., Rangel, A., Camerer, C. F., and O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature* 463, 1089–1091. doi: 10.1038/nature08785

Tusche, A., Böckler, A., Kanske, P., Trautwein, F. M., and Singer, T. (2016). Decoding the charitable brain: empathy, perspective taking, and attention shifts differentially predict altruistic giving. *J. Neur.* 36, 4719–4732. doi: 10.1523/JNEUROSCI.3392-15.2016

van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A., and Crone, E. A. (2009). What motivates repayment? Neural correlates of reciprocity in the Trust Game. *Soc. Cogn. Affect. Neur.* 4, 294–304. doi: 10.1093/scan/nsp009

van Overwalle, F., and Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage* 48, 564–584. doi: 10.1016/j.neuroimage.2009.06.009

van't Wout, M., Kahn, R. S., Sanfey, A. G., and Aleman, A. (2005). Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making. *Neuroreport* 16, 1849–1852. doi: 10.1097/01.wnr.0000183907.08149.14

van't Wout, M., Kahn, R. S., Sanfey, A. G., and Aleman, A. (2006). Affective state and decision-making in the ultimatum game. *Exp. Brain. Res.* 169, 564–568. doi: 10.1007/s00221-006-0346-5

Wagner, D. D., Kelley, W. M., and Heatherton, T. F. (2011). Individual differences in the spontaneous recruitment of brain regions supporting mental state understanding when viewing natural social scenes. *Cereb. Cortex.* 21, 2788–2796. doi: 10.1093/cercor/bhr074

Wang, G., Li, J., Yin, X., Li, S., and Wei, M. (2016). Modulating activity in the orbitofrontal cortex changes trustees' cooperation: a transcranial direct current stimulation study. *Behav. Brain. Res.* 303, 71–75. doi: 10.1016/j.bbr.2016.01.047

Watanabe, T., Takezawa, M., Nakawake, Y., Kunimatsu, A., Yamasue, H., Nakamura, M., et al. (2014). Two distinct neural mechanisms underlying indirect reciprocity. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3990–3995. doi: 10.1073/pnas.1318570111

Waytz, A., Zaki, J., and Mitchell, J. P. (2012). Response of dorsomedial prefrontal cortex predicts altruistic behavior. *J. Neurosci.* 32, 7646–7650. doi: 10.1523/JNEUROSCI.6193-11.2012

Wexler, B. E. (2006). *Brain and Culture: Neurobiology, Ideology, and Social Change*. Cambridge, MA: MIT Press.

White, S. F., Brislin, S. J., Sinclair, S., and Blair, J. R. (2014). Punishing unfairness: rewarding or the organization of a reactively aggressive response? *Hum. Brain. Mapp.* 35, 2137–2147. doi: 10.1002/hbm.22316

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., and Rizzolatti, G. (2003). Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron* 40, 655–664. doi: 10.1016/S0896-6273(03)00679-2

Winston, J. S., Strange, B. A., O'Doherty, J., and Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nat. Neurosci.* 5, 277–283. doi: 10.1038/nn816

Yamada, M., Camerer, C. F., Fujie, S., Kato, M., Matsuda, T., Takano, H., et al. (2012). Neural circuits in the brain that are activated when mitigating criminal sentences. *Nat. Commun.* 3:759. doi: 10.1038/ncomms1757

Yamagishi, T., Mifune, N., Li, Y., Shinada, M., Hashimoto, H., Horita, Y., et al. (2013). Is behavioral pro-sociality game-specific? Pro-social preference and expectations of pro-sociality. *Organ. Behav. Hum. Decis. Process.* 120, 260–271. doi: 10.1016/j.obhdp.2012.06.002

Yang, Y., and Raine, A. (2009). Prefrontal structural and functional brain imaging findings in antisocial, violent, and psychopathic individuals: a meta-analysis. *Psychiat. Res-Neuroim.* 174, 81–88. doi: 10.1016/j.pscychresns.2009.03.012

Ye, H., Chen, S., Huang, D., Zheng, H., Jia, Y., and Luo, J. (2015). Modulation of neural activity in the temporoparietal junction with transcranial direct current stimulation changes the role of beliefs in moral judgment. *Front. Hum. Neurosci.* 9:659. doi: 10.3389/fnhum.2015.00659

Young, L., Cushman, F., Hauser, M., and Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proc. Natl. Acad. Sci. U. S. A.* 104, 8235–8240. doi: 10.1073/pnas.070140 8104

Zaki, J., Weber, J., Bolger, N., and Ochsner, K. (2009). The neural bases of empathic accuracy. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11382–11387. doi: 10.1073/pnas.0902666106

Zheng, H., Huang, D., Chen, S., Wang, S., Guo, W., Luo, J., et al. (2016). Modulating the activity of ventromedial prefrontal cortex by anodal tDCS enhances the trustee's repayment through altruism. *Front. Psychol.* 7:1437. doi: 10.3389/fpsyg.2016.01437

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author alone is responsible for the content and writing of this paper.

frontiers in Psychology

REVIEW
published: 19 September 2017
doi: 10.3389/fpsyg.2017.01592

The Influence of Emotion on Fairness-Related Decision Making: A Critical Review of Theories and Evidence

Ya Zheng[1], *Zhong Yang*[2,3]*, *Chunlan Jin*[4], *Yue Qi*[2,3] *and Xun Liu*[2,3]*

[1] Department of Psychology, Dalian Medical University, Dalian, China, [2] CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China, [3] Department of Psychology, University of Chinese Academy of Sciences, Beijing, China, [4] School of Foreign Languages, East China University of Science and Technology, Shanghai, China

Fairness-related decision making is an important issue in the field of decision making. Traditional theories emphasize the roles of inequity aversion and reciprocity, whereas recent research increasingly shows that emotion plays a critical role in this type of decision making. In this review, we summarize the influences of three types of emotions (i.e., the integral emotion experienced at the time of decision making, the incidental emotion aroused by a task-unrelated dispositional or situational source, and the interaction of emotion and cognition) on fairness-related decision making. Specifically, we first introduce three dominant theories that describe how emotion may influence fairness-related decision making (i.e., the wounded pride/spite model, affect infusion model, and dual-process model). Next, we collect behavioral and neural evidence for and against these theories. Finally, we propose that future research on fairness-related decision making should focus on inducing incidental social emotion, avoiding irrelevant emotion when regulating, exploring the individual differences in emotional dispositions, and strengthening the ecological validity of the paradigm.

Keywords: emotion, emotion regulation, fairness-related decision making, fairness theory, neural mechanisms

OPEN ACCESS

Edited by:
Chao Liu,
Beijing Normal University, China

Reviewed by:
Xiuyan Guo,
East China Normal University, China
Jing Luo,
Capital Normal University, China

****Correspondence:***
Xun Liu
liux@psych.ac.cn
Zhong Yang
yangzhong@psych.ac.cn

Specialty section:
This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Psychology

Received: *10 May 2017*
Accepted: *30 August 2017*
Published: *19 September 2017*

Citation:
Zheng Y, Yang Z, Jin C, Qi Y and Liu X (2017) The Influence of Emotion on Fairness-Related Decision Making: A Critical Review of Theories and Evidence. Front. Psychol. 8:1592. doi: 10.3389/fpsyg.2017.01592

INTRODUCTION

Researchers of decision-making typically regard emotion as impulsive and irrational and neglect its role in decision making (Kahneman and Tversky, 1979; Von Neumann and Morgenstern, 2007). In "normative decision theory," economic decision making is based on "cold" mathematical calculation, and decision makers are idealized as perfect "rational machines." However, studies increasingly show that emotion is one of the most important factors in the irrational decision-making process (Hastie, 2001; Sanfey et al., 2006). For example, emotion may guide people's decision making under conditions of risk and uncertainty and with regard to intertemporal choices, social decisions, and moral decision making (Loewenstein and Lerner, 2003; Rilling and Sanfey, 2011).

Fairness-related decision making is an important issue in the field of psychological decision making (Güth and Kocher, 2014). Experiments on fairness-related decision making have usually been conducted using the classic "Ultimatum Game" (UG) paradigm (Güth et al., 1982). An increasing number of UG studies have revealed that responders tended to sacrifice their own payoffs to decline an unfair offer, especially when they receive an offer that is less than 20% of the total (Güth et al., 1982; Thaler, 1988; Camerer and Thaler, 1995). These irrational rejection behaviors cannot be captured by the economic rationality of utility, in which the responder should accept all offers since receiving at least some money is always preferable to receiving no money.

Some theories, such as "inequity aversion" theory (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) and "reciprocity equilibrium" theory (Rabin, 1993; Falk and Fischbacher, 2006), have attempted to explain irrational behaviors in fairness-related decision making. "Inequity aversion" means that people prefer equitable outcomes: they are willing to forego a material payoff to work toward more equitable outcomes (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). However, it is difficult to explain why unfair offers from computer partners were accepted at higher rates than human partners if people were pursuing only fairness in terms of their own material payoff relative to the payoff of others (Blount, 1995; Knoch et al., 2006). According to "reciprocity equilibrium" theory, the rejection in the UG with human partners is social punishment to promote fair offers in subsequent bargaining, establish a good reputation, or enforce fairness norms (Rabin, 1993; Falk and Fischbacher, 2006). Thus, people will reject unfair offers from human partners, but accept unfair offers from computer partners to maximize personal gains. One study found that players would reject unfair offers when rejection reduced only their own earning to 0, and even when they cannot communicate their anger to the proposers through rejection (Yamagishi et al., 2009). The rejection of unfair offers that increase inequity and fail to punish proposers cannot be explained by the "inequity aversion" and "reciprocity equilibrium" theories. Such studies have increased awareness of the fact that emotion may be an important reason for irrational behaviors in fairness-related decision making (Sanfey et al., 2003; Ferguson et al., 2014). They propose that rejection is used to express the negative emotions such as anger or disgust aroused by unfair offers (Xiao and Houser, 2005). Although the two classical theories do not deny the existence of emotion, they nevertheless do not clearly explain the role of emotion and its mechanism. A new perspective on emotion is required to explain behavior in fairness-related decision making. Many studies have explored the influence of emotion on fairness-related decision making using behavioral, electrophysiological and neuroimaging approaches and supported these theories.

The influence of emotion on decision making concerns integral emotions (i.e., task-driven) and incidental emotions (i.e., task-unrelated) (Loewenstein and Lerner, 2003). The Wounded Pride/Spite Model suggests that integral emotion, such as negative emotions provoked by unfair offers, prompt rejections (Straub and Murnighan, 1995; Pillutla and Murnighan, 1996). However, this model only focuses on the influence of emotional response aroused by fairness-related decision making; it does not consider the influence of emotion aroused by dispositional or situational sources objectively unrelated to the task. To address this gap, the Affect Infusion Model investigated how incidental emotion (emotion aroused by emotional videos or images) influence fairness-related decision making (Forgas et al., 2003; Bless et al., 2006). These two models emphasized the role of emotion in fairness-related decision making, but ignored the regulation of emotion by cognition in modulating behavior. The Dual-Process System claims that the rational system and the emotional system are dual subsystems in fairness-related decision making, with the former prompting an adaptive response to different situations by regulating the latter (Loewenstein and O'Donoghue, 2004; Sanfey and Chang, 2008; Feng et al., 2015). This review summarizes these models of the impact of emotions on fairness-related decision making and the corresponding behavioral and neural evidence.

WOUNDED PRIDE/SPITE MODEL AND ITS EVIDENCE

Wounded Pride/Spite Model

The Wounded Pride/Spite Model proposes that the integral emotion aroused by a task itself may change fairness-related decision making. The model claims that if responders perceive that offers are unfair, feelings of wounded pride and anger may be aroused (Straub and Murnighan, 1995; Pillutla and Murnighan, 1996). When direct channels for expressing emotions are either impossible or undesirable, individuals are willing to incur the costs of rejection to retaliate against perceived unfairness (Gross and Levenson, 1993; Gross, 1999). Even when the responder has no way to punish the proposers, the responder still wants to reject the unfair offer (Yamagishi et al., 2009), suggesting that rejection may be not only a strategy to enlarge future potential payoffs but also an effective means of emotional release. However, if responders can convey their feelings of unfairness to proposers, the acceptance rates (ARs) of unfair offers could be increased substantially (Xiao and Houser, 2005).

Evidence from Integral Emotion

According to a large number of recent studies, the integral negative emotions aroused by unfair offers can increase the punishment for violating fairness norms.

First, previous studies found that fairness-related decision making can evoke strong emotions, demonstrating the existence of integral emotion in fairness-related decision making. From the responders' self-reports, the researchers found that when responders received an unfair offer, their negative affective responses, such as anger, contempt, irritation, envy and sadness, increased, whereas positive affective responses, such as pleasure and happiness, decreased (Pillutla and Murnighan, 1996; Bosman et al., 2001; Xiao and Houser, 2005; Osumi and Ohira, 2009; Voegele et al., 2010; Hewig et al., 2011; Bediou and Scherer, 2014; Gilam et al., 2015). Researchers used the UG to examine the affective correlates of decision making and found that the decision to reject is positively related to more negative emotional reactions, increased autonomic nervous system and skin conductance activity (van't Wout et al., 2006; Hewig et al., 2011), and decelerated heart rate (Osumi and Ohira, 2009; Dunn et al., 2012). Furthermore, similar facial motor activities were evoked by unfair treatment, unpleasant tastes, and photographs of contaminants, suggesting that unfairness elicits the same disgust as bad tastes and disease vectors (Chapman et al., 2009).

Second, the affective response to unfairness offers is one possible reason for rejection in fairness-related decision making. Psychophysiological studies have shown that increased ARs of offers correlate with greater resting heart rate variability (Osumi and Ohira, 2009; Dunn et al., 2012). EEG studies found that feedback-related negativity (FRN) could predict the likelihood of rejection in the UG and that rejection was associated with negative emotion (van't Wout et al., 2006; Hewig et al., 2011). By using the dipole localization method, EEG studies showed that unfair offers could arouse the activation of the insula, which is associated with negative emotion, and the anterior cingulate cortex (ACC), which is associated with conflict monitoring (Guclu et al., 2012). Neuroimaging studies also showed a negative correlation between the activation of the insula specifically involved in aversive emotion and the ARs of unfair offers (Sanfey et al., 2003; Takagishi et al., 2009). The above findings indicate that negative emotion aroused by perceptions of unfairness play an important role in rejection behaviors, supporting the Wounded Pride/Spite Model.

Although the Wounded Pride/Spite Model proposes that negative emotion in fairness-related decision making is an important factor in the rejection of an unfair offer (van't Wout et al., 2006; Hewig et al., 2011) and can explain many behaviors in fairness-related decision making (Harle and Sanfey, 2007; Grecucci et al., 2013b), this model is only concerned with the responders' emotional reaction that is aroused by fairness-related decision making. It ignores the impact of the responders' emotional state and other contextual factors.

AFFECT INFUSION MODEL AND EVIDENCE

Affect Infusion Model

The Affect Infusion Model proposes that incidental emotion aroused by task-unrelated sources can significantly influence fairness-related decision making by priming mood-congruent concepts and dispositions (Forgas et al., 1990; Forgas, 2002). For instance, in fairness-related decision making, people must integrate negative (unfair social signals) and positive (financial benefits) information. Positive incidental emotion makes responders more concerned about their own benefits, thus increasing ARs. By contrast, negative incidental emotion makes responders more concerned about unfair offers, thus decreasing ARs (Harle et al., 2012). That is, acceptance or rejection decisions represent the internal rewards and external fairness principles in fairness-related decision making. Positive emotion can enhance cooperation by recruiting a more assimilative, internally focused processing style that promotes selfishness (Forgas et al., 1990). Negative emotion is an alert signal that requires accommodative processing and increases monitoring of the external environment to process potential threats and hazardous stimulation, increasing concern with social norms (Forgas et al., 2003; Bless et al., 2006). For example, sadness provokes pessimistic framing and increases the processing of threatening information, making responders more concerned about the negative consequences of unfairness and the punishment of those who violate the fairness norm (Harle and Sanfey, 2007).

Evidence of Incidental Emotion

To explore the influence of incidental emotion, many studies have manipulated the affective state by evoking different valences and arousal levels with images and videos. The results showed that participants in a negative emotional state will reject a greater number of unfair offers (Moretti and Di Pellegrino, 2010; Fabiansson and Denson, 2012; Harle et al., 2012; Liu et al., 2016; Riepl et al., 2016), whereas a positive emotional state may reduce or exert no influence on ARs (Harle and Sanfey, 2007; Andrade and Ariely, 2009; Forgas and Tan, 2013a,b; Liu et al., 2016).

Behavioral studies found that on the one hand, when the participants were responders, compared with a neutral group, sad participants reported more negative emotions, such as anger and disgust, when faced with unfair offers and subsequently made more rejections. However, participants who were induced to experience happy emotions accepted more unfair offers (Riepl et al., 2016), with no discernible impact on their decisions (Harle and Sanfey, 2007; Forgas and Tan, 2013a,b; Liu et al., 2016). On the other hand, when the participants were proposers, inducing amusement (compared with sadness) made them more selfish; they also allocated a greater number of points to themselves and had shorter response times (Forgas and Tan, 2013a,b). Neuroimaging studies indicate that incidental sad emotions are regulated by the three main brain regions for emotions, namely, the insula, ACC and striatum. First, compared with participant responses under neutral conditions, the ARs of unfair offers were associated with higher bilateral insula activations in participants who were sad. Insula is typically associated with negative emotions (Paulus et al., 2003; Knutson et al., 2007), suggesting that this region may indicate an aversive response, which may reduce ARs (Harle et al., 2012). Consequently, some researchers have speculated that insula activation can predict the influence of sadness on decision making (Sanfey et al., 2003). Increasing evidence suggests the important role of the anterior insula (AI) in detecting norm violations (Civai et al., 2012; Xiang et al., 2013). Researchers speculated that a sad participant with increased AI activity may experience high sensitivity to norm violation. Thus, sad incidental emotion could activate the insula involved in negative emotion (or detection of norm violation) and bias behavior accordingly. Second, receiving unfair offers in a sad vs. neutral mood resulted in greater activation in the ACC linked to error and decision conflict monitoring, suggesting that sad individuals may experience an enhanced perception of social norm violation (Harle et al., 2012). Furthermore, a moderating effect of mood was found in the left ventral striatum, which is associated with reward processing. Individuals who experienced a neutral mood showed stronger activation for fair offers relative to unfair offers, while individuals who were sad did not exhibit such a pattern of activation, implying decreased reward responsiveness to reward stimuli (Harle et al., 2012). Overall, both behavioral and neural studies have shown that negative emotions enhance participants' negative responses to behaviors that violate fairness norms and reduce reward activation for fair offers, thus decreasing ARs. These studies demonstrate that emotion plays a role in changing participants' decisions by altering their cognitive processing, supporting the Affect Infusion Model.

However, some researchers have noted that the dimension of emotional motivation, rather than emotional valence, is the key factor that influences fairness decision making. Emotional valence refers to the intrinsic attractiveness (positive valence) or averseness (negative valence) of an event, an object, or a situation (Frijda, 1986). Emotional motivation refers to the aversive and appetitive apparatuses, which, respectively, promote withdrawal and approach behavior (Schneirla, 1959; Lang et al., 1997). Two emotions with similar valences may have different motivations, and vice versa. For instance, amusement and serenity are positive emotions, whereas anger and disgust are negative emotions. However, amusement and anger are classified as approach-based emotions, whereas serenity and disgust are withdraw-based emotions. Therefore, researchers have suggested that compared with a valence framework, partitioning affective states based on motivational tendency could more accurately explain the changes in ARs in fairness-related decision. The results of a study that explored the influence of positive emotions (amusement and serenity) and negative emotions (anger and disgust) on fairness-related decision making, indicate that emotional valence did not predict ARs. However, the approach-based emotional states (amusement, anger) increased ARs, whereas withdrawal-based emotional states (disgust, serenity) decreased ARs (Harle and Sanfey, 2010). Thus, emotional motivation may help explain fairness-related decision making. Many researchers have explored the emotional influence of fairness-related decision making in terms of approach-based states (anger) and withdrawal-based emotional states (disgust) (Andrade and Ariely, 2009; Moretti and Di Pellegrino, 2010; Liu et al., 2016; Riepl et al., 2016).

Studies have shown that anger influences fairness-related decision making and leads responders to reject more unfair offers. On the one hand, anger functions as a negative emotion after unfair treatment (Pillutla and Murnighan, 1996) and thus decreases the ARs of unfair offers. Prior to a decision, the responders' anger elicited by watching the video clip made them reject more unfair offers compared with responders who watched a pleasant video clip (Andrade and Ariely, 2009; Riepl et al., 2016). When manipulating the facial expressions of the proposers, the same results were found: responders facing angry proposers provided the most rejections, whereas the least rejections were from those who faced pleasant proposers (Mussel et al., 2013; Liu et al., 2016). When the responder's anger was provoked by the controlled proposer's negative appraisal of the responder's speech, decreased ARs resulted (Fabiansson and Denson, 2012). To the best of our knowledge, only one study used an EEG and explored the neural mechanism of the influence of incident emotion on fairness-related decision making. That study induced anger, fear and happiness via short movie clips. The results showed that responders with high trait negative affect in aversive mood states had increased FRN amplitudes when they were in an angry mood but not when they experienced fear or happiness (Riepl et al., 2016). On the other hand, whether the proposer or the responder is the angry party leads to different perceptions of fairness and judgments of the proposer's offer. If the proposers are angry, more unfair offers are given. For example, if the proposer's anger is aroused by the responder, the proposer is more likely to split unfair offers (Fabiansson and Denson, 2012). In contrast, if the responder feels angry, more fair offers are given. For example, proposers will make more fair offers when they know that the responders watched an angry video clip in contrast with the knowledge that the responders watched a happy clip (Andrade and Ho, 2007). The above results may relate to the proposers' attribution of anger. Anger is a kind of high-arousal and approach-based negative emotion (Berkowitz and Harmon-Jones, 2004; Carver and Harmon-Jones, 2009), and it may cause antisocial behaviors related to revenge (Carnevale and Isen, 1986; Pillutla and Murnighan, 1996; Allred et al., 1997). Therefore, when the responder is the one to irritate the proposer, the proposer proposes more unfair offers in return. Second, anger may make people tougher and more dominant (Knutson, 1996; Tiedens, 2001). People know that angry people are impulsive and act irrationally (Bacharach and Lawler, 1981), so they may make more fair offers to reduce the possibility of being rejected instead of irritating the responder to maximize the profits in bargaining when they play as proposers (Andrade and Ho, 2007; Andrade and Ariely, 2009).

In addition, disgust aroused prior to a decision can increase the responder's punishment for unfair offers, whereas the idea of misattributing the disgust induced by the unfair offer to incidental disgust will reduce the responder's punishment. When responders have viewed emotional pictures or faces to arouse aversion prior to a decision, lower ARs to unfair offers are caused by the disgust (Moretti and Di Pellegrino, 2010; Liu et al., 2016). In a comparison of the influence of disgust and sadness on fairness decisions, disgust caused obviously lower ARs (Moretti and Di Pellegrino, 2010). However, another study using disgusting smells showed that participants misattributed the disgust induced by an unfair offer to the disgusting smell, which led to higher ARs (Bonini et al., 2011). These results indicate that the arousal of disgust prompts people's maintenance of social norms because disgust is a type of withdrawal-based emotion (Harle and Sanfey, 2010) and may be extended to moral and social violations (Rozin et al., 2000). As an indicator of the judgment of others' behavior as either right or wrong, feelings of disgust can function better than sadness as moral intuition (Haidt, 2001) to decrease the ARs of unfair offers. To an extent, disgust aroused prior to a task overlapped with disgust in the distribution, whereas the attribution of the latter to the former resulted in a subtraction of the emotion.

From the above, we may conclude that the valences of anger and aversion are the same; however, due to the different induction manipulations and attributions, they may have different impacts on fairness-related decision making. Consequently, the Affect Infusion Model takes the motivational direction of emotion as an important factor to interpret the emotional process of fairness-related decision making within a wider range.

DUAL-PROCESS SYSTEMS AND THE EMPIRICAL STUDY

Dual-Process Systems

The above two models focused on the function of emotional arousal and appraisal in fairness-related decision making but ignored the regulation of emotion by cognition to change decision making. The Dual-process System claims that there are dual subsystems in fairness-related decision making: one is automatic, with an immediate response and an emotional system with no cognitive effort, whereas the other is controlled and comparatively slow, with a rational system of cognitive effort. The emotional system represents the intuitive response; however, after learning and calculation, the rational system requires an adaptive response to different situations by regulating the emotional system (Loewenstein and O'Donoghue, 2004; Sanfey and Chang, 2008; Feng et al., 2015). Fairness-related decision making is influenced by systematically and effectively regulating responders' fairness perceptions via rational cognitive control (Rilling and Sanfey, 2011). For example, the model suggests that all types of emotional regulation strategies can change fairness-related decision making through the interaction of cognition and emotion.

The Empirical Study of Dual-Process Systems

Researchers have employed different emotion regulation strategies and compared their effectiveness. The results support the influence of emotion regulation on fairness-related decision making.

First, responders may spontaneously regulate the negative emotions induced by unfair offers in the UG. After decision making, responders are requested to report their own opinions on the offer and to write down the shift of their decisions as follows: "At the very beginning, I thought of. . ., then I considered.. . ." Some responders may remain angry, reject the unfair offer and refuse to report, whereas others may spontaneously employ cognitive reappraisal to reduce their own negative emotions and then accept more unfair offers (Voegele et al., 2010; Gilam et al., 2015). In physiological arousal, responders who employed reappraisal showed higher vagal activation and attenuated heart rate deceleration after accepting unfair offers (Voegele et al., 2010). Neuroimaging studies have revealed that increased ARs of unfair offers are associated with increased activity in the ventrolateral prefrontal cortex (vlPFC), a region involved in emotion regulation, and decreased activity in the AI, which is linked to negative affect (Tabibnia et al., 2008). Individuals with high monetary gains showed increased ventromedial prefrontal cortex (vmPFC) activity but also decreased AI activity (Tabibnia et al., 2008; Gilam et al., 2015). Furthermore, patients with vmPFC damage had lower ARs than control groups (Koenigs and Tranel, 2007). The studies suggested that brain areas associated with emotion regulation, such as vlPFC and vmPFC, may be engaged to diminish the aversion-related AI's response (Tabibnia et al., 2008; Gilam et al., 2015) and increase the ARs of unfair offers.

Second, multiple emotion regulation strategies can change decisions by regulating emotions. Researchers have employed two strategies for emotion regulation in fairness-related decision making: reappraisal and expressive suppression. The results showed that although the two strategies could reduce the negative emotions of responders to unfair offers, though compared with expressive suppression, the reappraisal strategy was more effective in changing responders' emotions and making them accept more unfair offers (Kirk et al., 2006; van't Wout et al., 2010; Fabiansson and Denson, 2012). In addition, reappraisal strategies may continue to reduce participants' negative emotions and make them propose more fair offers during a second interaction with partners who treated them unfairly in a previous interaction, whereas the expressive suppression strategy may reduce participants' previous negative emotions with no effect of ridding themselves of negative treatment, resulting in the proposal of unfair offers (van't Wout et al., 2010; Fabiansson and Denson, 2012). The results showed that to change emotions and behaviors using an emotion regulation strategy and to avoid previous negative impact, the reappraisal strategy is considerably more effective than expressive suppression and can extend beyond a single encounter to influence future interaction. Grecucci furthered the study of reappraisal strategies by discussing up- and down-regulation (Grecucci et al., 2013b). The former refers to the interpretation of intentions and behaviors of unfair offers as more negative (i.e., the player is a selfish person and wants to keep all the monetary gains), whereas the latter refers to these as less negative (i.e., the proposers' debt problems leading them to gain more). The results showed that responders with an up-regulation strategy rejected more unfair offers in contrast with down-regulation, demonstrating that reappraisal strategies may change the way responders understand others' intentions and affect their emotional reaction, resulting in changed decisions. Overall, the reappraisal strategy can modulate the impact of emotional stimuli, contributing to our decisions flexibly (Grecucci et al., 2013b). Neuroimaging studies revealed that the dorsolateral prefrontal cortex (DLPFC) and bilateral ACC play vital roles in the reappraisal process. The DLPFC is associated with cognitive control and inhibition (Miller and Cohen, 2001) as the basis of the generation and maintenance of reappraisal strategies (Ochsner et al., 2002; Ochsner and Gross, 2005). Additionally, Buckholtz et al. (2008, 2015), Buckholtz and Marois (2012) proposed the integrative model of DLPFC function, which suggested the role of DLPFC in the representational integration of the distinct information streams used to make punishment decisions. When applying cognition reappraisal in fairness-related decision making, the evaluation of fairness and the information concerning harm and blame changed. Therefore, the DLPFC activated to integrate the information from emotional response, regulation strategy, fairness evaluation and other sources to make punishment decisions. Furthermore, the ACC monitors and evaluates conflicting responses or motives (Yeung and Sanfey, 2004; Ochsner and Gross, 2005).

In addition to reappraisal and expressive suppression, expected emotion is an effective way of regulating fairness-related decision making. With regard to changing a decision,

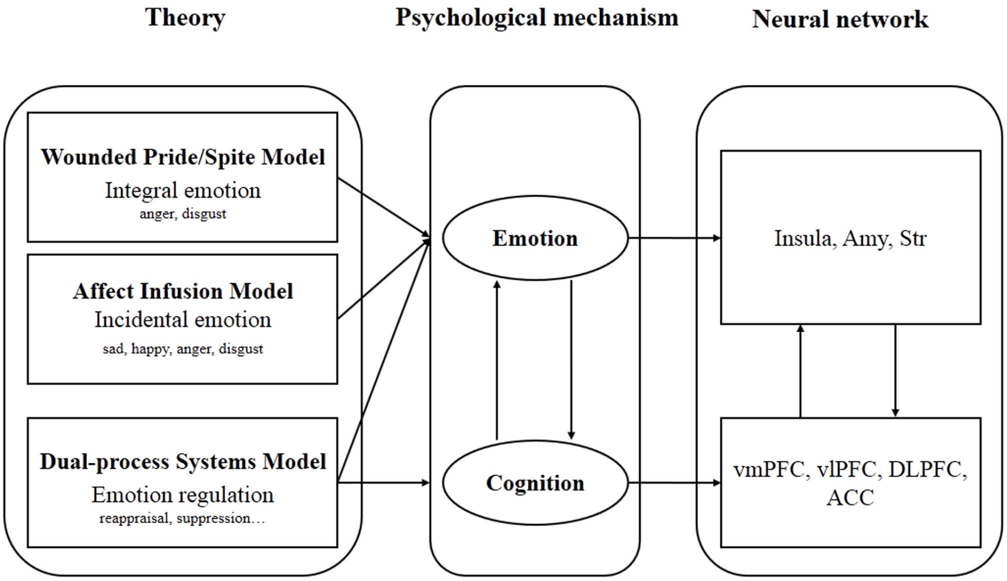

FIGURE 1 | A schematic illustration of the influence of emotion on fairness-related decision making. The Wounded Pride/Spite Model and the Affect Infusion Model describe the influence of emotion, whereas the Dual-Process Systems Model proposes that cognition could regulate emotion to regulate fairness-related decision making. The related brain regions include the insula, amygdala (Amy), striatum (Str), ventrolateral prefrontal cortex (vlPFC), ventromedial prefrontal cortex (vmPFC), dorsolateral prefrontal cortex (DLPFC), and anterior cingulate cortex (ACC).

some studies have investigated the regulation of individuals' expected emotion induced by the decision outcome. In the decision stage, responders will attempt to predict the probabilities of different outcomes and the emotional consequences associated with alternative actions. To minimize negative emotion and maximize positive emotion, responders will adjust their decisions (Loewenstein and Lerner, 2003; Rick and Loewenstein, 2007). If they predict they will be proud of their fair offers, more fair offers will be given, whereas if they predict that they will feel regretful, less fair offers will be chosen. The expected emotion helps them to anticipate future outcomes and modify their behaviors to evoke desirable emotions and avoid undesirable results. When an individual can expect a positive outcome, it is likely that a current offer will be supported. In contrast, an expected negative outcome will lead to modification of the current activities (Baumeister et al., 2007). Some researchers have manipulated the expected emotion using the autobiographical recall task and found that anticipated pride about fair behavior increased levels of fairness, whereas anticipated pride about unfair behavior decreased levels of fairness. Similarly, anticipated regret about fair behavior reduced levels of fairness, whereas anticipated regret about unfairness increased levels of fairness (van der Schalk et al., 2012). If the proposers were required to observe pride or regret after making fair or unfair offers in the UG, they made fewer fair offers if they had seen the responder's regret about a fair offer, whereas they made more fair offers if they had seen the responder's regret about unfair offers (van der Schalk et al., 2014). The results showed that past emotional experience make people reflect on and modify the outcome of their behavior because they pursue not only maximized benefits but also positive emotional experiences (Mellers et al., 1999; Loewenstein and Lerner, 2003). Other studies on regulating strategies of delay or distraction revealed that the delay of a decision did not change the emotional experience or behavior (Bosman et al., 2001), whereas distraction only decreased anger but did not change fairness-related or other decisions when anger was induced again by the same stimulus (Gross and Levenson, 1993; Gross, 1999; Xiao and Houser, 2005; Fabiansson and Denson, 2012).

Neural mechanism studies on the emotion regulation of fairness-related decision making have supported Dual-process Systems. The interaction of the automatic processing emotional system and the controlled cognitive system affects people's behavior. The emotional system includes the insula, which is associated with aversion to violating norms (Sanfey et al., 2003; Guo et al., 2013); the amygdala, which is associated with negative emotions (Haruno and Frith, 2010; Haruno et al., 2014); and the vmPFC, which is associated with encoding subjective values of perceived offers and emotion regulation (Tabibnia et al., 2008; Baumgartner et al., 2011; Gilam et al., 2015). In addition, the controlled cognitive system involves the dorsal ACC, which regulates the conflict of norm enforcement and self-interest and DLPFC (Knoch et al., 2006, 2008; Baumgartner et al., 2011) related to executive control.

Dual-process Systems focus on the function of emotions and involve the interaction of emotion and cognition for fairness-related decision making. This model has been supported by many behavioral and neuroimaging studies (Sanfey et al., 2003; Baumgartner et al., 2011). This model also proposes strategies for regulating emotion that provide a new way of changing fairness-related decision making (Knoch et al., 2006, 2008). However, current evidence is limited to the regulation of negative emotion induced by an offer (Grecucci et al., 2013a,b). Little is known about the regulation of incidental emotion in fairness-related decision making.

A SCHEMATIC ILLUSTRATION OF THE INFLUENCE OF EMOTION ON FAIRNESS-RELATED DECISION MAKING

In complex social environments, both the emotion and cognition systems are involved in processing the fairness perception of resource distribution (see **Figure 1**). The Wounded Pride/Spite Model and the Affect Infusion Model describe the influence of integral emotion aroused by task and incidental emotion aroused by task-unrelated resources, respectively. For instance, compared with fair offers, unfair offers have been associated with greater activation of the insula, which is involved in aversion emotion (Sanfey et al., 2003; Takagishi et al., 2009), whereas fair offers have been linked to the activation of reward regions, such as the ventral striatum (Tabibnia et al., 2008). Additionally, individuals in sad or angry moods showed an enhanced perception of unfairness, with a greater activation of the insula and amygdala (Harle et al., 2012). The Dual-process Systems perspective proposes that the rational system could regulate emotion to both up- and down-regulate fairness-related decision making. For example, the ACC monitors and evaluates conflicts between norm enforcement and financial benefit (Yeung and Sanfey, 2004; Ochsner and Gross, 2005). The vlPFC and vmPFC associated with emotion regulation could decrease the activation of AI to diminish conflicts (Tabibnia et al., 2008; Gilam et al., 2015). The DLPFC is associated with cognitive control and inhibition (Miller and Cohen, 2001) and influences generation and maintenance reappraisal strategies (Ochsner et al., 2002; Ochsner and Gross, 2005). It can integrate the information from emotional response, regulation strategy, fairness evaluation and other sources to make punishment decisions (Buckholtz and Marois, 2012).

SUMMARY AND PROSPECTS

In the history of studies on fairness-related decision making, the hypothesis has changed from viewing responders as completely rational with no influence from emotion to regarding both emotion and cognition as important factors in Dual-process Systems. Many studies have revealed that emotion plays an important role in fairness-related decision making. Based on the review of the theoretical and empirical studies, we conclude that the future research scope of the influence of emotion in fairness-related decision making can be furthered in the following ways.

First, recent studies that have induced incidental emotions are limited to several basic emotions, such as happiness, sadness, anger or disgust. However, as a social animal, humans have complicated, delicate and vast social structures and interpersonal relations. Among these, social emotions are one of the important motivations for human behavior. Since fairness is one of the basic norms in human society, it is influenced by many social emotions (Takahashi et al., 2009). As a result, future research should explore the impact of social emotions, including both positive social emotions (empathy, gratitude) and negative social emotions (envy, indignation), on fairness-related decision making.

Second, reappraisal is a common strategy to regulate emotional response, but this strategy involves reinterpreting the meaning of a stimulus. In studies on fairness-related decision making, responders can adopt an up-regulation strategy or a down-regulation strategy. Responders must evaluate the motivations and behaviors of proposers to decrease the anger or disgust caused by unfair offers (Grecucci et al., 2013b). However, reappraisal may induce other emotions, such as empathy from down-regulation (Gross, 2013). Future studies should aim to identify the irrelevant emotions aroused by the regulation strategy that may influence fairness-related decision making.

Third, some personal traits, such as emotional dispositions (Dunn et al., 2010), social value orientation (Karagonlar and Kuhlman, 2013; Haruno et al., 2014), and personality characteristics (Spitzer et al., 2007; Osumi et al., 2012), may influence personal emotional response and regulation, thus affecting fairness-related decision making. For this reason, we suggest that future studies should explore the possible interaction of personality traits, emotion and unfair offers.

Finally, the standard UG paradigm has been widely used in studies on the influence of emotions on fairness-related decision making. Some complex, modified versions of the UG may complicate the context of fairness-related decision making, but may nevertheless be accurate models of real-world situations. For instance, we can put fairness-related decision making in the more complex background of social comparison (Wu et al., 2011; Alexopoulos et al., 2012; McDonald et al., 2013), the loss context (Buchan et al., 2005; Zhou and Wu, 2011; Guo et al., 2013), or making responders perceive the intentions of the proposer (Radke et al., 2012; Ma et al., 2015). As a result, future studies on the influence of emotions on fairness-related decision making should consider ecological validity to make the studies more realistic.

AUTHOR CONTRIBUTIONS

Conceived and designed the study: ZY, YZ, and XL. Literature search and synthesis: CJ and YQ. Wrote the paper: ZY and YZ.

FUNDING

This research was supported by the National Natural Science Foundation of China (Grants 31571161, 31200782, 31500872, and 31640039), the China Scholarship Council, and the National Social Science Foundation of China (Grant 14ZDB161).

REFERENCES

Alexopoulos, J., Pfabigan, D. M., Lamm, C., Bauer, H., and Fischmeister, F. P. (2012). Do we care about the powerless third? An ERP study of the three-person ultimatum game. *Front. Hum. Neurosci.* 6:59. doi: 10.3389/fnhum.2012.00059

Allred, K. G., Mallozzi, J. S., Matsui, F., and Raia, C. P. (1997). The influence of anger and compassion on negotiation performance. *Organ. Behav. Hum. Decis. Process.* 70, 175–187. doi: 10.1006/obhd.1997.2705

Andrade, E. B., and Ariely, D. (2009). The enduring impact of transient emotions on decision making. *Organ. Behav. Hum. Decis. Process.* 109, 1–8. doi: 10.1016/j.obhdp.2009.02.003

Andrade, E. B., and Ho, T. H. (2007). How is the boss's mood today? I want a raise. *Psychol. Sci.* 18, 668–671. doi: 10.1111/j.1467-9280.2007.01956.x

Bacharach, S. B., and Lawler, E. J. (1981). *Bargaining: Power, Tactics and Outcomes*. San Francisco, CA: Jossey-Bass Inc.

Baumeister, R. F., Vohs, K. D., DeWall, C. N., and Zhang, L. (2007). How emotion shapes behavior: feedback, anticipation, and reflection, rather than direct causation. *Pers. Soc. Psychol. Rev.* 11, 167–203. doi: 10.1177/1088868307301033

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., and Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nat. Neurosci.* 14, 1468–1474. doi: 10.1038/nn.2933

Bediou, B., and Scherer, K. R. (2014). Egocentric fairness perception: emotional reactions and individual differences in overt responses. *PLOS ONE* 9:e88432. doi: 10.1371/journal.pone.0088432

Berkowitz, L., and Harmon-Jones, E. (2004). Toward an understanding of the determinants of anger. *Emotion.* 4, 107–130. doi: 10.1037/1528-3542.4.2.107

Bless, H., Fiedler, K., and Forgas, J. (2006). "Mood and the regulation of information processing and behavior," in *Hearts and Minds: Affective Influences on Social Cognition and Behavior*, ed. J. P. Forgas (New York, NY: Psychology Press), 65–84.

Blount, S. (1995). When social outcomes aren't fair: the effect of causal attributions on preferences. *Organ. Behav. Hum. Decis. Process.* 63, 131–144. doi: 10.1006/obhd.1995.1068

Bolton, G. E., and Ockenfels, A. (2000). ERC: a theory of equity, reciprocity, and competition. *Am. Econ. Rev.* 90, 166–193. doi: 10.1257/aer.90.1.166

Bonini, N., Hadjichristidis, C., Mazzocco, K., Dematte, M. L., Zampini, M., Sbarbati, A., et al. (2011). Pecunia olet: the role of incidental disgust in the ultimatum game. *Emotion* 11, 965–969. doi: 10.1037/a0022820

Bosman, R., Sonnemans, J., and Zeelenberg, M. (2001). *Emotions, Rejections, and Cooling off in the Ultimatum Game. Working Paper*. Amsterdam: University of Amsterdam.

Buchan, N., Croson, R., Johnson, E., and Wu, G. (2005). "Gain and loss ultimatums," in *Experimental and Behavioral Economics*, ed. J. Morgan (Bingley: Emerald Group Publishing Limited), 1–23.

Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., et al. (2008). The neural correlates of third-party punishment. *Neuron* 60, 930–940. doi: 10.1016/j.neuron.2008.10.016

Buckholtz, J. W., and Marois, R. (2012). The roots of modern justice: cognitive and neural foundations of social norms and their enforcement. *Nat. Neurosci.* 15, 655–661. doi: 10.1038/nn.3087

Buckholtz, J. W., Martin, J. W., Treadway, M. T., Jan, K., Zald, D. H., Jones, O., et al. (2015). From blame to punishment: disrupting prefrontal cortex activity reveals norm enforcement mechanisms. *Neuron* 87, 1369–1380. doi: 10.1016/j.neuron.2015.08.023

Camerer, C., and Thaler, R. H. (1995). Anomalies: ultimatums, dictators and manners. *J. Econ. Perspect.* 9, 209–219. doi: 10.1257/jep.9.2.209

Carnevale, P. J. D., and Isen, A. M. (1986). The influence of positive affect and visual access on the discovery of integrative solutions in bilateral negotiation. *Organ. Behav. Hum. Decis. Process.* 37, 1–13. doi: 10.1016/0749-5978(86)90041-5

Carver, C. S., and Harmon-Jones, E. (2009). Anger is an approach-related affect: evidence and implications. *Psychol. Bull.* 135, 183–204. doi: 10.1037/a0013965

Chapman, H. A., Kim, D. A., Susskind, J. M., and Anderson, A. K. (2009). In bad taste: evidence for the oral origins of moral disgust. *Science* 323, 1222–1226. doi: 10.2307/25471595

Civai, C., Crescentini, C., Rustichini, A., and Rumiati, R. I. (2012). Equality versus self-interest in the brain: differential roles of anterior insula and medial prefrontal cortex. *Neuroimage* 62, 102–112. doi: 10.1016/j.neuroimage.2012.04.037

Dunn, B. D., Evans, D., Makarova, D., White, J., and Clark, L. (2012). Gut feelings and the reaction to perceived inequity: the interplay between bodily responses, regulation, and perception shapes the rejection of unfair offers on the ultimatum game. *Cogn. Affect. Behav. Neurosci.* 12, 419–429. doi: 10.3758/s13415-012-0092-z

Dunn, B. D., Makarova, D., Evans, D., and Clark, L. (2010). "I'm worth more than that": trait positivity predicts increased rejection of unfair financial offers. *PLOS ONE* 5:e15095. doi: 10.1371/journal.pone.0015095

Fabiansson, E. C., and Denson, T. F. (2012). The effects of intrapersonal anger and its regulation in economic bargaining. *PLOS ONE* 7:e51595. doi: 10.1371/journal.pone.0051595

Falk, A., and Fischbacher, U. (2006). A theory of reciprocity. *Games Econ. Behav.* 54, 293–315. doi: 10.1016/j.geb.2005.03.001

Fehr, E., and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Q. J. Econ.* 114, 817–868. doi: 10.1162/003355399556151

Feng, C., Luo, Y.-J., and Krueger, F. (2015). Neural signatures of fairness-related normative decision making in the ultimatum game: a coordinate-based meta-analysis. *Hum. Brain Mapp.* 36, 591–602. doi: 10.1002/hbm.22649

Ferguson, E., Maltby, J., Bibby, P. A., and Lawrence, C. (2014). Fast to forgive, slow to retaliate: intuitive responses in the ultimatum game depend on the degree of unfairness. *PLOS ONE* 9:e96344. doi: 10.1371/journal.pone.0096344

Forgas, J. P. (2002). Feeling and doing: affective influences on interpersonal behavior. *Psychol. Inq.* 13, 1–28. doi: 10.1207/S15327965PLI1301-01

Forgas, J. P., Bower, G. H., and Moylan, S. J. (1990). Praise or blame? Affective influences on attributions for achievement. *J. Pers. Soc. Psychol.* 59, 809–819. doi: 10.1037/0022-3514.59.4.809

Forgas, J. P., Davidson, R., Scherer, K., and Goldsmith, H. (2003). "Affective influences on attitudes and judgments," in *Handbook of Affective Sciences*, eds R. J. Davidson, K. R. Scherer, and H. H. Goldsmith (New York, NY: Oxford University Press), 596–618.

Forgas, J. P., and Tan, H. B. (2013a). Mood effects on selfishness versus fairness: affective influences on social decisions in the ultimatum game. *Soc. Cogn.* 31, 504–517. doi: 10.1521/soco-2012-1006

Forgas, J. P., and Tan, H. B. (2013b). To give or to keep? Affective influences on selfishness and fairness in computer-mediated interactions in the dictator game and the ultimatum game. *Comput. Hum. Behav.* 29, 64–74. doi: 10.1016/j.chb.2012.07.017

Frijda, N. H. (1986). *The Emotions*. Cambridge: Cambridge University Press.

Gilam, G., Lin, T., Raz, G., Azrielant, S., Fruchter, E., Ariely, D., et al. (2015). Neural substrates underlying the tendency to accept anger-infused ultimatum offers during dynamic social interactions. *Neuroimage* 120, 400–411. doi: 10.1016/j.neuroimage.2015.07.003

Grecucci, A., Giorgetta, C., Bonini, N., and Sanfey, A. G. (2013a). Reappraising social emotions: the role of inferior frontal gyrus, temporo- parietal junction and insula in interpersonal emotion regulation. *Front. Hum. Neurosci.* 7:523. doi: 10.3389/fnhum.2013.00523

Grecucci, A., Giorgetta, C., van't Wout, M., Bonini, N., and Sanfey, A. G. (2013b). Reappraising the ultimatum: an fMRI study of emotion regulation and decision making. *Cereb. Cortex* 23, 399–410. doi: 10.1093/cercor/bhs028

Gross, J. J. (1999). Emotion regulation: past, present, future. *Cogn. Emot.* 13, 551–573. doi: 10.1080/026999399379186

Gross, J. J. (2013). *Handbook of Emotion Regulation*. New York City, NY: Guilford publications.

Gross, J. J., and Levenson, R. W. (1993). Emotional suppression: physiology, self-report, and expressive behavior. *J. Pers. Soc. Psychol.* 64:970. doi: 10.1037//0022-3514.64.6.970

Guclu, B., Ertac, S., Hortacsu, A., and List, J. A. (2012). Mental attributes and temporal brain dynamics during bargaining: EEG source localization and neuroinformatic mapping. *Soc. Neurosci.* 7, 159–177. doi: 10.1080/17470919.2011.586902

Guo, X., Zheng, L., Zhu, L., Li, J., Wang, Q., Dienes, Z., et al. (2013). Increased neural responses to unfairness in a loss context. *Neuroimage* 77, 246–253. doi: 10.1016/j.neuroimage.2013.03.048

Güth, W., and Kocher, M. G. (2014). More than thirty years of ultimatum bargaining experiments: motives, variations, and a survey of the recent literature. *J. Econ. Behav. Organ.* 108, 396–409. doi: 10.1016/j.jebo.2014.06.006

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental-analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* 3, 367–388. doi: 10.1016/0167-2681(82)90011-7

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol. Rev.* 108, 814–834. doi: 10.1037/0033-295X.108.4.814

Harle, K. M., Chang, L. J., van 't Wout, M., and Sanfey, A. G. (2012). The neural mechanisms of affect infusion in social economic decision-making: a mediating role of the anterior insula. *Neuroimage* 61, 32–40. doi: 10.1016/j.neuroimage.2012.02.027

Harle, K. M., and Sanfey, A. G. (2007). Incidental sadness biases social economic decisions in the ultimatum game. *Emotion* 7, 876–881. doi: 10.1037/1528-3542.7.4.876

Harle, K. M., and Sanfey, A. G. (2010). Effects of approach and withdrawal motivation on interactive economic decisions. *Cogn. Emot.* 24, 1456–1465. doi: 10.1080/02699930903510220

Haruno, M., and Frith, C. D. (2010). Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nat. Neurosci.* 13, 160–161. doi: 10.1038/nn.2468

Haruno, M., Kimura, M., and Frith, C. D. (2014). Activity in the nucleus accumbens and amygdala underlies individual differences in prosocial and individualistic economic choices. *J. Cogn. Neurosci.* 26, 1861–1870. doi: 10.1162/jocn_a_00589

Hastie, R. (2001). Problems for judgment and decision making. *Annu. Rev. Psychol.* 52, 653–683. doi: 10.1146/annurev.psych.52.1.653

Hewig, J., Kretschmer, N., Trippe, R. H., Hecht, H., Coles, M. G., Holroyd, C. B., et al. (2011). Why humans deviate from rational choice. *Psychophysiology* 48, 507–514. doi: 10.1111/j.1469-8986.2010.01081.x

Kahneman, D., and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–291. doi: 10.2307/1914185

Karagonlar, G., and Kuhlman, D. M. (2013). The role of social value orientation in response to an unfair offer in the ultimatum game. *Organ. Behav. Hum. Decis. Process.* 120, 228–239. doi: 10.1016/j.obhdp.2012.07.006

Kirk, D., Carnevale, P. J., and Gollwitzer, P. M. (2006). *Self-Regulation in Ultimatum Bargaining: Controlling Emotion with Binding Goals*. Available at SSRN: https://ssrn.com/abstract=539843. doi: 10.2139/ssrn.913732

Knoch, D., Nitsche, M. A., Fischbacher, U., Eisenegger, C., Pascual-Leone, A., and Fehr, E. (2008). Studying the neurobiology of social interaction with transcranial direct current stimulation - The example of punishing unfairness. *Cereb. Cortex* 18, 1987–1990. doi: 10.1093/cercor/bhm237

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., and Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832. doi: 10.1126/science.1129156

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *J. Nonverbal Behav.* 20, 165–182. doi: 10.1007/bf02281954

Knutson, B., Rick, S., Wimmer, G. E., Prelec, D., and Loewenstein, G. (2007). Neural predictors of purchases. *Neuron* 53, 147–156. doi: 10.1016/j.neuron.2006.11.010

Koenigs, M., and Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: evidence from the ultimatum game. *J. Neurosci.* 27, 951–956. doi: 10.1523/jneurosci.4606-06.2007

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1997). "Motivated attention: affect, activation and action," in *Attention and Orienting: Sensory and Motivational Processes*, eds P. J. Lang, R. F. Simons, and M. T. Balaban (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc), 97–135.

Liu, C., Chai, J. W., and Yu, R. (2016). Negative incidental emotions augment fairness sensitivity. *Sci. Rep.* 6:24892. doi: 10.1038/srep24892

Loewenstein, G., and Lerner, J. S. (2003). "The role of affect in decision making," in *Handbook of Affective Science*, eds R. J. Davidson, K. R. Scherer, and H. H. Goldsmith (New York, NY: Oxford University Press), 3.

Loewenstein, G., and O'Donoghue, T. (2004). *Animal Spirits: Affective and Deliberative Processes in Economic Behavior*. Available at SSRN: https://ssrn.com/abstract = 539843. doi: 10.2139/ssrn.539843

Ma, Q., Meng, L., Zhang, Z., Xu, Q., Wang, Y., and Shen, Q. (2015). You did not mean it: perceived good intentions alleviate sense of unfairness. *Int. J. Psychophysiol.* 96, 183–190. doi: 10.1016/j.ijpsycho.2015.03.011

McDonald, I. M., Nikiforakis, N., Olekalns, N., and Sibly, H. (2013). Social comparisons and reference group formation: Some experimental evidence. *Games Econ. Behav.* 79, 75–89. doi: 10.1016/j.geb.2012.12.003

Mellers, B., Schwartz, A., and Ritov, I. (1999). Emotion-based choice. *J. Exp. Psychol.* 128, 332–345. doi: 10.1037/0096-3445.128.3.332

Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167

Moretti, L., and Di Pellegrino, G. (2010). Disgust selectively modulates reciprocal fairness in economic interactions. *Emotion* 10, 169–180. doi: 10.1037/a0017826

Mussel, P., Goritz, A. S., and Hewig, J. (2013). The value of a smile: facial expression affects ultimatum-game responses. *Judgm. Decis. Mak.* 8, 381–385.

Ochsner, K. N., Bunge, S. A., Gross, J. J., and Gabrieli, J. D. E. (2002). Rethinking feelings: an fMRI study of the cognitive regulation of emotion. *J. Cogn. Neurosci.* 14, 1215–1229. doi: 10.1162/089892902760807212

Ochsner, K. N., and Gross, J. J. (2005). The cognitive control of emotion. *Trends Cogn. Sci.* 9, 242–249. doi: 10.1016/j.tics.2005.03.010

Osumi, T., Nakao, T., Kasuya, Y., Shinoda, J., Yamada, J., and Ohira, H. (2012). Amygdala dysfunction attenuates frustration-induced aggression in psychopathic individuals in a non-criminal population. *J. Affect. Disord.* 142, 331–338. doi: 10.1016/j.jad.2012.05.012

Osumi, T., and Ohira, H. (2009). Cardiac responses predict decisions: an investigation of the relation between orienting response and decisions in the ultimatum game. *Int. J. Psychophysiol.* 74, 74–79. doi: 10.1016/j.ijpsycho.2009.07.007

Paulus, M. P., Rogalsky, C., Simmons, A., Feinstein, J. S., and Stein, M. B. (2003). Increased activation in the right insula during risk-taking decision making is related to harm avoidance and neuroticism. *Neuroimage* 19, 1439–1448. doi: 10.1016/S1053-8119(03)00251-9

Pillutla, M. M., and Murnighan, J. K. (1996). Unfairness, anger, and spite: emotional rejections of ultimatum offers. *Organ. Behav. Hum. Decis. Process.* 68, 208–224. doi: 10.1006/obhd.1996.0100

Rabin, M. (1993). Incorporating fairness into game theory and economics. *Am. Econ. Rev.* 83, 1281–1302. doi: 10.2307/2117561

Radke, S., Guroglu, B., and de Bruijn, E. R. (2012). There's something about a fair split: intentionality moderates context-based fairness considerations in social decision-making. *PLOS ONE* 7:e31491. doi: 10.1371/journal.pone.0031491

Rick, S., and Loewenstein, G. (2007). "The role of emotion in economic behavior," in *Handbook of Emotions*, eds M. Lewis, J. M. Haviland-Jones, and L. Feldman Barrett (New York, NY: Oxford University Press).

Riepl, K., Mussel, P., Osinsky, R., and Hewig, J. (2016). Influences of state and trait affect on behavior, feedback-related negativity, and P3b in the ultimatum game. *PLOS ONE* 11:e0146358. doi: 10.1371/journal.pone.0146358

Rilling, J. K., and Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annu. Rev. Psychol.* 62, 23–48. doi: 10.1146/annurev.psych.121208.131647

Rozin, P., Haidt, J., and McCauley, C. R. (2000). "Disgust," in *Handbook of Emotions*, eds M. Lewise and J. M. Haviland (New York, NY: Guilford Press), 637–653.

Sanfey, A. G., and Chang, L. J. (2008). Multiple systems in decision making. *Ann. N. Y. Acad. Sci.* 1128, 53–62. doi: 10.1196/annals.1399.007

Sanfey, A. G., Loewenstein, G., McClure, S. M., and Cohen, J. D. (2006). Neuroeconomics: cross-currents in research on decision-making. *Trends Cogn. Sci.* 10, 108–116. doi: 10.1016/j.tics.2006.01.009

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755–1758. doi: 10.1126/science.1082976

Schneirla, T. C. (1959). "An evolutionary and developmental theory of biphasic processes underlying approach and withdrawal," in *Nebraska Symposium on Motivation*, ed. M. R. Jones (Lincoln, NE: University of Nebraska Press), 1–42.

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., and Fehr, E. (2007). The neural signature of social norm compliance. *Neuron* 56, 185–196. doi: 10.1016/j.neuron.2007.09.011

Straub, P. G., and Murnighan, J. K. (1995). An experimental investigation of ultimatum games: information, fairness, expectations, and lowest acceptable offers. *J. Econ. Behav. Organ.* 27, 345–364. doi: 10.1016/0167-2681(94)00072-M

Tabibnia, G., Satpute, A. B., and Lieberman, M. D. (2008). The sunny side of fairness - Preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychol. Sci.* 19, 339–347. doi: 10.1111/j.1467-9280.2008.02091.x

Takagishi, H., Takahashi, T., Toyomura, A., Takashino, N., Koizumi, M., and Yamagishi, T. (2009). Neural correlates of the rejection of unfair offers in the impunity game. *Neuroendocrinol. Lett.* 30, 496–500.

Takahashi, H., Kato, M., Matsuura, M., Mobbs, D., Suhara, T., and Okubo, Y. (2009). When your gain is my pain and your pain is my gain: neural correlates of envy and schadenfreude. *Science* 323, 937–939. doi: 10.1126/science.1165604

Thaler, R. H. (1988). Anomalies: the ultimatum game. *J. Econ. Perspect.* 2, 195–206. doi: 10.1257/jep.2.4.195

Tiedens, L. Z. (2001). Anger and advancement versus sadness and subjugation: the effect of negative emotion expressions on social status conferral. *J. Pers. Soc. Psychol.* 80, 86–94. doi: 10.1037/0022-3514.80.1.86

van der Schalk, J., Bruder, M., and Manstead, A. (2012). Regulating emotion in the context of interpersonal decisions: the role of anticipated pride and regret. *Front. Psychol.* 3:513. doi: 10.3389/fpsyg.2012.00513

van der Schalk, J., Kuppens, T., Bruder, M., and Manstead, A. S. (2014). The social power of regret: the effect of social appraisal and anticipated emotions on fair and unfair allocations in resource dilemmas. *J. Exp. Psychol.* 144, 151–157. doi: 10.1037/xge0000036

van't Wout, M., Chang, L. J., and Sanfey, A. G. (2010). The influence of emotion regulation on social interactive decision-making. *Emotion* 10, 815–821. doi: 10.1037/a0020069

van't Wout, M., Kahn, R. S., Sanfey, A. G., and Aleman, A. (2006). Affective state and decision-making in the ultimatum game. *Exp. Brain Res.* 169, 564–568. doi: 10.1007/s00221-006-0346-5

Voegele, C., Sorg, S., Studtmann, M., and Weber, H. (2010). Cardiac autonomic regulation and anger coping in adolescents. *Biol. Psychol.* 85, 465–471. doi: 10.1016/j.biopsycho.2010.09.010

Von Neumann, J., and Morgenstern, O. (2007). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton university press.

Wu, Y., Zhou, Y., van Dijk, E., Leliveld, M. C., and Zhou, X. (2011). Social comparison affects brain responses to fairness in asset division: an ERP study with the ultimatum game. *Front. Hum. Neurosci.* 5:131. doi: 10.3389/fnhum.2011.00131

Xiang, T., Lohrenz, T., and Montague, P. R. (2013). Computational substrates of norms and their violations during social exchange. *J. Neurosci.* 33, 1099–1108. doi: 10.1523/jneurosci.1642-12.2013

Xiao, E., and Houser, D. (2005). Emotion expression in human punishment behavior. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7398–7401. doi: 10.1073/pnas.0502399102

Yamagishi, T., Horita, Y., Takagishi, H., Shinada, M., Tanida, S., and Cook, K. S. (2009). The private rejection of unfair offers and emotional commitment. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11520–11523. doi: 10.1073/pnas.0900636106

Yeung, N., and Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *J. Neurosci.* 24, 6258–6264. doi: 10.1523/jneurosci.4537-03.2004

Zhou, X., and Wu, Y. (2011). Sharing losses and sharing gains: increased demand for fairness under adversity. *J. Exp. Soc. Psychol.* 47, 582–588. doi: 10.1016/j.jesp.2010.12.017

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

MINI REVIEW
published: 10 July 2018
doi: 10.3389/fnins.2018.00461

Neuroimaging Studies Reveal the Subtle Difference Among Social Network Size Measurements and Shed Light on New Directions

Xiaoming Liu[1,2], Shen Liu[1], Ruiqi Huang[3], Xueli Chen[3], Yunlu Xie[3], Ru Ma[3], Yuzhi Luo[1], Junjie Bu[3] and Xiaochu Zhang[1,3,4,5]**

[1] School of Humanities and Social Science, University of Science and Technology of China, Hefei, China, [2] School of Foreign Languages, Anhui Jianzhu University, Hefei, China, [3] CAS Key Laboratory of Brain Function and Disease and School of Life Sciences, University of Science and Technology of China, Hefei, China, [4] Hefei Medical Research Center on Alcohol Addiction, Anhui Mental Health Center, Hefei, China, [5] Centers for Biomedical Engineering, University of Science and Technology of China, Hefei, China

OPEN ACCESS

Edited by:
Hang Ye,
Zhejiang University, China

Reviewed by:
James P. Curley,
University of Texas at Austin,
United States
Xiaoxia Wang,
Third Military Medical University
(TMMU), China

***Correspondence:**
Xiaochu Zhang
zxcustc@ustc.edu.cn
Shen Liu
liushenpsy@ustc.edu.cn

Specialty section:
This article was submitted to
Decision Neuroscience,
a section of the journal
Frontiers in Neuroscience

Received: 15 March 2018
Accepted: 18 June 2018
Published: 10 July 2018

Citation:
Liu X, Liu S, Huang R, Chen X, Xie Y, Ma R, Luo Y, Bu J and Zhang X (2018) Neuroimaging Studies Reveal the Subtle Difference Among Social Network Size Measurements and Shed Light on New Directions. Front. Neurosci. 12:461. doi: 10.3389/fnins.2018.00461

Social network size is a key feature when we explore the constructions of human social networks. Despite the disparate understanding of individuals' social networks, researchers have reached a consensus that human's social networks are hierarchically organized with different layers, which represent emotional bonds and interaction frequency. Social brain hypothesis emphasizes the significance of complex and demanding social interaction environments and assumes that the cognitive constraints may have an impact on the social network size. This paper reviews neuroimaging studies on social networks that explored the connection between individuals' social network size and neural mechanisms and finds that Social Network Index (SNI) and Social Network Questionnaires (SNQs) are the mostly-adopted measurements of one's social network size. The two assessments have subtle difference in essence as they measure the different sublayers of one's social network. The former measures the relatively outer sub-layer of one's stable social relationship, similar to the sympathy group, while the latter assesses the innermost layer—the core of one's social network, often referred to as support clique. This subtle difference is also corroborated by neuroimaging studies, as SNI-measured social network size is largely correlated with the amygdala, while SNQ-assessed social network size is closely related to both the amygdala and the orbitofrontal cortex. The two brain regions respond to disparate degrees of social closeness, respectively. Finally, it proposes a careful choice among the measurements for specific purposes and some new approaches to assess individuals' social network size.

Keywords: social network size, brain regions, social brain hypothesis, SNI, SNQ

INTRODUCTION

Exploration of the features and constructions of human's social networks has a long history in both the sociological and social anthropological research fields (Lev-Ari, 2018). In contrast to the traditional ecological approaches, the recent attempt to explain the evolution of sociality in primates, known as social brain hypothesis emphasizes the significance of complex social environments in which primates live and assumes that the cognitive constraints may have an impact on social grouping patterns (Liu et al., 2018).

For many species, particularly for primates, living in groups is a major adaptive advantage. But living in a social group also presents its own challenges. To get along while getting ahead, it is necessary to learn who is who, who is friend and who is foe (Bickart et al., 2011). Accordingly, maintaining a stable social group is quite cognitively demanding (Dunbar, 2012). Thus, primate brain evolution was driven by the need to acquire the competence to manage complex social relationships effectively (Dunbar and Shultz, 2007; Liu et al., 2018).

Researches on social networks have concentrated on two major issues. One is to find the limiting size of social networks, while the other is to explain the possible factors that result in the individual difference in social network size. They inevitably raised fundamental questions about the nature of social networks and how they should be defined (Stiller and Dunbar, 2007).

In spite of the disparate definitions of social network among researches, at least one consensus has been reached, that is social networks are hierarchically organized, consisting of different layers, which reflect emotional connection and interaction frequency among individuals (Carmona and Gomila, 2016; Dunbar, 2016; Kardos et al., 2017; Spiegel et al., 2018). The innermost layer (often referred to as support clique) is the core of one's social network, and is understood as the number of individuals from whom one would seek personal advice or help in case of emotional and financial difficulties (Parkinson et al., 2018). Support cliques are embedded in a larger network that is often discerned as the sympathy group which is a set of individuals one contacts at least once a month and has special ties to. The above-mentioned different levels of social networks constitute an individual's stable social relationships maintained over a period of time (Stiller and Dunbar, 2007). The outer layers are rather unstable, including all kinds of acquaintances that one would not consider as friends or family, but know well enough to have a conversation with or put names to their faces (Dunbar, 2016).

In this paper, we briefly review studies concerning the connection between people's social network size and its underlying neural mechanisms. Throughout studies we find two social network size measurements that are widely-used in different studies. Upon closer examination, however, we spot the subtle difference between them, and subsequently find evidence from the underlying brain mechanisms to corroborate our findings.

SOCIAL NETWORK SIZE MEASUREMENTS

At the operational level, how to measure the size of individuals' multi-layer social networks is of practical importance. Throughout existing literature, two major types of measurements of social network size are frequently used.

Social Network Index (SNI) contains 12 different roles of people playing in their social networks. For each role, respondents are supposed to identify whether they have the particular relationship in the first place, and then choose the number of people they see or talk to on a regular basis (i.e., at least once every 2 weeks). Thus, the social network size can be computed by summing the total number of people in the 12 roles (Cohen et al., 1997; Peng et al., 2018).

Social Network Questionnaire (SNQ) and Norbeck Social Support Questionnaire (NSSQ) are the other frequently used types of measurements of social network size. In effect, the two questionnaires share many similarities. In SNQ, respondents are required to write down the names of their frequent contacts, and then they should identify those whom they would seek advice or comfort for a major personal problem like serious accidents or death of loved ones (Stiller and Dunbar, 2007). NSSQ requests respondents to list the names of network members who provide them with personal support and then rate each of their network members on a Likert scale by answering nine questions, such as "How much does this person make you feel liked or loved?" (Norbeck et al., 1981; Hampton et al., 2016).

COMPARISON OF THE MEASUREMENT TOOLS

Comparing the above-mentioned two types of measurements about social network size (SNI and SNQ) in the perspective of social network organization, it can be found that they both focus on the primary inner layer of individuals' social networks. However, the two assessments have subtle differences in essence as they measure the disparate sub-layers of social network size.

To be specific, what has been assessed in SNI is similar to the size of one's sympathy group within a social network. Faced with SNI questions like "How many people at work do you talk to at least once every 2 weeks?" Respondents are very likely to include acquaintances without such a strong emotional bond, such as seeking personal advice or help in times of severe distress. However, what has been tested through SNQ can be regarded as the size of the innermost layer—the support clique of one's social network, since one would always turn to those people for material and emotional support (Stiller and Dunbar, 2007; Ramirez and Palacios, 2016). In brief, compared with SNI, SNQ would arouse such a stronger affective feeling of being supported and adored as to remind respondents of the people at the core of their social network.

MRI FINDINGS

Advances in MRI analytics now provide tools to study brain–behavior relationships at the level of circuits and networks. Throughout studies on the connection between individuals' social network size and brain mechanisms, the seemingly disparate findings can be pulled together to corroborate the subtle difference among social network size measurements.

Studies With SNI Measurement

Bickart et al. (2011) performed a quantitative morphometric analysis of T1 weighted MRI data from 58 participants. The linear regression analysis reveals that individuals' social network size is positively correlated with their amygdala volume.

Jasper (2013) investigated the correlation between the amygdala activation and social network size in HIV patients. In the research, emotional pictures of angry and fearful faces were displayed to illicit robust amygdala responses during an fMRI task. The result shows that there is a significant correlation between right amygdala activation and individuals' social network size.

Dziura and Thompson (2014) recorded the motion of sensors attached to the limbs and torso of an actor and turned them into point-light arrays to present to the participants in an fMRI scanning. It is shown that the posterior superior temporal sulcus (pSTS), the amygdala and the fusiform gyrus are significantly activated in the perception of biological motions. Further exploration of the relationship between these cortical areas and social networks demonstrates that the amygdala and the pSTS activation are closely correlated with social network size. In addition, Lewis et al. (2011) conducted intentionality and memory tasks using a series of five short stories to test subjects' ability to correctly infer the mind states like the beliefs of the characters in the story. The result shows that the ventromedial prefrontal cortex (vmPFC) volume predicts understanding of others and social network size.

Bickart et al. (2012) used three seed regions—the lateral orbitofrontal cortex (lOFC), the vmPFC and the caudual anterior cingulate cortex (cACC) to identify voxels within the amygdala with the strongest connectivity, thus, three subregions within the amygdala were parceled, being ventrolateral, dorsal, medial amygdala, respectively. In addition, they used these three subregions as seeds to conduct a whole-brain exploration, and built a network sharing functional connectivity with each amygdala seed. The result demonstrates that the ventrolateral amygdala and the medial amygdala networks can predict social network size.

Studies With SNQ or NSSQ Measurement

Powell et al. (2012) focused on the PFC region and further divided the region into sub-regions, as dorsal and orbital prefrontal regions. The path analysis indicates that the orbital PFC volume is the best predictor of social network size.

Kanai et al. (2012) examined the correlation between gray matter density and social network size. The right amygdala density stood itself out to be significantly correlated with individuals' social network size.

Heide et al. (2014) directed the participants to view their friends' pictures and unfamiliar faces during an fMRI task. The result shows a significant BOLD activation in bilateral amygdala. In addition, the following VBM result indicates a positive correlation between gray matter volume in bilateral amygdala and bilateral OFC and social network size.

Similarly, Preller et al. (2014) conducted a social gaze task in an fMRI scanning, and found out a significant positive correlation between social network size and the medial OFC activation in the healthy control group.

Apart from that, white matter connectivity among different brain regions could also predict social network size as demonstrated in Hampton's research. The diffusion-weighted imaging (DWI) result showed that the amygdala-OFC and the amygdala-ATL (anterior temporal lobes) white matter microstructure as well as age factor accounted for 69% of the variability in social network size (Hampton et al., 2016).

A careful examination can elicit an explicit tendency of all research results—SNI measured social network size is largely correlated with the amygdala, while SNQ assessed social network size is closely related to both the amygdala and the OFC.

DISCUSSION

The above-mentioned studies show that individuals with larger social networks have more gray matters and better function in brain regions implicated in adaptive social behaviors. The main goal of this review is to test whether the measurements of social network size are equivalent through neuroimaging study results. Taken together, our findings showed that structures and functions of the amygdala, the orbitofrontal cortex (OFC), the pSTS, and the vmPFC could predict the size of one's social network. However, the OFC region was more saliently correlated to one's innermost layer of social network, which revealed the subtle difference between the two measurements.

The fact that individuals with larger social network size have larger amygdala volume provides plausible evidence to the social brain hypothesis that primates evolved under the pressure of increasingly complex social life. The larger amygdala enables us to perceive social cues, and allow us to devise complex strategies to cooperate or compete with others more efficiently.

It is widely accepted that the amygdala is important for the recognition and processing of negative and positive emotions (Baxter and Murray, 2002; Dennison et al., 2015). When the pleasurable social cues are identified, the activation of the amygdala promotes social affiliation behaviors, adjusts social aversion behavior and improves interpersonal relationship in a larger social network (Preller et al., 2014). In addition, individuals with stronger intrinsic amygdala connectivity within other networks, for instance the perception network is better at decoding the meaning of social cues and dealing with larger amount of people in more complex social contexts (Bickart et al., 2012). To sum up, the amygdala enlargement or its activation or its structural and functional connectivity could predict the inner layer (both sympathy group and the support clique) of individuals' social network size, regardless of the measurements being used. In contrast, another brain region—the OFC correlated with social network size is mostly identified by SNQ.

The OFC has been proved to be involved in a range of social functions. Intentionality is the ability to explain and predict the behavior of others. This social cognitive capacity has been demonstrated to be connected to the volume of the OFC (Powell et al., 2012). The result shows that a greater volume of the OFC means a better understanding of others, which contributes to maintaining a larger size of social network. Apart from that, previous studies found that the OFC volume or thickness could predict olfactory sensitivity (Frasnelli et al., 2010; Seubert et al., 2013), which in turn positively correlates to social network size (Zou et al., 2016). Even though the causal relationship is not clear, this result suggests that individuals with higher olfactory sensitivity are more sensitive to others' body odor and can obtain more social chemical signals which facilitate social communication. Furthermore, empathy is the critical social skill in understanding what another person is experiencing (Preller et al., 2014); and the OFC activations were observed in empathetic behaviors (Matsudaira et al., 2017), which were more frequent among close or loved ones than unfamiliar companions (Romero et al., 2010). Those findings imply the connection, even though not causal relationship between the OFC activation and the innermost layer of one's social network.

In addition, the anatomical location of the OFC is in the front end of the mesolimbic reward circuit (Rushworth et al., 2011), and it receives signals directly from visual, olfactory, taste, and somatosensory areas (Tanaka et al., 2016). Neuroimaging studies also found that the OFC was activated by pleasant touch, rewarding and aversive taste, and damage to the OFC impaired the learning of stimulus-reinforcement associations (Rolls, 2000; Dixon et al., 2017). As for the different types of rewards, social rewards like improving feelings of self-worth and importance through praise and the attention from others are the extremely important motivators for social interaction (Elliot et al., 2006; Izuma et al., 2008). Besides, recent studies showed that increased social interaction would enhance social reward, represented by the activation in the OFC, the mPFC, and the striatum of the reward system (Fareri et al., 2015; Kawamichi et al., 2016). Therefore, when assessed with SNQ which measures the most frequently interacted social network, the OFC activation would be more salient than that measured by SNI.

Apart from the ROI regions like the amygdala and the OFC, other brain regions are also found significantly related to one's social network size, like the pSTS, the vmPFC, etc. It is widely acknowledged that the cognitive capacity of inferring the mental states of others is crucial to human sociality, according to the theory of mind (ToM; Frith and Frith, 2003). Neuroimaging studies have associated this ability with specific brain regions, as above-mentioned the pSTS and the vmPFC (Frith and Frith, 2003; Gallagher and Frith, 2003; Saxe and Kanwisher, 2003; Mahy et al., 2014). During ToM processing, the vmPFC is frequently active in identification of goals and intentions in a wide range of tasks (Gallagher et al., 2000; German et al., 2004; Lewis et al., 2011). In addition, the vmPFC is also proved to be involved in decoupling the perspectives of other people from one's own (Gallagher and Frith, 2003). Equipped with theses capacities, individuals can infer the other person's intention, separate out various layers between their acquaintances and themselves and maintain a large and stable social network. Turning to the pSTS, the recent study shows that this brain region responds strongly when perceiving social interactions (Isik et al., 2017). Besides, it is believed that the pSTS is involved in the perception of non-verbal social signals (Kanai et al., 2012; Dziura and Thompson, 2014), which can help reduce ambiguity and uncertainty. Therefore, greater functioning of the pSTS permits individuals to detect social cues and keep a larger size of social network (Goldin-Meadow and Beilock, 2010).

FUTURE DIRECTIONS

Human's social networks are hierarchically organized with different layers, which represent emotional bonds and interaction frequency among individuals. In terms of the measurement of social network size, SNI and SNQ are most frequently used. A careful examination of these two measurements in view of the hierarchical organization of social networks reveals that the two assessments are dissimilar in effect. Neuroimaging researches shed light on a new perspective as they uncover the underlying neural mechanisms of human's social networks. Throughout the existing literature, social network size measured by SNI is largely correlated with the amygdala, while social network size assessed by SNQ is closely related to both the amygdala and the OFC, which provides evidence to the subtle difference between the two measure tools. This finding sheds new light on the understanding of the subtle distinctions among various social network assessments and suggests that we should choose the most suitable one for specific research purpose, since our brain would react distinctively to social interactions with dissimilar emotional closeness.

In recent years, the rise of the Internet has provided an opportunity to study social networks on a larger scale (Hayat et al., 2017). A key element of social networks is the ability for individuals to simultaneously interact in multiple social contexts by maintaining different types of social ties. The overlay of several networks on the same set of nodes (individuals) is called a multiplex network (MPN). The MPN facilitates the description, quantification, and analysis of complex sets of relationships among individuals (Hayat et al., 2017; Bilecen et al., 2018). Lately, Parkinson et al. (2018) proposed a new approach to characterize individuals' social network. They recruited an entire cohort of students in a graduate program and asked them to complete an online survey in which they indicated the individuals in the program with whom they were friends. Given that a mutually reported tie is a stronger indicator of the presence of a friendship than an unreciprocated tie, a graph consisting only of reciprocal social ties was used to estimate social distances between individuals.

Future studies could also focus on other dimensions of social network like its diversity and embeddedness. In SNI, social network diversity is represented by the number of social roles in which the participants have regular contact with at least one person; and social network embeddedness is represented by the number of different network domains in which a participant is active (Dziura and Thompson, 2014; Molesworth et al., 2015). Which layer do these measurements exactly focus on? Do they assess the same thing in essence? Answers to these questions would provide us with better comprehension of the essence of measurements and a new perspective of balancing the advantages of each measurement against their shortcomings.

AUTHOR CONTRIBUTIONS

XZ, XL, and SL conceived and designed the writing frame. XL and SL wrote the paper. XL, SL, RH, XC, YX, RM, YL, JB, and XZ revised the manuscript.

FUNDING

This work was supported by grants from the National Key Basic Research Program (2016YFA0400900) and the National Natural Science Foundation of China (31471071, 31771221, and 61773360) and the Fundamental Research Funds for the Central Universities of China, and the MOE-Mircrosoft Key Laboratory of University of Science and Technology of China.

REFERENCES

Baxter, M. G., and Murray, E. A. (2002). The amygdala and reward. *Nat. Rev. Neurosci.* 3, 563–573. doi: 10.1038/nrn875

Bickart, K. C., Hollenbeck, M. C., Barrett, L. F., and Dickerson, B. C. (2012). Intrinsic amygdala-cortical functional connectivity predicts social network size in humans. *J. Neurosci.* 32, 14729–14741. doi: 10.1523/JNEUROSCI.1599-12.2012

Bickart, K. C., Wright, C. I., Dautoff, R. J., Dickerson, B. C., and Barrett, L. F. (2011). Amygdala volume and social network size in humans. *Nat. Neurosci.* 14, 163–164. doi: 10.1038/nn.2724

Bilecen, B., Gamper, M., and Lubbers, M. J. (2018). The missing link: social network analysis in migration and transnationalism. *Soc. Netw.* 53, 1–3. doi: 10.1016/j.socnet.2017.07.001

Carmona, C. A., and Gomila, A. (2016). A critical review of Dunbar's social brain hypothesis. *Rev. Int. Soc.* 74:e037. doi: 10.3989/ris.2016.74.3.037

Cohen, S., Doyle, W. J., Skoner, D. P., Rabin, B. S., and Gwaltney, J. M. Jr. (1997). Social ties and susceptibility to the common cold. *JAMA* 277, 1940–1944. doi: 10.1001/jama.1997.03540480040036

Dennison, M., Whittle, S., Yücel, M., Byrne, M. L., Schwartz, O., Simmons, J. G., et al. (2015). Trait positive affect is associated with hippocampal volume and change in caudate volume across adolescence. *Cogn. Affect. Behav. Neurosci.* 15, 80–94. doi: 10.3758/s13415-014-0319-2

Dixon, M. L., Thiruchselvam, R., Todd, R., and Christoff, K. (2017). Emotion and the prefrontal cortex: an integrative review. *Psychol. Bull.* 143, 1033–1081. doi: 10.1037/bul0000096

Dunbar, R. I. M. (2012). The social brain meets neuroimaging. *Trends Cogn. Sci.* 16, 101–102. doi: 10.1016/j.tics.2011.11.013

Dunbar, R. I. (2016). Do online social media cut through the constraints that limit the size of offline social networks? *R. Soc. Open Sci.* 3:150292. doi: 10.1098/rsos.150292

Dunbar, R. I. M., and Shultz, S. (2007). Evolution in the social brain. *Science* 317, 1344–1347. doi: 10.1126/science.1145463

Dziura, S. L., and Thompson, J. C. (2014). Social-network complexity in humans is associated with the neural response to social information. *Psychol. Sci.* 25, 2095–2101. doi: 10.1177/0956797614549209

Elliot, A. J., Gable, S. L., and Mapes, R. R. (2006). Approach and avoidance motivation in the social domain. *Pers. Soc. Psychol. Bull.* 32, 378–391. doi: 10.1177/0146167205282153

Fareri, D. S., Chang, L. J., and Delgado, M. R. (2015). Computational substrates of social value in interpersonal collaboration. *J. Neurosci.* 35, 8170–8180. doi: 10.1523/JNEUROSCI.4775-14.2015

Frasnelli, J., Lundstrom J. N., Boyle, J. A., Djordjevic, J., Zatorre, R. J., and Jones, G. M. (2010). Neuroanatomical correlates of olfactory performance. *Exp. Brain Res.* 201, 1–11. doi: 10.1007/s00221-009-1999-7

Frith, U., and Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 358, 459–473. doi: 10.1098/rstb.2002.1218

Gallagher, H. L., and Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends Cogn. Sci.* 7, 77–83. doi: 10.1016/S1364-6613(02)00025-6

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., and Frith, C. D. (2000). Reading the mind in cartoons and stories: an fmri study of 'theory of the mind' in verbal and nonverbal tasks. *Neuropsychologia* 38, 11–21. doi: 10.1016/S0028-3932(99)00053-6

German, T. P., Niehaus, J. L., Roarty, M. P., Giesbrecht, B., and Miller, M. B. (2004). Neural correlates of detecting pretense: automatic engagement of the intentional stance under covert conditions. *J. Cogn. Neurosci.* 16, 1805–1817. doi: 10.1162/0898929042947892

Goldin-Meadow, S., and Beilock, S. L. (2010). Action's influence on thought: the case of gesture. *Perspect. Psychol. Sci.* 5, 664–674. doi: 10.1177/1745691610388764

Hampton, W. H., Unger, A., Von Der Heide, R. J., and Olson, I. R. (2016). Neural connections foster social connections: a diffusion-weighted imaging study of social networks. *Soc. Cogn. Affect. Neurosci.* 11, 721–727. doi: 10.1093/scan/nsv153

Hayat, T. Z., Lesser, O., and Samuel-Azran, T. (2017). Gendered discourse patterns on online social networks: a social network analysis perspective. *Comput. Hum. Behav.* 77, 132–139. doi: 10.1016/j.chb.2017.08.041

Heide, R. V. D., Vyas, G., and Olson, I. R. (2014). The social network-network: size is predicted by brain structure and function in the amygdala and paralimbic regions. *Soc. Cogn. Affect. Neurosci.* 9, 1962–1972. doi: 10.1093/scan/nsu009

Isik, L., Koldewyn, K., Beeler, D., and Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proc. Natl. Acad. Sci. U.S.A.* 114, E9145–E9152. doi: 10.1073/pnas.1714471114

Izuma, K., Saito, D. N., and Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron* 58, 284–294. doi: 10.1016/j.neuron.2008.03.020

Jasper, C. (2013). Amygdalae enlargement and activation are associated with social network complexity in individuals with Human Immunodeficiency Virus (HIV). *Undergrad. Rev.* 9, 68–74.

Kanai, R., Bahrami, B., Roylance, R., and Rees, G. (2012). Online social network size is reflected in human brain structure. *Proc. Biol. Sci.* 279, 1327–1334. doi: 10.1098/rspb.2011.1959

Kardos, P., Leidner, B., Pléh, C., Soltész, P., and Unoka, Z. (2017). Empathic people have more friends: empathic abilities predict social network size and position in social network predicts empathic efforts. *Soc. Netw.* 50, 1–5. doi: 10.1016/j.socnet.2017.01.004

Kawamichi, H., Sugawara, S. K., Hamano, Y. H., Makita, K., Kochiyama, T., and Sadato, N. (2016). Increased frequency of social interaction is associated with enjoyment enhancement and reward system activation. *Sci. Rep.* 6:24561. doi: 10.1038/srep24561

Lev-Ari, S. (2018). Social network size can influence linguistic malleability and the propagation of linguistic change. *Cognition* 176, 31–39. doi: 10.1016/j.cognition.2018.03.003

Lewis, P. A., Rezaie, R., Brown, R., Roberts, N., and Dunbar, R. I. (2011). Ventromedial prefrontal volume predicts understanding of others and social network size. *Neuroimage* 57, 1624–1629. doi: 10.1016/j.neuroimage.2011.05.030

Liu, Y., Wu, B., Petti, C., Wu, X. H., and Han, S. H. (2018). Self-construals moderate associations between trait creativity and social brain network. *Neuropsychologia* 111, 284–291. doi: 10.1016/j.neuropsychologia.2018.02.012

Mahy, C. E. V., Moses, L. J., and Pfeifer, J. H. (2014). How and where: theory-of-mind in the brain. *Dev. Cogn. Neurosci.* 9, 68–81. doi: 10.1016/j.dcn.2014.01.002

Matsudaira, I., Kawashima, R., and Taki, Y. (2017). Structural brain development in healthy children and adolescents. *Brain Nerve* 69, 539–545. doi: 10.11477/mf.1416200780

Molesworth, T., Sheu, L. K., Cohen, S., Gianaros, P. J., and Verstynen, T. D. (2015). Social network diversity and white matter microstructural integrity in humans. *Soc.Cogn. Affect. Neurosci.* 10, 1169–1176. doi: 10.1093/scan/nsv001

Norbeck, J. S., Lindsey, A. M., and Carrieri, V. L. (1981). The development of an instrument to measure social support. *Nurs. Res.* 30, 264–269. doi: 10.1097/00006199-198109000-00003

Parkinson, C., Kleinbaum, A. M., and Wheatley, T. (2018). Similar neural responses predict friendship. *Nat. Commun.* 9:332. doi: 10.1038/s41467-017-02722-7

Peng, S. C., Zhou, Y. M., Cao, L. H., Yu, S., Niu, J. W., and Jia, W. J. (2018). Influence analysis in social networks: a survey. *J. Netw. Comput. Appl.* 106, 17–32. doi: 10.1016/j.jnca.2018.01.005

Powell, J., Lewis, P. A., Roberts, N., Garcia-Finana., M., and Dunbar, R. I. (2012). Orbital prefrontal cortex volume predicts social network size: an imaging study of individual differences in humans. *Proc. Biol. Sci.* 279, 2157–2162. doi: 10.1098/rspb.2011.2574

Preller, K. H., Herdener, M., Schilbach, L., Stämpfli, P., Hulka, L. M., Vonmoos, M., et al. (2014). Functional changes of the reward system underlie blunted response to social gaze in cocaine users. *Proc. Natl. Acad. Sci. U.S.A.* 111, 2842–2847. doi: 10.1073/pnas.1317090111

Ramirez, L., and Palacios, X. (2016). Stereotypes about old age, social support, aging anxiety and evaluations of one's own health. *J. Soc. Issues* 72, 47–68. doi: 10.1111/josi.12155

Rolls, E. T. (2000). The orbitofrontal cortex and reward. *Cereb. Cortex* 10, 284–294. doi: 10.1093/cercor/10.3.284

Romero, T., Castellanos, M. A., and de Waal, F. B. (2010). Consolation as possible expression of sympathetic concern among chimpanzees. *Proc. Natl. Acad. Sci. U.S.A.* 107, 12110–12115. doi: 10.1073/pnas.1006991107

Rushworth, M. F. S., Noonan, M. A. P., Boorman, E. D., Walton, M. E., and Behrens, T. E. (2011). Frontal cortex and reward-guided learning and decision-making. *Neuron* 70, 1054–1069. doi: 10.1016/j.neuron.2011.05.014

Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage* 19, 1835–1842. doi: 10.1016/S1053-8119(03)00230-1

Seubert, J., Freiherr, J., Frasnelli, J., Hummel, T., and Lundström, J. N. (2013). Orbitofrontal cortex and olfactory bulb volume predict distinct aspects of olfactory performance in healthy subjects. *Cereb. Cortex* 23, 2448–2456. doi: 10.1093/cercor/bhs230

Spiegel, O., Sih, A., Leu, S. T., and Bull, C. M. (2018). Where should we meet? Mapping social network interactions of sleepy lizards shows sex-dependent social network structure. *Anim. Behav.* 136, 207–215. doi: 10.1016/j.anbehav.2017.11.001

Stiller, J., and Dunbar, R. I. M. (2007). Perspective-taking and memory capacity predict social network size. *Soc. Netw.* 29, 93–104. doi: 10.1016/j.socnet.2006.04.001

Tanaka, S. C., Doya, K., Okada, G., Ueda, K., Okamoto, Y., and Yamawaki, S. (2016). Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat. Neurosci.* 7, 887–893. doi: 10.1038/nn1279

Zou, L. Q., Yang, Z. Y., Yi, W., Lui, S. S. Y., Chen, A. T., Cheung, E. F. C., et al. (2016). What does the nose know? Olfactory function predicts social network size in human. *Sci. Rep.* 6:25026. doi: 10.1038/srep25026

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ORIGINAL RESEARCH
published: 19 September 2018
doi: 10.3389/fpsyg.2018.01683

Social Interaction Patterns of the Disabled People in Asymmetric Social Dilemmas

Shen Liu[1,2]†, Wenlan Xie[1,3]†, Shangfeng Han[1], Zhongchen Mou[1,4], Xiaochu Zhang[2,5,6,7] and Lin Zhang[1]**

[1] Department of Psychology and Institute of Psychology, Ningbo University, Ningbo, China, [2] School of Humanities and Social Science, University of Science and Technology of China, Hefei, China, [3] Ningbo Institute of Education, Ningbo, China, [4] School of Psychology, Nanjing Normal University, Nanjing, China, [5] Hefei Medical Research Center on Alcohol Addiction, Anhui Mental Health Center, Hefei, China, [6] CAS Key Laboratory of Brain Function and Disease, and School of Life Sciences, University of Science and Technology of China, Hefei, China, [7] Academy of Psychology and Behavior, Tianjin Normal University, Tianjin, China

The social participation of the disabled people is unsatisfactory and low, one of the reasons often overlooked but of great importance may lie in the disparate patterns of social interaction between the disabled people and the abled people. The current study respectively recruited 41 and 80 disabled people in two experiments and adopted give-some games and public good dilemma to explore social interaction patterns between the disabled abled people. The results were as follows: (1) the disabled people preferred to interact with the disabled people and the abled people preferred to interact with the abled people. (2) The disabled abled people had higher cooperation, satisfaction and sense of justice when interacting with the disabled people than interacting with the abled people. (3) Advantage in the number of the disabled people could reverse their disadvantage in the identity. These results are of important practical value, which provides related theoretical support for the disabled people's federation and communities when carrying out activities for the disabled people.

Keywords: social interaction, social dilemmas, asymmetry, cooperation, the disabled people

OPEN ACCESS

Edited by:
Chao Liu,
Beijing Normal University, China

Reviewed by:
Pei Wang,
Shanghai Normal University, China
Lumei Tian,
Shandong Normal University, China

****Correspondence:***
Lin Zhang
zhanglin1@nbu.edu.cn
Xiaochu Zhang
zxcustc@ustc.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Psychology

Received: *15 March 2018*
Accepted: *21 August 2018*
Published: *19 September 2018*

Citation:
Liu S, Xie W, Han S, Mou Z, Zhang X and Zhang L (2018) Social Interaction Patterns of the Disabled People in Asymmetric Social Dilemmas. Front. Psychol. 9:1683. doi: 10.3389/fpsyg.2018.01683

INTRODUCTION

With the continuous improvement of social security for the disabled people, the objective conditions such as income and living conditions of the disabled people have improved remarkably while the quality of social life of the disabled people has not. In particular, the social participation of the disabled people is still unsatisfactory and low (Zhang et al., 2014). The reasons for the current social participation of the disabled people are various, such as their limited physical abilities (Lu et al., 2017), perceived discrimination resulting from physical disabilities (Zhang et al., 2014). However, one of the reasons often overlooked but of great importance may lie in the disparate patterns of social interaction between the disabled and abled people. Because of their own and social reasons, the disabled people are at a disadvantage in their social status and resource distribution. Therefore, they are in an unequal position in their interactions with abled people, which results in their low level of social participation.

The social interactions of the disabled people are not optimistic with simple social network, simple social interactive object and low social interaction willingness as the main manifestations of their difficulties in social interactions. Reasons for the disabled people's difficulties in social interactions may lie in the following aspects: (1) one is that the abled people tend to show negative attitudes (e.g., social stigma) and behaviors to the disabled people in daily lives (Zhang et al., 2015). For example, it is reported that the disabled people claimed discriminations from their peers (Moore et al., 2011) and families (O'Reilly et al., 2016). The abled people have an inherent prejudice against the disabled people that results in the formation of public stigma (Forber-Pratt et al., 2017). While on the other hand, the disabled people internalize the public stigma. They recognize and accept the cultural stereotype of the group they live in and apply it to themselves, and thus self-stigma forms (Ditchman et al., 2013); (2) the other possible reason is the unequal status in social interactions between the disabled and abled people. The disabled are on the fringes of society, both in terms of accessing to social resources and the distribution of living environment as well as working opportunities, socio-economic status and quality of life, compared with abled people (Riddell and Weedon, 2014). Such an unequal status may be the core reason why the disabled people do not want to participate in social interactions and establish good relationships with other people, especially the abled people. However, the studies aiming directly at the social interaction patterns between the disabled and abled people in unequal status are still rare.

Cooperation and competition are basic forms of social interactions. People need to have social interactions and exchanges of resources with others for survival and development in society (Wang et al., 2016). Cooperative behavior is an important factor in maintaining good social interactions as well as the redistribution of interests of social subjects. Therefore, cooperative behavior is of particular importance for the disabled people who are at a disadvantage of resource distribution and social status. For example, a possible reason why the disabled people show low cooperation is that they may take into account their own economic conditions or status when interacting with the abled people in the hope of making up for these deficiencies in the distribution of resources so as to distribute more resources to themselves and less resources to the abled people. Even worse, the disabled people's low cooperation and unfriendliness may in turn affect the abled people's attitudes and behaviors toward them, such as indifference or avoidance (unwillingness to interact with the disabled people). In summary, the different patterns of social interactions between the disabled and abled people may be the important reasons for the difficulties the disabled people encounter when integrating into the society. In addition, the disabled people's disadvantaged status caused by physical or mental defects are unlikely to change in a short period and are irreversible (Brown and Brunell, 2017). Based on the theories of cooperation and competition under unequal status and social dilemmas paradigms in the field of decision making, the current study intended to explore social interaction patterns between the disabled and abled people under the conditions of unequal resources and status.

Social dilemma is a situation in which the interests of an individual and groups conflict. The benefits of members choosing not to cooperate in this situation are higher than those of choosing cooperation, but the overall benefit of all members choosing to cooperate is greater than the benefit of defection (Roch and Samuelson, 1997; Wang and Chen, 2011; Liu and Hao, 2014). Based on the hypothesis of "rational man," the classical game theory holds that the two parties of a game will make their own decisions in accordance with the maximization of their own interests in a dilemma situation. However, in many social dilemmas, the two parties of a game still choose to cooperate which seems irrational and the distribution tend to be fair (Roth, 1991; Martínez-Cánovas et al., 2016). In addition, individuals are constrained by their own resources and interactive situations when weighing their own and others' interests. In social dilemma models, researchers often assume that participants have equal resources or status prior to distribution. However, in real lives, the two competing parties often have strengths and weaknesses due to a variety of reasons, which in turn will lead to different levels of dominance. When it is projected into social interactions, asymmetric social dilemmas are formed (Liu and Hao, 2015). In asymmetric social dilemmas, the existence of dominance gives part of the population more opportunities than others to access resources that contribute to the survival and reproduction of them. It is also for this reason that the dominance level induces inevitable conflicts between the dominating and the dominated individuals. For example, inequalities in resources, income, power and the like can hinder cooperation from taking place (Liu and Hao, 2014; Hao et al., 2016). However, some studies showed that a certain degree of inequality had a positive effect on cooperative behaviors, namely the theory of "disadvantage makes people more cooperative" (Zitek and Tiedens, 2012; He et al., 2014). Therefore, the effect of unequal status on cooperative behaviors is very complicated. However, the social dilemmas confronted by the disabled people in the current study were different from those by general population. The disadvantaged situation of the disabled people caused by physical or psychological defects was irreversible. However, in the previous studies, disadvantaged status resulting from inequality in resources, interests and power could be somewhat altered. Therefore, the effect of the irreversible inequalities on the cooperative behaviors of the disabled people may be more complex than that of the general population. Hence, exploring the patterns of social interaction between the disabled and abled people can not only help us to understand the difficulties the disabled people encounter in social interactions but also enhance the social participation of them. On the other hand, it can help us to understand the characteristics of social interactions between vulnerable groups represented by the disabled people and advantaged groups represented by the abled people.

There are many theories explaining cooperation under unequal status. For example, Trivers put forward the reciprocity theory and thought that the essence of cooperative behaviors was the exchange of interests among individuals, namely they can choose either to "cooperate" or "defect" (Trivers, 1971). Reciprocity is divided into strong and weak reciprocity, while weak reciprocity is manifested as direct reciprocity and indirect reciprocity (Tsvetkova and Buskens, 2013). Direct reciprocity occurs between two persons and its principle is "you help me and I will help you" (Sigmund, 2012). The indirect reciprocity is to gain mutual benefits from others through reputation and its principle is "you help me and others will help you" (Sigmund, 2012). However, neither direct nor indirect reciprocity can explain individuals' cooperative behaviors when they face threats such as war, plague or famine that threaten the survival of the community (Rao et al., 2011). In these situations, as the probability of group disintegration increases, the probability of survival of the entire population declines. As a result, time was extremely valuable and those who could not wait for the third parties' reward often chose to defect. However, once the defection spreads among the group, the destruction of it will soon follow. Therefore, in order to ensure that the group will not disintegrate, the members of the group will punish the betrayals at the expense of their own interests, which is called "strong reciprocity." The existence of strong reciprocity individuals ensures more benefits of the group than the price paid by the individuals. What's more, as the group is also more inclined to favor those who are willing to bear the costs and protect the interests of the masses, they are more likely to survive in the group and thus strong reciprocity evolves. Rao et al. (2011) put forward the theory of "disadvantage makes people more cooperative" and thought there was a set of system in human genes to improve the probability of reproduction and survival of individuals. Disadvantaged individuals need to increase their chances of survival and reproduction through cooperation due to their weak competiveness. Compared with the reciprocity theory, this theory can explain the mechanism of pro-social behaviors more succinctly and effectively. While the "fairness theory" is a competing assumption that indicates individuals have a perception of justice or averseness to injustice in making decisions (Xiao, 2013). During the game, both parties of the interaction will have "consensus" on "recognition of justice." In other words, they tend to reduce the unfairness during the distribution process in cooperative decision-making (Vandello et al., 2011). Hence, the reciprocity theory, the theory of "disadvantage makes people more cooperative" and the fairness theory all could explain cooperative behaviors under unequal situations to some extent. However, as a special group, the asymmetric situations formed by the interactions of the disabled people with the abled are different from those formed by laboratory manipulation. Therefore, it is also one of the issues the current study tended to explore that whether the interaction patterns between the disabled and abled people followed the above theories or assumptions.

Social dilemmas often involve two or more people (Liu and Hao, 2014). According to the number of people involved, social dilemma can be divided into two-person dilemma and multiple-person dilemma (Rand, 2017). The social interactions of individuals are not just one-to-one interactions, but interactions involving different groups, such as the disabled/abled people may interact with the abled/disabled people or mixed group (including the disabled and abled). Two-person interaction is the simplest social relationship. Although it also has the characteristics of social dilemma, social dilemma is often manifested as multiple-person social interactions (Liu and Hao, 2014). Individuals' psychological and behavioral performances in a two-person dilemma are different from those of multiple-person dilemma (EL-Seidy et al., 2016; Płatkowski, 2016). Therefore, it is necessary to study the social interaction patterns between the disabled people and the abled people in two-person and multiple-person interactions. The current study doubted whether there were changes in psychological feelings caused by the changes in the number of the disabled and abled people in multiple-person interactions. Another question the current study intents to answer is that whether there is any difference in social interaction patterns between the disabled and the abled people.

The current study carried out two experiments to investigate two-person and multiple-person social interaction patterns by using give-some games and public good dilemma. According to the previous studies (Radke et al., 2014; Gilam et al., 2015), two indexes are used to investigate individuals' social interaction behaviors and results. One is objective index, namely cooperative behaviors (distribution of resources in social dilemmas). The other is subjective index, which includes initial social interaction tendencies and psychological feelings during interaction processes. In general, the current study tried to answer four questions. Firstly, which group (the disabled people or the abled people) does the disabled people prefer to interact with? Secondly, does the disabled people have the same social interaction patterns (including cooperation and psychological feelings) as that of the abled people? Thirdly, is the social interaction patterns of the disabled people influenced by the change in status in multiple-person interactions? Fourthly, which theory of asymmetric game could better explain social interaction patterns between the disabled people and the abled people? Based on these, the current study put forward the following hypotheses. Hypothesis one: The disabled people will prefer to interact with the disabled people and the abled people will prefer to interact with the abled people. Hypothesis two: The disabled people will have higher cooperation when interacting with the disabled people than with the abled people, the abled people will have higher cooperation when interacting with the disabled people than with the abled people. Hypothesis three: The disabled people will have higher satisfaction and sense of justice when interacting with the disabled people than with the abled people. Hypothesis four: Asymmetric status will affect the disabled and abled people's cooperation and the disabled people can use their superiority in number to make up their inferiority in the status.

The current study aimed to reveal the disabled people's strategies of selection in the face of conflicts between personal interests and others' interests under social dilemmas to further explore human nature and to promote altruistic behaviors of human beings and social development.

EXPERIMENT 1

Participants

The current experiment randomly recruited 41 disabled people including 23 males and eighteen females with an average age of 51.65 years old (*SD* = 10.55) and forty abled people were randomly recruited including twenty-two males and eighteen females with an average age of 50.25 years old (*SD* = 10.59). All disabled participants met the national standard for disabled people and they were mostly Grade II or Grade III of physical disability (mainly as disabilities in arms of legs) with the mean time of disability for 23.6 years. All participants had normal or corrected-to-normal vision, with no partial tritanopia or achromatopsia, and could skillfully operate the computer. The present study was approved by the Ethics Committee of the authors' University in accordance with the ethical principles of the Declaration of Helsinki. All subjects gave written informed consent in accordance with the ethical principles of the Declaration of Helsinki.

Materials

Social Interaction Tendency

The measurement of social interaction tendency was based on previous studies (Radke et al., 2014; Gilam et al., 2015). We respectively adopted a question to measure social interaction tendency including "Which group do you prefer to contact/communicate with in daily life? 1 = Disabled people, 2 = Abled people."

Cooperation

The measurement of cooperation level was based on Liu and Hao (2014). Repetitive give-some games were adopted to set up the social dilemmas of two-person interactive situations. Before the experiment, participants owned certain amount of initial monetary resource, which was 100 RMB. Then, participants could distribute initial monetary resource to the others. The amount of the distribution represented the participants' cooperation level. Subsequently, participants distributed initial monetary resource according to the instructions (see details in the **Supplementary Information**).

Psychological Feelings

The measurement of psychological feelings (satisfaction and justice) during social interactions was based on previous studies (Radke et al., 2014; Gilam et al., 2015). For satisfaction, the question was "Are you satisfied with the previous round of interaction, including the performance of the other and yours and overall experience?" For sense of justice, the question was "do you think the amount your partner distributed to you is fair during the 10 rounds of distribution?" The two questions are all rated with five-point scale ranging from 1 (extremely unsatisfied or extremely unjustified) to 5 (extremely satisfied or justified).

Experimental Design

A two factor between-subject design with types of participants and interactive objects both including two levels as the disabled people and the abled people was adopted. Dependent variables included cooperation and psychological feelings. The cooperation referred to the amount of money the participants distributed to the other person (experiment one) or the public (experiment two) and psychological feelings referred to participants' satisfaction and sense of justice during social interactions.

Task and Procedure

All the materials were presented on the computer screen, and participants were ordered to conduct 10 transactions with the interactive objects randomly selected via the computer. In addition, all interactive objects were virtual. Computers were connected to the Internet and participants could obtain information they needed at any time. In each round of transaction, participants and interactive objects each owned gifts worth 0–100 RMB. They had to offer to each other corresponding gifts. When each interaction was finished, they all received gifts offered by virtual interactive objects. The current experiment let virtual interactive objects imitate actual interactive objects' general distribution. Individuals often tend to offer resource equally (Martínez-Cánovas et al., 2016). Therefore, the mean of feedback of the ten rounds were 51.4 and five rounds were higher than 50 and five rounds were lower than 50. This number came from ten numbers selected randomly, which was pseudo-random. In other words, feedback-based numbers were randomly selected as higher and lower than 50. Participants' cooperation in each trial could be compared by a fixed sequence order. The participants were all notified that the total value they would receive after ten-round investment was the true value of gifts after the experiment (but in fact the true value was a fixed amount irreverent to the amount of the experiment). Only when the participants correctly answered and offered the money could they enter the formal experiment. Before the experiment, social interaction tendency of the disabled and abled people needed to be measured and satisfaction and sense of justice were also needed to be rated after every round.

Results

Social Interaction Tendency

Among 41 disabled people, 56.1% of them preferred to interact with the disabled people. Among 40 abled people, 97.5% of them preferred to interact with abled people. A Chi-square test found that the disabled people preferred to interact with the disabled people [$\chi^2(12) = 36.41$, $p = 0.052$] and abled people preferred to interact with abled people [$\chi^2(15) = 58.99$, $p < 0.01$].

Cooperation

In experiment one, the main effect of the types of participants was not significantly different [$F_{(1,77)} = 0.40$, $p = 0.530$] indicating that there was no difference in the amount given to the peers between the two different types of participants. There was a significant difference of the main effect of interactive objects [$F_{(1,77)} = 6.65$, $p < 0.05$, $\eta_p^2 = 0.08$] indicating that there was a significant difference in the investment given during the interaction. The cooperation during the interaction between the disabled people and the disabled people ($M = 58.72$) was significantly higher than that between the disabled people and the abled people ($M = 49.87$; $p < 0.01$) while the cooperation during the interaction between the abled people and the disabled people ($M = 68.15$) was significantly higher than that between the abled people and the abled people ($M = 44.15$; $p < 0.001$). Taking the average feedback value ($M = 51.4$) from the 10-round interactions with

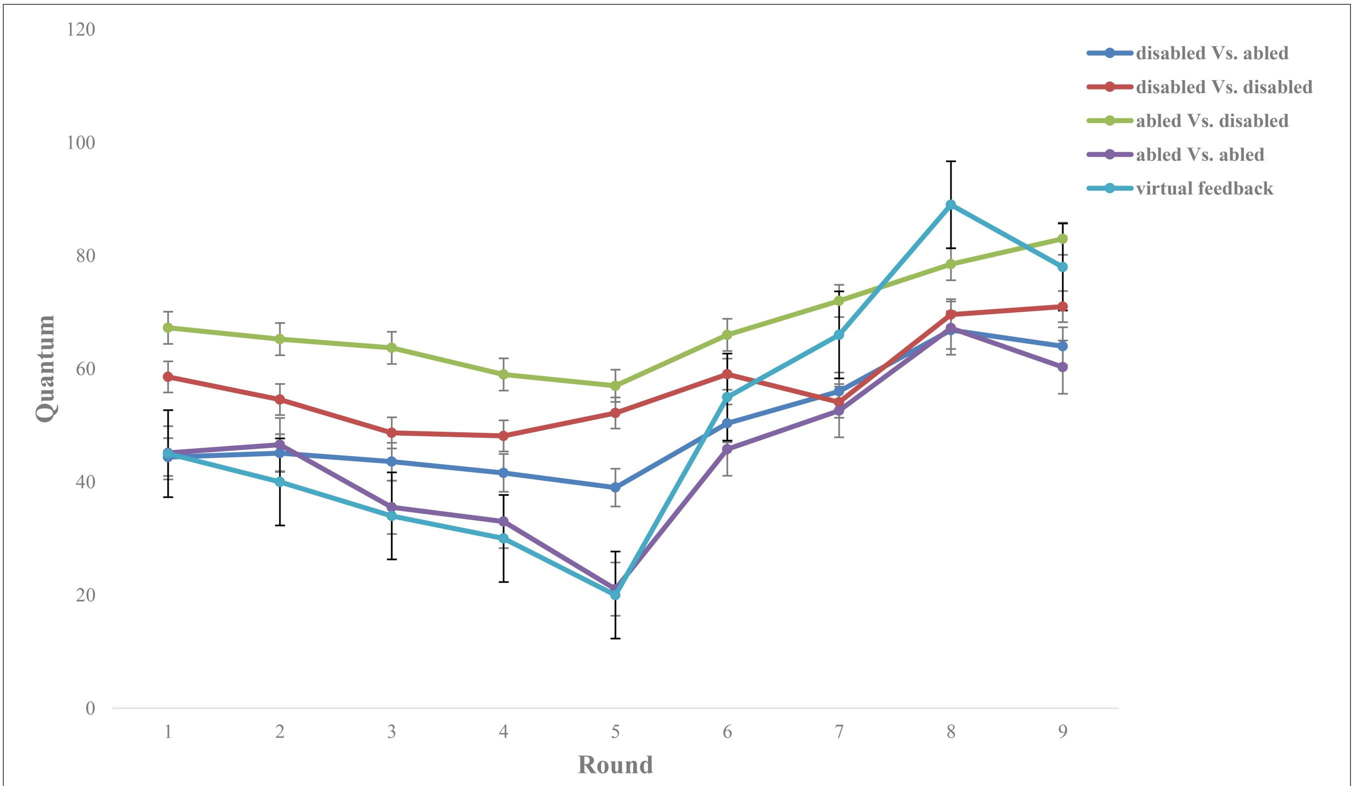

FIGURE 1 | Feedback and different participants' cooperation in each round during ten rounds. The amount given in the interaction between the disabled people and the disabled people was significantly higher than the fair baseline ($p < 0.05$) while there was no significant difference between the amount given in the interaction between the disabled people and the abled people and the fair baseline ($p = 0.645$). Moreover, the investment given in the interaction between the abled people and the disabled people was significantly higher than the fair baseline ($p < 0.001$) while the investment given in the interaction between the abled people and the abled people was significantly lower than the fair baseline ($p < 0.001$). It indicated that there was a high level of cooperation in the interaction between the disabled people and the disabled people while the disabled people tended to distribute equally when interacting with the abled people. Moreover, there was a high level of cooperation in the interaction between the abled people and the disabled people while the disabled people showed comparatively rational selfishness when interacting with the abled people.

the virtual interactive object as the fair baseline, we analyzed whether the investment of the disabled people and the abled people would be higher or lower than the fair baseline (see **Figure 1**).

Psychological Feelings

In experiment one, for the results of satisfaction, there was no significant main effect of the types of participants [$F_{(1,77)} = 1.15$, $p = 0.704$] indicating that there was no difference of satisfaction of social interactions for the two groups. There was a significant main effect of interactive objects [$F_{(1,77)} = 24.66$, $p < 0.001$, $\eta_p^2 = 0.24$] indicating that there was a significant difference in participants' satisfaction with different interactive objects. Moreover, the interaction was significant [$F_{(1,77)} = 24.66$, $p < 0.001$, $\eta_p^2 = 0.24$]. Satisfaction in the interaction between the disabled people and the disabled people ($M = 4.33$) was significantly higher than that between the disabled people and the abled people ($M = 2.90$; $p < 0.001$) while there was no difference in satisfaction in the interaction between the abled people and the disabled people ($M = 3.4$) and the abled people and the abled people ($M = 3.7$; $p = 0.203$). For the results of justice, there was a significant main effect of types of participants [$F_{(1,\ 77)} = 4.59$, $p < 0.05$, $\eta_p^2 = 0.06$] and the disabled people's justice perception ($M = 3.66$, $SD = 0.88$) was higher than that of the abled people ($M = 3.25$, $SD = 0.84$). There was a significant main effect of interactive objects [$F_{(1,77)} = 4.59$, $p < 0.05$, $\eta_p^2 = 0.06$] and the cooperation of the same group ($M = 3.66$, $SD = 0.73$) was significantly higher than that of different groups ($M = 3.25$, $SD = 0.98$). Moreover, the interaction was not significant [$F_{(1,77)} = 1.16$, $p = 0.286$].

EXPERIMENT 2

Participants

The current experiment randomly recruited eighty disabled people including 44 males and 36 females whose mean age was 52.67 years ($SD = 9.38$) and eighty abled people including forty-one males and thirty-nine females whose mean age was 48.21 ($SD = 9.47$). All disabled participants met the national standard for the disabled people and they were mostly Grade II or Grade III of physical disability (mainly as disabilities in arms of legs) with the mean time of disability for 22.8 years. All participants had normal or corrected-to-normal vision, with no partial tritanopia or achromatopsia, and could skillfully operate the computer. The present study was approved by the Ethics Committee of the authors' University in accordance with the ethical principles of the Declaration of Helsinki. All subjects gave written informed consent in accordance with the ethical principles of the Declaration of Helsinki.

Materials

Social Interaction Tendency

The measurement of social interaction tendency was based on previous studies (Radke et al., 2014; Gilam et al., 2015). We respectively adopted a question to measure social interaction tendency including "Which group do you prefer to contact in daily life? (A) Three disabled people; (B) Two disabled people and one abled people; (C) One disabled people and two abled people; (D) Three abled people." Participates sort the order according to the range from the most willingly to participate to the most unwillingly to participate with the highest ranked as four points and the lowest ranked as one point" to show their social interaction tendency.

Cooperation

The measurement of cooperation level was based on Liu and Hao (2014). Public good dilemma was adopted to set up social dilemmas of multiple-person interactions of the current experiment. It was assumed that four people completed a decision task together and each one had a personal and a group account. The personal account was only used by participants and the group account was used by all members of the group. Everyone needed to distribute their initial resource to the personal and group account and the amount distributed to the group account stood for the cooperation level of participants (see details in the **Supplementary Information**).

Psychological Feelings

The measurement of psychological feelings (satisfaction and justice) during social interactions in experiment two was the same as in experiment one.

Experimental Design

A two factor between-subject design with types of participants and interactive situations was adopted. Types of participants included two levels: the disabled people and the abled people, and the interactive situations included four levels: the single identity group, the advantage group, the peer group and the disadvantage group. The constitutions of these four groups were shown in **Table 1**. Except the single identity group, the other three groups were all mixed group, which contained both the disabled and abled people. Dependent variables in experiment two were the same as in experiment one.

Task and Procedure

Participants needed to complete a ten-round investment task on the computer with other randomly selected (virtual) participants. Everyone had a personal account and the group had a public account. The personal account belonged to the participants and contained an initial amount of 100 RMB. Participants could distribute any amount of the personal account (0–100 RMB) to the public group. When the amount of the public account reached or exceeded 200 RMB, the amount of the public account would double and was averagely distributed to the group members. Whether investment or not, the investment would be confiscated when the amount of the public account did not reach 200 RMB. The total value after 10-round investment was the true value of the gifts after the experiment. When participants read instructions, they needed to complete some task-related computations. Then, they needed to complete five practicing trials to familiar themselves with the experiment. Only when participants correctly answered and distributed the money could they enter the formal experiment. The disabled and abled people were randomly assigned to different multiple-person interactive situations.

TABLE 1 | Four groups of the interactive situations in experiment two.

Interactive situations	Constitutions
The single identity group	(I) One disabled people interacting with three virtual disabled people (II) One abled people interacting with three virtual abled people
The advantage group	(I) One disabled people interacting with two virtual disabled people and one virtual abled people (II) One abled people interacting with two virtual abled people and one virtual disabled people
The peer group	(I) One disabled people interacting with one disabled people and two abled virtual people (II) One abled people interacting with one abled people and two virtual disabled people
The disadvantage group	(I) One disabled people interacting with three virtual abled people (II) one abled people interacting with three virtual three disabled people

Results

Social Interaction Tendency

There was a significant difference in the preference of the disabled people for the different types of interactive situations [$F_{(3,237)} = 5.36$, $p < 0.001$, $\eta_p^2 = 0.06$] and the range from high to low was the single identity group ($M = 2.86$, $SD = 1.05$), the peer group ($M = 2.58$, $SD = 0.87$), the advantage group ($M = 2.49$, $SD = 1.04$) and the disadvantage group ($M = 2.08$, $SD = 1.34$). The planned t-test found that the social interaction tendency of the disadvantage group was significantly lower than that of the single identity group ($p < 0.001$) and the peer group ($p < 0.05$), which indicated that the disabled people preferred to interact with the same type of individuals and preferred not to interact with the disadvantage group in the multiple-person interactive situations. In addition, there was a significant difference in the preference of the abled people for different types of interactive situations ($F_{(3,237)} = 34.45$, $p < 0.001$, $\eta_p^2 = 0.30$) and the range from high to low was the single identity group ($M = 3.48$, $SD = 1.03$), the advantage group ($M = 2.48$, $SD = 0.95$), the peer group ($M = 2.15$, *SD* = 0.83) and the disadvantage group (*M* = 1.85, *SD* = 0.92). The planned *t*-test found that there was a significant difference among each interactive situations ($ps < 0.05$), which indicated that the abled people preferred to interact with the abled people and the preference tended to decrease with the decrease in the proportion of the abled people in the group.

Cooperation

There was no significant main effect of the types of the participants [$F_{(1,152)} = 0.24$, $p = 0.622$] indicating that there was no difference in the public goods investment in the multiple-person interactions between the two groups. There was a significant main effect of interactive situations [$F_{(1,152)} = 24.64$, $p < 0.001$, $\eta_p^2 = 0.33$] indicating that there was a significant difference in the public goods investment in different interactive situations. Moreover, the interaction was significant [$F_{(1,152)} = 7.63$, $p < 0.001$, $\eta_p^2 = 0.13$]. In addition, we set 50 as the fair baseline to analyze the investment of the disabled people and the abled people (see **Figure 2**).

Psychological Feelings

For the satisfaction, there was no significant main effect of the types of participants [$F_{(1,152)} = 0.012$, $p = 0.911$] indicating that there was no difference of satisfaction of social interactions for the two groups. There was a significant main effect of interactive situations [$F_{(1,152)} = 11.42$, $p < 0.001$, $\eta_p^2 = 0.18$] indicating that there was a significant difference in participants' satisfaction in different interactive situations. Moreover, the interaction was significant [$F_{(1,152)} = 9.26$, $p < 0.001$, $\eta_p^2 = 0.16$]. For the justice, there was no significant main effect of types of participants [$F_{(1,152)} = 1.62$, $p = 0.205$] indicating that there was no difference in justice of social interactions for the two groups. There was a significant main effect of interactive situations [$F_{(1,152)} = 3.20$, $p < 0.05$, $\eta_p^2 = 0.06$] indicating that there was a significant difference in participants' justice in different interactive situations. Moreover, the interaction was significant [$F_{(1,152)} = 8.49$, $p < 0.001$, $\eta_p^2 = 0.14$].

DISCUSSION

The current experiment revealed possible social interaction patterns of the disabled abled people in social interactions. The disabled people preferred to interact with the disabled people and the abled people preferred to interact with the abled people, which was consistent with the previous studies (Zeedyk et al., 2014). These results confirmed the hypothesis one. This asymmetry might indicate that disabled people's preference for interactive objects was lower than the abled people. In the social interactions, the abled people might make a more explicit distinction between these two groups compared with the disabled people. The disabled people had a higher cooperation when interacting with the disabled people than interacting with the abled people and the abled people had higher cooperation when interacting with the disabled people than interacting with the abled people. These results confirmed the hypothesis two. The interaction between the disabled abled people appears to be more compliant with the fairness theory. In other words, the cooperation level during the interactions between the disabled abled people was lower than that between the abled disabled people. It indicated that the disabled people at a disadvantage were more sensitive to the equality of the distribution (Zeng et al., 2016) for there was no significant difference between the distribution of the disabled abled people and the fair baseline. It also indicated that the abled people's "unequal averseness" made them tend to narrow the gap in the distribution to distribute more to the opposite, namely the high distribution of the abled people could be perceived by the disabled as unreasonable respect. During the whole interactions, although there were differences in participants' average distribution, participants' cooperation levels in every round were all influenced by the opposite. It might indicate that no matter the disabled abled people, their social behaviors and social attitudes were all influenced by interactive objects in daily lives.

The current experiment also found that the change in the member of groups did affect the cooperation between the disabled abled people. For the disabled people, although they preferred to interact with their own groups and the peer group, their cooperation level in the single identity group and in the peer group was low while their cooperation level in the advantage group was high. These results confirmed the hypothesis four. Previous studies found that the disabled people entering the integrated environment, which is comprised of the disabled people, can promote their participation and interactions in physical activities (Bossaert et al., 2013). However, other studies found that the integrated environment may restrict some psychological factors and put forward the reverse integration (RI) environment. It was thought that the disabled people under this environment had lower desire to integrate (Rao et al., 2011). The current experiment also confirmed that the RI environment could relieve the disabled people's psychological disadvantage in asymmetric status to some extent. Moreover, the cooperation level declined as the number of disabled people decreased in multiple-person interactions. These results supported the justice theory and clarified that "disadvantage makes people more cooperative" might only be feasible in the two-person interaction and in the peer group. When the disabled people's advantage in the number reversed their disadvantage in the status, social interaction patterns of the disabled people were the same as those of the abled people. For the abled people, the cooperation level was higher in the advantage group and in the disadvantage group. The former was that the abled people were in advantage in number and status and they needed high devotion to narrow the gap, which supported the justice theory. The latter might be that one single abled people in the group would highlight his advantage in status or might be that disadvantage in number stimulated higher cooperation. However, the current study could not interpret individuals' inclination of decision when their identity and number were at disadvantage simultaneously. In addition,

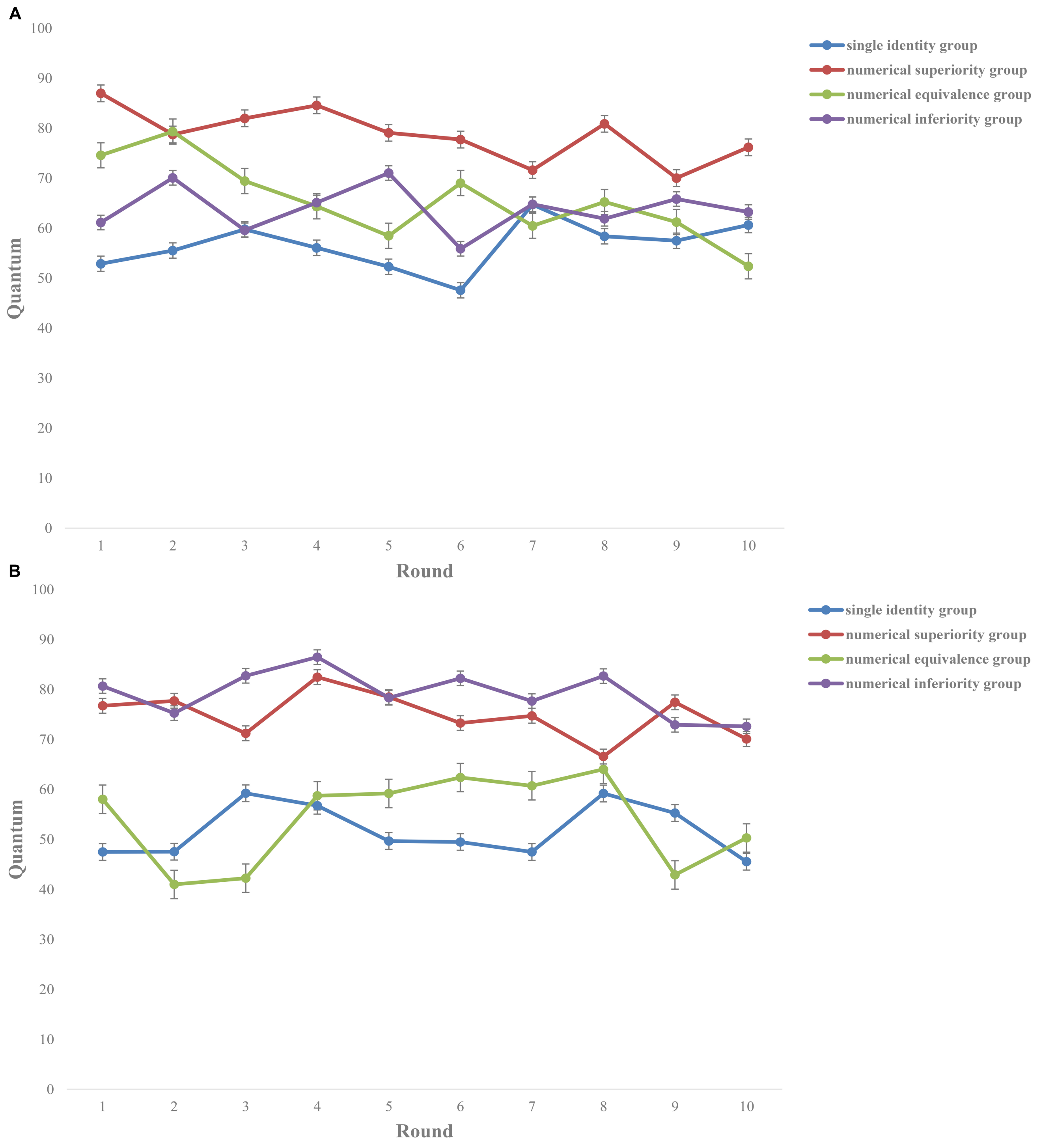

FIGURE 2 | Cooperation of **(A)** the disabled people and **(B)** the abled people in ten rounds, respectively. The results showed that the amount the disabled people invested in the single identity group, in the advantage group, in the peer group and in the disadvantage group were all higher than the fair baseline (*ps* < 0.05). There was no difference of the amount the abled people invested in the single identity group (*p* = 0.175) and in the peer group (*p* = 0.079) compared with the fair baseline while the investment in the advantage group and in the disadvantage group was significantly higher than the fair baseline (*ps* < 0.05). It indicated that the advantage group and the disadvantage group highlighted the "individuals' unequal status," which resulted in their higher level of cooperation than the fair baseline.

there was higher level of cooperation of the disabled people in two-person and multiple-person interactions of the single identity group and the peer group compared with the abled people, which was consistent with the results of the two-person interactions and further supported "disadvantage makes people more cooperative."

As for the psychological feelings, the current experiment found that although interactive objects' feedback was the same, the disabled people had a higher level of satisfaction and sense of justice when interacting with the disabled people. These results confirmed the hypothesis three. It explains to some extent the reason why the disabled people do not want to interact with the abled people. In the meanwhile, the abled people's high distribution did not improve the disabled people's satisfaction, which might be related to the fact that the disabled people did not regard the high distribution as respect. Chen and Shu (2012) also found that the disabled students' disabled identity could on the one hand gives them extra help, but on the other hand be regarded as one of the source of stigma. In daily lives, the disabled people may have misunderstanding and prejudice against the abled people so that they don't want or evade interacting with the abled people.

Based on the social game theory and its paradigms, the current study explored social interaction patterns of the disabled people in asymmetric dilemmas and has some significant meanings. Firstly, exploring social interaction patterns between the disabled people and the abled people in unequal situations of resource and status is conducive to deepening the publics' understanding of the disabled people's social interaction patterns and feelings, and encouraging more the disabled people to participate social interaction. Secondly, it is both of great theoretical and practical significance to understand the social interaction dilemmas of the disabled people, to improve social participation of the disabled people, to strengthen publics' understanding of the disabled people's social behaviors, and to deepen and extend researches of vulnerable groups. It also enhances the awareness of the disabled people about their and other people's behaviors, improves their cognition of self-stigma and social interaction. Thirdly, it provides theoretical and practical evidence for the government, the community and other organization to establish policies or hold activities. Fourthly, it will help the relevant departments of the government, community service organizations for the disabled and other relevant organizations to formulate policies and regulations or carry out activities that are beneficial to the physical and mental health of the disabled people, as well as to provide theoretical and empirical evidence for caring for and interacting with the disable people effectively, scientifically and rationally.

However, there are some limitations of the current study that need to be improved in the future studies. Firstly, the participants of the current study were special groups and experimental procedure was comparatively complex. Therefore, the sample size may be small and not representative enough. In particular, the sample collection was mainly concentrated in urban areas, the lack of samples in other areas such as rural areas, may affect the generalization of the findings. If conditions permit, a larger sample size and expanded sample collection area will be required in future studies. Secondly, the current study not only focused on intragroup cooperation of the disabled people, but also on intergroup cooperation between the disabled people and the abled people, which expanded researches of cooperation in asymmetric social dilemmas. However, the current study adopted simplified real-life dilemmas, which reflected abstract social dilemmas. Although the simplified real-life dilemmas in the current study also included some real-life factors (e.g., multiple interactions, feedbacks), the behavioral index was too simple. Other behavioral variables need to be combined in future studies in order to carry out more comprehensive researches. Thirdly, the psychological indexes in the current study did not correspond well to the behavioral indexes due to the measurement of only using a single or twofold items. Therefore, the psychological indicators on cooperative behaviors of the disabled still need to be improved in future researches.

ETHICS STATEMENT

The current study was implemented in conformity to the recommendations of the Ethical Committee of Ningbo University. Informed consent of all participants was obtained in line with the Declaration of Helsinki. The protocol was approved by the Ethical Committee of Ningbo University.

AUTHOR CONTRIBUTIONS

LZ and XZ designed and implemented the study. SL, WX, SH, and ZM analyzed the data. SL, LZ, and XZ interpreted the data. SL, WX, SH, ZM, LZ, and XZ wrote the manuscript.

FUNDING

This work was supported by grants from the National Social Science Fund of China (12BSH055) and the K. C. Wong Magna Fund at Ningbo University. This work was also supported by grants from the National Key Basic Research Program (2016YFA0400900 and 2018YFC0831101), the National Natural Science Foundation of China (31471071, 31771221, 61773360, and 71874170), and the Fundamental Research Funds for the Central Universities of China.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01683/full#supplementary-material

REFERENCES

Bossaert, J. E., Colpin, H., Pijl, S. J., and Petry, K. (2013). Truly included? A literature study focusing on the social dimension of inclusion in education. *Int. J. Inclusive Educ.* 17, 60–79. doi: 10.1080/13603116.2011.580464

Brown, A. A., and Brunell, A. B. (2017). The "modest mask"? An investigation of vulnerable narcissists' implicit self-esteem. *Pers. Individ. Dif.* 119, 160–167. doi: 10.1016/j.paid.2017.07.020

Chen, C. H., and Shu, B. C. (2012). The process of perceiving stigmatization: perspectives from taiwanese young people with intellectual disability. *J. Appl. Res. Intellect. Disabil.* 25, 240–251. doi: 10.1111/j.1468-3148.2011.00661.x

Ditchman, N., Werner, S., Kosyluk, K., Jones, N., Elg, B., and Corrigan, P. W. (2013). Stigma and intellectual disability: potential application of mental illness research. *Rehabil. Psychol.* 58, 206–216. doi: 10.1037/a0032466

EL-Seidy, E., Elshobaky, E. M., and Soliman, K. M. (2016). Two population three-player prisoner's dilemma game. *Appl. Math. Comput.* 277, 44–53. doi: 10.1016/j.amc.2015.12.047

Forber-Pratt, A. J., Lyew, D., Mueller, C., and Samples, L. B. (2017). Disability identity development: a systematic review of the literature. *Rehabil. Psychol.* 62, 198–207. doi: 10.1037/rep0000134

Gilam, G., Lin, T., Raz, G., Azrielant, S., Fruchter, E., Ariely, D., et al. (2015). Neural substrates underlying the tendency to accept anger-infused ultimatum offers during dynamic social interactions. *Neuroimage,* 120, 400–411. doi: 10.1016/j.neuroimage.2015.07.003

Hao, F., Gong, Q. B., and Liu, C. J. (2016). Impact of peer behavior and cooperative belief on cooperative changes in social dilemmas. *J. Psychol. Sci.,* 39, 448–453. doi: 10.16719/j.cnki.1671-6981.20160230

He, J. Z., Wang, R. W., and Li, Y. T. (2014). Evolutionary stability in the asymmetric volunteer's dilemma. *PLoS One* 9:e103931. doi: 10.1371/journal.pone.0103931

Liu, C. J., and Hao, F. (2014). Social dilemmas: theoretical framework and experimental research. *Adv. Psychol. Sci.* 22, 1475–1484. doi: 10.3724/SP.J.1042.2014.01475

Liu, C. J., and Hao, F. (2015). Decision making in asymmetric social dilemmas: a dual mode of action. *Adv. Psychol. Sci.* 23, 1–10. doi: 10.3724/SP.J.1042.2015.00001

Lu, N., Liu, J. Y., Wang, F., and Lou, V. W. Q. (2017). Caring for disabled older adults with musculoskeletal conditions: a transactional model of caregiver burden, coping strategies, and depressive symptoms. *Arch. Gerontol. Geriatr.* 69, 1–7. doi: 10.1016/j.archger.2016.11.001

Martínez-Cánovas, G., Val, E. D., Botti, V., Hernández, P., and Rebollo, M. (2016). A formal model based on game theory for the analysis of cooperation in distributed service discovery. *Inf. Sci.* 326, 59–70. doi: 10.1016/j.ins.2015.06.043

Moore, M. E., Konrad, A. M., Yang, Y., Ng, E. S. W., and Doherty, A, J. (2011). The vocational well-being of workers with childhood onset of disability: life satisfaction and perceived workplace discrimination. *J. Vocat. Behav.* 79, 681–689. doi: 10.1016/j.jvb.2011.03.019

O'Reilly, M., Bowlay-Williams, J., Svirydzenka, N., and Vostanis, P. (2016). A qualitative exploration of how adopted children and their parents conceptualise mental health difficulties. *Adopt. Fostering,* 40, 60–76. doi: 10.1177/0308575915626383

Płatkowski, T. (2016). Egalitarian solutions to multiperson social dilemmas in populations. *Appl. Math. Comput.* 284, 226–233. doi: 10.1016/j.amc.2016.03.011

Radke, S., Güths, F., André, J. A., Müller, B. W., & de Bruijn, E. R. A. (2014). In action or inaction? Social approach–avoidance tendencies in major depression. *Psychiatry Res.* 219, 513–517. doi: 10.1016/j.psychres.2014.07.011

Rand, D. G. (2017). Social dilemma cooperation (unlike dictator game giving) is intuitive for men as well as women. *J. Exp. Soc. Psychol.* 73, 164–168. doi: 10.1016/j.jesp.2017.06.013

Rao, L. L., Han, R., Ren, X. P., Bai, X. W., Zheng, R., Liu, H., et al. (2011). Disadvantage and prosocial behavior: the effects of the wenchuan earthquake. *Evol. Hum. Behav.* 32, 63–69. doi: 10.1016/j.evolhumbehav.2010.07.002

Riddell, S., and Weedon, E. (2014). Disabled students in higher education: discourses of disability and the negotiation of identity. *Int. J. Educ. Res.* 63, 38–46. doi: 10.1016/j.ijer.2013.02.008

Roch, S. G., and Samuelson, C. D. (1997). Effects of environmental uncertainty and social value orientation in resource dilemmas. *Organ. Behav.Hum. Decis. Process.* 70, 221–235. doi: 10.1006/obhd.1997.2707

Roth, A. E. (1991). Game theory as a part of empirical economics. *Econ. J.* 101, 107–114. doi: 10.2307/2233845

Sigmund, K. (2012). Moral assessment in indirect reciprocity. *J. Theor. Biol.* 299, 25–30. doi: 10.1016/j.jtbi.2011.03.024

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Q. Rev. Biol.* 46, 35–57. doi: 10.1086/406755

Tsvetkova, M., and Buskens, V. (2013). Coordination on egalitarian networks from asymmetric relations in a social game of chicken. *Adv. Complex Syst.* 16:1350005. doi: 10.1142/S0219525913500057

Vandello, J. A., Michniewicz, K. S., and Goldschmied, N. (2011). Moral judgments of the powerless and powerful in violent intergroup conflicts. *J. Exp. Soc. Psychol.* 47, 1173–1178. doi: 10.1016/j.jesp.2011.04.009

Wang, P., and Chen, L. (2011). The effects of sanction and social value orientation on trust and cooperation in public goods dilemmas. *Acta Psychol. Sin.* 43, 52–64. doi: 10.3724/SP.J.1041.2011.00052

Wang, Y., Wei, Z. H., Shen, S. C., Wu, B., Cai, X. H., Guo, H. F., et al. (2016). The response of chinese scholars to the question of "how did cooperative behavior evolve?" *Chin. Sci. Bull.* 61, 20–33. doi: 10.1360/N972015-00613

Xiao, E. (2013). Profit seeking punishment corrupts norm obedience. *Games Econ. Behav.* 77, 321–344. doi: 10.1016/j.geb.2012.10.010

Zeedyk, S. M., Rodriguez, G., Tipton, L. A., Baker, B. L., and Blacher, J. (2014). Bullying of youth with autism spectrum disorder, intellectual disability, or typical development: victim and parent perspectives. *Res. Autism Spectr. Disord.* 8, 1173–1183. doi: 10.1016/j.rasd.2014.06.001

Zeng, W. J., Li, M. Q., and Chen, F. Z. (2016). Cooperation in the evolutionary iterated prisoner's dilemma game with risk attitude adaptation. *Appl. Soft Comput.* 44, 238–254. doi: 10.1016/j.asoc.2016.03.025

Zhang, L., Li, W. T., Liu, B. B., and Xie, W. L. (2014). Self-esteem as mediator and moderator of the relationship between stigma perception and social alienation of chinese adults with disability. *Disabil. Health J.* 7, 119–123. doi: 10.1016/j.dhjo.2013.07.004

Zhang, L., Liu, S., Xie, W. L., and Li, W. T. (2015). Attentional bias of individuals in adulthood with disabilities for different types of social cues. *Psychol. Dev. Educ.* 31, 676–684. doi: 10.16187/j.cnki.issn1001-4918.2015.06.06

Zitek, E. M., and Tiedens, L. Z. (2012). The fluency of social hierarchy: the ease with which hierarchical relationships are seen, remembered, learned, and liked. *J. Pers. Soc. Psychol.* 102, 98–115. doi: 10.1037/a0025345

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 04 April 2018
doi: 10.3389/fpsyg.2018.00407

Empathy Modulates the Evaluation Processing of Altruistic Outcomes

*Xin Liu, Xinmu Hu, Kan Shi and Xiaoqin Mai**

Department of Psychology, Renmin University of China, Beijing, China

Empathy plays a central role in social decisions involving psychological conflict, such as whether to help another person at the cost of one's own interests. Using the event-related potential (ERP) technique, the current study explored the neural mechanisms underlying the empathic effect on the evaluation processing of outcomes in conflict-of-interest situations, in which the gain of others resulted in the performer's loss. In the high-empathy condition, the beneficiaries were underprivileged students who were living in distress (stranger in need). In the low-empathy condition, the beneficiaries were general students without miserable information (stranger not in need). ERP results showed that the FRN was more negative-going for self no-gain than self gain, but showed reversed pattern for other's outcome (i.e., more negative for gain than no-gain) in the low-empathy condition, indicating that participants interpreted the gain of others as the loss of themselves. However, the reversed FRN pattern was not observed in the high-empathy condition, suggesting that the neural responses to one's own loss are buffered by empathy. In addition, the P3 valence effect was observed only in the self condition, but not in the two stranger conditions, indicating that the P3 is more sensitive to self-relevant information. Moreover, the results of subjective rating showed that more empathic concern and altruistic motivation were elicited in the high-empathy condition than in the low-empathy condition, and these scores had negative linear correlations only with the FRN, but not with the P3. These findings suggest that when outcomes following altruistic decisions involve conflict of interest, the early stage of the processing of outcome evaluation could be modulated by the empathic level.

Keywords: empathy, outcome evaluation, event-related potential (ERP), feedback-related negativity (FRN), P3, altruism

Edited by:
Hang Ye,
Zhejiang University, China

Reviewed by:
Ruolei Gu,
Institute of Psychology (CAS), China
Jianhui Wu,
Institute of Psychology (CAS), China

Correspondence:
Xiaoqin Mai
maixq@ruc.edu.cn

Specialty section:
This article was submitted to
Decision Neuroscience,
a section of the journal
Frontiers in Psychology

Received: *23 November 2017*
Accepted: *12 March 2018*
Published: *04 April 2018*

Citation:
Liu X, Hu X, Shi K and Mai X (2018)
Empathy Modulates the Evaluation
Processing of Altruistic Outcomes.
Front. Psychol. 9:407.
doi: 10.3389/fpsyg.2018.00407

INTRODUCTION

In our daily life, humans are sometimes required to make difficult social decisions involving benefit conflict between themselves and other social agents, such as whether they are willing to sacrifice personal benefit on behalf of a stranger's welfare (Rilling and Sanfey, 2011). Numerous studies have focused on the inner mechanisms underlying the processing of such altruistic decisions which defined as increasing the welfare of others at a cost of the self (Batson and Shaw, 1991; de Waal, 2008), and found that multiple motivational and emotional factors, such as kin selection (Hamilton, 1964), reciprocal relation (Trivers, 1971), and empathic concern (Batson, 2008), could give rise to prosocial decisions. However, little is known about how people evaluate the consequent outcomes after they made altruistic decisions. Given that humans use positive or negative feedback to guide their next behaviors (Nieuwenhuis et al., 2004; Yang et al., 2015), it is necessary to understand the neural mechanisms underlying the processing of evaluating altruistic outcomes when self-interests are sacrificed.

Previous studies using the event-related potential (ERP) have found two ERP components related to the processing of outcome evaluation: the feedback-related negativity (FRN) and P3 (Schupp et al., 2000; Gehring and Willoughby, 2002; Miltner et al., 2014). The FRN, sometimes also called medial frontal negativity (MFN), originates from the medial-frontal cerebral regions (Holroyd and Coles, 2002; Nieuwenhuis et al., 2004; Wu et al., 2017), especially the anterior cingulate cortex (ACC) a brain area playing a central role in empathic responses for other person's pain (Bernhardt and Singer, 2012). Accumulating studies have found that the FRN is more negative for the unfavorable outcomes than for the favorable outcomes, and reaches maximum between 200 and 300 ms following the onset of feedback stimuli (Gehring and Willoughby, 2002; Hajcak et al., 2005; Hauser et al., 2014; Paul and Pourtois, 2017). Furthermore, an enhanced FRN indicates the result being worse than expected (Holroyd and Coles, 2002; Nieuwenhuis et al., 2004) and reflects stronger motivational impact of the current stimuli (Masaki et al., 2006; Itagaki and Katayama, 2008; Luo et al., 2015). The P3 is a positive, large-amplitude potential with typical peak in the period of 300–600 ms after the onset of stimuli. It is larger for the positive feedback than for the negative feedback and for a large reward than for a small reward (Holroyd et al., 2006; Hewig et al., 2011; Peterburs et al., 2017). The P3 is generally believed to be related to the allocation of cognitive resources and the processing of attentional distribution (Polich, 1987, 2007; Yang et al., 2015; Hu et al., 2017), especially self-relevant attentional allocation (Gray et al., 2004; Linden, 2005). Extensive research regarding outcome evaluation suggests that the two ERP components could represent not only the evaluating processes of self-related outcomes but also those of other-related outcomes (e.g., Kang et al., 2010; Leng and Zhou, 2010; Ma et al., 2011; Wang et al., 2014; Hu et al., 2017). When the outcomes of other people have nothing to do with participants' own benefit, the similar neural responses were observed in both self and other outcome conditions (Yu and Zhou, 2006; Leng and Zhou, 2014; Zhu et al., 2016). For example, in a pioneering work, Yu and Zhou (2006) asked participants to earn money in a gambling task for themselves and observe the reward/punishment feedback of others in which other's outcomes were irrelevant to participants' own interests. The results showed that the FRN was more negative-going to the loss outcome whenever outcomes related to self or to others, indicating that the FRN effect was elicited not only in self-evaluation condition, but also in other-evaluation condition. In other words, when there was no conflict of interests between oneself and others, comparable neural activities of outcome evaluation were observed in both self and others' losing situations.

However, when there are benefit conflict between performers and beneficiaries, the ERPs of outcome evaluation change in a reverse way (Fukushima and Hiraki, 2006; Itagaki and Katayama, 2008; Marco-Pallares et al., 2010). Marco-Pallares et al. (2010) compared the ERP responses to outcomes of gambling in different situations across three groups. In the neutral group, individuals simply observed the performer's action and their own benefit was not affected by others. In the parallel group, observers gained or lost the same amount of money as the performer. Finally, in the reverse group, competing motivation was aroused because the gain of others led to a loss of the observer and vice versa. The results showed that the ERPs of evaluators in the reverse situation showed an inverse pattern compared to the neutral and parallel conditions, indicating that the neural responses of evaluators translated the gain of others into the loss for themselves. However, an interesting study by Fukushima and Hiraki (2006) suggested that the inversed neural responses in competing situation are probably modulated by the empathetic processes. In their study, participants were required to perform a gambling task with their friends in which the friends' loss resulted in the gain of themselves. The results showed that the inversed FRN effect for loss trials was only elicited for participants with less empathic tendency, whereas the neural discrepancy between gain and loss vanished in individuals with more empathic trait. The author proposed that the individual difference in the FRN is probably based on the allocation between empathetic and utilitarian processing. It is further confirmed by some studies suggesting that there are individual differences in the capacity for empathy and which links to the differences in the brain structure (Rueckert and Naybar, 2008; Banissy et al., 2012; Christov-Moore et al., 2014).

In addition, evidence from behavioral studies has indicated that the decision-making in competing situation (i.e., interest conflict with other social agents) can be influenced by the level of empathy (Batson and Moran, 1999; Batson and Ahmad, 2001). In their studies, they manipulated the individual's empathic emotions in the prisoner's dilemma (PD) task to induce the high altruistic motivation and found that participants increased their prosocial behaviors to cooperate with others, even though the best strategy was defecting the other partner to guarantee the maximized personal gain. Taking these studies together, we can conclude that empathy has great impact on altruistic decision-making and may play an important role in evaluating processes.

The present study aimed to examine whether the level of empathy could modulate the ERP responses to outcome evaluation when there was conflict between self-interest and other-interest. We revised the classical gambling task (Gehring and Willoughby, 2002) and required each participant to perform it in three conditions: gambling for themselves (self condition) and for two strangers. One of the strangers was described as an underprivileged student living in distress (stranger-in-need condition), while the other one was depicted as a general student who was studying in a regular urban school (stranger-not-in-need condition). Based on the empathy-altruism hypothesis that people would feel strong empathy for others in need and in distress (Batson and Moran, 1999; Batson and Ahmad, 2001), we considered the stranger-in-need scenario as the high-empathy condition while the stranger-not-in-need scenario as the low-empathy condition. One point should be noted that psychological conflict was settled in two strangers' situations in which participants had to pay the same amount of money from their remuneration as the amount they gained for others. Our hypotheses were that the FRN effect would inverse in the low-empathy condition, but would not inverse in the high-empathy condition. Further, the P3 effect would be observed only in the self condition given that P3 is more sensitive to self-related stimuli (Gray et al., 2004; Linden, 2005).

MATERIALS AND METHODS

Participants

Thirty undergraduate and graduate students (15 females; mean age 21.27 ± 2.1 years) at Renmin University of China were recruited in the present study. All participants were right handed, had normal or corrected to normal vision, and reported no history of neurological or psychiatric diagnoses. The data of two male participants were excluded because there were not enough trials (less than 30 trials) after artifacts were removed (Marco-Pallares et al., 2011). Written informed consent was obtained from all participants. The study was approved by the Institutional Review Board of Department of Psychology at Renmin University of China.

Procedure

At the beginning of the experiment, each participant was instructed to play the gambling game three times for different beneficiaries, including himself/herself and two strangers. In the high-empathy condition, the beneficiary would be an underprivileged student who came from a school in remote poverty regions (stranger in need). In the low-empathy condition, the reward receiver would be a general student who was studying in a normal urban school (stranger not in need). All the participants were informed that they would get the amount of money they gained when they played for themselves. However, when the participant played games for two strangers, the beneficiaries would receive the money they won in the game as the prize, and the participant would lose the same amount of money. All participants were informed how much money they earned for themselves and strangers after the experiment was over. Ultimately, they were paid an amount of money between 60 and 65 Chinese yuan.

Participants were seated comfortably in front of a computer screen in an electrically isolated room. They were asked to play the gambling game adapted from the task designed by Gehring and Willoughby (2002). As illustrated in **Figure 1**, each trial began with a white fixation cross presented for 500 ms on a black background. Then, two gray cards were presented on either side of the fixation point with no numeral cue on them. Participants were required to choose between the two alternatives by pressing a corresponding response button (F or J key on the keyboard) with their left or right index finger. When the participant responded, the chosen card was highlighted by a thickening of a yellow border for 600–800 ms, and then the outcome (5 or 0) behind the chosen card shown centrally was displayed for 1000 ms. The inter-trail interval was 600–800 ms. To increase the salience of the valence of the outcome, the chosen card turned red/green color to indicate gain/no-gain outcomes and the colors of cards were counterbalanced among participants. In the situation that participants played the game for themselves, the numeral 5 means that participants gained 5 points and 0 means that participants gained no points. In the situation that participants played the game for strangers, 5 means that strangers gained 5 points but participants themselves lost 5 points; 0 means that strangers gained no points and participants did not lost points either. According to previous research, the FRN is determined by the value of the outcome relative to the range of other possible outcomes in the task, rather than by the objective value of the outcome (Holroyd et al., 2004). We thus expected that no-gain feedback could elicit the FRN effect as same as loss feedback did.

There were 270 trials in total, divided into three blocks with 90 trails and only one of three beneficiary conditions in each block. At the beginning of each block, participants were informed that which beneficiary they would play for in this block, and were emphasized to notice the meaning of winning money in this block. Unknown to the participants, the gain/no-gain feedback was manipulated according to a random sequence, and each participant received equal times of each feedback condition. The order of the three blocks was counterbalanced over participants.

The stimuli were presented by E-prime 2.0 software package (PST, Pittsburgh, PA, United States). The formal experiment started after 5 trials of practice for each participant. After finishing the gambling task, the participants firstly filled out the Chinese version of the Self-Report Altruism Scale (C-SRA scale) (Rushton et al., 1981; Chou, 1996), a paper questionnaire that contains 20 statements to measure altruism in a behaviorally concrete manner. Then, they were asked to complete a 5-point scale to rate their subjective "motivation" to win the game and "empathic feeling" about the outcome. Specifically, they were asked to rate how much they were willing to play the game (1 = "not at all" to 5 = "very much"), how much they were willing to win in the game (1 = "not at all" to 5 = "very much"), and what they felt about the winning outcomes (1 = "very unhappy" to 5 = "very happy") for themselves, the stranger in need, and the stranger not in need, respectively. The first question measured the general "motivation" of participants to make efforts on this task and the second one measured the specific "motivation" to increase welfare of self and others. The scores of the former two questions were clumped together to create a composite measure for the "motivation" to win for each beneficiary. The last question measured whether participants felt positive or negative emotions when gaining money for themselves and for strangers, regarding as "empathic emotion" to others.

EEG Recording and Analysis

EEG was recorded with NeuroScan synamp2 amplifier (Neuroscan Inc., Sterling, VA, United States), using an elastic cap with 64 tin electrodes according to the international 10/20 system. The signals were amplified with a band-pass filter of 0.01–100 Hz and continuously sampled at 1000 Hz/channel for the offline analysis. All rows of electrode recordings were referenced to an electrode placed over the left mastoid, and were re-referenced offline to the average of the left and right mastoids. The vertical and horizontal electrooculograms (EOGs)

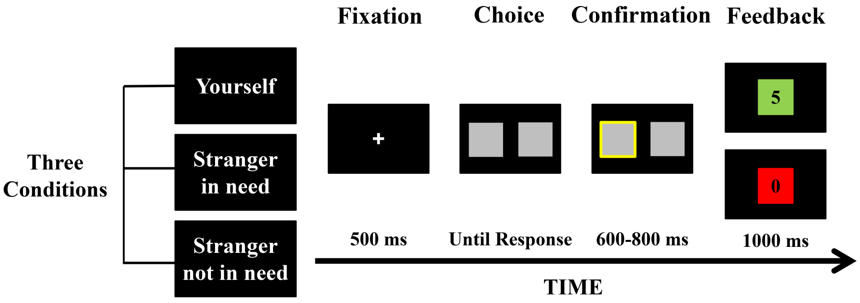

FIGURE 1 | An illustration of a single trial in the gambling task. Each trail began with a fixation cross. Participants viewed two gray cards without numeral cue and were required to choose one of them by pressing the corresponding key. Their choice was then highlighted for 600–800 ms. After that, the outcome feedback was presented for 1000 ms.

were collected with electrodes placed on the left supraorbital and infraorbital, and on the outer canthi of the left and right eyes respectively. All the interelectrode impedances were less than 5 kΩ.

The EEG data were processed offline using the Neuroscan 4.5 software. Ocular artifacts were corrected using a regression procedure implemented in the Neuroscan software (Semlitsch et al., 1986). Raw EEG data were segmented into epochs from 200 ms before to 800 ms after the onset of outcome feedback. The 200 ms preceding the feedback stimulus served as baseline. Epochs containing artifacts exceeding ± 75 μV were rejected from the analysis. The data were digitally low-pass filtered below 30 Hz and were then averaged for each condition.

The present analyses focused on the FRN and P3 elicited by outcome feedback. The FRN was measured as the mean amplitudes in the time window of 210–300 ms following the feedback presentation. The P3 was defined as the most positive peak in the window of 330–430 ms after the onset of feedback stimuli. Based on the topographical distribution of each ERP component and previous research (e.g., Yeung and Sanfey, 2004; Leng and Zhou, 2014), the FRN was preliminary calculated across 3 electrodes (Fz, FPz and Cz) and the P3 was quantified across 2 electrodes (CPz and Pz). The results indicated that the effect of FRN was greatest at the FCz site, and the effect of P3 was largest at the CPz site. Hence, we focused on the FCz and CPz electrodes for more detailed analyses at which the ERP effects were maximal.

The FRN and P3 data were each subjected to repeated measures analysis of variance (ANOVA) with two within-subject's factors: Beneficiary (self vs. stranger-in-need vs. stranger-not-in-need) and Reward Valence (gain vs. no-gain). The significance level was set at 0.05 for all the statistical analyses. Bonferroni-corrected method was performed for *post hoc* testing of significant main effects, while simple effect analysis was using for testing significant interactions. Greenhouse–Geisser correction of the ANOVA assumption of sphericity was applied where appropriate. Effect size in all ANOVA analyses were reported by partial eta-squared (η_p^2), where 0.05 represents a small effect, 0.10 represents a medium effect, and 0.20 represents a large effect (Cohen, 1973). All the statistical analyses were performed by SPSS (23.0; SPSS, Inc., Chicago, IL, United States).

RESULTS

Behavioral Results

A few trails with reaction time (RT) greater than 2000 ms were deleted as extreme value. In the gambling task, the mean (±SD) RTs for choice responses in three conditions were 431 ± 113 ms (self), 467 ± 110 ms (stranger-in-need), and 456 ± 137 ms (stranger-not-in-need), respectively. One-way ANOVA was used to compare the RTs among three beneficiaries. No significant difference was found among them [$F(2,81) = 0.341$, $p = 0.7$].

Subjective Ratings

Figure 2 shows the subjective ratings of feelings about win and motivation to win for each beneficiary. One-way ANOVA on the subjective rating of the feeling of empathy toward winning money for different beneficiaries (self vs. stranger-in-need vs. stranger-not-in-need) was conducted. The results revealed a significant effect of beneficiary, [$F(2,81) = 44.26$, $p < 0.001$, $\eta_p^2 = 0.84$]. Bonferroni-corrected *post hoc* test showed that participants felt happier when they getting reward for underprivileged students than general students ($p < 0.001$), while a similar positive feeling was found toward gaining money for themselves and for underprivileged students ($p = 0.62$). It indicated that participants experienced more empathic emotion in the high-empathy condition rather than in the low-empathy condition. One-way ANOVA on the subjective rating of motivation to win for the beneficiary (self vs. stranger-in-need vs. stranger-not-in-need) revealed a significant effect of beneficiary, [$F(2,81) = 132.37$, $p < 0.001$, $\eta_p^2 = 0.93$]. Bonferroni-corrected *post hoc* test showed that the motivation to win for the self (4.61) was higher than that for two strangers, (*ps* < 0.001), whereas the motivation to win for the stranger-in-need (3.84) was higher than that for stranger-not-in-need (1.87), ($p < 0.001$).

The FRN Results

Figure 3A shows grand-average ERP waveforms at the FCz site. The mean amplitude of FRN was analyzed by a 3 (Beneficiary: self vs. stranger-in-need vs. Stranger-not-in-need) × 2 (Reward Valence: gain vs. no-gain) repeated measures ANOVA. The results showed that the main effect of the beneficiary was

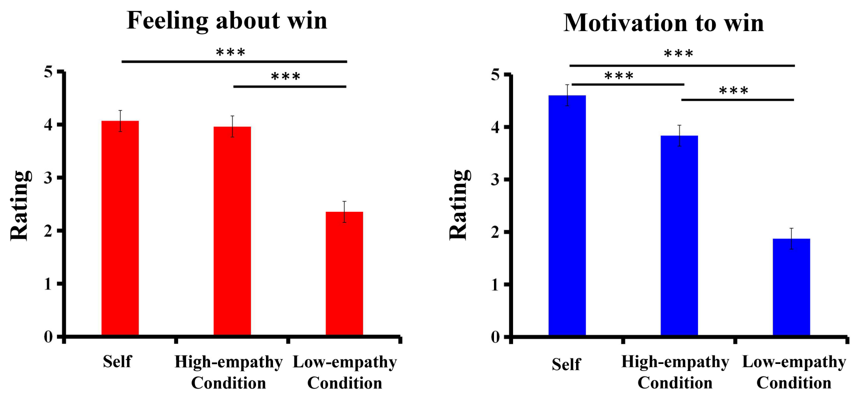

FIGURE 2 | Subjective ratings for motivation to win and feelings about win. Error bars indicate SEM (standard error of the mean). ***$p < 0.001$.

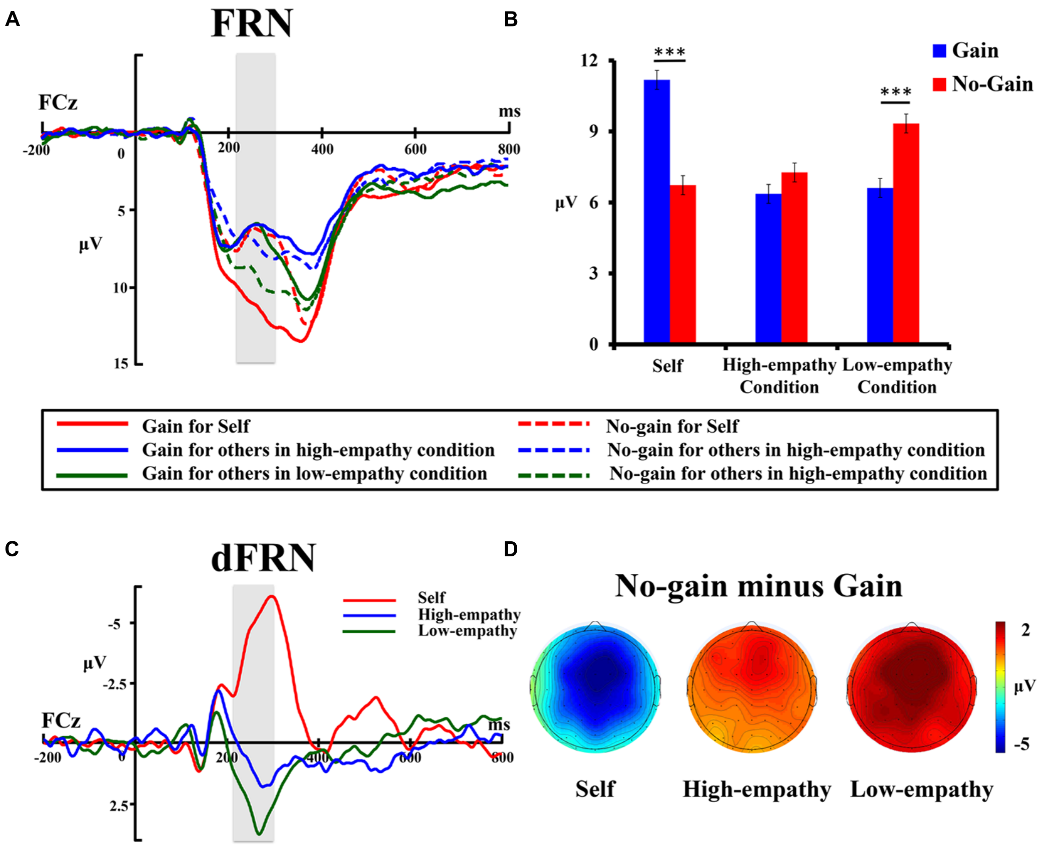

FIGURE 3 | (A) Grand-average ERP waveforms from the FCz electrode site. The gray areas highlight the time window of the FRN (210–300 ms) used for statistical analysis. **(B)** The bar graphs show the mean value of the FRN amplitude for each condition. Error bars indicate standard error of the mean (SEM). ***$p < 0.001$. **(C)** Difference waveforms of no-gain minus gain. The gray areas highlight the time window of the dFRN (210–300 ms) used for statistical analysis. **(D)**Topographic maps of different waveforms (no-gain minus gain) in the 210–300 ms time window for self, high-empathy, and low-empathy conditions.

significant [$F(2,26) = 6.52$, $p < 0.01$, $\eta_p^2 = 0.33$], indicating that the size of the FRN effect was different among the three beneficiary conditions. The main effect of reward valence was not significant [$F(1,27) = 0.43$, $p = 0.5$]. Moreover, the ANOVA revealed a significant interaction between Beneficiary and Valence [$F(2,26) = 23.14$, $p < 0.001$, $\eta_p^2 = 0.64$].

Further simple effect analyses were conducted to investigate the interaction. As we can see in **Figure 3B**, no-gain trials showed greater negativity than gain trials only for self condition ($p < 0.001$), while the typical pattern was reversed in trails for the outcomes of strangers. In the stranger-in-need condition, the FRN differentiation between gain and no-gain was remarkably diminished and no significant FRN difference was found between gain and no-gain ($p = 0.31$). On the other hand, in the stranger-not-in-need condition, the FRN difference between gain and no-gain outcomes was reversed, with more negative-going FRN for gain than no-gain outcomes ($p < 0.001$), indicating that participants regarded the gain for others as negative outcome (i.e., loss) for themselves only in the low-empathy condition.

In addition, we measured the mean amplitude of the FRN on the difference waves of no-gain minus gain (dFRN) for further repeated measures ANOVA. The dFRN for the participant's personal performance (self-dFRN) was calculated as self-no-gain minus self-gain, while the dFRN for the strangers (other-dFRN) was calculated as the other's no-gain minus the other's gain. As **Figure 3C** showed, a significant main effect of beneficiary was found [$F(2,26) = 23.14$, $p < 0.001$, $\eta_p^2 = 0.640$]. Bonferroni-corrected *post hoc* test showed that the self-dFRN (-4.47 μV) was significantly more negative than two other-dFRNs (*p*s < 0.001), whereas the other-dFRN in the high-empathy condition (0.91 μV) was smaller than that in the low-empathy condition (2.73 μV), though only marginally significant ($p = 0.06$). Scalp topographies of the dFRN also revealed these differences among three conditions (**Figure 3D**).

Pearson correlation analysis was conducted between the FRN amplitudes and subjective assessment scores. The results showed that the FRN was negatively correlated with subjective scores of motivation ($r = -0.282$; $p < 0.001$) and empathic emotion ($r = -0.336$; $p < 0.001$), indicating that the more the participants motivated to win or felt affect to the other's outcomes, the more the FRN enhanced. However, no correlation was found between the FRN amplitude and self-report altruism scale ($p = 0.518$).

The P3 Results

Figure 4A shows grand-average ERP waveforms at CPz electrode site. The peak amplitude of P3 at CPz was analyzed by a 3 (Beneficiary: self vs. stranger-in-need vs. Stranger-not-in-need) × 2 (Reward Valence: gain vs. no-gain) repeated measure ANOVA. The main effect of beneficiary was significant [$F(2,26) = 15.19$, $p < 0.001$, $\eta_p^2 = 0.54$], but the effect of reward valence was not found [$F(1,27) = 2.98$, $p = 0.09$]. Bonferroni-corrected *post hoc* test showed that the P3 was larger in the self condition than in both stranger conditions (*p*s < 0.001), while the P3 amplitude was the smallest in the high empathy condition (*p*s < 0.05). More importantly, the ANOVA revealed a significant interaction between Beneficiary and Valence [$F(2,26) = 7.29$, $p < 0.001$, $\eta_p^2 = 0.36$]. Further simple effect analysis was conducted to examine this interaction. As we can see in **Figure 4B**, the results revealed that the gain feedback induced a larger P3 than the no-gain did only in the self condition ($p < 0.001$), but this P3 difference between gain and no-gain feedback was not observed in the other two conditions. Scalp topographies of the P3 also revealed these differences among three conditions (**Figure 4C**).

Pearson correlation analysis was also conducted between the P3 amplitudes and subjective assessment scores. However, no significant correlation was found between P3 with either subjective scores of motivation ($p = 0.11$), empathic emotion ($p = 0.15$), or the rating of self-report altruism ($p = 0.29$).

DISCUSSION

In this study, using the gambling task in which participants made money for themselves and two strangers, we examined the neural correlates of empathy modulating the evaluation of outcomes that involved benefit conflict. The ERP results showed that an inversed FRN effect occurred when evaluating another person's outcomes in the low-empathy condition, but did not appear in the high-empathy condition. Further, the P3 was larger for the gain outcome than the no-gain outcome in the self condition, but did not show the valence effect in the two stranger conditions. The results of the present study suggest that empathy could modulate the neural responses to altruistic outcomes in which increasing welfare of others could result in a cost of the self.

The FRN was more negative-going to no-gain than to gain when gambling for self, but reversed in opposite polarity when gambling for others in the low-empathy condition. This finding is consistent with previous studies in which they found a negative-going FRN for antagonist's gain, as if gains of others were interpreted as losses of oneself (Fukushima and Hiraki, 2006; Itagaki and Katayama, 2008; Marco-Pallares et al., 2010). Given that the FRN elicited by self-outcome (self-FRN) represented the motivational/affectional impact of the outcomes (Gehring and Willoughby, 2002), our results provide a direct evidence to the theory that the FRN elicited by other's outcome (other-FRN) also reflects the response of inner meanings of positive/negative stimuli. Previous studies have reported that when other's outcomes did not relate to one's own benefit, the other-FRN showed the same polarity as the self-FRN (Yu and Zhou, 2006; Fukushima and Hiraki, 2009; Kang et al., 2010; Ma et al., 2011; Leng and Zhou, 2014), indicating that the neural activities of evaluating other's outcomes are comparable with those of evaluating one's own. However, the circumstances become complicated when the interests of self conflict with that of others. Based on the ideally defined hypothesis in traditional economics that people are generally maximizing their own interests, it was not surprising that the FRN was more positive-going to no-gain than to gain when gambling for strangers in low empathy condition, indicating that individuals evaluated the outcomes of decisions depending on their own motivation, and regarded the gain of others as the loss of self in the interest-competing context.

Critically, as we expected, the other-FRN was not reversed in the high-empathy condition, and showed no difference between other's gain and no-gain. It might suggest that the neural activities in the low-empathy condition are sensitively elicited by other's gains, while the neural responses to other's outcomes are inhibited in the high-empathy condition. We believe that the different patterns of FRN between the low- and high-empathy condition may be attributed to the buffer function of empathy. Behavioral studies have found that empathy, an ability to infer and share the mental and emotional states of others (Preston and de Waal, 2002; Lamm et al., 2011), can induce altruistic motivation to increase other's welfare and improve more prosocial behaviors (Eisenberg and Fabes, 1990; Batson, 2008). Subsequently, the findings in neuro-imaging studies provided a neural substrate perspective to understand the

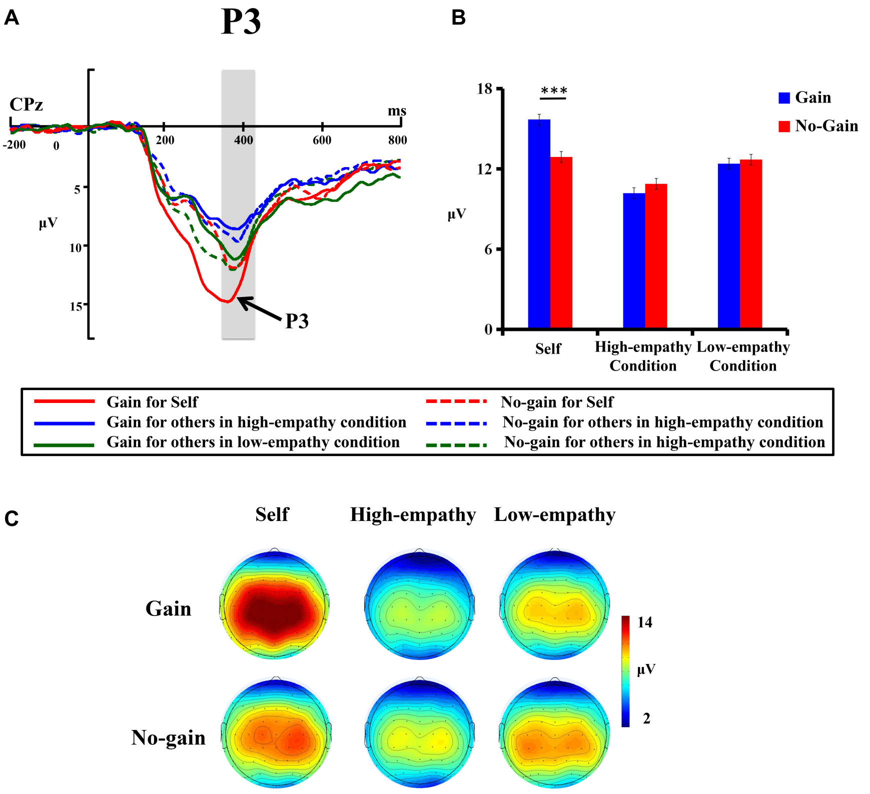

FIGURE 4 | (A) Grand-average ERP waveforms from the CPz electrode site. The gray areas highlight the time window of the P3 (330–430 ms) in which the peak amplitude was measured. **(B)** The bar graphs show the mean value of the P3 amplitude for each condition. Error bars indicate standard error of the mean (SEM). $^{***}p < 0.001$. **(C)** Topographic maps of the P3 for the self, high-empathy, and low-empathy conditions.

effect of empathy on altruistic decisions. Using the functional magnetic resonance imaging (fMRI), a number of studies have found that perceiving others' affective states would activate neural network involving in the first-hand experience of these states called "shared representative network" (Singer et al., 2004; Cacioppo and Decety, 2009; Lamm et al., 2011; Marsh et al., 2014). For example, Singer et al. (2004) asked volunteers to observe their lovers who could elicit their highest level of empathy for suffering pain. The results showed that the brain areas, such as the anterior insula (AI) and dorsal-anterior midcingulate cortex (dACC), were activated in both direct pain and vicarious pain situations. Later, Mobbs et al. (2009) extended pain empathy to social emotions by contrasting the neural responses to the socially desirable others getting reward vs. to directly gaining money for themselves. They found a similar reward mechanism employed in both situations, confirming that the "shared representation network" could apply to complex social emotions elicited by favorable or unfavorable outcomes. Taking these findings together, the corresponding neural network would be evoked in individuals who are induced high empathy, which makes them be more likely to experience the other's feeling. Therefore, in the high-empathy condition of the present study, individuals would feel internal pain for needy students and have a strong altruistic motivation to help them, which could counteract the suffering of their own loss. We thus observed, a decreased FRN when participants evaluated other's gain that led to the loss of themselves.

The finding of P3 showed a main effect of beneficiary in which the P3 amplitudes were larger in the self condition than in the two stranger's conditions. This is consistent with previous finding (Ma et al., 2011) that the mean amplitude of P3 was larger for the self-execution than for friends or strangers. Interestingly, the finding that P3 was larger in the low empathy condition than in the high empathy condition did not congruent with the recent works of Leng and Zhou (2010, 2014) who found that P3 was more positive for the friends than for the strangers. Since the P3 reflects the allocation of cognitive resources (Polich, 1987, 2007; Yang et al., 2015; Hu et al., 2017), these results suggest that the larger P3 indicates that more resources are allocated to the ongoing task. As we expected, the most cognitive resources were used to evaluate self-related feedback in order to maximize one's own profits. However, when the interests were conflict between oneself and others, cognitive load was increased to balance two competing motivations, egoistic motives and altruistic motives. In other words, the cognitive resources of outcome evaluation were affected by the processing of empathy. Therefore, the P3 was smallest in high empathy condition than in low empathy condition indicating that more cognitive resources were occupied by processes of empathic concern and conflict management.

In addition, the valence effect of P3 was only observed in the self-condition, but disappeared in both high-empathy and low-empathy conditions. Such inapparent valence effect on other's feedback was consistent with the findings of previous studies on neural processes of outcome evaluation when the interest of oneself was conflict with that of others (Fukushima and Hiraki, 2006; Itagaki and Katayama, 2008; Leng and Zhou, 2014). Moreover, the P3 amplitudes did not covary with subjective scores of empathy nor motivation, suggesting that different from the FRN, the P3 effect of outcomes was not modulated by empathy nor motivation. Given that the P3 effect was only observed in self-related feedback rather than in other-related feedback, we thus suggest that the P3 reflects an allocation of attentional resources that may distinguish between "self" and "others." This interpretation can also be supported by the previous studies which found that P3 was larger for self-relevant stimuli relative to control stimuli (Gray et al., 2004; Tacikowski and Nowicka, 2010), suggesting that P3 is an index of the allocation of attentional resources, and evokes by autobiographical stimuli, instead of empathic emotion.

Moreover, we found that the subjective score of empathic emotion correlated with FRN, but did not covary with P3, indicating that empathy play a central role in the early stage of neural processes when we evaluating other's outcomes. However, no significant correlation was found between the rating of self-report altruism with either the FRN or the P3, suggesting that the individual difference in altruistic trait have no effect on the processes of outcome evaluation. These results together may support the hypothesis that the neural mechanism underlying empathy could be independent of that underlying altruistic tendencies (Tankersley et al., 2007).

In sum, the current study investigated the neural mechanism of how empathy modulates outcome evaluation toward others in a gambling task involving conflict between self and other interest. A reversed FRN effect was elicited for strangers only in the low-empathy condition, whereas such FRN pattern was not observed in the high-empathy condition. These findings indicate that the neural processes for other's outcomes are modulated by individuals' empathy levels. Specifically, the high level of empathy could let people think from the perspective of others and induce a stronger altruistic motivation which counteracts with the egoistic motivation. These findings support previous studies showing that empathy could promote prosocial decision-making and cooperative behaviors (Batson and Ahmad, 2001; Smith, 2006; Christov-Moore et al., 2014) and provide the underlying neural evidence to help us understand prosocial behaviors better. In addition, there was the P3 valence effect only in the self condition, but not in the two stranger conditions, regardless of the levels of empathy, indicating that P3 is more sensitive to the distribution of attention resource in self-relevant information.

There are limitations in the present study. We manipulated the level of empathy through impoverishing strangers, which might result in the activation of an altruistic motivation. Thus it is hard to exclude the influence of motivation on the evaluation processing of other's outcomes in the present study. In the future studies, it would be worthwhile to separate the two important factors: altruistic motivation and empathy, and differentiate their influences on the evaluation of other's outcomes. In addition, accumulating evidence has shown that there are differences in the capacity of empathy between females and males (Schirmer et al., 2007; Gardner et al., 2012; Christov-Moore et al., 2014) and among individuals with different social value orientations (Declerck and Bogaert, 2008). Fukushima and Hiraki (2006) also found that the discernable MFN to the opponent's outcomes only emerged for female participants, but not for males. Therefore, the individual difference of empathy modulating outcome evaluation is a very interesting issue, which is worth further research in the future. Moreover, the ecological validity of the current experimental design may need to be improved. In our daily life, people usually make decisions and evaluate outcomes in more complex social contexts. Other individual's attitudes and behaviors also have impacts on how we evaluate other's outcomes. These factors should also be considered in the future studies.

AUTHOR CONTRIBUTIONS

XL and XM designed the study. XL and XH collected and analyzed the data. XL wrote the manuscript. XM, KS, and XH edited the manuscript. All authors reviewed the manuscript.

FUNDING

This work was supported by grants from the Major Project of the National Social Science Foundation (13&ZD155), the National Natural Science Foundation of China (31771206 and 31400888), and the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China (15XNLQ05).

REFERENCES

Banissy, M. J., Kanai, R., Walsh, V., and Rees, G. (2012). Inter-individual differences in empathy are reflected in human brain structure. *Neuroimage* 62, 2034–2039. doi: 10.1016/j.neuroimage.2012.05.081 doi: 10.1016/j.neuroimage.2012.05.081

Batson, C. D. (2008). "Empathy-induced altruistic motivation," in *Prosocial Motives, Emotions, and Behavior*, eds P. R. Shaver and M. Mikulincer (Washington, DC: American Psychological Association), 15–34.

Batson, C. D., and Ahmad, N. (2001). Empathy-induced altruism in a prisoner's dilemma II: what if the target of empathy has defected? *Eur. J. Soc. Psychol.* 31, 25–36. doi: 10.1002/ejsp.26

Batson, C. D., and Moran, T. (1999). Empathy-induced altruism in a prisoner's dilemma. *Eur. J. Soc. Psychol.* 29, 909–924. doi: 10.1002/(SICI)1099-0992(199911)29:7<909::AID-EJSP965>3.0.CO;2-L

Batson, C. D., and Shaw, L. L. (1991). Evidence of prosocial motives toward a pluralism for altruism. *Psychol. Inq.* 2, 107–122. doi: 10.1207/s15327965pli0202_1

Bernhardt, B. C., and Singer, T. (2012). The neural basis of empathy. *Annu. Rev. Neurosci.* 35, 1–23. doi: 10.1146/annurev-neuro-062111-150536

Cacioppo, J. T., and Decety, J. (2009). What are the brain mechanisms on which psychological processes are based? *Perspect. Psychol. Sci.* 4, 10–18. doi: 10.1111/j.1745-6924.2009.01094.x

Chou, K. L. (1996). The rushton, chrisjohn and fekken self-report altruism scale: a chinese translation. *Pers. Individ. Diff.* 21, 297–298. doi: 10.1016/0191-8869(96)00040-2

Christov-Moore, L., Simpson, E. A., Coude, G., Grigaityte, K., Iacoboni, M., and Ferrari, P. F. (2014). Empathy: gender effects in brain and behavior. *Neurosci. Biobehav. Rev.* 46, 604–627. doi: 10.1016/j.neubiorev.2014.09.001 doi: 10.1016/j.neubiorev.2014.09.001

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educ. Psychol. Meas.* 33, 107–112. doi: 10.1177/001316447303300111

de Waal, F. B. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annu. Rev. Psychol.* 59, 279–300. doi: 10.1146/annurev.psych.59.103006.093625

Declerck, C. H., and Bogaert, S. (2008). Social value orientation: related to empathy and the ability to read the mind in the eyes. *J. Soc. Psychol.* 148, 711–726. doi: 10.3200/SOCP.148.6.711-726

Eisenberg, N., and Fabes, R. A. (1990). Empathy: conceptualization, measurement, and relation to prosocial behavior. *Motiv. Emot.* 14, 131–149. doi: 10.1007/BF00991640

Fukushima, H., and Hiraki, K. (2006). Perceiving an opponent's loss: gender-related differences in the medial-frontal negativity. *Soc. Cogn. Affect. Neurosci.* 1, 149–157. doi: 10.1093/scan/nsl020

Fukushima, H., and Hiraki, K. (2009). Whose loss is it? Human electrophysiological correlates of non-self reward processing. *Soc. Neurosci.* 4, 261–275. doi: 10.1080/17470910802625009

Gardner, M. R., Sorhus, I., Edmonds, C. J., and Potts, R. (2012). Sex differences in components of imagined perspective transformation. *Acta Psychol.* 140, 1–6. doi: 10.1016/j.actpsy.2012.02.002

Gehring, W. J., and Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science* 295, 2279–2282. doi: 10.1126/science.1066893

Gray, H. M., Ambady, N., Lowenthal, W. T., and Deldin, P. (2004). P300 as an index of attention to self-relevant stimuli. *J. Exp. Soc. Psychol.* 40, 216–224. doi: 10.1016/S0022-1031(03)00092-1

Hajcak, G., Holroyd, C. B., Moser, J. S., and Simons, R. F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology* 42, 161–170. doi: 10.1111/j.1469-8986.2005.00278.x

Hamilton, W. D. (1964). The genetical evolution of social behavior. *J. Theor. Biol.* 7, 1–16. doi: 10.1016/0022-5193(64)90038-4

Hauser, T. U., Iannaccone, R., Stampfli, P., Drechsler, R., Brandeis, D., Walitza, S., et al. (2014). The feedback-related negativity (FRN) revisited: new insights into the localization, meaning and network organization. *Neuroimage* 84, 159–168. doi: 10.1016/j.neuroimage.2013.08.028

Hewig, J., Kretschmer, N., Trippe, R. H., Hecht, H., Coles, M. G., Holroyd, C. B., et al. (2011). Why humans deviate from rational choice. *Psychophysiology* 48, 507–514. doi: 10.1111/j.1469-8986.2010.01081.x

Holroyd, C. B., and Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* 109, 679–709. doi: 10.1037/0033-295X.109.4.679

Holroyd, C. B., Hajcak, G., and Larsen, J. T. (2006). The good, the bad and the neutral: electrophysiological responses to feedback stimuli. *Brain Res.* 1105, 93–101. doi: 10.1016/j.brainres.2005.12.015

Holroyd, C. B., Larsen, J. T., and Cohen, J. D. (2004). Context dependence of the event-related brain potential associated with reward and punishment. *Psychophysiology* 41, 245–253. doi: 10.1111/j.1469-8986.2004.00152.x

Hu, X., Xu, Z., and Mai, X. (2017). Social value orientation modulates the processing of outcome evaluation involving others. *Soc. Cogn. Affect. Neurosci.* 12, 1730–1739. doi: 10.1093/scan/nsx102

Itagaki, S., and Katayama, J. (2008). Self-relevant criteria determine the evaluation of outcomes induced by others. *Neuroreport* 19, 383–387. doi: 10.1097/WNR.0b013e3282f556e8

Kang, S. K., Hirsh, J. B., and Chasteen, A. L. (2010). Your mistakes are mine: self-other overlap predicts neural response to observed errors. *J. Exp. Soc. Psychol.* 46, 229–232. doi: 10.1016/j.jesp.2009.09.012

Lamm, C., Decety, J., and Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* 54, 2492–2502. doi: 10.1016/j.neuroimage.2010.10.014

Leng, Y., and Zhou, X. (2010). Modulation of the brain activity in outcome evaluation by interpersonal relationship: an ERP study. *Neuropsychologia* 48, 448–455. doi: 10.1016/j.neuropsychologia.2009.10.002

Leng, Y., and Zhou, X. (2014). Interpersonal relationship modulates brain responses to outcome evaluation when gambling for/against others: an electrophysiological analysis. *Neuropsychologia* 63, 205–214. doi: 10.1016/j.neuropsychologia.2014.08.033

Linden, D. E. (2005). The p300: where in the brain is it produced and what does it tell us? *Neuroscientist* 11, 563–576.

Luo, Y., Feng, C., Wu, T., Broster, L. S., Cai, H., Gu, R., et al. (2015). Social comparison manifests in event-related potentials. *Sci. Rep.* 5:12127. doi: 10.1038/srep12127

Ma, Q., Shen, Q., Xu, Q., Li, D., Shu, L., and Weber, B. (2011). Empathic responses to others' gains and losses: an electrophysiological investigation. *Neuroimage* 54, 2472–2480. doi: 10.1016/j.neuroimage.2010.10.045

Marco-Pallares, J., Cucurell, D., Münte, T. F., Strien, N., and Rodriguez-Fornells, A. (2011). On the number of trials needed for a stable feedback-related negativity. *Psychophysiology* 48, 852–860. doi: 10.1111/j.1469-8986.2010.01152.x

Marco-Pallares, J., Kramer, U. M., Strehl, S., Schroder, A., and Munte, T. F. (2010). When decisions of others matter to me: an electrophysiological analysis. *BMC Neurosci.* 11:86. doi: 10.1186/1471-2202-11-86

Marsh, A. A., Stoycos, S. A., Brethel-Haurwitz, K. M., Robinson, P., VanMeter, J. W., and Cardinale, E. M. (2014). Neural and cognitive characteristics of extraordinary altruists. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15036–15041. doi: 10.1073/pnas.1408440111

Masaki, H., Takeuchi, S., Gehring, W. J., Takasawa, N., and Yamazaki, K. (2006). Affective-motivational influences on feedback-related ERPs in a gambling task. *Brain Res.* 1105, 110–121. doi: 10.1016/j.brainres.2006.01.022

Miltner, W. H. R., Braun, C. H., and Coles, M. G. H. (2014). Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a "generic" neural system for error detection. *J. Cogn. Neurosci.* 9, 788–798. doi: 10.1162/jocn.1997.9.6.788

Mobbs, D., Yu, R., Meyer, M., Passamonti, L., Seymour, B., Calder, A. J., et al. (2009). A key role for similarity in vicarious reward. *Science* 324:900. doi: 10.1126/science.1170539

Nieuwenhuis, S., Holroyd, C. B., Mol, N., and Coles, M. G. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neurosci. Biobehav. Rev.* 28, 441–448. doi: 10.1016/j.neubiorev.2004.05.003

Paul, K., and Pourtois, G. (2017). Mood congruent tuning of reward expectation in positive mood: evidence from FRN and theta modulations. *Soc. Cogn. Affect. Neurosci.* 12, 765–774. doi: 10.1093/scan/nsx010

Peterburs, J., Voegler, R., Liepelt, R., Schulze, A., Wilhelm, S., Ocklenburg, S., et al. (2017). Processing of fair and unfair offers in the ultimatum game under social observation. *Sci. Rep.* 7:44062. doi: 10.1038/srep44062

Polich, J. (1987). Comparison of p300 from a passive tone sequence paradigm and an active discrimination task. *Psychophysiology* 24, 41–46. doi: 10.1111/j.1469-8986.1987.tb01859.x

Polich, J. (2007). Updating p300: an integrative theory of p3a and p3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Preston, S. D., and de Waal, F. B. (2002). Empathy: its ultimate and proximate bases. *Behav. Brain Sci.* 25, 1–20.

Rilling, J. K., and Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annu. Rev. Psychol.* 62, 23–48. doi: 10.1146/annurev.psych.121208.131647

Rueckert, L., and Naybar, N. (2008). Gender differences in empathy: the role of the right hemisphere. *Brain Cogn.* 67, 162–167. doi: 10.1016/j.bandc.2008.01.002

Rushton, J. P., Chrisjohn, R. D., and Fekken, G. C. (1981). The altruistic personality and the self-report altruism scale. *Pers. Individ. Diff.* 2, 293–302. doi: 10.1016/0191-8869(81)90084-2

Schirmer, A., Simpson, E., and Escoffier, N. (2007). Listen up! Processing of intensity change differs for vocal and nonvocal sounds. *Brain Res.* 1176, 103–112. doi: 10.1016/j.brainres.2007.08.008

Schupp, H. T., Cuthbert, B. N., Bradley, M. M., Cacioppo, J. T., Ito, T., and Lang, P. J. (2000). Affective picture processing: the late positive potential is modulated by motivational relevance. *Psychophysiology* 37, 257–261. doi: 10.1111/1469-8986.3720257

Semlitsch, H. V., Anderer, P., Schuster, P., and Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology* 23, 695–703. doi: 10.1111/j.1469-8986.1986.tb00696.x

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 1157–1162. doi: 10.1126/science.1093535

Smith, A. (2006). Cognitive empathy and emotional empathy in human behavior and evolution. *Psychol. Rec.* 56, 3–21. doi: 10.1007/BF03395534

Tacikowski, P., and Nowicka, A. (2010). Allocation of attention to self-name and self-face: an ERP study. *Biol. Psychol.* 84, 318–324. doi: 10.1016/j.biopsycho.2010.03.009

Tankersley, D., Stowe, C. J., and Huettel, S. A. (2007). Altruism is associated with an increased neural response to agency. *Nat. Neurosci.* 10, 150–151. doi: 10.1038/nn1833

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Q. Rev. Biol.* 46, 35–57. doi: 10.1086/406755

Wang, Y., Qu, C., Luo, Q., Qu, L., and Li, X. (2014). Like or dislike? affective preference modulates neural response to others' gains and losses. *PLoS One* 9:e105694. doi: 10.1371/journal.pone.0105694

Wu, J., Sun, X., Wang, L., Zhang, L., Fernandez, G., and Yao, Z. (2017). Error consciousness predicts physiological response to an acute psychosocial stressor in men. *Psychoneuroendocrinology* 83, 84–90. doi: 10.1016/j.psyneuen.2017.05.029

Yang, Q., Tang, P., Gu, R., Luo, W., and Luo, Y. J. (2015). Implicit emotion regulation affects outcome evaluation. *Soc. Cogn. Affect. Neurosci.* 10, 824–831. doi: 10.1093/scan/nsu124

Yeung, N., and Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *J. Neurosci.* 24, 6258–6264. doi: 10.1523/JNEUROSCI.4537-03.2004

Yu, R., and Zhou, X. (2006). Brain responses to outcomes of one's own and other's performance in a gambling task. *Neuroreport* 17, 1747–1751. doi: 10.1097/01.wnr.0000239960.98813.50

Zhu, X., Wang, L., Yang, S., Gu, R., Wu, H., and Luo, Y. (2016). The motivational hierarchy between the personal self and close others in the Chinese brain: an ERP study. *Front. Psychol.* 7:1476. doi: 10.3389/fpsyg.2016.01467

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ORIGINAL RESEARCH
published: 23 May 2018
doi: 10.3389/fpsyg.2018.00749

Advantageous Inequity Aversion Does Not Always Exist: The Role of Determining Allocations Modulates Preferences for Advantageous Inequity

Ou Li[1,2,3†], *Fuming Xu*[4†] *and Lei Wang*[1,2*]

[1] School of Management, Zhejiang University, Hangzhou, China, [2] Neuromanagement Lab, Zhejiang University, Hangzhou, China, [3] School of Psychology, Central China Normal University, Wuhan, China, [4] School of Psychology, Jiangxi Normal University, Nanchang, China

OPEN ACCESS

Edited by:
Yefeng Chen,
Zhejiang University, China

Reviewed by:
Anthony John Porcelli,
Marquette University, United States
Zhixian Cheng,
Yale University, United States

***Correspondence:**
Lei Wang
wang_lei@zju.edu.cn

†These authors have contributed equally to this work and co-first author.

Specialty section:
This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Psychology

Received: *31 December 2017*
Accepted: *30 April 2018*
Published: *23 May 2018*

Citation:
Li O, Xu F and Wang L (2018) Advantageous Inequity Aversion Does Not Always Exist: The Role of Determining Allocations Modulates Preferences for Advantageous Inequity. Front. Psychol. 9:749. doi: 10.3389/fpsyg.2018.00749

Previous studies have shown that people would like to sacrifice benefits to themselves in order to avoid inequitable outcomes, not only when they receive less than others (disadvantageous inequity aversion) but also when they receive more (advantageous inequity aversion). This feature is captured by the theory of inequity aversion. The present study was inspired by what appears to be asymmetry in the research paradigm toward advantageous inequity aversion. Specifically, studies that supported the existence of advantageous inequity aversion always relied on the paradigm in which participants can *determine* allocations. Thus, it is interesting to know what would occur if participants could not determine allocations or simply passed judgment on *predetermined* allocations. To address this, a behavioral experiment ($N = 118$) and a skin conductance response (SCR) experiment ($N = 29$) were adopted to compare participants' preferences for advantageous inequity directly when allocations were *determined* and when allocations were *predetermined* in an allocating task. In the *determined* condition, participants could divide by themselves a sum of money between themselves and a matched person, whereas in the *predetermined* condition, they could simply indicate their satisfaction with an equivalent program-generated allocation. It was found that, compared with those in the *determined* condition, participants in the *predetermined* condition behaved as if they liked the advantageous inequity and equity to the same degree (Experiment One) and that the SCRs elicited by advantageous inequity had no differences from those elicited by equity, suggesting that participants did not feel negatively toward advantageous inequity in this situation (Experiment Two). The present study provided mutual corroboration (behavioral and electrophysiological data) to document that advantageous inequity aversion may differ as a function of the individual's role in determining allocations, and it would disappear if individual cannot determine allocations.

Keywords: inequity aversion, fairness decision-making, advantageous inequity, SCRs, sense of agency, responsibility

INTRODUCTION

Equity is a fundamental concern in people's interactions that influences many aspects of daily life, from how people share their resources with partners to how policymakers shape income distribution policy. A key component of equity is related to inequity aversion, which means that individuals resist inequitable outcomes; that is, they are willing to give up some material payoff to move in the direction of more equitable outcomes (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), not only when they receive less than others (i.e., disadvantageous inequity aversion, DI) but also when they receive more (i.e., advantageous inequity aversion, AI). It is well-accepted that inequity aversion captures the critical feature of humans' fairness in decision-making (Fehr and Schmidt, 2006; Tricomi and Sullivan-Toole, 2015). Its empirical applicability has been confirmed not only by several subsequent experiments conducted by Ernst Fehr et al. (Falk et al., 2003; Knoch et al., 2006; Fehr et al., 2008) but also by other researchers from the fields of psychology (Blake and McAuliffe, 2011; Güroglu et al., 2011), economics (Eckel and Grossman, 2001; Fershtman et al., 2012), anthropology (Henrich et al., 2001), neuroscience (Sanfey et al., 2003; Tricomi et al., 2010; Tricomi and Sullivan-Toole, 2015), and other disciplines.

Given that inequity aversion is the main theory for understanding humans' fairness behaviors and can even be seen as the preferred approach to explore this issue (Xu et al., 2016), an in-depth analysis of inequity aversion seems essential. We believe that at least one deficiency has remained unsolved in the current research; that is, the research paradigm toward AI is asymmetric. Currently, studies on AI always have the participants themselves decide how to divide some resources between themselves and others and use the proportion that they share as the measure of their degree of AI (Tricomi and Sullivan-Toole, 2015; Xu et al., 2016)[1]. Studies using this paradigm have found that the majority of participants would offer 40–50% of the total sum to others (see a meta-analysis: Oosterbeek et al., 2004; or a review: Güth and Kocher, 2014). Therefore, they claimed that people have a strong preference for equity instead of for self-interest. However, when considering this paradigm, we can easily find an inherent feature that may weaken the reliability of such a conclusion; that is, all of the final offers in this paradigm are the results of the self-executed actions of participants. That is, participants can determine allocations on their own initiative. For example, a participant considers how to divide a sum of 10 RMB; he can keep all for himself (10, 0), divide the sum equitably (5, 5), or choose any amount *x* in the range of 10 (*x*, 10-*x*) (henceforth, the number on the left is given to the participant, while the number on the right is given to the other). For this, it is implied that current studies, most of which support the existence of AI, have relied solely on the paradigm in which participants can *determine* allocations while ignoring the paradigm in which they cannot.

Given this asymmetry, it is interesting to know what would occur if participants could not determine allocations or simply passed judgment on *predetermined* allocations. Indeed, to date, few studies have involved *predetermined* allocations or inactions. For example, in Albrecht et al.'s (2013) design, participants were required to indicate their satisfaction with a series of allocations that were assigned by experimenters [including an advantageous one, i.e., (20, 30)]. Furthermore, in Moser et al.'s (2014) and Lamichhane et al.'s (2014) ultimatum game task, participants were placed in the role of the responder instead of in the general role of proposer, such that they could merely say "yes" or "no" to an allocation (advantageous, equitable, or disadvantageous) imposed by the opponent but could not determine how to divide the offer. All of these studies found that people in such a situation appeared to prefer advantageous inequity, which conflicted with the theory of inequity aversion. Although these studies gave some insights on the open questions, the paradigm feature of *determined* or *predetermined* was not at their center[2], and they also failed to manipulate it. Therefore, it is still unclear whether the individual's role in determining allocations could affect their AI degree. To address this, the present study may be the first to manipulate the paradigm feature of *determined*/*predetermined* and investigate its effect on AI.

In the area of decision-making, the individual's reaction to outcomes following their actual actions may be different from reactions to outcomes following inactions. For example, Kahneman and Tversky (1982) found that negative outcomes resulting from actions induced more regret than the same outcomes resulting from inactions. Such an effect can also be manifested by the omission bias, i.e., the tendency that people are more likely to judge harmful actions as worse or less moral than equally harmful inactions (Ritov and Baron, 1990). Subsequently, Choshen-Hillel and Yaniv (2011, 2012) extended this action effect to the area of prosocial preference by showing that the sense of agency, which was defined as "a person's degree or level of control over her or his outcomes and those of other parties" in their publication, can increase one's concern with another's welfare. Considering outcomes (11, 10) and (10, 11), in one of their experiments, 26.7% of the participants in the high-agency group chose the other-dominated outcome (10, 11), even at a financial cost to themselves. In contrast, only 6.5% of the participants in the low-agency group chose the same outcome. For a decision-maker, since an equitable allocation (vs. an advantageous allocation) is more in the interest of others, he may less frequently maintain equality when he cannot control outcomes than when he has the ability to do so. Choshen-Hillel and Yaniv (2011) considered that those who had a higher agency might view the others' outcomes as evidence of their own effectiveness and generosity and derive positive utilities from these outcomes. This idea is connected with the finding that the fact of having a choice itself can activate the subjective reward

[1] A routine way is to use the ultimatum game (Güth et al., 1982) or its modified version, the dictator game (Kahneman et al., 1986), as the design, in which participants act as the role of proposer who can divide the offer by himself (Fehr and Schmidt, 2006).

[2] The research objects of these studies are various. Albrecht et al. (2013) focused on the effect of status on satisfaction with relative rewards; Moser et al. (2014) focused on how social information and personal interests change fairness in decision-making; and Lamichhane et al. (2014) focused on the comparison of the neural basis between advantageous and disadvantageous inequity.

processing (for a review, see Leotti et al., 2010). For example, Leotti and Delgado (2011) found that merely anticipating an opportunity for choice could recruit the reward-related brain circuity, particularly the striatum. It is possible that the internal reward resulting from actual actions can partly offset the cost of giving to others, making individuals who have control more likely to be kind to others. From this review, it is suggested that the actual actions (i.e., *determining* an allocation on one's own initiative) could make people's focus change from self-interest to the other's welfare. Inaction (i.e., passively receiving a *predetermined* allocation), in contrast, would lead to the opposite effect. Taken together, the first hypothesis is that individuals' role in determining allocations can modulate their preference for advantageous inequity:

> *Hypothesis 1:* Participants would show a strong tendency of AI in the *determined* condition, as previous studies claimed, whereas their tendency of AI would diminish or even disappear in the *predetermined* condition.

Previous studies have indicated the importance of negative emotions in inequity aversion (Xu et al., 2016). When participants received an inequitable allocation, their self-reported negative emotional responses, such as anger, spite, or sadness, increased (Pillutla and Murnighan, 1996; Bosman et al., 2001). These correlations were also manifested by neuroimaging studies. For example, Harlé et al. (2012) and Sanfey et al. (2003) found that the anterior insula, a brain region specifically involved in representing negative emotional states, played a critical role in processing inequitable outcomes. Its activation degree could also be used to predict the likelihood of someone's acceptance or rejection of inequitable allocations (Tricomi and Sullivan-Toole, 2015). Therefore, the arousal of negative emotions can be an indicator of the aversion toward advantageous inequity in the present study. The skin conductance response (SCR) is a measurement of the electrical conductance of the skin. It is related to physiological arousal elicited by the cognitive inhibition system (Fowles, 1980), which, in turn, is supposed to be the biological basis of negative emotions (Gray, 1994) and is commonly used as an electrophysiological indication to evaluate the feeling of inequity/unfairness (Tricomi and Sullivan-Toole, 2015). Indeed, there is growing evidence that an increased SCR is positively correlated with an increased degree of inequity that one is exposed to and an increased negative feeling that one is experiencing (van't Wout et al., 2006; Civai et al., 2010; Hewig et al., 2011; Dunn et al., 2012). Analyzing SCRs to advantageous inequity in the *determined/predetermined* conditions can provide more convincing evidence on the present issue. Based on the aforementioned reviews, we formed the second hypothesis that SCRs elicited by advantageous inequity can be modulated by the *determined/predetermined* feature during allocation:

> *Hypothesis 2:* Advantageous inequity (vs. equity) might elicit a higher SCR in the *determined* condition, (i.e., a strong feeling of inequity), whereas the SCR elicited by the equivalent advantageous inequity may be the same as that elicited by equity.

To sum up, the main object of the present study was to investigate whether individual's preferences for advantageous inequity was affected by their role in determining allocations. More specifically, would they still resist advantageous inequity, or would they like it if this inequity did not result from their actions but was *predetermined*? To test this, we conducted a behavioral study (Experiment One) and a SCR study (Experiment Two) in the money distribution setting. In the *determined* condition, participants could decide by themselves how to divide a sum of money between themselves and a matched person, following the same procedure adopted by most studies (Fehr and Schmidt, 2006). In the *predetermined* condition, participants were asked to indicate whether an equivalent program-generated allocation between themselves and the match would satisfy them; in particular, they could not determine the allocation. Across the *determined/predetermined* conditions, the difference between participants' role in determining allocations was salient, while all other aspects between conditions were constant.

EXPERIMENT ONE

Materials and Methods

Participants

In total, 141 college students, who were anonymous to each other, were recruited in this experiment. Seven participants were excluded because they did not believe that they had performed the task together with a real person simultaneously (they rated below 4 points on a 7-point Likert scale presented after the experiment; the remaining participants rated 5.856 ± 0.989 points on average). In addition, participants who majored in psychology or economics (seven and nine, respectively) were also excluded. With this inclusion criterion, 118 participants aged 18–23 years old (on average, 19.57 ± 0.70 years) were finally left (45 males, 73 females). This experiment was approved by the research ethics board of Central China Normal University. Informed consent was obtained from all participants before the experiment.

Study Design

A 2 (Condition: *determined* vs. *predetermined*) × 5 [Allocation: (8, 2) vs. (7, 3) vs. (5, 5) vs. (3, 7) vs. (2, 8)] mixed design was employed, with the Condition referring to a between-subject factor and the Allocation referring to a within-subject factor. Thus, participants were randomly assigned to either the *determined* or the *predetermined* condition. Advantageous offers could be (8, 2) or (7, 3), the equitable offer was (5, 5), and disadvantageous offers could be (3, 7) or (2, 8). To make the distributions continuous, the present design also included the offers (6, 4) and (4, 6). In the past, a large body of studies have found that people tend to view an offer that is 10% over or under the median (i.e., 40–60%) as being reasonable (Camerer, 2003; Güth and Kocher, 2014). In other words, whether it is (6, 4) or (4, 6), people view it as a form of marginally equitable offers, although there is still some objective disparity. Because of this, (6, 4) and (4, 6) are not clear-cut: some decision-makers may see they as equitable, while others may disagree, producing mixed results overall. To clarify the difference between equity and inequity, studies of inequity aversion commonly exclude those confused offers from their design or analysis, as did Sanfey et al. (2003) and (van't Wout et al., 2006). The offers (6, 4) and (4, 6) therefore were included as filler tasks in the present study. Furthermore, (10, 0), (9, 1), (1, 9), and (0, 10) were excluded because these cases were too extreme to be chosen by real people and were unusual in daily life (Güth et al., 1982; Falk et al., 2003). The outcome factor was participants' preference for different Allocations, namely, the degree of inequity aversion.

Experimental Procedure

The participants completed experimental tasks collectively in a standard laboratory. Since the capacity of the laboratory was up to 43, we had to conduct four experiments successively, three of which were held in January 2017 and one in May. The participants were randomly assigned to one of four sessions. Using the collective measure has two advantages: first, it can equalize the external situation (such as time, temperature, and brightness) imposed on each participant; second, this allowed us to easily manipulate participants' belief that "I am completing the task with someone else simultaneously." We wanted to make participants believe that they played the task with a real person randomly selected from the same room at the same time because this could arouse their real motivation in decision-making. Actually, this was a deceptive operation, and the response of the alleged partners was set by experimenters (we debriefed participants at the close of the experiment about the true nature of the research).

All tasks were conducted by computer. As illustrated in **Figure 1**, at the beginning of each trial, a fixation appeared as a cue for 3,000 ms on the black screen, which was followed by a pre-task. As the very definition of inequity is receiving uneven outcomes despite investing the same effort (Adams, 1965), we decided to use a pre-task so that the effort of both sides could be balanced out. Before the experiments, five psychology graduates were recruited to select pre-tasks from Raven Matrices, which ensured that the chosen pre-task was simple enough and could not impact the following tasks. According to the check beforehand, all pre-tasks ($n = 14$) were easy to solve (on average, 2.017 ± 0.913 points on a 7-point Likert scale for difficulty), and there was no significant difference between scores [$F_{(13,\ 117)} = 1.004$, $p = 0.452$]. A further test showed that the pre-tasks had no effect on the later responses [$F_{(5,\ 928)} = 0.294$, $p = 0.917$], and neither did the interaction effect of the pre-tasks and Condition [$F_{(2,\ 928)} = 0.15, p = 0.861$]. After the pre-task, the participant and the matched person could receive a reward of 10 RMB for their completion of the task, and then the participant was asked to consider the scheme on how to divide this reward between himself and the matched person.

In the *determined* condition, the participant had to make a two-alternative forced choice between an always equitable offer (5, 5) and an always inequitable offer, which was either advantageous or disadvantageous (Allocation Stage and Alternative Choice). Participants' preference was counted as their choice rate for each offer. Contrary to the *determined* condition, in the *predetermined* condition, a random program-generated offer from one of the five Allocations (Allocation Stage) was first presented, and then participants were asked to make a two-alternative forced judgement of whether they were satisfied with the offer (Alternative Judgment). Thus, participants' preference was counted as their satisfaction rate for each offer. Importantly, the revealed preference theory assumed that the preference of a decision-maker could be revealed by his actual decision-making, suggesting that one would choose the thing that satisfies him most (Samuelson, 1938). From this perspective, the choice

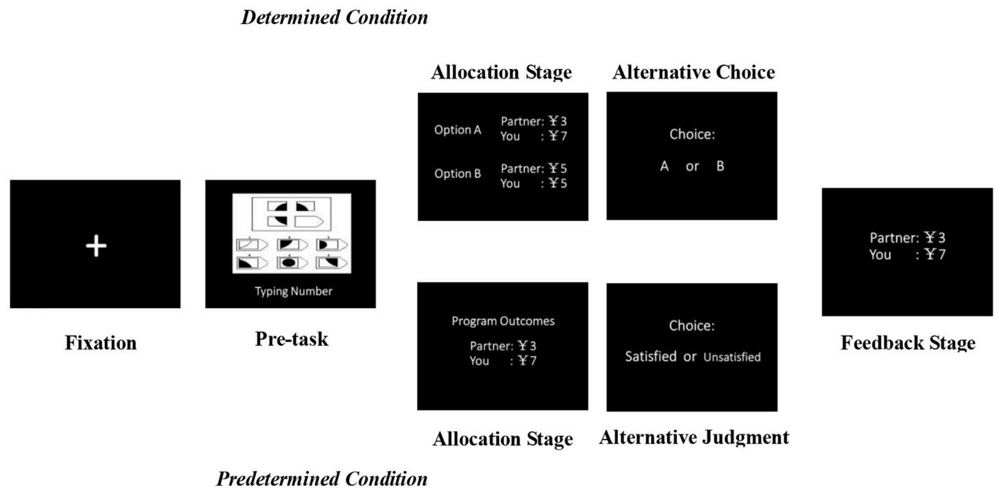

FIGURE 1 | Experiment One procedure. Participants were randomly assigned to either the *determined* or the *predetermined* condition. One participant and a matched person attended the experiment, and all of them first completed a pre-task picked from Raven Matrices. Then, the participant was asked to consider the scheme on how to divide a reward of 10 RMB. In the *determined* condition, he had to make a two-alternative forced choice to divide the reward. In the *predetermined* condition, he had to make an alternative satisfaction judgment to a program-generated offer. Afterwards, payoffs for both the participant and the match in that round were presented at the Feedback Stage.

that people makes is whatever they are satisfied with, thus establishing a connection between the alternative choice of the *determined* condition and the alternative judgment of the *predetermined* condition. Afterwards, the Feedback Stage lasted until participants pressed the Enter key at the end of each trial.

Participants' final income was related to the actual outcome of their decision-making in each trial, which was paid in the ratio of 12:1 (each participant gained 5.75 ± 0.73 RMB on average, added to a show-up fee 5 RMB). This allowed us to simulate the real-life situation in which individuals are remunerated for their work, providing the participant with a meaningful basis for comparing their own and the matched partner's incomes.

The design offer and the filler task offer were repeated twice, preceded by a practice session, and the presentation order of all trials was randomized by the program. Stimuli, recording triggers, and responses were presented adopting E-Prime 1.0 software package (Psychology Software Tools, Pittsburgh, PA, USA).

Results

Statistical analysis was conducted in SPSS 23.0. The Greenhouse-Geisser correction for violation of the assumption of sphericity was applied when necessary. The Bonferroni correction was used for pairwise comparisons. We excluded (4, 6) and (6, 4) from the analysis because they were filler tasks. Nevertheless, we still take them into consideration in an additional test; for results see Appendix A.

Preferences for each Allocation across Conditions are presented in **Table 1**. Gender had no effect on the preference [$F_{(1,\,114)} = 0.092$, $p = 0.762$], and neither did the interaction effect of gender and Condition [$F_{(1,\,114)} = 1.460$, $p = 0.229$]. Similarly, there were no effects of the order of experiments [$F_{(3,\,110)} = 0.213$, $p = 0.887$] or of the interaction between the order and Condition [$F_{(3,\,110)} = 0.200$, $p = 0.896$].

The repeated measure ANOVA for preferences yielded main effects of both Condition [$F_{(1,\,116)} = 140.04$, $p < 0.001$, $\eta^2 = 0.547$] and Allocation [$F_{(2,\,273)} = 147.18$, $p < 0.001$, $\eta^2 = 0.559$], and further yielded an interaction effect of the two factors [$F_{(1,\,116)} = 39.22$, $p < 0.001$, $\eta^2 = 0.253$]. The results of the simple effect analysis for Condition showed that participants were more satisfied with all inequitable offers in the *predetermined* condition than in the *determined* condition [(8, 2): $p < 0.001$; (7, 3): $p < 0.001$; (3, 7): $p < 0.01$; (2, 8): $p < 0.001$, respectively]. However, preferences for the equitable offer (5, 5) were not significantly different ($p = 0.119$). More importantly, the results of the simple effect analysis for Allocation indicated that this factor had significant effects in both the *determined* [$F_{(4,\,464)} = 92.26$, $p < 0.01$] and the *predetermined* [$F_{(4,\,464)} = 94.13$, $p < 0.001$] conditions. Because of this, we would examine them respectively below.

In the *determined* condition, as shown in **Figure 2A**, pairwise comparisons showed that participants were more willing to choose (5, 5) compared to inequitable offers, regardless of whether the inequitable offers were advantageous or disadvantageous. For inequitable offers, only one pair reached statistical significance, with (7, 3) having a higher response than (2, 8) ($p < 0.05$). In addition, no other pairs had a significant difference. However, in the *predetermined* condition, the case was totally different. As illustrated in **Figure 2B**, pairwise comparisons showed that the satisfaction judgment rates for advantageous offers (8, 2) and (7, 3) were not different from the rates for the equitable offer (5, 5), while the rates for both were significantly higher than disadvantageous offers (3, 7) and (2, 8).

Discussion

The results of Experiment One supported Hypothesis 1. More specifically, participants resisted receiving more than others only when they could determine allocations, which was consistent with previous studies (Fehr and Schmidt, 1999, 2006). However, once they simply passed judgment on *predetermined* allocations, they became satisfied to find that they had a higher payoff than others, which implied that their tendency of AI might disappear. It is noted that the overall preferences for inequitable offers were higher in the *predetermined* condition (vs. *determined* condition), which suggested a facilitating effect of making participants become more accepting of inequitable outcomes with the change of the task feature from *determined* to *predetermined* (see **Table 1**). By conducting a *t*-test for this between (dis)advantages, we found that its influence on advantageous inequity was nearly four times as great as that on disadvantageous inequity [72.46% vs. 19.49%, $t_{(234)} = 8.953$, $p < 0.001$, Cohen's $d = 1.426$], suggesting that participants would sharply turn from resisting advantages to seeking advantages as long as their control in the allocating was removed.

TABLE 1 | The descriptive data of the preference for each Allocation in the *determined* and *predetermined* conditions.

Inequity types	Allocations	Preference in the *determined* condition (%) (*N* = 59)	Preference in the *predetermined* condition (%) (N = 59)	Facilitating Effects (%)
Advantageous	(8,2)	13.56 ± 31.94	88.14 ± 31.26	72.6
	(7,3)	18.64 ± 38.17	88.98 ± 29.46	
Equitable	(5,5)	88.56 ± 19.32	94.07 ± 18.77	5.51
Disadvantageous	(3,7)	7.62 ± 20.94	25.42 ± 39.80	19.49
	(2,8)	5.93 ± 20.92	27.11 ± 39.74	

Facilitating effects are the variations of preferences with the Condition changing from determined to predetermined, which was calculated as the positive change between conditions.

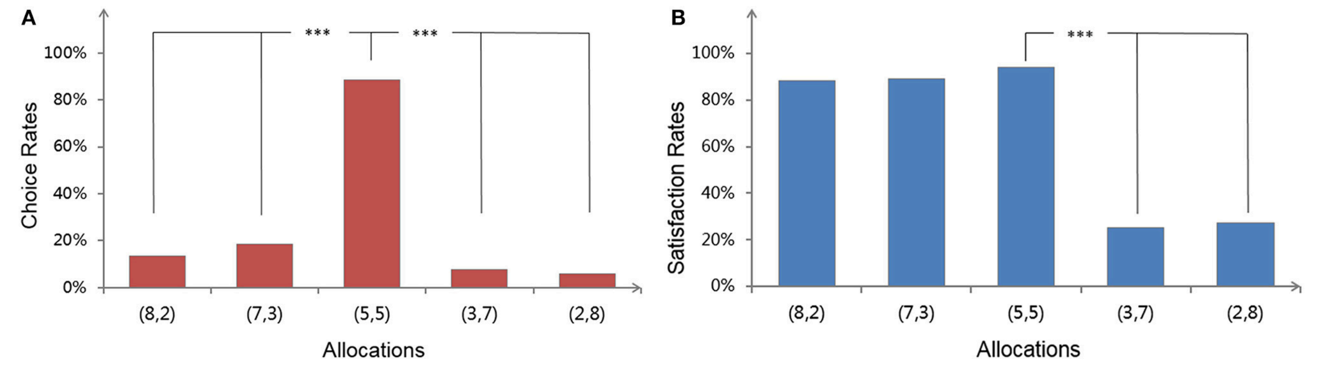

FIGURE 2 | Effects of Allocation on preferences in the *determined* (left, **A**) and *predetermined* condition (right, **B**). Significant differences ($p < 0.001$) between Allocations are marked with ***.

EXPERIMENT TWO

Materials and Methods

Participants

In total, 31 healthy right-handed college students (10 males, 21 females), whose major was neither psychology nor economics, were recruited in this experiment; 27 of them had never joined a psychological experiment before. The ages of the participants ranged from 18 to 23 years (on average, 19.52 ± 0.65 years). All of the participants were included in behavior analysis, while three participants were excluded from SCRs analysis because they were outliers according to the boxplot. The final income a participant could earn equaled the sum of a show-up fee of 15 RMB and a performance-based fee, which was similar to that of Experiment One, although the paid ratio increased to 8:1. This experiment was approved by the research ethics board of Central China Normal University. Informed consent was obtained from all participants before the experiment.

Study Design

Similar to Experiment One, a 2 (Condition: determined vs. predetermined) × 7 [Allocation: (5, 1) vs. (5, 2) vs. (5, 3) vs. (5, 5) vs. (5, 7) vs. (5, 8) vs. (5, 9)] mixed design was adopt, with the Condition referring to a between-subject factor and the Allocation referring to a within-subject factor. Thus, participants were randomly assigned to either the *determined* or the *predetermined* condition. Importantly, in a prior test, we found that SCRs were susceptible to the absolute payoff that one received in an allocation. Thus, the design was changed from amount-constant to payoff-constant, whereby participants' payoffs were kept constant at five units in each trial. Correspondingly, (5, 1), (5, 2), and (5, 3) were assigned to advantageous inequity, (5, 5) was designated to equity, and (5, 7), (5, 8), and (5, 9) were assigned to disadvantageous inequity. To control the confounding effect of the variation of amount, we did not present allocations in the form of showing payoffs for two players directly, such as "You: 7 RMB, Partner: 3 RMB." Instead, we told participants the positive or negative difference between their payoff and their partner's payoff, such as "You get 5 RMB, Your partner gets 2 RMB less (more) than you." The outcome factors were preferences and SCRs elicited by different Allocations.

Experimental Procedure

Participants completed experimental tasks with a matched person in a quiet laboratory. In this experiment, we told the participants that our research focused on their numerical ability, preventing them from conjecturing our real purposes. Therefore, the pre-task in this experiment was replaced by a correlation judgment task, in which participants needed to evaluate the correlation coefficient from a scatter plot. Before the experiments, five psychology graduates were recruited to design the correlation judgement task, which ensured that the chosen pre-task was simple enough and could not impact the following tasks. According to the check beforehand, all pre-tasks ($n = 16$) were easy to solve (on average, 1.875 ± 0.815 points on a 7-point Likert scale of difficulty), and there was no significant difference between the scores [$F_{(15,\ 135)} = 1.153$, $p = 0.316$]. A further test showed that the pre-tasks had no effect on the following responses [behavior data: [$F_{(8,\ 435)} = 0.305$, $p = 0.964$], SCRs data: ($F_{(8,\ 405)} = 0.374$, $p = 0.934$)], and neither did the interaction effect of the pre-tasks and Condition [behavior data: [$F_{(8,\ 928)} = 0.319$, $p = 0.959$], SCRs data: ($F_{(8,\ 405)} = 0.336$, $p = 0.952$)]. Indeed, the matched person in this experiment was an experimental confederate, who was a female graduate student and a stranger to all of the participants. After the real participant came to the laboratory, followed by the confederate, he/she and the confederate were told that they would sit in front of two computers face to face and perform a task together through the computer network. The real participant was ostensibly selected by lot to the position required for the experiment. According to the survey after the experiment, no participants doubted this manipulation, and on average, they rated 8.656 ± 0.135 points on a 9-point Likert scale.

The main procedure was similar to Experiment One. All tasks were conducted by computer. As illustrated in **Figure 3**, at the beginning of each trial, a fixation appeared as a cue for 3,000 ms on the black screen, followed by the pre-task. After completing the pre-task, the participant and the confederate entered the main session to divide a reward. In the *determined* condition,

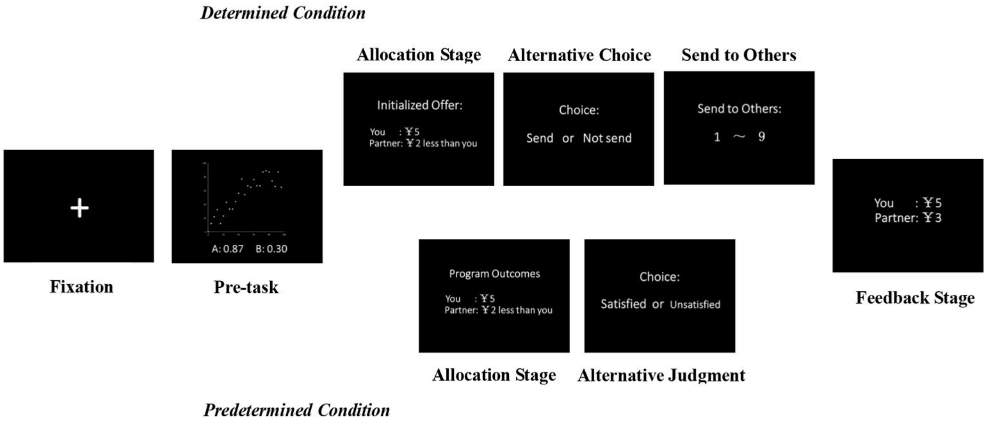

FIGURE 3 | Experiment Two procedure. Participants were randomly assigned to either the *determined* or the *predetermined* condition. A participant and a matched person attended the experiment, and both of them first completed a pre-task (correction judgment task). Then, the participant was asked to consider the scheme on how to divide a reward. In the *determined* condition, he had to make a two-alternative forced choice of whether to send a presented offer to the match. If he approved, then the final distribution was assigned as the offer; otherwise, he could determine the unit he wanted to send to the match. In the *predetermined* condition, participants had to make an alternative satisfaction judgment to a program-generated offer. Afterwards, payoffs for both the participant and the match in that round were presented at the Feedback Stage. SCRs were recorded from the participant throughout the experiment.

the participant had to make a two-alternative forced choice of whether to send (accept) an offer, which was presented by the program in the name of initialization and varied between seven Allocations to the matched person (Allocation Stage and Alternative Choice). If the participant accepted the presented offer, then each player received the payoff assigned by this offer. If he rejected sending the presented offer, he could determine any unit of money in the range of 1–9 to send to his match (Send to Others). The *predetermined* condition was similar to that in Experiment One; a program-generated offer, which was one of the seven Allocations, was presented first (Allocation Stage). Then, participants were asked to indicate whether they were satisfied with the offer (Alternative Judgment). In both conditions, the Allocation Stage lasted for 5000 ms. Afterwards, the feedback Stage lasted for 3,000 ms before the end of each trial.

All of the inequitable offers were repeated twice, and equitable offers were repeated three times. The presentation order of all trials was counterbalanced. Before the formal session, participants joined a practice session. Stimuli, recording triggers, and responses were presented adopting E-Prime 1.0 software package (Psychology Software Tools, Pittsburgh, PA, USA).

Skin Conductance Recording

While the participants were involved in the task, SCRs were continuously recorded using a BIOPAC MP150 system (Biopac Systems Inc., Goleta, CA) acquiring data at 1,000 samples per second in another computer. SCRs were recorded using two grounded Ag-AgCl electrodes (BIOPAC TSD203 transducer) that were secured medially on the distal ring and index finger of the non-dominant hand, with BIOPAC SCR paste (with a NaCl concentration of 0.05 m) as the electrolyte. Values of SCRs were baseline corrected and transformed to microsiemens (μS) values using AcqKnowledge 4.3 software. SCR amplitudes were quantified as the maximum positive change between 1 and 5 s after the start of the Allocation Stage, excluding data that did not exceed a threshold of 0.02 μs (van't Wout et al., 2006; Benedek and Kaernbach, 2010)[3]. Before each Allocation Stage, we also set a time of 6,000 ms to buffer the SCRs and made the values regress to baseline. To normalize the data, a square transformation was used, as Dunn et al. (2012) suggested.

Results

Statistical analysis was conducted in SPSS 23.0. The Greenhouse-Geisser correction for violation of the assumption of sphericity was applied when necessary. The Bonferroni correction was used for pairwise comparisons of behavior results, while Fisher's least significant difference (LSD) test was used for SCR results[4]. Gender had no effect on behavior data [$F_{(1,\,27)} = 1.214$, $p = 0.280$] or on SCR data [$F_{(1,\,25)} = 0.124$, $p = 0.728$].

Behavior Results

In the *determined* condition, the overall acceptance rates for advantageous, equitable and disadvantageous offers were 64.44% (± 38.25%), 91.11% (± 19.79%), and 58.89% (± 40.76%),

[3] Participants displayed SCRs to 12.28 out of 15 trials on average (mean nonresponses = 2.72 ± 2.90). To examine whether the proportion of nonresponses varied as a function of Allocation, two Cochran's Q tests were conducted, showing that nonresponses were evenly distributed (for the *determined* condition, $\chi^2 = 1.714$, $p = 0.424$; for the *predetermined* condition, $\chi^2 = 5.750$, $p = 0.452$). A further M-W U test showed that nonresponse rates were relatively comparable across Conditions (Z = 1.680, $p = 0.092$).

[4] We did not use Bonferroni correction to correct the *p*-value of multiple comparisons for SCRs data because it made almost all of the comparisons non-salient. The within-factor of Allocation had 7 levels, which required carrying out 21 comparisons to complete the *post-hoc* test. This would make the corrected significance level too low (i.e., 0.05/21 = 0.00238) to be detected and then increase the likelihood of making a β-error. Actually, the data difference between experimental treatments in the SCR tasks was not as obvious as that in the behavior tasks; thus, the actual effect might be covered under such a low significance level if we applied this correction.

respectively. A repeated measure ANOVA revealed a significant effect of Allocation [$F_{(2,\,28)} = 7.164$, $p < 0.01$, $\eta^2 = 0.338$], with equitable offers receiving more favorable responses than both advantageous ($p < 0.01$) and disadvantageous ($p < 0.01$) offers. However, the latter two had no differences between each other ($p = 0.591$). In the *predetermined* condition, the overall satisfaction rates for advantageous, equitable and disadvantageous offers were 81.25% (± 8.72%), 93.75% (± 3.36%), and 55.21% (± 11.15%), respectively. There was also a significant effect of Allocation [$F_{(2,\,30)} = 8.48$, $p < 0.001$, $\eta^2 = 0.361$], with preferences for advantageous offers being the same as those for equitable offers ($p = 0.158$), but the preferences for both of them were significantly higher than those for disadvantageous offers ($p = 0.020$ and $p = 0.002$, respectively), which was in accordance with the finding of Experiment One.

Skin Conductance Response Results

SCRs elicited by each Allocation across Conditions are presented in **Table 2**. The repeated measure ANOVA for SCRs yielded a main effect of Allocation [$F_{(4,\,109)} = 6.482, p < 0.001, \eta^2 = 0.194$] but did not find a main effect of Condition [$F_{(1,\,27)} = 1.813$, $p = 0.189$]. However, the interaction effect of Allocation and Condition was significant [$F_{(4,\,109)} = 2.744, p < 0.05, \eta^2 = 0.092$]. We adopted a simple effect analysis for Condition and found that only (5, 5) and (5, 7) produced different SCRs between Conditions [$F_{(1,\,27)} = 7.48, p < 0.05$ and $F_{(1,\,27)} = 7.71, p < 0.05$, respectively], with both SCRs being greater in the *predetermined* condition (vs. *determined* condition). However, SCRs elicited by the other offers had no significant differences. More importantly, we further adopted a simple effect analysis for Allocation, which showed that the effect of Allocation on SCRs was significant in both the *determined* [$F_{(6,\,162)} = 3.53$, $p < 0.01$] and the *predetermined* conditions [$F_{(6,\,162)} = 5.77$, $p < 0.001$]. In the following, thus, we examine this effect in the two Conditions separately.

In the *determined* condition, as shown in **Figure 4A**, pairwise comparisons showed that advantageous offer (5, 2) and disadvantageous offers (5, 8) and (5, 9) elicited a greater SCR than did the equitable offer (5, 5), and the SCR elicited by another advantageous offer (5, 1) was marginally greater than that elicited by the equitable one ($p = 0.070$). Although there is not a statistically significant difference, the actual SCRs of (5, 3) and (5, 7) were still higher than those of (5, 5). In addition, we also found that the SCRs to (5, 9) were higher than those to (5, 3) ($p < 0.05$), and SCRs to (5, 2) were higher than those to (5, 8) ($p < 0.05$), while the rest of the comparisons were not significant.

In contrast, in the *predetermined* condition, SCRs elicited by advantageous offers (5, 1), (5, 2), and (5, 3) had no differences from those elicited by (5, 5) (see **Figure 4B**). Although the difference was not significant, the actual SCRs of these advantageous offers were all lower than (5, 5). On the other side, SCRs elicited by disadvantageous offers (5, 9) and (5, 8) were significantly higher than those elicited by (5, 5), while those for the offer (5, 7) were marginally higher ($p = 0.087$). We also found that the SCRs for all of the disadvantageous offers were significantly higher than those for the advantageous offers, except for one comparison between (5, 8) and (5, 1).

Discussion

Experiment Two replicated the outcome of Experiment One in the behavior analysis and further supported Hypothesis 2 by electrophysiological data. In the *determined* condition, the SCRs to both advantageous and disadvantageous inequity were higher than those to equity, whereas in the *predetermined* condition, the SCRs elicited by advantageous inequity had no significant differences from those elicited by equity. Since SCRs can be used as an electrophysiological indicator to evaluate the feeling of inequity/unfairness (van't Wout et al., 2006; Civai et al., 2010; Hewig et al., 2011; Dunn et al., 2012), we can infer that (1) if participants could determine allocations, they felt negatively toward the two types of inequity, which was consistent with previous studies, and (2) if participants passively received a program-generated allocation, however, they did not feel negatively toward receiving more than others and even felt more satisfied. Consequently, the individual's tendency toward AI may be modulated by their role in determining allocations and may disappear if they have no chance to determine the allocation.

GENERAL DISCUSSION

We began this paper with the hypothesis that individuals' preferences for advantageous inequity might differ as a function of their role in determining allocations. Participants showed a

TABLE 2 | The descriptive data of SCRs elicited by each Allocation in the *determined* and *predetermined* conditions.

Inequity type	Allocations	SCRs in the *determined* condition (μS) (*N* = 14)	SCRs in the *predetermined* condition (μS) (*N* = 15)
	(5,1)	0.2433 ± 0.9748	0.2556 ± 0.1530
Advantageous	(5,2)	0.3029 ± 0.1175	0.2425 ± 0.0681
	(5,3)	0.2266 ± 0.0713	0.2361 ± 0.0987
Equitable	(5,5)	0.2017 ± 0.0539	0.2687 ± 0.0753
	(5,7)	0.2197 ± 0.0793	0.3372 ± 0.1385
Disadvantageous	(5,8)	0.2971 ± 0.1342	0.3513 ± 0.1332
	(5,9)	0.3146 ± 0.1348	0.3710 ± 0.1404

Values of SCR were baseline corrected and transformed to microsiemens (μS) values. To normalize the data, the square transformation was used, as Dunn et al. (2012) suggested.

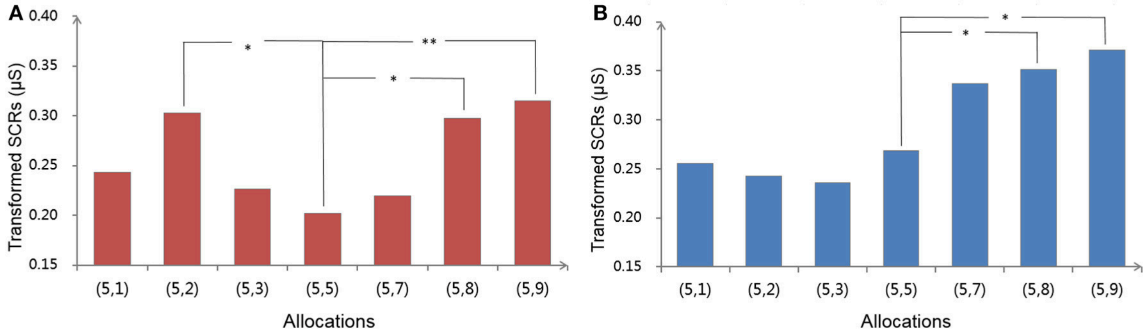

FIGURE 4 | Effects of Allocation on SCRs in the *determined* (left, **A**) and *predetermined* conditions (right, **B**). Significant differences ($p < 0.05$, $p < 0.01$) between Allocations are marked with *, ** respectively.

far lower preference for equitable offers than for advantageous offers if they could *determine* allocations in the money distribution setting. However, when participants simply passed judgment on *predetermined* allocations, their preferences for advantageous offers were as high as those for equitable offers (Experiment One). We replicated this pattern of results in a further electrophysiological experiment. The SCR, an indicator of the feeling of inequity/unfairness, elicited by advantageous offers had no difference from that elicited by equitable offers in the *predetermined* condition, providing evidence that individuals did not feel negatively toward advantages in this situation (Experiment Two). Taken together, the present studies provided mutual corroboration from behavioral and electrophysiological data to document the dramatic impact of the *determined*/*predetermined* feature on AI and further noted that AI would disappear if the distribution paradigm was merely based on the *predetermined* feature.

It should be noted that, in the existing literature, it was not always true that individuals resisted receiving more than others. Some studies, which did not take the paradigm feature of *determined*/*predetermined* as their center, have found that participants appeared to prefer to receive more than others, rather than the opposite. For example, Moser et al. (2014) showed that in their ultimatum game individuals who played the role of responders (therefore, they could not decide how to divide an offer) were more likely to accept advantageous offers compared to equitable offers (acceptance rates: 97.9 vs. 91.4%, respectively, $p < 0.05$). The same pattern was also observed on the side of responders' rejection behavior. In a study conducted by Lamichhane et al. (2014), participants' rejection rates of advantageous offers (80–100% of the amount to participants) were as low as their rejection rates of equitable offers (40–60% of the amount to participants) (rejection rates: 6.5 vs. 11.3%, respectively, $p > 0.05$). Similar patterns can also be observed in Wu et al.'s (2012) and Albrecht et al.'s (2013) studies. Obviously, these studies were in conflict with the previous studies on AI, which claimed that people would resist receiving more than others (Fehr and Schmidt, 1999, 2006). We suggest that a clue for understanding this conflicting result can be found by investigating the paradigm feature of *determined*/*predetermined*, which is still unclear. Unlike those in previous studies, the allocations in these studies were not proposed by participants, as was the case of the *predetermined* condition in the present study. Therefore, it may have been due to the fact that their research paradigms were mainly based on the *predetermined* feature, then they failed to reveal the tendency of AI and were in conflict with previous studies. Although these studies did not take the effect of the *determined*/*predetermined* feature as their focus, they provide additional evidence to support our hypothesis that AI would diminish or even disappear if advantageous inequity is *predetermined*.

One of the potential causes for individuals' different behaviors between the *determined* and *predetermined* conditions may be the sense of agency, referring to the subjective experience of controlling one's own actions, and through these actions, controlling external events (Gallagher, 2000). Interestingly, the notion of agency in the cognitive literature refers mostly to a person's control over the outcomes of his actions (Caspar et al., 2016), which is innately related to the paradigm feature of *determined*/*predetermined*. According to Choshen-Hillel and Yaniv (2011, 2012), there is a causal relationship between the sense of agency and one's concern with others' well-being. More specifically, in settings in which people have a high agency, their concern with others' welfare is prominent, whereas in settings in which people have a low agency, their concern with self-interest (in their studies, this means avoiding receiving less than others) figures prominently. In addition, this effect even exists in children from 3 to 4 years old, with children given a sense of agency becoming happier to share more with a new individual (Chernyak and Kushnir, 2013). It is implied that individuals who have a higher agency could derive some internal rewards from being kind to others, and the gained positive utilities, in turn, could partly offset the cost of the benevolence (Choshen-Hillel and Yaniv, 2011, 2012). In our terms, the *determined* condition mostly refers to a high-agency condition, and the *predetermined* condition mostly refers to a low-agency condition. Due to the fact that the sense of agency could make people's focus change from self-interest to the other's welfare, the result that those who were in the *determined* condition were more likely to keep offers equitable than those who were in the *predetermined* condition is to be expected. We suggest that the sense of agency may serve as an approach motivation to push people in the *determined* condition to behave as the theory of inequity aversion expects. Conversely, it also provides an explanation for why AI would diminish or even disappear when people are in the *predetermined* condition. On the other hand, the responsibility for negative consequences may also play a role. In our experiments, the linkage between one's actions and outcomes was stronger in the *determined* condition than in the *predetermined* condition. The allocations of the *determined* condition were totally decided by participants, and because of this, participants needed to take responsibility for the final distribution outcomes. Nevertheless, the outcomes of the *predetermined* condition were not due to participants' actions, and thus, they were free to be responsible for the final outcomes. To date, a body of studies have demonstrated that, as the linkage between actions and outcomes becomes stronger, decision-makers would not only show greater prosocial preferences toward others (Hamman et al., 2010; Bartling and Fischbacher, 2011) but also be more in compliance with social norms (Andreoni and Gee, 2012; Kamei et al., 2014). That is, because being responsible means being blameworthy for potentially negative outcomes and the prospect of blame for immoral behaviors (e.g., selfish or greedy) would make people avoid doing the things inconsistent with social expectations (Bartling and Fischbacher, 2011). Since seeking advantageous inequity is commonly viewed as a behavior that is inconsistent with social expectations (Spitzer et al., 2007; Fershtman et al., 2012), participants in the *determined* condition would avoid showing this behavior and conversely choice to be equitable with others. Because of this, responsibility may work as an avoidance motivation to pull people in the *determined* condition to acquire advantageous inequity. In contrast, people in the *predetermined* condition may be free from blame for receiving more than others because they need not to take negative actions in the allocating. As a result, they would feel less negatively toward advantageous inequity and thus be willing to accept it.

Taken together, maybe both the approach motivation of concerning with others (sense of agency) and the avoidance motivation of avoiding blame (responsibility) work together to inhibit people's preferences for advantageous inequity in the *determined* condition. However, the *predetermined* condition is the routine case without the sense of agency and the responsibility. Individuals in this condition may be free to receive advantageous inequity. Further testing is needed to tease apart these possible interpretations of the participants' behavior.

Theoretically, there is a more in-depth discussion referring to the question of whether AI is an authentic behavioral tendency in human beings. Notice that, in the present study, the tendency of AI occurs only in the situation where participants can determine allocations, while it disappears if they cannot. If AI, according to the definition of inequity aversion, is focused on avoiding receiving more than others, why would it be observed in one condition (*determined*) and not in the other (*predetermined*)? Maybe the reactions to advantageous inequity have different psychological mechanisms between the *determined* and *predetermined* conditions. Alternatively, maybe the tendency of AI does not have a solid foundation, but other motivations are preventing individuals from showing satisfaction for advantageous inequity in the *determined* condition. This question needs to be examined in future works.

To our knowledge, the present study might be the first to investigate correlates of the impact of individuals' role in determining allocations on their preferences for advantageous inequity and to prove that if individuals cannot determine allocations, their tendency of AI would disappear. The present study has at least three contributions to the current research. First, our questions directly concern the structural integrity of the theory of inequity aversion (especially on advantageous inequity) and help extend the current research from the *determined* domain to the *predetermined* domain. If we make the paradigm feature (*determined* vs. *predetermined*) and types of inequity aversion (AI vs. DI) intersect with each other, we can get four connections: *determined*—AI, *predetermined*—AI, *determined*—DI, and *predetermined*—DI. The first and the fourth connections are used as the general method to investigate AI and DI by the current researches (Sanfey et al., 2003; Fehr and Schmidt, 2006; Fehr et al., 2008), and the third connection is commonly regarded as a form of the altruism preference (Batson and Powell, 2003), leaving the second connection still unclear. As shown, the present study sheds light on this gap, suggesting that the tendency of AI is weak in this connection. Second, the present study reveals that AI has a boundary condition: only those who are in the *determined* condition would show AI, whereas those who are in the *predetermined* condition would not. This further implies that the current literature related to AI might be biased because almost all of the studies are based only on the *determined* feature, ignoring the situation of the *predetermined* feature. Thus, caution should be used in generalizing to other situations the conclusion that individuals resist advantageous inequity. Further studies, however, are required to explore the mechanism behind this boundary effect. Third, the present finding is important methodologically because it may help reconcile why most of the past studies found a robust tendency for humans to resist advantageous inequity (Fehr and Schmidt, 1999, 2006), whereas another group of studies mentioned above instead demonstrated that individuals are happy to receive advantageous inequity (Wu et al., 2012; Albrecht et al., 2013; Lamichhane et al., 2014; Moser et al., 2014). We suggest that differences concerning the *determined*/*predetermined* feature cause these ostensible conflicts. Thus, the existent conflicting results can be unified into a common theoretical framework by the present study.

One limitation of the present study is that we carried out comparisons between choice (as the dependent variable of the *determined* condition) and satisfaction (as the dependent variable of the *predetermined* condition). As mentioned above, the revealed preference theory implied that one would choose the thing that satisfies him most (Samuelson, 1938). In this case, to a certain extent, the choice that people makes is whatever they are content with, thus establishing a connection between choice and satisfaction. Actually, in two studies conducted by Choshen-Hillel and Yaniv (2011, 2012), conditions were very similar to that of the present study, and they also carried out direct comparison between choice and satisfaction. Similarly, in the field of equity aversion, researchers often make direct comparisons between and conduct subsequent discussions on AI and DI (noting that AI is based on choice and DI is based on satisfaction) (Güth and Kocher, 2014; Tricomi and Sullivan-Toole, 2015; Xu et al., 2016). Although there are reasonable arguments for comparing choice and satisfaction, it does indeed have its limits. Therefore, further works should develop a better research paradigm to overcome the problem of direct comparisons.

CONCLUSION

The current study demonstrated that individual's tendency of AI might differ as a function of their role in determining allocations. Both behavioral and electrophysiological data showed that, in the situation in which participants could determine allocations, they seemed to dislike advantageous inequity, which is consistent with the prediction of the theory of inequity aversion. However, in the situation in which participants could not determine allocations, they appeared to prefer advantageous inequity. This finding suggests the possibility that the tendency of AI may have a different mechanism between the two situations, or more strictly, it does not have a solid foundation, and the preference for advantageous inequity that would exist in the *determined* condition may have been prevented by other factors.

ETHICS STATEMENT

This study was carried out in accordance with the requirements of the Ethics Committee at Central China Normal University. All subjects gave a written informed consent according to the Declaration of Helsinki. The protocol was approved by the Ethics Committee at Central China Normal University.

AUTHOR CONTRIBUTIONS

OL had made contributions to the conception, design, analysis, acquisition and interpretation of the data. OL also wrote the first manuscript. LW and FX revised the manuscript and gave excellent advice. OL, LW, and FX finally approved the version to be submitted.

FUNDING

This work was supported by the National Natural Science Foundation of China to LW (Grant Number: 71471163), and FX (Grant Number: 71571083).

REFERENCES

Adams, J. S. (1965). Inequity in social exchange. *Adv. Exp. Soc. Psychol.* 2, 267–299. doi: 10.1016/S0065-2601(08)60108-2

Albrecht, K., von Essen, E., Fliessbach, K., and Falk, A. (2013). The influence of status on satisfaction with relative rewards. *Front. Psychol.* 4:804. doi: 10.3389/fpsyg.2013.00804

Andreoni, J., and Gee, L. K. (2012). Gun for hire: delegated enforcement and peer punishment in public goods provision. *J. Public Econ.* 96, 1036–1046. doi: 10.1016/j.jpubeco.2012.08.003

Bartling, B., and Fischbacher, U. (2011). Shifting the blame: on delegation and responsibility. *Rev. Econ. Stud.* 79, 67–87. doi: 10.1093/restud/rdr023

Batson, C. D., and Powell, A. A. (2003). "Altruism and prosocial behavior," in *Handbook of Psychology, Vol. 5: Personality and Social Psychology*, eds T. Millon and M. J. Lerner (Hoboken, NJ: Wiley), 463–484.

Benedek, M., and Kaernbach, C. (2010). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology* 47, 647–658. doi: 10.1111/j.1469-8986.2009.00972.x

Blake, P. R., and McAuliffe, K. (2011). "I had so much it didn't seem fair": eight-year-olds reject two forms of inequity. *Cognition* 120, 215–224. doi: 10.1016/j.cognition.2011.04.006

Bolton, G. E., and Ockenfels, A. (2000). ERC: a theory of equity, reciprocity, and competition. *Am. Econ. Rev.* 90, 166–193. doi: 10.1257/aer.90.1.166

Bosman, R., Sonnemans, J., and Zeelenberg, M. (2001). *Emotions, Rejections, and Cooling off in the Ultimatum Game. Working Paper*. Amsterdam: University of Amsterdam.

Camerer, C. (2003). *Behavioral Game Theory*. New Jersey, NJ: Princeton University Press.

Caspar, E. A., Christensen, J. F., Cleeremans, A., and Haggard, P. (2016). Coercion changes the sense of agency in the human brain. *Curr. Biol.* 26, 585–592. doi: 10.1016/j.cub.2015.12.067

Chernyak, N., and Kushnir, T. (2013). Giving preschoolers choice increases sharing behavior. *Psychol. Sci.* 24, 1971–1979. doi: 10.1177/0956797613482335

Choshen-Hillel, S., and Yaniv, I. (2011). Agency and the construction of social preference: between inequality aversion and prosocial behavior. *J. Pers. Soc. Psychol.* 101, 1253–1261. doi: 10.1037/a0024557

Choshen-Hillel, S., and Yaniv, I. (2012). Social preferences shaped by conflicting motives: when enhancing social welfare creates unfavorable comparisons for the self. *Judgm. Decis. Mak.* 7, 618–627. doi: 10.13140/2.1.1134.6561

Civai, C., Corradi-Dell'Acqua, C., Gamer, M., and Rumiati, R. I. (2010). Are irrational reactions to unfairness truly emotionally-driven? Dissociated behavioural and emotional responses in the Ultimatum Game task. *Cognition* 114, 89–95. doi: 10.1016/j.cognition.2009.09.001

Dunn, B. D., Evans, D., Makarova, D., White, J., and Clark, L. (2012). Gut feelings and the reaction to perceived inequity: the interplay between bodily responses, regulation, and perception shapes the rejection of unfair offers on the ultimatum game. *Cogn. Affect. Behav. Neurosci.* 12, 419–429. doi: 10.3758/s13415-012-0092-z

Eckel, C. C., and Grossman, P. J. (2001). Chivalry and solidarity in ultimatum games. *Econ. Inq.* 39, 171–188. doi: 10.1111/j.1465-7295.2001.tb00059.x

Falk, A., Fehr, E., and Fischbacher, U. (2003). Reasons for conflict: lessons from bargaining experiments. *J. Inst. Theor. Econ.* 159, 171–187. doi: 10.1628/0932456032974925

Fehr, E., and Schmidt, K. M. (1999). A theory of fairness, competition and cooperation. *Q. J. Econ.* 114, 817–868. doi: 10.1162/003355399556151

Fehr, E., Bernhard, H., and Rockenbach, B. (2008). Egalitarianism in young children. *Nature* 454, 1079–1083. doi: 10.1038/nature07155

Fehr, E., and Schmidt, K. M. (2006). "The economics of fairness, reciprocity and altruism: experimental evidence and new theories," in *Handbook of the Economics of Giving, Altruism and Reciprocity*, eds S.-C. Kolm and J. M. Ythier (Amsterdam: Elsevier), 615–691.

Fershtman, C., Gneezy, U., and List, J. A. (2012). Equity aversion: social norms and the desire to be ahead. *Am. Econ. J. Microecon.* 4, 131–144. doi: 10.1257/mic.4.4.131

Fowles, D. C. (1980). The three arousal model: implications of gray's two-factor learning theory for heart rate, electrodermal activity, and psychopathy. *Psychophysiology* 17, 87–104. doi: 10.1111/j.1469-8986.1980.tb00117.x

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends Cogn. Sci.* 4, 14–21. doi: 10.1016/s1364-6613(99)01417-5

Gray, J. A. (1994). "Personality dimensions and emotion systems," in *The Nature of Emotion: Fundamental Questions*, eds P. Ekman and R. Davidson (New York, NY: Oxford University Press), 329–331.

Güroglu, B., van den Bos, W., van Dijk, E., Rombouts, S. A., and Crone, E. A. (2011). Dissociable brain networks involved in development of fairness considerations: understanding intentionality behind unfairness. *Neuroimage* 57, 634–641. doi: 10.1016/j.neuroimage.2011.04.032

Güth, W., and Kocher, M. G. (2014). More than thirty years of ultimatum bargaining experiments: motives, variations, and a survey of the recent literature. *J. Ecno. Behav. Organ.* 108, 396–409. doi: 10.1016/j.jebo.2014.06.006

Güth, W., Schmittberger, R., and Schwarz, B. (1982). An experimental analysis of ultimatum bargaining. *J. Ecno. Behav. Organ.* 3, 367–388. doi: 10.1016/0167-2681(82)90011-7

Hamman, J. R., Loewenstein, G., and Weber, R. A. (2010). Self-interest through delegation: an additional rationale for the principal-agent relationship. *Am. Econ. Rev.* 100, 1826–1846. doi: 10.1257/aer.100.4.1826

Harlé, K. M., Chang, L. J., Van, W. M., and Sanfey, A. G. (2012). The neural mechanisms of affect infusion in social economic decision-making: a mediating role of the anterior insula. *Neuroimage* 61, 32–40. doi: 10.1016/j.neuroimage.2012.02.027

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., et al. (2001). In search of homo economicus: behavioral experiments in 15 small-scale societies. *Am. Econ. Rev.* 91, 73–78. doi: 10.1257/aer.91.2.73

Hewig, J., Kretschmer, N., Trippe, R. H., Hecht, H., Coles, M. G. H., Holroyd, C. B., et al. (2011). Why humans deviate from rational choice. *Psychophysiology* 48, 507–514. doi: 10.1111/j.1469-8986.2010.01081.x

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1986). Fairness and the assumptions of economics. *J. Bus.* 59, S285–S300. doi: 10.1086/296367

Kahneman, D., and Tversky, A. (1982). The psychology of preferences. *Sci. Am.* 246, 160–173. doi: 10.1038/scientificamerican0182-160

Kamei, K., Putterman, L., and Tyran, J.-R. (2014). State or nature? Endogenous formal versus informal sanctions in the voluntary provision of public goods. *Exp. Econ.* 18, 38–65. doi: 10.1007/s10683-014-9405-0

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., and Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832. doi: 10.1126/science.1129156

Lamichhane, B., Adhikari, B. M., Brosnan, S. F., and Dhamala, M. (2014). The neural basis of perceived unfairness in economic exchanges. *Brain Connect.* 4, 619–630. doi: 10.1089/brain.2014.0243

Leotti, L. A., and Delgado, M. R. (2011). The inherent reward of choice. *Psychol. Sci.* 22, 1310–1318. doi: 10.1177/0956797611417005

Leotti, L. A., Iyengar, S. S., and Ochsner, K. N. (2010). Born to choose: the origins and value of the need for control. *Trends Cogn. Sci.* 14, 457–463. doi: 10.1016/j.tics.2010.08.001

Moser, A., Gaertig, C., and Ruz, M. (2014). Social information and personal interests modulate neural activity during economic decision-making. *Front. Hum. Neurosci.* 8:31. doi: 10.3389/fnhum.2014.00031

Oosterbeek, H., Sloof, R., and van de Kuilen, G. (2004). Cultural differences in ultimatum game experiments: evidence from a meta-analysis. *Exp. Econ.* 7, 171–188. doi: 10.2139/ssrn.286428

Pillutla, M. M., and Murnighan, J. K. (1996). Unfairness, anger, and spite: emotional rejections of ultimatum offers. *Organ. Behav. Hum. Decis. Process.* 68, 208–224. doi: 10.1006/obhd.1996.0100

Ritov, I., and Baron, J. (1990). Reluctance to vaccinate: omission bias and ambiguity. *J. Behav. Decis. Making.* 3, 263–277. doi: 10.1002/bdm.3960030404

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Econimica* 5, 61–71. doi: 10.2307/2548836

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science* 300, 1755–1758. doi: 10.1126/science.1082976

Spitzer, M., Fischbacher, U., Herrnberger, B., Gron, G., and Fehr, E. (2007). The neural signature of social norm compliance. *Neuron* 56, 185–196. doi: 10.1016/j.neuron.2007.09.011

Tricomi, E., Rangel, A., Camerer, C. F., and O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature* 463, 1089–1091. doi: 10.1038/nature08785

Tricomi, E., and Sullivan-Toole, H. (2015). "Fairness and inequity aversion," in *Brain Mapping: An Encyclopedic Reference, Vol. 3*, ed A. W. Toga (New York, NY: Elsevier), 3–8.

van't Wout, M., Kahn, R. S., Sanfey, A. G., and Aleman, A. (2006). Affective state and decision-making in the Ultimatum Game. *Exp. Brain Res.* 169, 564–568. doi: 10.1007/s00221-006-0346-5

Wu, Y., Zhang, D., Elieson, B., and Zhou, X. (2012). Brain potentials in outcome evaluation: when social comparison takes effect. *Int. J. Psychophysiol.* 85, 145–152. doi: 10.1016/j.ijpsycho.2012.06.004

Xu, F., Li, O., Deng, Y., Liu, C., and Shi, Y. (2016). The inequity aversion in behavioral economics. *Adv. Psychol. Sci.* 24:1613. doi: 10.3724/sp.j.1042.2016.01613

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor declared a shared affiliation, though no other collaboration, with the authors OL and LW.

APPENDIX A

The Analyses Incorporated Filler Tasks of (6, 4) and (4, 6)

Statistics analysis was conducted in SPSS 23.0. The Greenhouse-Geisser correction for violation of the assumption of sphericity was applied when necessary. The Bonferroni correction was used for pairwise comparisons.

The preference for (6, 4) and (4, 6) were 15.25 and 8.47%, respectively. After incorporated (6, 4) and (4, 6), the main effects of Condition [$F_{(1,\,116)} = 757.5$, $p < 0.001$, $\eta^2 = 0.867$] and Allocation [$F_{(3,\,381)} = 110.2$, $p < 0.001$, $\eta^2 = 0.487$] were still significant, the same as the interaction effect of the two factors [$F_{(3,\,381)} = 35.7$, $p < 0.001$, $\eta^2 = 0.235$]. The results of simple effect analyses for Allocation showed that this factor still had a significant effect in both the *determined* [$F_{(6,\,696)} = 68.62$, $p < 0.001$] and *predetermined* [$F_{(6,\,696)} = 77.29$, $p < 0.001$] conditions. Considering to (6, 4) and (4, 6), although they were the lower magnitude (dis)advantageous inequity, participants were still less likely to choose them compared to (5, 5) ($p < 0.001$ for both) in the *determined* condition. In contrast, in the *predetermined* condition, the preference for (6, 4) was not different from (5, 5) ($p > 0.05$), while both of them were significantly higher than (4, 6) ($p < 0.001$ for both).

From these results, we can find that statistical trend for (6, 4) and (4, 6) were the same as that for other advantageous and disadvantageous offers, respectively, in the Experiment One. Even if we incorporated (6, 4) and (4, 6) into analysis, the results remained unchanged.

ORIGINAL RESEARCH
published: 20 February 2018
doi: 10.3389/fpsyg.2018.00182

Social Support Modulates Neural Responses to Unfairness in the Ultimatum Game

Chunli Wei[1], *Li Zheng*[1,2,3,4]*, *Liping Che*[5], *Xuemei Cheng*[6], *Lin Li*[1,4]* and *Xiuyan Guo*[1,2,3,4]

[1] School of Psychology and Cognitive Science, East China Normal University, Shanghai, China, [2] Shanghai Key Laboratory of Magnetic Resonance, East China Normal University, Shanghai, China, [3] Shanghai Key Laboratory of Brain Functional Genomics, Ministry of Education, East China Normal University, Shanghai, China, [4] National Demonstration Center for Experimental Psychology Education, East China Normal University, Shanghai, China, [5] Business School, University of Shanghai for Science and Technology, Shanghai, China, [6] Department of Mechanical and Electrical Engineering, Beijing Polytechnic College, Beijing, China

The current functional MRI study aimed to investigate how responders' fairness considerations and related decision-making processes were affected by social support in the ultimatum game (UG). During scanning, responders either played the standard UG with proposers (control condition) or played the modified UG in which three unknown observers showed social support for responders by acknowledging proposers' norm violation. Results revealed that participants reported higher unfairness feelings and rejection rates of unfair offers in the social support condition relative to the control condition. At the neural level, compared to the control condition, perception of social support from others induced greater activations of anterior cingulate gyrus and right anterior insula when receiving unfair (vs. fair) offers. The medial prefrontal cortex and right anterior insula were more active when the unfair offers were rejected (vs. accepted) in the social support condition than the control condition. These results highlighted the modulation effect of social support on responders' fairness considerations and related decision-making processes.

Keywords: unfairness, ultimatum game (UG), social support, decision-making, fMRI

OPEN ACCESS

Edited by:
Qi Li,
University of Chinese Academy of Sciences (UCAS), China

Reviewed by:
Jamie Ward,
University of Sussex, United Kingdom
Yiping Zhong,
Hunan Normal University, China

****Correspondence:***
Li Zheng
zhengli@psy.ecnu.edu.cn
Lin Li
lilin@psy.ecnu.edu.cn

Specialty section:
This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Psychology

Received: *01 October 2017*
Accepted: *02 February 2018*
Published: *20 February 2018*

Citation:
Wei C, Zheng L, Che L, Cheng X, Li L and Guo X (2018) Social Support Modulates Neural Responses to Unfairness in the Ultimatum Game. Front. Psychol. 9:182. doi: 10.3389/fpsyg.2018.00182

INTRODUCTION

Fairness-related decision-making has attracted much attention in the past decades and been widely studied by employing the Ultimatum Game (UG) (Güth et al., 1982; Camerer and Thaler, 1995; Sanfey et al., 2003; Civai et al., 2012; Guo et al., 2014; Hu et al., 2016). This game was developed by Güth et al. (1982), in which two players have to divide a sum of money according to the simple rule. One player proposes how to split and the other player responds (i.e., the proposer and the responder). The responder can either accept or reject the proposal. If the proposal is accepted, both players get the amount specified in the proposal. If the proposal is rejected, none of them receives any money. It has been documented in previous studies that responders accepted all fair offers, but often rejected extremely unfair offers (Güth et al., 1982; Camerer and Thaler, 1995). This results appeared in contradiction to the standard economic models, which idealized individuals as completely rational cognitive agents aiming to maximize their own payoff and assumed that the responder should accept any offer as long as it is larger than zero. The reason why people make such irrational decisions has been attributed to the negative emotion caused by perception of unfairness, people's preference for fairness and tendency to maintain fairness norms (Bolton and Rami, 1995; Nowak et al., 2000; Sanfey et al., 2003; Yamagishi et al., 2009).

Over the past few years, a large body of neuroimaging studies have investigated the neural basis underlying the fairness-related decision-making processes and identified the engagement of several brain regions, including anterior insula (AI), anterior cingulate cortex (ACC), amygdala and prefrontal cortex (Sanfey et al., 2003; Haruno and Frith, 2010; Güroğlu et al., 2011; Civai et al., 2012; Corradi-Dell'Acqua et al., 2013). It has been proved that the activations of AI and ACC observed during receiving and rejecting unfair offers are associated with detecting and responding to fairness norm violations (Sanfey et al., 2003; Chang and Sanfey, 2013; Corradi-Dell'Acqua et al., 2013; Xiang et al., 2013; Guo et al., 2014). Amygdala has been found playing a key role in emotional processing (Scott et al., 1997; Rauch et al., 2003; Feinstein et al., 2011), and its activation in UG was suggested to be related to inequity aversion (Haruno and Frith, 2010). As for the prefrontal cortex, previous studies have observed the activation of the dorsal lateral prefrontal cortex (DLPFC) and medial prefrontal cortex (mPFC) during fairness-related decision processes (Sanfey et al., 2003; Baumgartner et al., 2011; Civai et al., 2012; Corradi-Dell'Acqua et al., 2013; Cheng et al., 2015). The activation of DLPFC was interpreted to be engaged in the integration of information and the selection of context-appropriate decisions to unfairness (Buckholtz et al., 2008; Buckholtz and Marois, 2012; Cheng et al., 2015). The mPFC has been thought to be involved in monitoring one's behavioral responses in social decision-making (Civai et al., 2012, 2015; Corradi-Dell'Acqua et al., 2013).

As a kind of complex social interactions, responders' fairness-related decision-making processes were not only determined by the proposal he or she received, but also influenced by various social contexts, such as the social distance between proposers and responders (Wu et al., 2011), the framing of distribution (Zhou and Wu, 2011; Guo et al., 2013), self-contribution to the income (Guo et al., 2014), proposers' economic status (Zheng Y. et al., 2017) and so on. The present study will investigate whether one of these contextual factors, social support, modulates peoples' fairness-related decision making, behaviorally and neurally. Social support refers to the mental and material resources which people obtained from the social network, including sympathy, caring, actions, advice, information (Cobb, 1976; Thoits, 1986). In UG, the responders were at relative disadvantage positions, hence resulted in negative emotional feelings in them (Sanfey et al., 2003; Wout et al., 2006). Social support has been identified as having a critical impact on people's psychological state and behaviors when they are under negative emotional states (Cohen and Wills, 1985; Sarason et al., 1997). It has been found that social support can help people cope with stress situations, cease smoking and alcohol consumption (Cohen and Wills, 1985; Steptoe et al., 1996; Burns et al., 2014), and has beneficial effects on one's well-being, physical and psychological health (Turner, 1981; Uchino et al., 2016). However, little researchers have discussed its impact on fairness-related decision-making behaviors, less for the underlying neural mechanisms.

To explore the modulation effect of social support on responders' fairness-related decision-making process and the underlying neural mechanisms, we designed the current functional magnetic resonance imaging (fMRI) study. During the experiment, the participants carried out both the standard version of UG (control condition) and a modified version of UG (social support condition) as responders in the scanner. According to the typical laboratory manipulation of social support, which employ support providers who deliver emotional support to participants by verbal comments, such as expressions of blaming the norm violators (Cutrona and Russell, 1990; Thorsteinsson and James, 1999; Cohen et al., 2000), in the present study, there are support providers delivering social support for the participants by acknowledging the proposer's fairness norm violations and having themselves at the participants' back.

Based on the emerging evidence, there might be two different hypothesizes on how social support would modulate responders' fairness-related decision-making processes. On the one hand, researchers have argued that responders' rejection is driven by the negative emotions evoked by unfair treatment (Sanfey et al., 2003; Wout et al., 2006; Yamagishi et al., 2009). Social support has been demonstrated as being able to alleviate people's negative emotions effectively (Cohen and Wills, 1985; Sarason et al., 1997), thus the responder's unfairness-related negative emotional feelings might decrease when receiving others' social support. Decreased rejection rates and amygdala activation might also be observed. We called it as the "negative emotion buffer" hypothesis. On the other hand, some prior studies have pointed out that people were easily infected by other's attitudes (Prislin and Wood, 2005; Huang et al., 2014). The social support supplied to the responder by verbal comments implied that the support providers confirmed the proposer's violations to fairness norm. In this case, the responder might be influenced by attitudes of supporters and be more sensitive to the violations of fairness norms, showing increased unfairness-related negative emotional feelings and rejection rates to unfair offers under the social support condition, accompanied with increased activations of AI and ACC. This was called as the "norm violation confirmation" hypothesis.

MATERIALS AND METHODS

Participants

Twenty-eight right-handed volunteers [15 females, mean age = 22.46 (years), *SD* = 2.62 (years)] from the university community participated in this experiment. None of the participants had an abnormal neurological history. All of them had normal or corrected-to-normal vision. Three participants were excluded from further statistical analyses. One participant was excluded due to a technical problem during scanning and the other two had severe head movements (>3 mm or 3°) (Cheng et al., 2015; Nebel et al., 2016). Written informed consent was acquired from all participants before scanning. This study was approved by the Ethics Committee on Human Experiments of East China Normal University.

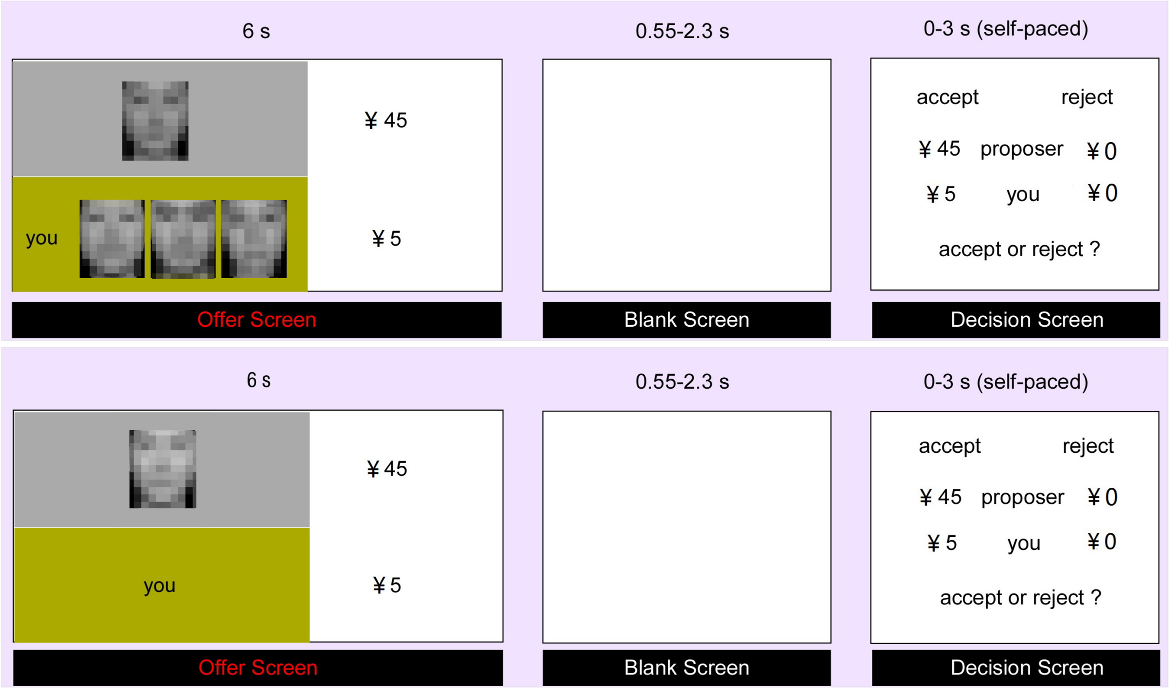

FIGURE 1 | Experimental Procedure. The participant firstly received the offer from the proposer (social support context in the upper part and control context in the lower part). After a jittered blank lasting for 0.55–2.3 s, the participant was asked to decide to accept or reject the offer within 3 s. During the experiment, all the face pictures presented to the participant were clear without mosaic.

Procedure

Before scanning, participants were told the rules of the game and that they would receive proposals about how to divide 50 RMB from 72 different proposers whose proposals were collected before the experiment. In half of the trials, participants acted as the responder and played the standard UG with the proposer. For the standard UG, the proposer gave her/his division schema about a sum of money and the responder decided to accept or reject it (control condition). While in the other half trials, participants were supported by three unknown observers when playing the UG with the proposer (social support condition). In the social support condition, participants were told that the observers acknowledged the proposer's fairness norm violations and had themselves at their back. As for the payment, participants were informed that several trials would be randomly selected and that both they and the proposers would be paid according to their decisions. Finally, participants would be paid with the amount of money obtained from a random selection of 5% trials in the game plus a 50RMB (approximately equal to 32 dollars) bonus. In fact, the proposals were manipulated by the experimenter and there were no real proposers or supporters. One hundred and eighty female or male neutral face pictures were randomly selected from the Chinese Affective Face Picture System (Gong et al., 2011) were used as the proposers and the supporters in different contexts.

Then, the participants completed 72 trials in the scanner. There were 36 trials in each context, including 12 fair trials (25:25) and 24 unfair trials. The unfair trials contained four types of proposals, i.e., 30:20, 35:15, 40:10 and 45:5, with each type having 6 trials. All the trials were presented randomly and functional images were acquired simultaneously. Each trial began with the presentation of the proposer's offer, which lasted for 6 s. At the same time, context information about whether participants had supporters or not would also be presented. Then a blank screen jittered from 0.55 ∼ 2.3 s was presented. After that, participants were required to decide (accept or reject) within 3 s. Each trial was jittered with inter-stimulus intervals (approximately 3 ∼ 8 s), during which a black fixation cross was presented (**Figure 1**). After scanning, the same stimuli as inside the scanner were presented again. Participants were asked to rate the extent of unfairness-related negative emotional feelings they felt for each offer (i.e., unfairness ratings) in a 9-point Likert-type scale (1 indicated extremely unfair and 9 indicated extremely fair).

fMRI Image Acquisition and Data Analyze

The scanning was carried out on a 3T Siemens scanner at the Shanghai Key Laboratory of Magnetic Resonance of East China Normal University. Anatomical images were acquired using a T1-weighted, multiplanar reconstruction sequence (MPR) (TR = 1900 ms, TE = 3.42 ms, 192 slices, slice thickness = 1 mm, FOV = 256 mm, matrix size = 256 * 256). After that, functional images were acquired using a gradient echo echo-planar imaging (EPI) sequence (TR = 2200 ms, TE = 30 ms, FOV = 220 mm, matrix size = 64 * 64, 35 slices, slice thickness = 3 mm, gap = 0.3 mm).

Participants' data were analyzed using the SPM8 software package (Wellcome Department of Imaging Neuroscience, London, United Kingdom). During data preprocessing, the first five volumes were discarded to allow for T1 equilibration effects. Then, the functional images were corrected for the delay in slice acquisition and were realigned to the first image to correct for interscan head movements. The individual structural image was co-registered to the mean EPI image generated after realignment. The co-registered structural image was then segmented into gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) using a unified segmentation algorithm (Ashburner and Friston, 2005). The functional images after slice timing correction and realignment procedures were spatially normalized to the Montreal Neurological Institute (MNI) space (resampled at 2 mm × 2 mm × 2 mm voxels) using the normalization parameters estimated during unified segmentation and then spatially smoothed with a Gaussian kernel of 8 mm full-width half-maximum (FWHM).

First-level analyses were then performed across the whole brain for each subject using two general linear models (GLM) implemented in SPM8. The fairness-related model was built to explore the impact of social pressure on unfairness-related neural responses, consisting of four types of events (Fair*ss*: fair offers in the social support condition, Unfair*ss*, unfair offers in the social support context; Fair*cc*: fair offers in the control condition, Unfair*cc*, unfair offers in the control condition). Events were convolved with a canonical hemodynamic response function (HRF). All the encoding trials were time-locked to the onset of the offers with null duration. Decision phase and trials with no response were also added into the model as additional covariates of no interest. Moreover, six realignment parameters and one overall mean during the whole phase were included in the design matrix as well. To filter the low-frequency noise, a cutoff of 128 s was applied. Contrast images for each type of event (Fair*ss*, Unfair*ss*, Fair*cc*, Unfair*cc*) were computed for each participant at the first-level analysis. At the second group level, these four first-level individual contrast images were fed into a 2 (Context: social support condition vs. control condition) × 2(Unfairness: Unfair vs. Fair) factorial design using a random-effects model (flexible factorial ANOVA in SPM8). The main effect of unfairness was defined using the (Unfair – Fair) and the reverse contrasts. The interaction between unfairness and social context was defined by the (Unfair*ss* – Fair*ss*) – (Unfair*cc* – Faircc) and the reverse contrasts. A cluster-level threshold of $p < 0.05$ (family wise error corrected) and a voxel-level threshold of $p < 0.001$ (uncorrected) were used to define activations.

To explore how the neural correlates underlying people's response to unfairness (rejection/acceptance) were modulated by social support, we built a response-related model in which unfair offers were further divided according to participants' responses (UAss, accepted unfair offers in the social support condition, UR*ss*, rejected unfair offers in the social support condition; UA*cc*, accepted unfair offers in the control condition, UR*cc*, rejected unfair offers in the control condition). The rest of the analyses were carried out in the same way as those in the first model. Contrast images for four types of event (UA*ss*, UR*ss,* UA*cc*, UR*cc*) were computed for each participant at the first-level analysis and then fed into a 2 (Context: social support vs. control) × 2 (Response: UA vs. UR) flexible factorial using a random-effects model (flexible factorial ANOVA in SPM8). The main effect of response to unfairness was defined using the (UR – UA) and the reverse contrasts. The interaction between response and social context was defined by the (UR*ss* – UA*ss*) – (UR*cc* – UA*cc*) and the reverse contrasts. A cluster-level threshold of $p < 0.05$ (family wise error corrected) and a voxel-level threshold of $p < 0.001$ (uncorrected) were used to define activations.

In addition, parametric analyses, an efficient statistical procedure to reveal voxels that shows a particular pattern of activation throughout several conditions (Büchel et al., 1998), was conducted at the first-level to assess how brain activities were modulated by unfairness. Specifically, unfairness ratings were used as the parametric regressor separately for two social contexts. The resulting subject-specific estimates of the parametric regressor at each voxel were then entered into a second-level one sample t-tests. A cluster-level threshold of $p < 0.05$ (family wise error corrected) and a voxel-level threshold of $p < 0.001$ (uncorrected) were used to define activations.

RESULTS

Behavior Results

The behavioral results were shown in **Table 1**. For rejection rates, participants accepted all the fair offers in both contexts. However, paired t-tests revealed higher rejection rates for unfair offers in the social support condition than those in the control condition [$t(27) = 8.25$, $p < 0.001$]. For unfairness ratings, a 2 (Fairness: Unfair vs. Fair) × 2 (Social context: social support vs. control) repeated-measure ANOVA revealed a significant main effects of fairness [$F(1,27) = 1848.50$, $p < 0.001$, $\eta_p^2 = 0.99$] and social context [$F(1,27) = 45.52$, $p < 0.001$, $\eta_p^2 = 0.63$], also a significant interaction [$F(1,27) = 29.54$, $p < 0.001$, $\eta_p^2 = 0.52$]. *Post hoc* analyses showed that unfairness ratings for unfair offers in the social support condition were lower relative to those in the control condition [$t(27) = 7.97$, $p < 0.001$], indicating participants' stronger unfairness feelings in the social support condition. The current results of rejection rates and unfairness ratings were contrary to the "negative emotion buffer" hypothesis, while in line with the "norm violation confirmation" hypothesis.

fMRI Results

Main Effects

The main effect of Unfairness was tested by the (Unfair – Fair) and the reverse contrasts. Results showed stronger activations

TABLE 1 | Mean (*SD*) for rejection rates (%) and unfairness ratings.

	Social support		Control	
	Fair	**Unfair**	**Fair**	**Unfair**
Rejection rates	0.00 (*0.00*)	77.00 (*0.08*)	0.00 (*0.00*)	62.00 (*0.10*)
Unfairness ratings	8.95 (*0.18*)	2.96 (*0.64*)	8.93 (*0.26*)	3.51 (*0.75*)

in right dACC, bilateral AI, left DLPFC, left supplementary motor area and left middle temporal gyrus during unfair compared to fair trials. No suprathreshold activation was detected in the reverse contrast. When contrasting trials in the social support condition with trials in the control condition, significant activations in right calcarine gyrus, right inferior frontal and left precentral were revealed. The reverse contrast revealed significant activations in left superior temporal gyrus, right superior temporal gyrus and left Cuneus. The main effect of response computed by the (UR – UA) contrast revealed significant activations in bilateral putamen, bilateral supramarginal gyrus and right supplementary motor area. The reverse contrast revealed no suprathreshod activations (**Table 2**).

Unfairness–Related Effects: Context × Unfairness Interaction

The interaction between context and unfairness computed by the (Unfair*ss* – Fair*ss*) – (Unfair*cc* – Fair*cc*) contrast showed stronger activations in right AI, dACC and pgACC. No significant activations were revealed in the reverse contrast. The activation of amygdala wasn't observed in these two contrasts even at the uncorrected threshold (**Table 3**). Beta values in different conditions were extracted from all the significant voxels in the 6 mm-radius spherical regions centered on AI (MNI 26 24 −10), dACC (MNI −4 32 28) and pgACC (MNI 12 40 0) (beta values were extracted in the same way throughout the paper). As shown in the **Figure 2**, the activations of AI and dACC were stronger in the social support condition than those in the control condition for unfair offers [AI (**Figure 2B**): $F(1,24) = 7.86$, $p < 0.05$, $\eta_p^2 = 0.25$; dACC (**Figure 2C**): $F(1,24) = 23.02$, $p < 0.001$, $\eta_p^2 = 0.49$], which was consistent with the "norm violation confirmation" hypothesis. The pgACC was more active for fair offers compared with unfair offers in the control condition [$F(1,24) = 10.81$, $p < 0.05$, $\eta_p^2 = 0.31$], while this pattern was almost reversed in the social support condition [$F(1,24) = 3.46$, $p > 0.05$, $\eta_p^2 = 0.13$] (**Figure 2D**).

TABLE 2 | Brain activities showing unfairness, context and response main effects.

Region	Side	Peak activation			*t*-value	Voxels
		X	*Y*	*Z*		
(Unfair – Fair)						
Supplementary motor area	L	−4	18	50	11.28	49083
Dorsal anterior cingulate cortex	R	8	26	36	9.88	
Insula lobe	L	−32	22	4	9.05	
Insula lobe	R	32	24	2	8.08	
Dorsolateral prefrontal cortex	L	−46	38	30	7.81	
Middle temporal gyrus	L	−46	2	−26	5.88	273
(Fair – Unfair)						
No regions						
(Social support – Control)						
Calcarine gyrus	R	16	−90	2	15.8	22274
Inferior frontal gyrus	R	42	10	34	8.58	3084
Precentral gyrus	L	−36	0	50	6.11	920
(Control – Social support)						
Superior temporal gyrus	L	−54	−32	12	5.06	802
Superior temporal gyrus	R	66	−10	4	4.51	736
Cuneus	L	−4	−90	24	5.36	316
(UR – UA)						
Putamen	L	−22	10	0	5.84	1082
Putamen	R	28	8	12	6.25	745
Supramarginal gyrus	R	60	−26	24	4.64	395
Supplementary motor area	R	14	−10	68	4.31	357
Supramarginal gyrus	L	−52	−36	28	4.48	268
(UA – UR)						
No regions						

Coordinates (mm) are in MNI space. L, left hemisphere; R, right hemisphere. Cluster–level, p < 0.05, family wise error corrected, voxel-level, p < 0.001, uncorrected.

Response–Related Effects: Context × Response Interaction

Significant activations in right mPFC and right AI were observed in the (UR*ss* – UA*ss*) –(UR*cc* – UA*cc*) contrast (**Table 3**). The reverse contrast revealed no significant activation. Further analyses on beta estimates revealed that right mPFC and right AI were more active during rejecting relative to accepting unfair offers in the social support condition [right mPFC (**Figure 3A**), $F(1,24) = 8.53$, $p < 0.05$, $\eta_p^2 = 0.26$; right AI (**Figure 3B**), $F(1,24) = 9.60$, $p < 0.05$, $\eta_p^2 = 0.29$], but not in the control condition ($ps > 0.05$). Actually, amygdala also showed stronger activations when unfair offers were rejected in the social support condition compared with the control condition, though the voxel size ($k = 107$) failed to survive the current corrected criterion.

Parametric Analyses on Unfairness Ratings

Parametric analyses on unfairness ratings revealed that left AI, right dACC (MNI 10 34 26) and left DLPFC (MNI −38 58 16) activations increased with the decrease of unfairness ratings in the social support condition and left AI (MNI −30 18 12) activation increased with the decreasing level of unfairness ratings in the control condition (**Table 4**). No suprathreshold activations were revealed with the increase of unfairness ratings.

DISCUSSION

The present study used a modified version of UG to explore how social support modulates responders' fairness-related decision-making processes and the underlying neural mechanisms. Behavioral results showed increased unfairness feelings and rejection rates for unfair offers in the social support condition compared to the control condition, suggesting that social support indeed impacted participants' fairness considerations and responses. These results helped to identify which is a more reasonable explanation between two possible hypotheses: the "negative emotion buffer" hypothesis that social support might buffer participants' negative emotional feelings elicited by unfairness and result in decreased rejection rates, or the "norm violation confirmation" hypothesis that social support might enhance participants' awareness of fairness

TABLE 3 | Brain activities showing context × unfairness interaction and context × response interaction.

Region	Side	Peak activation			*t*-value	Voxels
		X	*Y*	*Z*		
(Unfairss– Fairss) – (Unfaircc– Faircc)						
Pregenual anterior cingulate cortex	R	12	40	0	6.33	7631
Dorsal anterior cingulate cortex	L	−4	32	28	6.31	
Thalamus	R	4	−12	8	5.46	2460
Anterior insula	R	26	24	−10	4.31	
Rolandic operculum	R	56	−6	12	5.04	1397
Heschls gyrus	L	−38	−20	4	5.24	1047
Temporal pole	L	−54	12	−2	4.41	226
(Unfaircc – Faircc) –(Unfairss–Fairss)						
No regions						
(URss – UAss) – (URcc – UAcc)						
Inferior occipital gyrus	R	40	−72	−6	7.22	9966
Supplementary motor area	L	−8	0	62	6.54	4690
Medial prefrontal cortex	R	12	58	6	5.36	735
Precentral gyrus	R	30	−8	48	4.47	603
Inferior frontal gyrus	R	32	32	−8	4.91	302
Anterior insula	R	34	28	6	3.67	
(URcc – UAcc) – (URss – UAss)						
No regions						

Coordinates (mm) are in MNI space. L, left hemisphere; R, right hemisphere. Cluster–level, p < 0.05, family wise error corrected, voxel-level, p < 0.001, uncorrected.

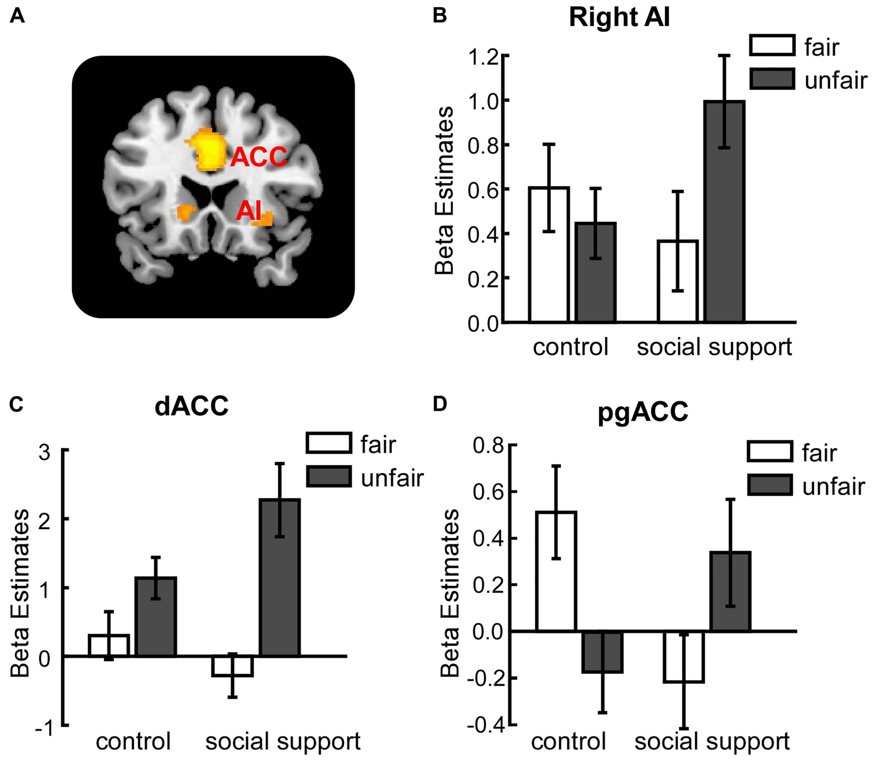

FIGURE 2 | Unfairness-related activations in right AI **(B)**, dACC **(C)** and pgACC **(D)** were modulated by social support. **(A)** The activation map. AI, anterior insula. dACC, dorsal anterior cingulate cortex. pgACC, pregenual anterior cingulate cortex. Error bars indicated 95% confidence intervals. Cluster level, $p < 0.05$, family wise error corrected; voxel level, $p < 0.001$, uncorrected.

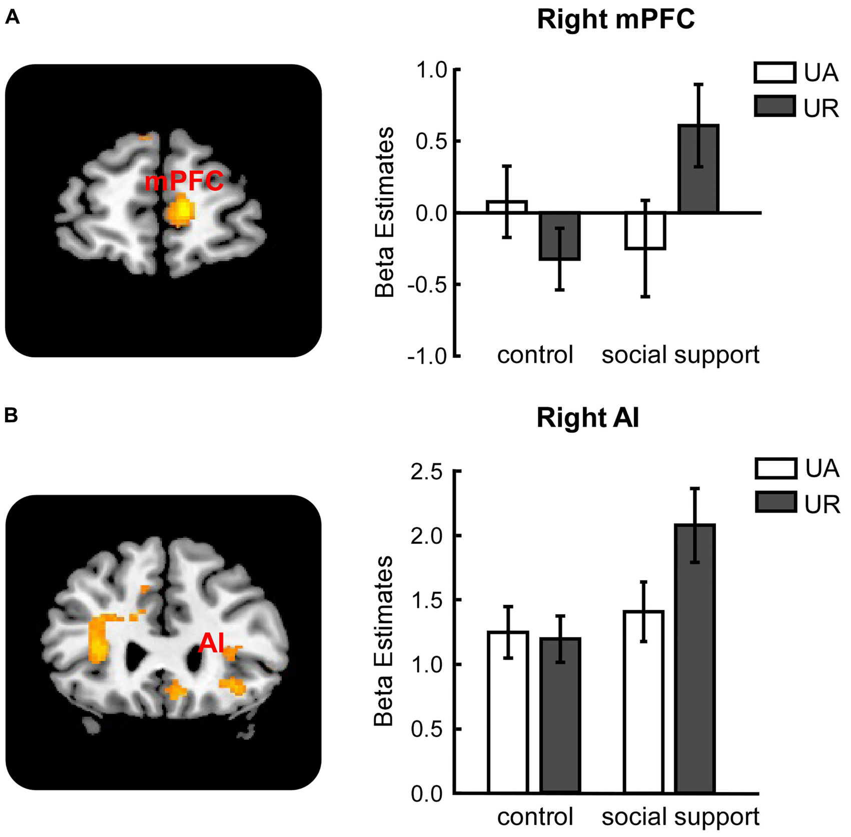

FIGURE 3 | Response-related activations in right mPFC **(A)**, right AI **(B)** were modulated by social support. mPFC, medial prefrontal cortex. AI, anterior insula. UA, accepted unfair offers. UR, rejected unfair offers. Error bars indicated 95% confidence intervals. Cluster level, $p < 0.05$, family wise error corrected; voxel level, $p < 0.001$, uncorrected.

TABLE 4 | Regions showing increased activations with the decrease of unfairness ratings in two conditions.

Region	Side	Peak activation			*t*-value	Voxels
		X	***Y***	***Z***		
Social support condition						
Anterior insula	L	−42	14	8	6.05	5446
Supplementary motor area	L	−6	20	48	5.59	2332
Anterior cingulate cortex	R	10	34	26	4.53	
Superior parietal lobe	R	16	−62	60	4.27	1871
Superior parietal lobe	L	−22	−64	52	4.69	1705
Precentral gyrus	L	−46	−2	56	4.93	681
Superior medial gyrus	L	−4	64	20	5.36	412
Dorsolateral prefrontal cortex	L	−38	58	16	5.51	383
Control condition						
Precuneus	L	−16	−62	32	4.68	787
Supplementary motor area	L	−8	10	58	5.23	474
Inferior frontal gyrus	L	−42	20	8	6.04	344
Anterior insula	L	−30	18	12	4.49	

Coordinates (mm) are in MNI space. L, left hemisphere; R, right hemisphere. Cluster-level, $p < 0.05$, family wise error corrected, voxel-level, $p < 0.001$, uncorrected.

norm violations hence resulting in increased rejection rates. The current results were in line with the "norm violation confirmation" hypothesis. Participants reported higher level of unfairness feelings when getting social support from others, indicating that they became more sensitive to fairness norm violations, the motivation for rejection was enhanced.

A recent UG study has found that self-affirmation can augment the responders' psychological resources and increase their rejection rates of unfair offers (Gu et al., 2016). Given the emerging evidence that people can gain enough psychological resources from social support to cope with problems (Wilcox, 1981; Cohen and Wills, 1985; Delongis et al., 1988), our findings demonstrated that others' support can act as powerful psychological resources and lead people a stronger a tendency to reject unfairness.

In fact, neither the behavioral results nor the neural results supported the "negative emotion buffer" hypothesis. This hypothesis would expect decreased activation in amygdala toward unfair offers being observed under the social support condition. However, decreased amygdala activation wasn't observed in the interaction between context and unfairness or the interaction between context and response. It was increased amygdala activation (although at an uncorrected threshold) that was found during rejecting unfair offers in the social support condition compared to the control condition. The overall neural results were consistent with the "norm violation confirmation" hypothesis. Stronger neural activations in right AI and left dACC in the social support condition compared to the control condition. Further parametric analysis revealed increased activations in left AI and right dACC with unfairness feelings under the social support condition. The activation of AI observed in UG studies have been considered to be associated with the detection of norm violations, supported by the evidence of its involvement in signaling deviations from people's expectations (Spitzer et al., 2007; Xiang et al., 2013; Zheng et al., 2015; Cheng et al., 2017; Vavra et al., 2017; Zinchenko and Arsalidou, 2017). The dACC was also suggested to be involved in detecting conflicts related to social expectation violations (Chang and Sanfey, 2013; Guo et al., 2014; Zheng et al., 2015; Vavra et al., 2017; Zinchenko and Arsalidou, 2017). The present behavioral results showed that unfair offers evoked participants' higher unfairness feelings, indicating their stronger perception of fairness norm violations. Taken together, these data suggested that the responders experienced higher level of unfairness and detected stronger norm violations in the social support condition, resulting in greater AI and ACC activations. The increased activation of pgACC was revealed during receiving unfair offers in the social support condition. Similar result has also been found in another study of our group which focused on the impact of social pressure (Zheng L. et al., 2017). The pgACC has been thought to be engaged in person perception and mentalizing (Amodio and Frith, 2006). Considering the similar activations of pgACC in these two studies, this region might not involve in the processing of specific social situation, but the common communicative intentions during general social situations, future work is needed to probe the exact function of pgACC during such social contexts.

Additionally, accompanied with the increased rejection rates of unfairness in the social support condition, significant activations of AI and mPFC were identified in the interaction between response and context. Both regions were more active during rejecting than accepting unfair offers in the social support condition compared to the control condition. Existing studies have proposed that the AI activation was linked to the rejection response to unfair offers (Sanfey et al., 2003; Tabibnia et al., 2008; Kirk et al., 2011). Our results also proved the critical role of AI in unfairness rejection, which might imply that the responders perceived a stronger norm violation signal when they made a rejection decision. This finding provided further evidence that social support let the responder be more concerned about fairness norm violations. Consistent with previous studies that mPFC was more activated during the rejection than acceptance of unfair offers, the activation of mPFC in the current study was considered to be engaged in monitoring individuals' behavioral responses (Civai et al., 2012; Corradi-Dell'Acqua et al., 2013).

To sum up, our findings provided both behavioral and neural evidence for the modulation of the responders' fairness-related decision-making processes by social support. Behaviorally, the responders reported higher level of unfairness feelings and rejection rates of unfair offers when they received social support from others, indicating that they were more sensitive to fairness norm violations. Neurally, with other's social support, increased activations were found in AI and dACC during processing unfairness, further implicating these two regions as being responsible for the detection of norm violations. The stronger activations in AI and mPFC were observed when rejecting unfair offers in the social support condition. In summary, the present study demonstrated that the fairness-related decision-making processes are context-dependent and are modulated by social support.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Ethical Committee of East China Normal University. Written informed consent was acquired from all subjects before the experiment. This study was approved by the Ethics Committee on Human Experiments of East China Normal University.

AUTHOR CONTRIBUTIONS

XG and LZ designed the experiments. LZ and XC programmed the experimental scenario and performed the experiments. CW, LZ, and XC analyzed the data. LL, CW, LZ, and LC joined in the interpretation of data. CW, LL, LZ, and LC carried out the writing. All authors read and approved the final version of the manuscript for submission.

FUNDING

This work was supported by the Key Program of National Social Science Foundation of China (Grant No. 14AZD106); the Outstanding Doctoral Dissertation Cultivation Plan of Action (Grant No. PY2015042), Shanghai Leading Talent Plan and Shanghai Education and Science Research: General Project (B11043).

REFERENCES

Amodio, D. M., and Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277. doi: 10.1038/nrn1884

Ashburner, J., and Friston, K. J. (2005). Unified segmentation. *Neuroimage* 26, 839–851. doi: 10.1016/j.neuroimage.2005.02.018

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., and Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nat. Neurosci.* 14, 1468–1474. doi: 10.1038/nn.2933

Bolton, G. E., and Rami, Z. (1995). Anonymity versus punishment in ultimatum bargaining. *Games Econ. Behav.* 10, 95–121. doi: 10.1006/game.1995.1026

Büchel, C., Holmes, A. P., Rees, G., and Friston, K. J. (1998). Characterizing stimulus–response functions using nonlinear regressors in parametric fMRI experiments. *Neuroimage* 8, 140–148. doi: 10.1006/nimg.1998.0351

Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., et al. (2008). The neural correlates of third-party punishment. *Neuron* 60, 930–940. doi: 10.1016/j.neuron.2008.10.016

Buckholtz, J. W., and Marois, R. (2012). The roots of modern justice: cognitive and neural foundations of social norms and their enforcement. *Nat. Neurosci.* 15, 655–661. doi: 10.1038/nn.3087

Burns, R. J., Rothman, A. J., Fu, S. S., Lindgren, B., and Joseph, A. M. (2014). The relation between social support and smoking cessation: revisiting an established measure to improve prediction. *Ann. Behav. Med.* 47, 369–375. doi: 10.1007/s12160-013-9558-7

Camerer, C., and Thaler, R. H. (1995). Anomalies: ultimatums, dictators and manners. *J. Econ. Perspect.* 9, 209–219. doi: 10.1257/jep.9.2.209

Chang, L. J., and Sanfey, A. G. (2013). Great expectations: neural computations underlying the use of social norms in decision-making. *Soc. Cogn. Affect. Neurosci.* 8, 277–284. doi: 10.1093/scan/nsr094

Cheng, X., Li, Z., Lin, L., Guo, X., Wang, Q., Lord, A., et al. (2015). Power to punish norm violations affects the neural processes of fairness-related decision making. *Front. Behav. Neurosci.* 9:344. doi: 10.3389/fnbeh.2015.00344

Cheng, X., Zheng, L., Li, L., Zheng, Y., Guo, X., and Yang, G. (2017). Anterior insula signals inequalities in a modified Ultimatum Game. *Neuroscience* 348, 126–134. doi: 10.1016/j.neuroscience.2017.02.023

Civai, C., Crescentini, C., Rustichini, A., and Rumiati, R. I. (2012). Equality versus self-interest in the brain: differential roles of anterior insula and medial prefrontal cortex. *Neuroimage* 62, 102–112. doi: 10.1016/j.neuroimage.2012.04.037

Civai, C., Miniussi, C., and Rumiati, R. I. (2015). Medial prefrontal cortex reacts to unfairness if this damages the self: a tDCS study. *Soc. Cogn. Affect. Neurosci.* 10, 1054–1060. doi: 10.1093/scan/nsu154

Cobb, S. (1976). Social support as a moderator of life stress. *Psychosom. Med.* 38, 300–314. doi: 10.1097/00006842-197609000-00003

Cohen, S., Underwood, L. G., and Gottlieb, B. H. (2000). *Social Support Measurement and Intervention*. New York, NY: Oxford University Press.

Cohen, S., and Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychol. Bull.* 98, 310–357. doi: 10.1037/0033-2909.98.2.310

Corradi-Dell'Acqua, C., Civai, C., Rumiati, R. I., and Fink, G. R. (2013). Disentangling self- and fairness-related neural mechanisms involved in the ultimatum game: an fMRI study. *Soc. Cogn. Affect. Neurosci.* 8, 424–431. doi: 10.1093/scan/nss014

Cutrona, C. E., and Russell, D. W. (1990). "Type of social support and specific (stress): toward a theory of optimal matching," in *Social Support: An Interactional View*, eds B. Sarason, I. G. Sarason, and G. Pierce (New York, NY: Wiley).

Delongis, A., Folkman, S., and Lazarus, R. S. (1988). The impact of daily stress on health and mood: psychological and social resources as mediators. *J. Pers. Soc. Psychol.* 54, 486–495. doi: 10.1037//0022-3514.54.3.486

Feinstein, J. S., Adolphs, R., Damasio, A., and Tranel, D. (2011). The human amygdala and the induction and experience of fear. *Curr. Biol.* 21, 34–38. doi: 10.1016/j.cub.2010.11.042

Gong, X., Huang, Y. X., Yan, W., and Luo, Y. J. (2011). Revision of the Chinese facial affective picture system. *Chin. Ment. Health J.* 25, 40–46.

Gu, R., Yang, J., Shi, Y., Luo, Y., Luo, Y. L., and Cai, H. (2016). Be strong enough to say no: self-affirmation increases rejection to unfair offers. *Front. Psychol.* 7:1824. doi: 10.3389/fpsyg.2016.01824

Guo, X., Zheng, L., Cheng, X., Chen, M., Zhu, L., Li, J., et al. (2014). Neural responses to unfairness and fairness depend on self-contribution to the income. *Soc. Cogn. Affect. Neurosci.* 9, 1498–1505. doi: 10.1093/scan/nst131

Guo, X., Zheng, L., Zhu, L., Li, J., Wang, Q., Dienes, Z., et al. (2013). Increased neural responses to unfairness in a loss context. *Neuroimage* 77, 246–253. doi: 10.1016/j.neuroimage.2013.03.048

Güroğlu, B., Bos, W. V. D., Dijk, E. V., Rombouts, S. A. R. B., and Crone, E. A. (2011). Dissociable brain networks involved in development of fairness considerations: understanding intentionality behind unfairness. *Neuroimage* 57, 634–641. doi: 10.1016/j.neuroimage.2011.04.032

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* 3, 367–388. doi: 10.1016/0167-2681(82)90011-7

Haruno, M., and Frith, C. D. (2010). Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nat. Neurosci.* 13, 160–161. doi: 10.1038/nn.2468

Hu, J., Blue, P. R., Yu, H., Gong, X., Xiang, Y., Jiang, C., et al. (2016). Social status modulates the neural response to unfairness. *Soc. Cogn. Affect. Neurosci.* 11, 1–10. doi: 10.1093/scan/nsv086

Huang, Y., Kendrick, K. M., and Yu, R. (2014). Conformity to the opinions of other people lasts for no more than 3 days. *Psychol. Sci.* 25, 1388–1393. doi: 10.1177/0956797614532104

Kirk, U., Downar, J., and Montague, P. R. (2011). Interoception drives increased rational decision-making in meditators playing the ultimatum game. *Front. Neurosci.* 5:49. doi: 10.3389/fnins.2011.00049

Nebel, M. B., Eloyan, A., Nettles, C. A., Sweeney, K. L., Ament, K., Ward, R. E., et al. (2016). Intrinsic visual-motor synchrony correlates with social deficits in autism. *Biol. Psychiatry* 79, 633–641. doi: 10.1016/j.biopsych.2015.08.029

Nowak, M. A., Page, K. M., and Sigmund, K. (2000). Fairness versus reason in the ultimatum game. *Science* 289, 1773–1775. doi: 10.1126/science.289.5485.1773

Prislin, R., and Wood, W. (2005). "Social influence in attitudes and attitude change," in *The Handbook of Attitudes*, eds D. Albarracín, B. T. Johnson, and M. P. Zanna (Mahwah, NJ: Lawrence Erlbaum Associates), 671–705.

Rauch, S. L., Shin, L. M., and Wright, C. I. (2003). Neuroimaging studies of amygdala function in anxiety disorders. *Ann. N. Y. Acad. Sci.* 985, 389–410. doi: 10.1111/j.1749-6632.2003.tb07096.x

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755–1758. doi: 10.1126/science.1082976

Sarason, B. R., Sarason, I. G., and Gurung, R. A. R. (1997). "Close personal relationships and health outcomes: a key to the role of social support," in *Handbook of Personal Relationships: Theory, Research, and Intervention*, ed. S. Duck (Chichester: Wiley).

Scott, S. K., Young, A. W., Calder, A. J., Hellawell, D. J., Aggleton, J. P., and Johnsons, M. (1997). Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature* 385, 254–257. doi: 10.1038/385254a0

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., and Fehr, E. (2007). The neural signature of social norm compliance. *Neuron* 56, 185–196. doi: 10.1016/j.neuron.2007.09.011

Steptoe, A., Wardle, J., Pollard, T. M., Canaan, L., and Davies, G. J. (1996). Stress, social support and health-related behavior: a study of smoking, alcohol consumption and physical exercise. *J. Psychosom. Res.* 41, 171–180. doi: 10.1016/j.neuron.2007.09.011

Tabibnia, G., Satpute, A. B., and Lieberman, M. D. (2008). The sunny side of fairness. *Psychol. Sci.* 19, 339–347. doi: 10.1111/j.1467-9280.2008.02091

Thoits, P. A. (1986). Social support as coping assistance. *J. Consult. Clin. Psychol.* 54, 416–423. doi: 10.1037/0022-006X.54.4.416

Thorsteinsson, E. B., and James, J. E. (1999). A Meta-analysis of the effects of experimental manipulations of social support during laboratory stress. *Psychol. Health* 14, 869–886. doi: 10.1080/08870449908407353

Turner, R. J. (1981). Social support as a contingency in psychological well-being. *J. Health Soc. Behav.* 22, 357–367. doi: 10.2307/2136677

Uchino, B. N., Bowen, K., and Kent, R. (2016). "Social support and mental health," in *Encyclopedia of Mental Health*, 2nd Edn, eds H. Friedman and K. Fingerman (Oxford: Elsevier), 189–195. doi: 10.1016/B978-0-12-397045-9.00117-8

Vavra, P., Baar, J. V., and Sanfey, A. (2017). "The neural basis of fairness," in *Interdisciplinary Perspectives on Fairness, Equity, and Justice*, eds M. Li and D. P. Tracer (Cham: Springer International Publishing), 9–31.

Wilcox, B. L. (1981). Social support, life stress, and psychological adjustment: a test of the buffering hypothesis. *Am. J. Community Psychol.* 9, 371–386. doi: 10.1007/BF00918169

Wout, M. V. T., Kahn, R. S., Sanfey, A. G., and Aleman, A. (2006). Affective state and decision-making in the Ultimatum Game. *Exp. Brain Res.* 169, 564–568. doi: 10.1007/s00221-006-0346-5

Wu, Y., Leliveld, M. C., and Zhou, X. (2011). Social distance modulates recipient's fairness consideration in the dictator game: an ERP study. *Biol. Psychol.* 88, 253–262. doi: 10.1016/j.biopsycho.2011.08.009

Xiang, T., Lohrenz, T., and Montague, P. R. (2013). Computational substrates of norms and their violations during social exchange. *J. Neurosci.* 33, 1099–1108. doi: 10.1523/JNEUROSCI

Yamagishi, T., Horita, Y., Takagishi, H., Shinada, M., Tanida, S., and Cook, K. S. (2009). The private rejection of unfair offers and emotional commitment. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11520–11523. doi: 10.1073/pnas.0900636106

Zheng, L., Guo, X., Zhu, L., Li, J., Chen, L., and Dienes, Z. (2015). Whether others were treated equally affects neural responses to unfairness in the Ultimatum Game. *Soc. Cogn. Affect. Neurosci.* 10, 193–243. doi: 10.1093/scan/nsu071

Zheng, L., Ning, R., Li, L., Wei, C., Cheng, X., Zhou, C., et al. (2017). Gender differences in behavioral and neural responses to unfairness under social pressure. *Sci. Rep.* 7:13498. doi: 10.1038/s41598-017-13790-6

Zheng, Y., Cheng, X., Xu, J., Li, Z., Li, L., Yang, G., et al. (2017). Proposers' economic status affects behavioral and neural responses to unfairness. *Front. Psychol.* 8:847. doi: 10.3389/fpsyg.2017.00847

Zhou, X., and Wu, Y. (2011). Sharing losses and sharing gains: increased demand for fairness under adversity. *J. Exp. Soc. Psychol.* 47, 582–588. doi: 10.1016/j.jesp.2010.12.017

Zinchenko, O., and Arsalidou, M. (2017). Brain responses to social norms: meta-analyses of fMRI studies. *Hum. Brain Mapp.* 39, 955–970. doi: 10.1002/hbm.23895

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 20 March 2018
doi: 10.3389/fpsyg.2018.00350

Higher Status Honesty Is Worth More: The Effect of Social Status on Honesty Evaluation

Philip R. Blue[1], *Jie Hu*[1] *and Xiaolin Zhou*[1,2,3,4]*

[1] *Center for Brain and Cognitive Sciences and School of Psychological and Cognitive Sciences, Peking University, Beijing, China,* [2] *Key Laboratory of Machine Perception, Ministry of Education, Peking University, Beijing, China,* [3] *Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China,* [4] *PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, China*

Promises are crucial for maintaining trust in social hierarchies. It is well known that not all promises are kept; yet the effect of social status on responses to promises being kept or broken is far from understood, as are the neural processes underlying this effect. Here we manipulated participants' social status before measuring their investment behavior as Investor in iterated Trust Game (TG). Participants decided how much to invest in their partners, who acted as Trustees in TG, after being informed that their partners of higher or lower social status either promised to return half of the multiplied sum (4 × invested amount), did not promise, or had no opportunity to promise. Event-related potentials (ERPs) were recorded when the participants saw the Trustees' decisions in which the partners always returned half of the time, regardless of the experimental conditions. Trustee decisions to return or not after promising to do so were defined as honesty and dishonesty, respectively. Behaviorally, participants invested more when Trustees promised than when Trustees had no opportunity to promise, and this effect was greater for higher status than lower status Trustees. Neurally, when viewing Trustees' return decisions, participants' medial frontal negativity (MFN) responses (250–310 ms post onset) were more negative when Trustees did not return than when they did return, suggesting that not returning was an expectancy violation. P300 responses were only sensitive to higher status return feedback, and were more positive-going for higher status partner returns than for lower status partner returns, suggesting that higher status returns may have been more rewarding/motivationally significant. Importantly, only participants in low subjective socioeconomic status (SES) evidenced an increased P300 effect for higher status than lower status honesty (honesty – dishonesty), suggesting that higher status honesty was especially rewarding/motivationally significant for participants with low SES. Taken together, our results suggest that in an earlier time window, MFN encodes return valence, regardless of honesty or social status, which are addressed in a later cognitive appraisal process (P300). Our findings suggest that social status influences honesty perception at both a behavioral and neural level, and that subjective SES may modulate this effect.

Keywords: social status, promise, trust, trust game, social hierarchy, ERP, MFN, P300

OPEN ACCESS

Edited by:
Chao Liu,
Beijing Normal University, China

Reviewed by:
Raoul Bell,
Heinrich-Heine-Universität Düsseldorf, Germany
Ruolei Gu,
Institute of Psychology (CAS), China

****Correspondence:***
Xiaolin Zhou
xz104@pku.edu.cn

Specialty section:
This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Psychology

Received: *28 November 2017*
Accepted: *02 March 2018*
Published: *20 March 2018*

Citation:
Blue PR, Hu J and Zhou X (2018) Higher Status Honesty Is Worth More: The Effect of Social Status on Honesty Evaluation. Front. Psychol. 9:350. doi: 10.3389/fpsyg.2018.00350

INTRODUCTION

Promises are crucial for creating trust in situations where trust does not yet exist (Malhotra and Murnighan, 2002; Friedrich and Southwood, 2011). As such, promises are particularly useful in social hierarchies by acting to decrease feelings of distrust between individuals of different social status (Fiske, 2010a; Lount and Pettit, 2012). Promises are ubiquitously used not only to signal/foster trustworthiness to the hierarchy (i.e., pledges or oaths), but are also critical in facilitating trust between individuals of different social ranks, from a high-ranking politician promising voters that she will increase the economy to a low-ranking employee assuring her manager that she will finish her work on time. Despite the importance of promises in facilitating trust between different members of a hierarchy, it is common knowledge that promises are not always kept, and broken promises can have large downstream effects on trust at both personal (Simpson, 2007) and economic levels (Zak and Knack, 2001), making the evaluation of promise outcomes (i.e., promise kept vs. promise broken) of utmost importance to understanding trust in social hierarchies. However, the effects of social status on the evaluation of promise outcomes is far from understood at both behavioral and neural levels.

Previous work on responses to promise outcome evaluation in social hierarchies is almost completely restricted to feedback related to high status promisors (e.g., politicians; Johnson and Ryu, 2010; Corazzini et al., 2014; Born et al., 2017). Here we turn to related work regarding the effects of social status on responses to social norm violations to inform our hypotheses regarding the potential differential effects of lower status and higher status on promise outcome evaluation. In particular, there is an ongoing debate regarding the effects of social status on the evaluation of social norm violations (Wahrman, 2010). One line of research suggests that high status norm violation is judged more harshly than that of their low status counterparts because people tend to have higher expectations of high status than low status others, attributing them with more intentionality and perceiving them as being more responsible for wrongdoing (i.e., "expectation violation" account; Wahrman, 1970; Hamilton and Sanders, 1981). In one study, participants rated the norm violation (i.e., underpayment of personal income taxes) of a wealthy and politically connected New Yorker (i.e., higher status) as being more intentional and recommended increased punishment severity for this norm violation than the same action committed by an immigrant to New York (i.e., low status; Fragale et al., 2009). Moreover, research has found that when a low status (employee) and high status (boss) individual enter into a formal agreement, people are more likely to break off the agreement if the boss breaks the terms of the agreement than if the employee breaks the terms of the agreement (Fiddick and Cummins, 2001).

A second line of research shows that people judge high status norm violation *less* harshly than that of their low status counterparts (Bowles and Gelfand, 2010; von Essen and Ranehill, 2013). In a field study, researchers measured people's responses to a low or high status individual (based on dress and perceived occupation) who accidentally knocked over a person's briefcase (Ungar, 1981); they found that while blame assigned to low or high status others was the same, people were more likely to derogate (i.e., judge in a negative way) low status than high status suitcase-kickers. Some researchers speculate that this decreased response to high status norm violations may be because high status individuals are given "idiosyncrasy credits" and "wiggle room" to engage in more creative and more beneficial, but sometimes unethical, behavior because they have more value to add to the group (i.e., "social value" account; Hollander, 1958; Polman et al., 2013). As long as the norm violator has not "used up" her/his credits and retains value to the group, then the norm violation of the high status individual goes unpunished. If, however, high status deviance results in group failure, then high status group members are punished even more severely than low status group members (Wiggins et al., 1965; Alvarez, 1968).

One reason for this divergence in the literature is that the magnitude of the social norm violation is inconsistent across studies. For example, accidentally kicking over one's briefcase (low magnitude social norm violation) and not paying federal income taxes (high magnitude social norm violation) lead to opposite effects of social status on social norm violation evaluation. Moreover, almost all studies mentioned above are restricted to the evaluation of social norm violations, overlooking the evaluation of social norm *adherence*. Another reason for the divergence in the above-mentioned literature is that the majority of these studies manipulate the socioeconomic status (SES) of the target individual being judged, which has two major disadvantages. The first is that they fail to account for the participant's own relative social status, which is problematic given that previous research on social status shows that self-status and other-status often interact to affect evaluation of norm violation (Cummins, 1999; Haselhuhn et al., 2015) and social interaction/related processing (Deaner et al., 2005; Ly et al., 2011; Blue et al., 2016). The second disadvantage is that SES is often confounded by feelings of power, making it difficult to distinguish which effects are uniquely driven by social status, and which effects are driven by power. In fact, some researchers speculate that what is driving the diminished punishment of high status norm violators is not "social value" *per se*, but instead a fear of retaliation from the high status norm violators (Homans, 1961; Aquino et al., 2006), which suggests that power may be confounding the effects of social status on norm violation evaluation.

Social status and power are similar but distinct and have been shown to have different effects on behavior and social cognition (Magee and Galinsky, 2008; Blader and Chen, 2012; Dubois et al., 2015; Blader et al., 2016). One common type of social status, socioeconomic status (i.e., SES), is composed of Objective SES and Subjective SES, where Objective SES refers to an individual's/parents' salary, vocation, and/or highest achieved level of education (Oakes and Rossi, 2003; Kraus et al., 2009), and Subjective SES refers to an individual's feelings regarding his/her relative level of salary, vocation, and education in comparison with a relevant population (Adler et al., 2000). In contrast, power (i.e., dominance-based status) refers to an individual's level of control over another individual's access to a valued resource or outcomes (Dépret and Fiske, 1993; Galinsky et al., 2003; Keltner et al., 2003; Fiske, 2010b). To disentangle these two types of status constructs, researchers often turn to prestige-based status measures and manipulations (Zink et al., 2008). Prestige-based status refers to the amount of deference, respect, or admiration an individual receives along a relevant domain (Adler et al., 2000; Henrich and Gil-White, 2001; Fiske, 2010b). This type of social status is particularly advantageous because it is distinct from power and wealth and is easily manipulated in a lab setting.

To address the above-mentioned limitations, in the current study we systematically analyze the behavioral and neural effects of both prestige-based status (manipulated at the beginning of the experiment) and SES (measured after the experiment) on promise feedback evaluation. We do so by manipulating participants' prestige-based status before playing as Investor in a modified version of iterated Trust Game (TG) with promises (Charness and Dufwenberg, 2006, 2010). Participants' prestige-based status was manipulated via performance ranking on a math quiz in comparison with six confederate players. This is a proven and established inducer of prestige-based status (Hu et al., 2016) with the advantage that it can control for other potential confounds such as power or dominance. In line with previous research (Albrecht et al., 2013), we also control for potential emotional confounds of achieving low-status or high-status ranking (Steckler and Tracy, 2014) by endowing participants with a middle-status ranking in comparison with the six other players and pair them with partners of lower or higher status. After receiving their ranking, participants played several trials of TG as Investor with these players (whose identity was kept anonymous) acting as Trustees. At the beginning of each TG trial, the participant first viewed the social status of the Trustee (lower vs. higher) who had been drawn randomly from the pool of six confederates. To prevent reputation effects and learning, no other personal information was given at this stage. Then, to measure the effects of social status on responses to promise-based feedback, in TG, Trustees either promised ("promise" condition) or did not promise ("no promise" condition, filler) to return half of the multiplied sum (i.e., half of the investment amount after it has been multiplied by 4) to the participant. To create a condition where promise information was not available, in certain trials, Trustees were not given the opportunity to make a promise decision (i.e., "unknown" condition). After viewing the promise information, participants decided whether or not to invest 2 yuan, which was endowed to the investor, in the Trustee. This amount was set at 2 yuan to control for potential magnitude effects of returning or not returning. Finally, the participant was given feedback regarding whether the Trustee had behaved in a trustworthy manner (i.e., return in the "unknown" condition) or in an honest manner (i.e., return in the "promise" condition) before beginning the next trial of TG. In this way, participants experienced both negative *and* positive outcome feedback. Feedback was given regardless of whether or not the participants invested in the Trustee (i.e., forced feedback). This measure was taken to ensure that all participants were made aware that lower and higher status Trustees were trustworthy and honest in half of the trials. We recorded event-related potentials (ERPs) time-locked to the TG feedback. The empirical question was whether and how social status modulates the behavioral and neural responses to honesty and trustworthiness feedback.

At the behavioral level, we focused on the investment rate in TG. Our previous work using a similar prestige-based status manipulation before measuring participants' behavior as Investor in iterated one-shot TG shows that participants tend to invest more in higher status Trustees than lower status Trustees, and that this effect is most pronounced in the "promise" condition (Blue et al., under review). Given these findings, and considering the two diverging accounts regarding the effects of social status on responses to norm violation, two hypotheses emerge: the "social value" hypothesis would predict that participants would be more likely to invest in higher status promises than in lower status promises, despite feedback showing that lower and higher status partners were equally honest ("social value" hypothesis). The "social value" hypothesis contrasts with an alternative "expectation violation" hypothesis, which would predict that participants would be more surprised by higher status partners' dishonesty and lack of trustworthiness in 50% of the trials than by that of lower status partners, and would thus invest less in higher status than lower status partners over time.

At the neural level, we focused on two ERP components time-locked to TG feedback which are known for their involvement in outcome evaluation: Medial-frontal negativity (i.e., MFN) and P300 (i.e., P3). The "social value" hypothesis would predict that P300 amplitudes would be more positive-going in response to higher status honesty than lower status honesty, whereas the "expectation violation" hypothesis would predict that MFN responses would be more negative-going to higher status dishonesty than lower status dishonesty. Below, we briefly introduce MFN and P300 and their role in outcome evaluation before we move on to the methodological details of the study.

MFN reflects a family of components related to negative performance feedback (i.e., feedback-related negativity, FRN) and error-related processing (i.e., error-related negativity; ERN). MFN is a negative deflection peaking between 200 and 350 ms post-onset and is found in the frontocentral electrodes. MFN is generated by activity in the anterior cingulate cortex (i.e., ACC). It is often described as reflecting whether the evaluation of events/feedback is good or bad (Gehring and Willoughby, 2002; Yeung and Sanfey, 2004; Sato et al., 2005). In particular, MFN amplitudes are more pronounced for negative feedback and unfavorable outcomes than for positive feedback and favorable outcomes. The reinforcement learning account of MFN states that the mesencephalic dopamine system sends reinforcement learning signals, which are manifest by changes in phasic dopamine (Schultz et al., 1997), via the basal ganglia to ACC, which then learns which decisions are best (Holroyd and Coles, 2002). The MFN thus reflects the reinforcement learning signals to ACC: negative prediction errors (i.e., outcome is worse than expected) elicit greater MFN amplitudes, reflecting decreases of phasic dopamine to ACC, whereas positive prediction errors (i.e., outcome is better than expected) elicit decreased MFN amplitudes, reflecting increases of phasic dopamine to ACC.

Apart from outcome valence, MFN has also been shown to be sensitive to outcome expectancy (Jia et al., 2007; Wu and Zhou, 2009), such that outcomes that are less expected elicit a more pronounced MFN response. Expectancy violation effects on

MFN are also found in social contexts (Wu et al., 2011b), such as social norm violations. MFN is sensitive to social norm violations, such as those related to fairness and generosity (Boksem and De Cremer, 2010; Wu et al., 2011a). Violation of trust, another form of social norm violation, also elicits enhanced MFN responses (Long et al., 2012). Moreover, relevant to the current study, previous research also shows that MFN is sensitive to Trustee honesty in TG (Ma et al., 2015). A few studies have found that social factors, such as social distance (Ma et al., 2011; Wu et al., 2011a), can modulate MFN responses to norm violations, suggesting that factors such as social status may also be capable of modulating the effect of MFN on TG feedback. If higher status dishonesty elicits greater expectation violation (i.e., is perceived as a greater social norm violation) than lower status dishonesty, then the MFN effect should be more pronounced for higher status than lower status dishonesty.

P300 (also referred to as P3; Sutton et al., 1965) represents later top-down attentional resources devoted to outcome evaluation and reward. P300 is the most positive peak 200–600 ms post-onset of feedback and is found in the medial parietal electrodes. The P300 was originally recognized as encoding the motivational significance of a stimulus (Duncan-Johnson and Donchin, 1977). Motivationally significant stimuli are those that “are either relevant to the current task or that have the potential to be associated with some form of utility (positive or negative)” (Nieuwenhuis et al., 2005, p. 511). In addition to the motivational significance of the stimulus, P300 is modulated by the probability of a stimulus occurrence (Donchin and Coles, 1988) and the amount of attention paid to the stimulus (Yeung et al., 2005; Wu and Zhou, 2009).

P300 is also involved in outcome evaluation. Certain studies measuring P300 responses to gambling outcomes have shown that P300 is sensitive to the magnitude of the outcome, but not its valence (Yeung and Sanfey, 2004; Sato et al., 2005; Yeung et al., 2005), causing these researchers to attribute P300 as being sensitive only to the motivational significance of an outcome. However, other research on P300 responses to gambling outcomes shows that P300 amplitude is sensitive to both the magnitude and the valence of the outcome (Hajcak et al., 2005, 2007; Wu and Zhou, 2009), suggesting that P300 reflects broad cognitive appraisal processing related to attention and reward (Gray et al., 2004; Linden, 2005). P300 amplitudes are also sensitive to social factors in outcome-processing (Li et al., 2010; Zhou et al., 2010; Ma et al., 2011). One study showed that when gambling outcomes were compared with those of friends and strangers, P300 was sensitive not only to the reward valence, but also to the social distance of the person with whom they were comparing gambling outcomes (Leng and Zhou, 2010), which suggests that social factors can modulate the effect of attention or motivational significance on P300.

In trust-related situations, P300 may also reflect social value above and beyond simple trustworthiness feedback. For example, Boudreau et al. (2009) measured participants’ trust behavior in a coin-toss game, in which participants guessed whether a coin tossed by a partner was heads or tails based on the partner’s indication. In one condition, participants had common interests with the partners (i.e., if the participant guessed correctly, they both received a monetary reward); in another condition, participants were told that partners would receive a penalty for misleading the participant. Findings showed that, despite the equal probability of participants trusting partners in the “common interests” and “penalty for lying” conditions (96 and 97%, respectively), participants’ P300 responses were more pronounced for recommendations given by partners with common interests than by partners who would receive a penalty for lying, which suggests that other factors, such as social value, may modulate P300 above and beyond simple perceived trustworthiness levels.

Similar work using TG shows that factors related to the Trustee can influence investment behavior in the Trustee and neural responses to Trustee TG outcome feedback (i.e., “return” vs. “no return”; Delgado et al., 2005). In particular, Fareri et al. (2015) found that participants acting as Investors were more likely to invest in Trustees who were their friends than Trustees who were strangers, and that activity in ventral striatum, a brain region known for reward processing, was greater when friends returned than when strangers returned, despite the fact that the reinforcement rate (i.e., return percentage) was equal for friends and for strangers. The authors found support showing that the neural and behavioral responses were influenced by the increased “social value” of the Trustee when they were friends, compared with when they were strangers (Fareri et al., 2015). Similarly, when viewing TG outcomes, return decisions by Trustees with good reputations elicit greater activity in ventral striatum than the same return responses from partners with bad reputations (Phan et al., 2010), which further demonstrates that certain Trustee characteristics may increase the perceived reward of return decisions in TG. Moreover, work simultaneously measuring fMRI and EEG during gain and loss anticipation shows that increased P300 amplitudes for gains over losses is positively correlated with ventral striatum activity during the anticipation of gains (Pfabigan et al., 2014), which suggests that P300 reward processing may be associated with reward processing in ventral striatum. Taken together, we suspect that P300 will not only detect the valence of TG feedback (i.e., “return” vs. “no return”), but that factors potentially related to perceived “social value,” such as social status and honesty, may also influence attentional resources devoted to TG feedback by interacting with P300 during TG outcome processing.

Finally, we test for the potential effects of SES on the evaluation of TG outcomes. SES has been shown to modulate attention allocation in social settings (Dietze and Knowles, 2016). Moreover, past research shows that, in comparison with lower status, interaction with higher status others elicits greater activity in the ventral striatum (Zink et al., 2008), and that activity in the ventral striatum can be modulated by participants’ SES when viewing low and high status others (Ly et al., 2011). Research on monkeys also shows that rhesus macaques will give up highly valued rewards (i.e., sugary liquid) in order to view high status others, and that this effect is modulated by rhesus macaques’ own social status (Deaner et al., 2005). Taken together, we suspect that participants’ SES may modulate the valuation of lower and higher status honesty.

MATERIALS AND METHODS

Participants

To determine the sample size, we used G*Power 3 software (Faul et al., 2007), which showed that we needed a sample size of at least 32 for this study to have adequate power ($1 - \beta > 0.95$) to detect a medium-size effect ($f = 0.30$). The power analysis (repeated-measures, within participants effect) was performed for the interaction between partner social status (lower vs. higher) and promise (promise vs. unknown). The correlation among repeated measures was set at 0.6, which was based off of the correlation among repeated measures in a previous behavioral pilot study [$r(28) = 0.594$; Blue et al., under review]. Among the 42 participants we tested, two were removed because alpha-wave artifacts, two failed the post-experiment questionnaire for understanding the experimental setup and task requirements, four were suspicious of the experimental setup, and one was removed due to a technical malfunction. These nine participants were removed from data analysis, leaving 33 participants (20 females) in the following analysis whose age was between 18 and 23 years (mean: 19.70 years, *SD* = 1.40). All participants were healthy, right-handed, had normal or corrected-to-normal vision, and no participants had a history of neurological or psychiatric disorders. Before the experiment, all participants gave their informed consent and were informed that the basic payment for participation was 80 Chinese yuan (about 12.5 USD) with a bonus of 0–15 yuan, which was based on performance in TG. The experiment was in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the School of Psychological and Cognitive Sciences, Peking University.

Design and Procedure

The experiment had a 2 (partner social status: lower vs. higher) × 2 (promise: promise vs. unknown) × 2 (return: return vs. no return) within-participants factorial design. An additional filler condition, in which the partner had an opportunity to promise but did not choose to promise ("no promise" condition) was included to increase the perceived agency in the "promise" condition. As in past experiments (Hu et al., 2014, 2016), we used a star system (Zink et al., 2008) to assign social status, with one star indicating low status, two stars indicating middle status, and three stars indicating high status. The investment decision was binary (i.e., invest vs. no invest). The investment amount was set at 2 yuan, making the multiplied sum 8 yuan.

Participants arrived alone to the laboratory for each experimental session, where they were told that six same-sex participants (confederates) were ostensibly waiting in another room. Participants then gave permission to have their photo taken, which would later be used in the math quiz ranking screen, along with the photos of the six confederates. The participants were told that the six confederates would also complete the math quiz and would later act as their partner in TG (**Figure 1**).

The math quiz task is an established inducer of social status (Hu et al., 2016). Participants were given 10 s to select which of two arithmetic expressions had a greater value by pressing the "F" or "J" key with the left or right index finger. If the participant had not selected a response after 7 s, he/she was given a reminder that time was running out on that particular question. Each problem was composed of either two-digit multiplication (e.g., 45*72) or complex fraction addition (e.g., $4\frac{7}{8} + 5\frac{5}{9}$). In total, there were 24 arithmetic problems (12 easy, 12 difficult). Half of the problems were solvable in the time allotted while the other half were extremely difficult to solve in the time allotted, which facilitated the participant's belief that they had achieved a two-star (middle status) ranking. Upon completion of the quiz, participants viewed their rank in comparison with the ranks of the six other confederates. All participants were assigned a middle (2-star) ranking in order to avoid the potential influences of emotion after gaining high or low status (Steckler and Tracy, 2014) and to test the effects of others' social status on participants' responses to promise outcome feedback (Albrecht et al., 2013).

In TG, participants acted as Investor, and the six confederates from the math quiz acted as Trustees. Participants were informed that they would only be paired with Trustees who had achieved rankings that were different from their own, so they only faced low (one-star) and high (three-star) status Trustees. This was meant to increase the number of trials in the critical conditions. At the start of each trial (264 trials in total), participants were given 2 yuan. Then, participants viewed the ranking of their anonymous partner for that particular trial. Next, participants viewed the partner's promise decision, with " ! " indicating that the partner promised to return 4 yuan (50% of the multiplied sum; "promise" condition), "- -" indicating that the partner did not promise to return 4 yuan ("no promise" condition; filler), and " | | " indicating that the partner was not given the opportunity to make a promise decision on that trial ("unknown" condition). Then the participant chose whether or not to invest the 2 yuan in the partner. The participant was given a maximum of 5 s and used the "F" and "J" keys on the keyboard to make this decision ("invest" and "keep" decision locations were counterbalanced over trials). If the participant did not make an investment decision within 5 s, the trial started again from the beginning. If the participant chose to invest the 2 yuan, then the partner received 8 yuan; if the participant chose to keep the 2 yuan, then the partner came away from that trial with nothing. Finally, the participant viewed the Trustee's feedback (i.e., decision to return or not to return) on that trial. Importantly, participants were told that Trustees made their return decisions at the same time as participants were making their decision to invest or not (i.e., before viewing the participant's investment decision). This point was emphasized to the participants because it was necessary to give feedback to the participants on each TG trial regardless of whether they invested or not. We used forced feedback for two reasons: (1) to ensure that lower and higher status partner trustworthiness and honesty levels were identical (i.e., lower status Trustees and higher status Trustees both returned on 50% of the "promise" trials and 50% of the "unknown" trials), and (2) to ensure that there were enough trials in the critical conditions for ERP data analysis. As filler trials, we also included certain trials where Trustees did not promise to return in 50% of the multiplied sum (i.e., "no promise" condition; 12 trials in total); in these trials, TG partners did not return.

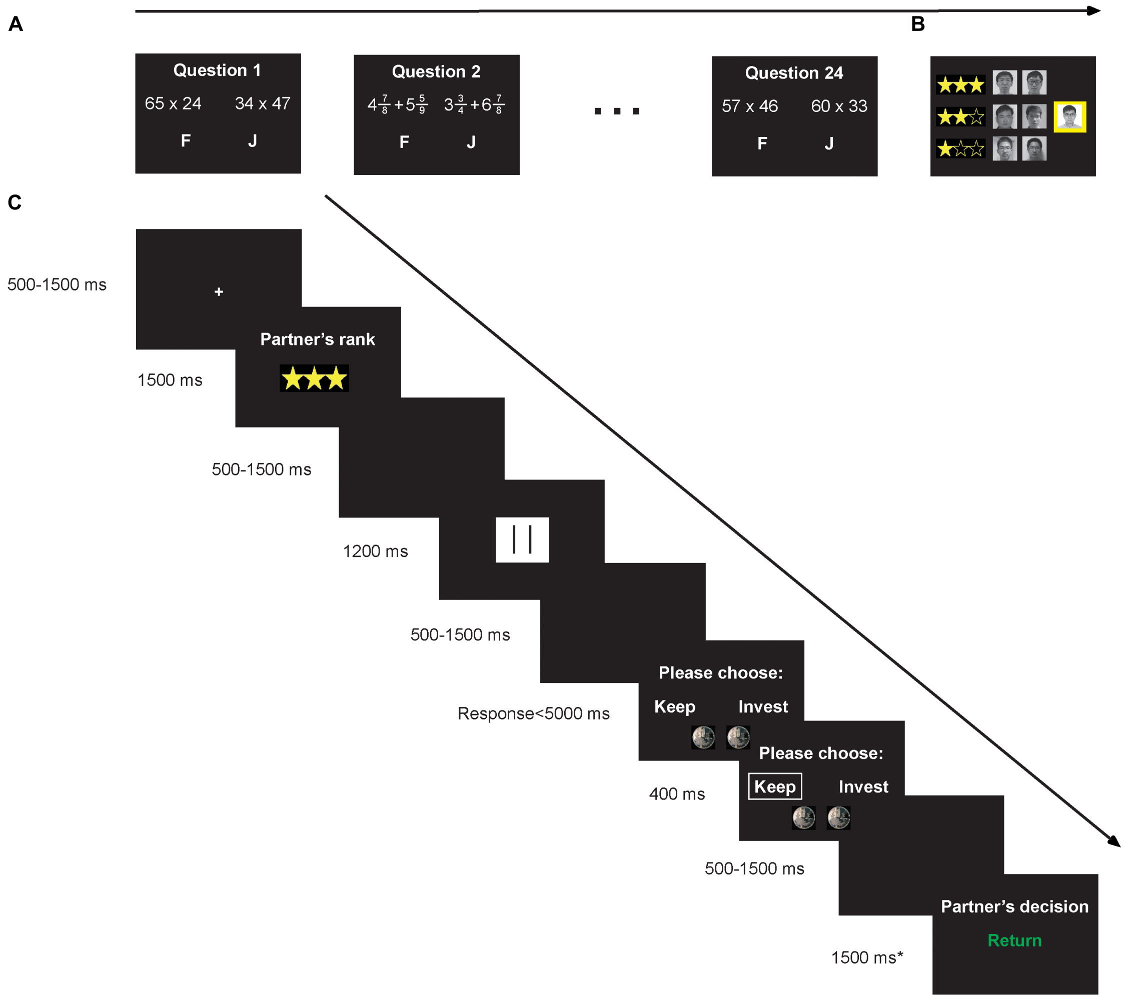

FIGURE 1 | Schematic diagram of the experiment. **(A)** In the rank-inducing task (i.e., math quiz), participants were given 10 s to select which of two arithmetic expressions had a greater value by pressing the "F" or "J" key with the left or right index finger. **(B)** Upon completion of the quiz (24 problems in total), participants were shown their performance ranking in comparison with six same-sex confederate players. All participants were assigned a middle (2-star) ranking, highlighted in yellow. **(C)** In TG, participants acted as Investor with Trustees of lower and higher status. On each trial, participants viewed the Trustee's ranking from the math quiz along with his/her promise information before deciding whether or not to invest in the Trustee. Regardless of whether the participant invested in the Trustee, participants were shown the return feedback. Return feedback was set at 50% regardless of the condition. We focused on the ERP responses time-locked to the feedback screen (noted with the asterisk).

Each trial of TG began with a fixation sign (white cross subtended 0.3° of visual angle) for either 500, 700, 900, 1100, 1300, or 1500 ms against a black background (**Figure 1C**). On the next screen, participants viewed the words "Your partner's rank:" in Chinese (white and Song font, size 32) above the star ranking (subtended 2° × 0.8°) for 1500 ms; the star ranking was composed of either a yellow filled star with two empty yellow stars (one-star, lower status rank) or three yellow filled stars (three-star, higher status rank). After the presentation of a blank screen for a jittered time between 500 and 1500 ms, participants then viewed the partner's promise information for 1200 ms. After the presentation of a blank screen for a jittered time between 500 and 1500 ms, the participants then viewed the words "Please choose" above the choices "Invest" and "Keep" (the locations of which were counterbalanced across trials) for a maximum of 5000 ms. After making their selection, a white box immediately highlighted the answer response for 400 ms. After the presentation of a blank screen for a jittered time between 500 and 1500 ms, the participants finally viewed the words "Partner's decision:" above the words "Return" or "No return" in Chinese in green and red, respectively, with colors counterbalanced across participants. The final screen appeared for 1500 ms.

EEG data were recorded throughout the experiment. We focused our analysis on the TG feedback screen. The participants were comfortably seated in a dimly lit and electromagnetically shielded room about 1.5 m in front of a computer screen. The experiment used Presentation software (Neurobehavioral System Inc.) to control the timing and presentation of stimuli and was displayed on a Visuosonic 22-in. CRT display. The experiment consisted of the status-inducing task (i.e., math quiz, 24 problems in total) followed by six blocks of TG (44 TG trials per block). There were 30 trials per condition (lower status "unknown" return; lower status "unknown" no return; lower status "promise" return; lower status "promise" no return; higher status "unknown" return; higher status "unknown" no return; higher status "promise" return; higher status "promise" no return) and 24 "no promise" filler trials (12 lower status; 12 higher status) in which Trustees did not return on any trial.

Before beginning the experiment, participants were told that bonus payments (0–15 yuan) were based on TG behavior from 10 randomly selected trials. In addition, all participant completed practice math quiz problems and TG trials until they were comfortable with the setup, with a minimum of 6 math problems and 10 TG trials in the practice session. Participants were also tested on their recognition of the promise symbols. No participants were allowed to begin the experiment without being able to consistently and accurately identify each symbol. No participants reported difficult with remembering the promise symbols.

After the experiment, participants reported on a 7-point Likert scale to what extent they felt superior or inferior (1 = very inferior; 7 = very superior) when facing partners of the two ranks. This measure served as a manipulation check of social status; each participant indicated this rating once for each social status. As an additional manipulation check, participants indicated their own Subjective SES and the Subjective SES of both lower and higher status partners. Subjective SES was measured using the MacArthur Subjective Social Status Scale (Adler et al., 2000), which asks participants to indicate the target's subjective status in Chinese society on a ladder, with the lowest rungs indicating individuals with the lowest level of income, occupation, and education, and the highest rungs indicating individuals with the highest level of income, occupation, and education. To test for potential effects of Objective SES, participants also indicated their parents' highest level of education (1 = middle school diploma; 2 = high school diploma/middle trade school certificate; 3 = trade school certificate; 4 = bachelor's degree; 5 = graduate degree) and their parents' annual income (1 = 0 – 10,000 yuan; 2 = 10,000 – 100,000 yuan; 3 = 100,000 – 300,000 yuan; 4 = 300,000 – 500,000 yuan; 5 = 500,000 – 1,000,000 yuan; 6 = 1,000,000 – 5,000,000 yuan; 7 $\geq$ 5,000,000 yuan. Note, for the sake of privacy, participants were allowed to select "8" which indicated that they did not want to respond to this question).

To test for potential differences in learning of lower and higher status trustworthiness and honesty, immediately after the experiment, participants were asked to recall lower and higher status behavior during TG and indicate the ratio of lower status and higher status trustworthiness (percentage of "return" decisions in the "unknown" condition) and honesty (percentage of "return" decisions in the "promise" condition). For each item, participants were asked to indicate any number from 0 to 100%. Finally, to more explicitly measure perceived trustworthiness, participants were asked to indicate perceived trustworthiness of lower/higher status partners based on their behavior in TG. To measure perceived trustworthiness of lower/higher status partners, after the experiment, we recorded participants' feelings of perceived ability, benevolence, and integrity of lower and higher status TG partners, which are three fundamental components of trustworthiness (Mayer et al., 1995). The perceived trustworthiness measures were the same measures as those used in similar research (Lount and Pettit, 2012), which were drawn from previous work in organizational psychology on trustworthiness perception (Mayer and Davis, 1999). The questions are aimed at addressing employees' feelings toward employers ("top management"); we adjusted the questions to be less work-oriented and more suitable for students. Participants rated each status of partners on each of the three dimensions. Ability was composed of 6 items (e.g., This individual is very efficient) (α = 0.822); Benevolence was composed of five items (e.g., "This individual is concerned about my welfare.") (α = 0.805); Integrity was composed of six items (e.g., This individual has a strong sense of justice) (α = 0.798). Participants recorded their responses using a 7-point Likert scale (1 = completely disagree, 7 = completely agree).

EEG Recording and Analysis

EEGs were recorded from 64 scalp sites using tin electrodes mounted in an elastic cap (Brain Products, Munich, Germany) according to the international 10–20 system. The horizontal electrooculogram (HEOG) was recorded from electrodes placed at the outer cantus of the left eye, and the vertical EOG (VEOG) was recorded supra-orbitally from the right eye. All EEGs and EOGs were referenced online to an external electrode on the tip of the nose; they were re-referenced off-line to the mean of the left and right mastoids. For all electrodes, electrode impedance was kept below 5 kΩ. Bio-signals were amplified with a band-pass from 0.016 to 100 Hz and digitized online with a sampling frequency of 500 Hz.

Offline, we extracted separate EEG epochs (200 ms pre-stimulus to 800 ms post-stimulus), which were time-locked to the onset of the TG feedback screen. The EEG data were high-pass filtered at 0.1 Hz and low-pass filtered at 30 Hz. Baseline correction for each epoch was done by subtracting the average activity of the channel during the baseline period from each sample. Trials in which EEG voltages exceeded ± 70 μV were excluded from further analysis. After artifact rejection, an average of 86% of trials (*SD* = 10%) of the epochs on the TG feedback screen were entered into statistical analysis.

For statistical analysis, electrodes were divided based on two three-level factors: Region (anterior vs. central vs. posterior) and Hemisphere (left vs. medial. vs. right), which resulted in 9 regional clusters: the left anterior cluster was composed of F3, F5, FC3, and FC5; the medial anterior cluster was composed of F1, Fz, F2, FC1, FCz, and FC2; the right anterior cluster was composed of F4, F6, FC4, and FC6; the left central cluster which was composed of C3, C5, CP3, and CP5; the medial central cluster was composed of C1, Cz, C2, CP1, CPz, and CP2; the right central cluster was composed of C4, C6, CP4, and CP6; the left posterior cluster was composed of P3, P5, and PO7; the medial poster cluster was composed of P1, Pz, P2, PO3, POz, and PO4; the right posterior cluster was composed of P4, P6, and PO8. This clustering method for analyzing the EEG data is similar to the method used in a related study analyzing P300 in response to feedback in a social dilemma game (Bell et al., 2016). For statistical purposes, we averaged the amplitude and/or peaks over electrodes in each regional cluster. Time windows were determined by visual inspection of the waveforms and preliminary analyses.

For ERP responses to the TG feedback, we focused our analysis on MFN (the mean amplitudes in the time window of 250–310 ms) and P300 (the peak values in the time window of 250–600 ms). For MFN, we focused our analysis on the medial anterior cluster. We selected these electrodes because the MFN effect was largest on these electrodes. We conducted ANOVA with three within-subjects factors: partner social status (lower vs. higher), promise condition (promise vs. unknown), and return (return vs. no return). For P300, we conducted ANOVAs with five within-subjects factors: partner social status (lower vs. higher), promise condition (promise vs. unknown), return (return vs. no return), region (anterior vs. central vs. posterior), and hemisphere (left vs. medial vs. right). In order to account for multiple comparisons, Bonferroni correction was used when appropriate. In cases of non-sphericity, we applied the Greenhouse–Geisser correction.

RESULTS

Manipulation Check of Social Status

In order to ensure that the social status manipulation elicited feelings of inferiority and superiority, we conducted a one-factor (star-ranking: one vs. three) repeated-measures ANOVA, which confirmed the social status manipulation, $F(1,31) = 57.923$, $p < 0.001$, $\eta_p^2 = 0.651$. Participants reported higher feelings of superiority when facing a lower status partner (5.313 ± 0.171) than when facing a higher status partner (3.469 ± 0.149). The status manipulation also affected feelings of Subjective SES, $F(2,64) = 39.123$, $p < 0.001$, $\eta_p^2 = 0.550$, as participants rated three-star partners as having a higher Subjective SES (6.955 ± 0.224) than their own (6.015 ± 0.250), $p < 0.001$, and one-star partners as having lower Subjective SES (5.045 ± 0.231) than their own (6.015 ± 0.250), $p < 0.001$. Objective SES results are reported below (see *Objective SES*).

Behavioral Results

A repeated-measures analysis of variance (ANOVA) showed that the investment ratio varied as a function of promise, $F(1,32) = 52.019$, $p < 0.001$, $\eta_p^2 = 0.619$ (**Figure 2**). Participants were more likely to trust (i.e., invest) in "promise" trials (mean ± SE, 0.681 ± 0.027) than in "unknown" trials (0.453 ± 0.027). There was no main effect of partner social status, $p = 0.915$. Importantly, consistent with our previous studies (Blue et al., under review), there was a non-significant trend or tendency of an interaction between partner social status and promise conditions, $F(1,32) = 3.783$, $p = 0.061$, $\eta_p^2 = 0.106$. Further tests revealed that when interacting with higher status partners, participants tended to be more likely to invest in "promise" trials (0.69 ± 0.03) than in "unknown" trials (0.44 ± 0.03, $p < 0.001$, $\eta_p^2 = 0.64$), and this effect was smaller for participants when playing with lower status partners ("promise" condition: 0.67 ± 0.03, "unknown" condition: 0.46 ± 0.03, $p < 0.001$, $\eta_p^2 = 0.54$).

To evaluate the strength of the empirical evidence in favor of (or against) the interaction between partner social status and promise conditions, we also conducted a Bayes factor analysis (Dienes, 2014). Bayes factor analysis tests the strength of evidence between two theories (a null hypothesis theory and the proposed effect in the data), and its value ranges from 0 to infinity, with an increase in value indicating stronger support to reject the null hypothesis. The conventional cut-offs for Bayes factor sensitivity are 1/3 and 3, which means that any value outside of this range (less than 1/3 or greater than 3) provides strong evidence in support of the null hypothesis or the proposed effect in the data, respectively. Values between 1/3 and 3 are considered weak or "anecdotal" evidence (Jeffreys, 1939/1961). Our analysis was conducted using the BayesFactor (Morey et al., 2015) package in the R statistical language. We found a Bayes factor of 2.508 ± 7.65% which suggests that there is an interaction between partner social status and promise condition, but that it is a weak effect. This result indicates that independent confirmation is needed to confirm the interaction between partner social status and promise conditions.

To examine potential differences in learning of lower and higher status trustworthiness and honesty, after the experiment we tested participants' recall of lower and higher status trustworthiness (i.e., ratio of return in the "unknown" condition) and honesty (i.e., ratio of return in the "promise" condition). In particular, to measure recall of trustworthiness, we analyzed participants' responses to the prompts: "Please indicate what percentage of the time (lower) higher status partners returned half of the multiplied sum when they did not have an opportunity to promise to do so?"; to measure recall of honesty, we analyzed participants' responses to the prompts: "Please indicate what percentage of the time (lower) higher status partners returned half of the multiplied sum when they promised to do so?" There was no difference in recall of lower status and higher status trustworthiness, $t(32) = 0.376$, $p = 0.709$, or honesty, $t(32) = 1.491$, $p = 0.146$. As an additional check of learning, we

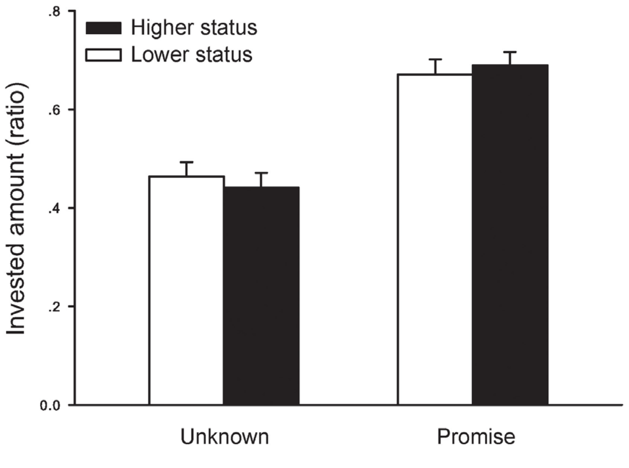

FIGURE 2 | Mean and standard error of the means for ratio of investment in partners of different social status across the two promise conditions ("unknown" vs. "promise").

entered the difference between higher and lower status honesty investment for each block (i.e., [(higher status "promise" – higher status "unknown") – (lower status "promise" – lower status "unknown")]) into a one-factor (block: 1 vs. 2 vs. 3 vs. 4 vs. 5 vs. 6) repeated-measures ANOVA, which was not significant, $F < 1$; $p = 0.997$. We also conducted this one-factor repeated-measures ANOVA separately for status differences in each block of the "unknown" condition, $F = 1.571$, $p = 0.171$, and each block of the "promise" condition, $F = 1.491$, $p = 0.195$. Regardless of the condition, our data show that participants' investment behavior showed no evidence of changing over time. Taken together, these results indicate that there is no evidence that participants learned or adjusted their behavior across the experiment.

Results regarding the post-experiment perceived trustworthiness measurements (i.e., ability, benevolence, and integrity) were as follows. Participants rated higher status partners (4.697 ± 0.117) as having greater ability than lower status partners (4.066 ± 0.110), $t(32) = -4.937$, $p < 0.001$. There was a non-significant trend or tendency for participants rating higher status partners (3.042 ± 0.149) as being more benevolent than lower status partners (3.430 ± 0.187), $t(32) = 1.954$, $p = 0.059$. There was no difference in participants' ratings of higher status (3.859 ± 0.147) and lower status (4.015 ± 0.164) partner integrity, $p = 0.473$. We also tested for the possibility that differences in these factors between lower and higher status may have correlated with the TG behavior interaction between partner social status and promise conditions [i.e., (Higher status "promise" – Higher status "unknown") – (Lower status "promise" – Lower status "unknown")]. No evidence was found for the role of perceived ability, benevolence, or integrity to predict the behavioral interaction between partner social status and promise conditions, perceived ability, $p = 0.699$; perceived benevolence, $p = 0.276$; perceived integrity, $p = 0.569$.

MFN in the 250–310 ms Time Window Following TG Feedback

For ERPs time-locked to the TG feedback (**Figure 3**), in the time window of 250–310 ms in the medial anterior cluster of electrodes, a 2 (partner social status: lower vs. higher) × 2 (promise: promise vs. unknown) × 2 (return: return vs. no return) repeated-measures ANOVA showed a significant main effect of return $F(1,32) = 6.147$, $p = 0.019$, $\eta_p^2 = 0.161$, indicating that participants evidenced more negative-going MFN in response to "no return" feedback (10.738 ± 0.982 μV) than to "return" feedback (11.803 ± 0.931 μV). There was also a significant main effect of promise $F(1,32) = 12.747$, $p = 0.001$, $\eta_p^2 = 0.285$, indicating that feedback in the "unknown" condition elicited more negative-going MFN (10.755 ± 0.912 μV) than feedback in the "promise" condition (11.787 ± 0.974 μV). There was no main effect of partner social status, $p = 0.289$. Moreover, there was no interaction between the three conditions (interaction between promise and partner social status, $p = 0.982$; interaction between promise and return, $p = 0.346$; interaction between partner social status and return, $p = 0.308$; interaction between partner social status, promise, and return, $p = 0.741$).

Given that forced feedback was given to the participant regardless of whether or not an investment decision was made, we tested for the potential effect of investment on MFN response. To test whether MFN responses were modulated by investment behavior, we compared average MFN responses from the same medial anterior cluster electrodes in the 250–310 ms time window time-locked to the TG feedback. In particular, we compared MFN amplitudes from only those trials in which participants invested (i.e., invest-only trials) with MFN amplitudes on all trials regardless of investment (i.e., all trials). Given the limited number of trials (30 trials/condition) and given that participants only invested in 45% of the trials in the "unknown" condition, we only analyzed trials from the "promise" condition (participants invested on 68% of "promise" condition trials). There were too few no-invest trials in both the "promise" and the "unknown" conditions to conduct a meaningful comparison between invest-only trials and no-invest trials, thus we compared invest-only trials with all trials. After removing "promise" condition trials in which the participant did not invest, 7 participants had less than 15 trials per condition. These 7 participants were removed from this supplementary analysis, leaving 26 participants in the analysis. A 2 (invest: yes vs. all trials) × 2 (partner social status: lower vs. higher) × 2 (return: return vs. no return) repeated-measures ANOVA revealed a main effect of return, $F(1,25) = 5.053$, $p = 0.034$, $\eta_p^2 = 0.168$, indicating that participants evidenced more negative-going MFN in response to "no return" feedback (10.133 ± 0.997 μV) than to "return" feedback (11.539 ± 0.972 μV). There was a significant interaction between invest and return, $F(1,25) = 6.891$, $p = 0.015$, $\eta_p^2 = 0.216$. Further tests showed that the main effect of return was stronger for invest-only trials, $F(1,25) = 6.692$, $p = 0.016$, $\eta_p^2 = 0.211$, than in trials that included both invest and no invest trials, $F(1,25) = 2.973$, $p = 0.097$, $\eta_p^2 = 0.106$. No other effects reach significance.

P300 Following TG Feedback

The peak amplitudes of the P300 time-locked to the TG feedback (**Figure 4**) were entered into a 2 (partner social status: lower vs. higher) × 2 (promise: promise vs. unknown) × 2 (return: return vs. no return) × 3 (region: anterior vs. central vs. posterior) × 3 (hemisphere: left vs. medial vs. right) repeated-measures ANOVA. There was a significant main effect of partner social status, $F(1,32) = 6.345$, $p = 0.017$, $\eta_p^2 = 0.165$, indicating that feedback related to higher status partners evoked a more positive-going P300 amplitude (14.351 ± 0.779 μV) than feedback related to lower status partners (13.945 ± 0.780 μV). There were no significant main effects of promise, $p = 0.142$, or return, $p = 0.172$. The interaction between partner social status, promise, and return was not significant, $p = 0.632$. Interestingly, the interaction between partner social status and return was significant, $F(1,32) = 4.819$, $p = 0.036$, $\eta_p^2 = 0.131$, such that P300 amplitudes were more positive-going for higher status "return" feedback (14.731 ± 0.833 μV) than higher status "no return" feedback (13.972 ± 0.752 μV), $p = 0.019$, while there was no difference in P300 amplitude for lower status return feedback (lower status "return:" 13.934 ± 0.798 μV; lower status "no return:" 13.955 ± 0.796 μV, $p = 0.949$) (**Figure 4D**). Moreover, P300 amplitudes were more positive-going in response to higher

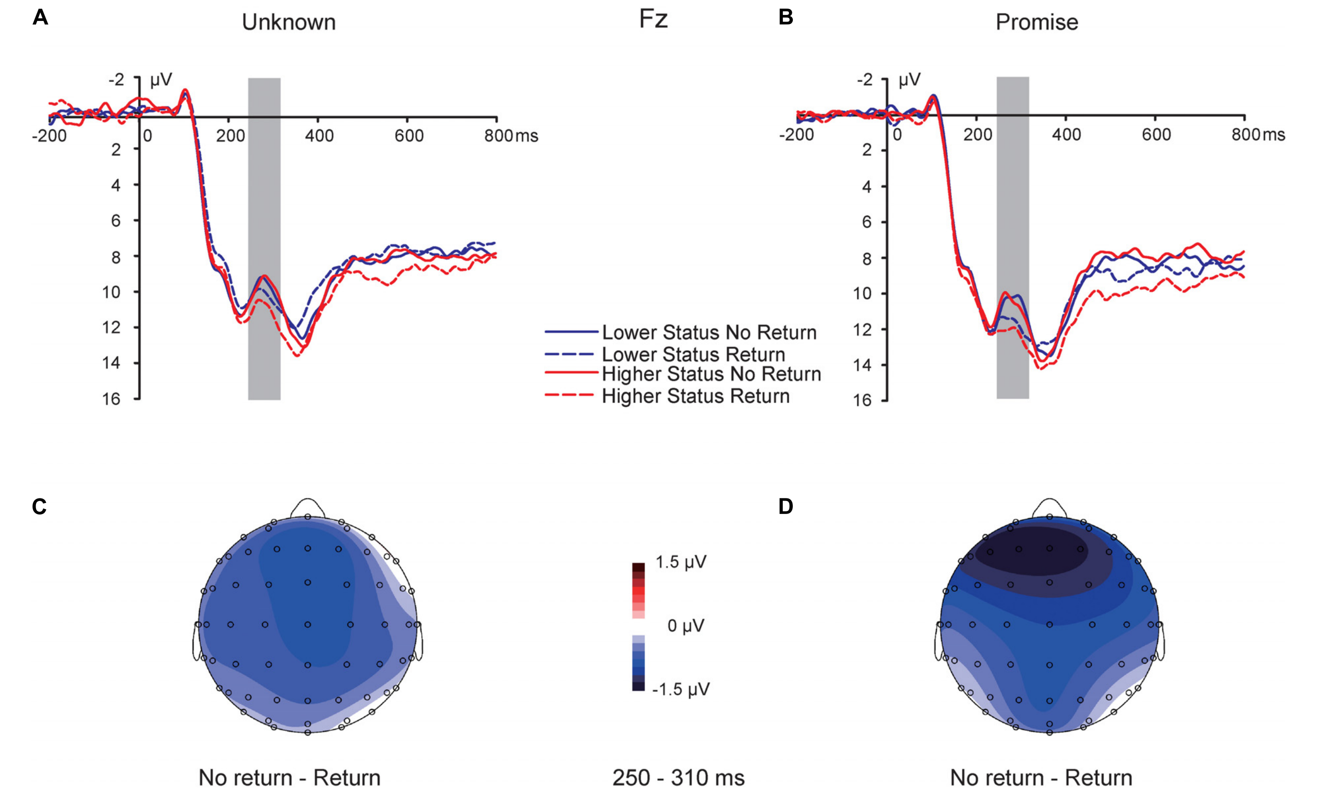

FIGURE 3 | Medial frontal negativity (MFN). ERP response time-locked to the onset of the TG return feedback at the Fz electrode in the "unknown" **(A)** and "promise" **(B)** conditions. Topographic maps for the MFN effect in the 250–310 ms time window in the "unknown" **(C)** and "promise" **(D)** conditions.

status "return" feedback (14.731 ± 0.833 µV) than to lower status "return" feedback (13.934 ± 0.798 µV), p = 0.001, while there was no difference in P300 amplitudes in response to "no return" feedback between higher status partners (13.972 ± 0.752 µV) and lower status partners (13.955 ± 0.796 µV), p = 0.949. If we restrict our analysis to the peak values on the electrode CPz, and enter these peak values into a 2 (partner social status: lower vs. higher) × 2 (promise condition: promise vs. unknown) × 2 (return: yes vs. no) repeated-measures ANOVA, the same pattern of effects was obtained, with the exception that there was a non-significant trend or tendency for a main effect of status, $F(1,32) = 3.763$, $p = 0.061$, $\eta_p^2 = 0.105$.

Regarding the influence of the electrode location, there was a main effect of hemisphere, $F(2,64) = 115.938$, $p < 0.001$, $\eta_p^2 = 0.784$, with P300 amplitudes being most positive-going in the medial hemisphere (16.866 ± 0.927 µV) than in the left (12.409 ± 0.721 µV) and right (13.170 ± 0.721 µV) hemispheres, $ps < 0.001$. P300 amplitudes were also more positive-going in the right hemisphere than in the left hemisphere, $p = 0.017$. There was a main effect of region, $F(2,64) = 15.891$, $p < 0.001$, $\eta_p^2 = 0.332$, with P300 amplitudes being more positive-going in the central region (16.006 ± 0.896 µV) than in the anterior (13.600 ± 0.907 µV) and posterior (12.839 ± 0.722 µV) regions, $ps < 0.001$. Importantly, there was a significant interaction between partner social status and region, $F(2,64) = 3.764$, $p = 0.048$, $\eta_p^2 = 0.105$, such that higher status feedback elicited a more positive-going P300 than lower status feedback in the anterior region (lower status: 13.322 ± 0.907 µV; higher status: 13.878 ± 0.919 µV, $p = 0.015$) and the central region (lower status: 15.754 ± 0.897 µV; higher status: 16.257 ± 0.906 µV, $p = 0.011$), whereas in the posterior region, there was no difference in P300 amplitude for lower and higher status feedback (lower status: 12.758 ± 0.739 µV; higher status: 12.920 ± 0.711 µV, $p = 0.273$). There was also an interaction between promise and region, $F(2,64) = 7.814$, $p = 0.004$, $\eta_p^2 = 0.196$, such that, in the anterior region, feedback in the "promise" condition elicited a more positive-going P300 amplitude (13.986 ± 0.942 µV) than feedback in the "unknown" condition (13.214 ± 0.899 µV), $p = 0.022$, whereas there was no difference in P300 amplitudes for "promise" and "unknown" conditions in the central region ("unknown": 15.815 ± 0.899 µV; "promise": 16.196 ± 0.922 µV, $p = 0.239$) or the posterior region ("unknown": 12.787 ± 0.732 µV; "promise": 12.891 ± 0.731 µV, $p = 0.665$).

There was a significant interaction between promise, return, and hemisphere, $F(2,64) = 4.140$, $p = 0.033$, $\eta_p^2 = 0.115$. In the "unknown" condition, there was a significant interaction between return and hemisphere, $F(2,64) = 5.314$, $p = 0.009$, $\eta_p^2 = 0.142$, whereas in the "promise" condition this interaction was not significant, $p = 0.974$. Tests for simple effects showed that, in the "unknown" condition, feedback indicating "no return" elicited more positive-going P300 amplitudes in

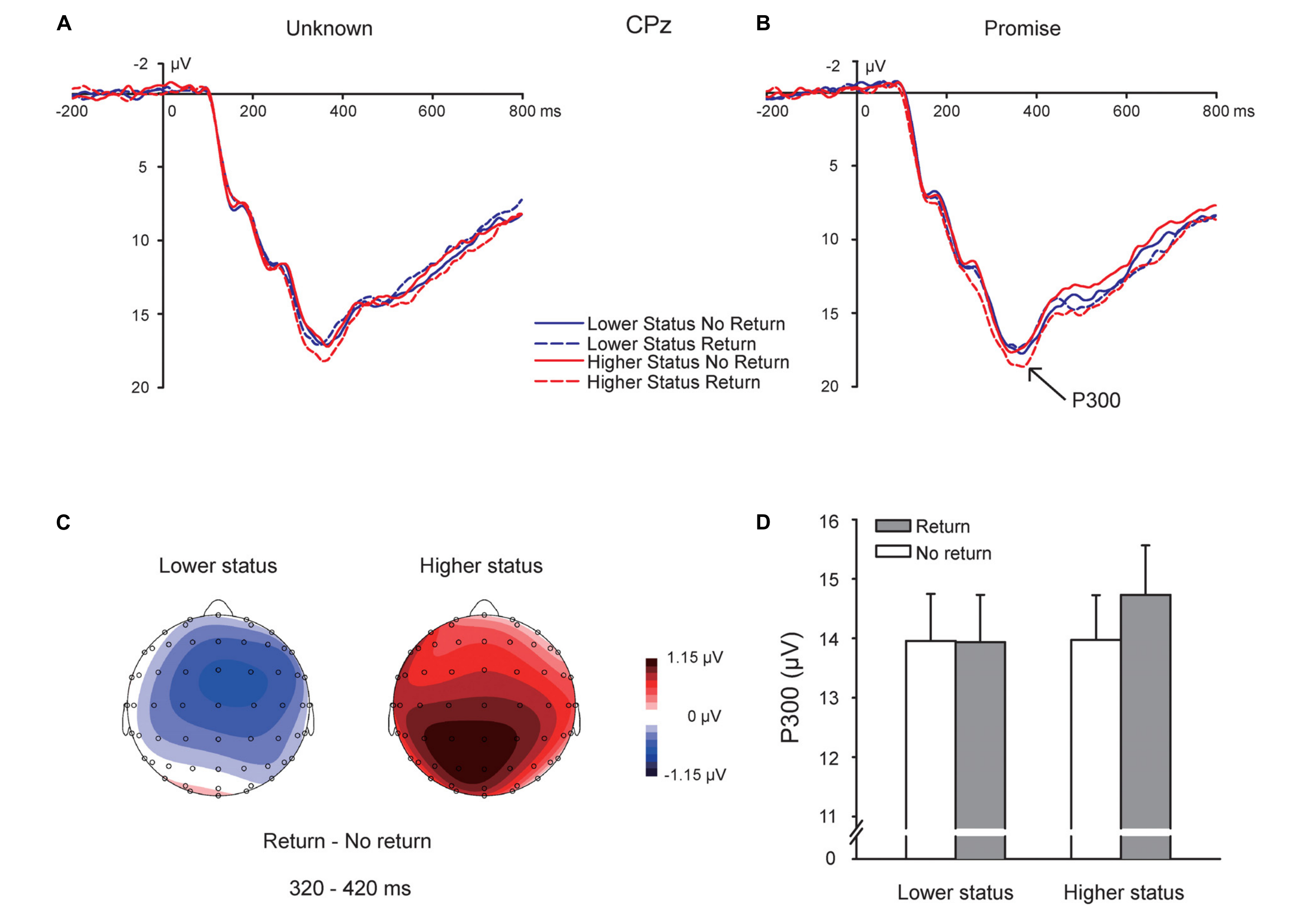

FIGURE 4 | P300. ERP response time-locked to the onset of the TG return feedback at the CPz electrode in the "unknown" **(A)** and "promise" **(B)** conditions. Topographic maps for the P300 effect in the 320–420 ms time window **(C)**. Bar plot depicting the whole-brain average peak values for P300 responses to "return" and "no return" feedback for lower status and higher status partners **(D)**.

the medial hemisphere (16.399 ± 0.902 µV) than in the left hemisphere (11.988 ± 0.699 µV) and right hemisphere (12.956 ± 0.699 µV), *ps* < 0.001; moreover, feedback indicating "no return" also elicited more positive-going P300 amplitudes in the right hemisphere (12.956 ± 0.699 µV) than in the left hemisphere (11.988 ± 0.699 µV), $p = 0.012$. Similarly, feedback indicating "return" elicited more positive-going P300 amplitudes in the medial hemisphere (16.806 ± 0.988 µV) than in the left hemisphere (12.508 ± 0.779 µV) and right hemisphere (12.974 ± 0.787 µV), *ps* < 0.001. There was no difference in P300 amplitudes in the left hemisphere and right hemisphere, $p = 0.583$.

To test whether P300 responses were modulated by investment behavior, we compared peak P300 responses from the medial posterior cluster electrodes time-locked to the TG feedback. In particular, we compared P300 peak amplitudes on only those trials in which participants invested (i.e., invest-only trials) with P300 peak amplitudes on all trials regardless of investment (i.e., all trials). Similar to the MFN analysis, we only analyzed trials from the "promise" condition. There were too few no-invest trials in both the "promise" and the "unknown" conditions to conduct a meaningful comparison between invest-only trials and no-invest trials, thus we compared invest-only trials with all trials. After removing "promise" condition trials in which the participant did not invest, 7 participants had less than 15 trials per condition. These 7 participants were removed from this supplementary analysis, leaving 26 participants in the analysis. A 2 (invest: yes vs. all trials) × 2 (partner social status: lower vs. higher) × 2 (return: return vs. no return) repeated-measures ANOVA revealed a main effect of invest, $F(1,25) = 10.847$, $p = 0.003$, $\eta_p^2 = 0.303$, indicating that participants evidenced more positive-going P300 when receiving feedback on invest-only trials (14.856 ± 0.949 µV) than on invest and no invest trials combined (14.436 ± 0.916 µV). No other effects reach significance.

Objective SES

Objective SES was measured using parents' highest attained level of education and parents' combined annual salary. Parents' highest level of education, ($M = 2.758$, $SE = 0.200$) on average ranged from high school diploma/middle trade school certificate to trade school certificate (a level slightly lower than a bachelor's

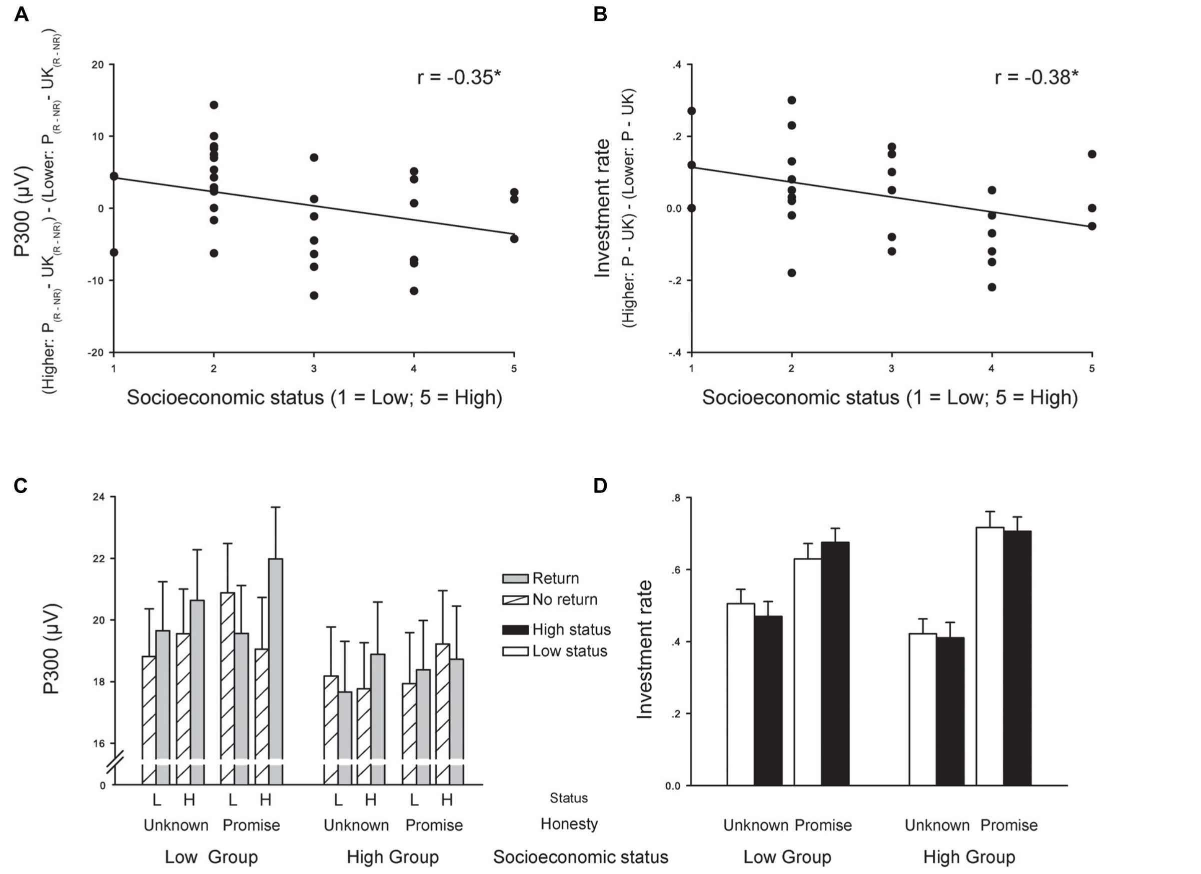

FIGURE 5 | Effects of Objective SES (i.e., parents' highest achieved level of education) on TG behavior and ERP response (P300 peak amplitudes in the medial central cluster time-locked to the TG feedback). **(A)** Correlation between Objective SES and the interaction between partner social status, promise, and return on P300 peak amplitudes. **(B)** Correlation between Objective SES and the interaction between partner social status and promise on investment behavior in TG. **(C)** Objective SES split-group analysis of P300 peak amplitudes plotted as a function of partner social status, promise, and return. **(D)** Objective SES split-group analysis of investment behavior plotted as a function of partner social status and promise. Groups based on median split: Low Objective SES group (n = 17) and High Objective SES group (n = 16). "P" indicates Promise; "UK" indicates Unknown; "L" indicates Lower Status Trustee; "H" indicates Higher Status Trustee.

degree); parents' average annual salary (M = 2.533, SE = 0.115) on average ranged from 10,000 yuan to a little over 100,000 yuan per year (i.e., ∼$1,500 – $20,000). Note that due to concerns over privacy, three participants did not report their parents' annual income; these participants were, however, willing to report their parents' highest level of education (n = 33). Objective SES based on parents' salary was positively correlated with Objective SES based on parents' highest attained level of education, $r(31)$ = 0.450, p = 0.012. Past research on Objective SES recommends the use of parents' highest attained level of education over parents' annual income as an index of student Objective SES (Rosenberg, 1965), given that, in comparison to salary levels, education levels tend to be better predictors of other social factors such as self-esteem (Twenge and Campbell, 2002). As a result, we used parents' highest attained level of education as our index of participants' Objective SES below.

Objective SES and TG Behavior

There was a negative correlation between Objective SES and the TG behavior interaction between partner social status and promise conditions [i.e., (Higher status "promise" – Higher status "unknown") – (Lower status "promise" – Lower status "unknown")], $r(31)$ = −0.381, p = 0.029 (**Figure 5B**). To better understand the effect of Objective SES on behavior, we used a median split (median = 2) to divide Objective SES into two groups: low Objective SES [education level of 1 and 2 (middle school – high school diploma), n = 17] and high Objective SES [education level of 3, 4, and 5 (trade school – graduate degree), n = 16]. We then entered Objective SES group (low vs. high) as a between-subjects factor along with two within-subjects factors [partner social status (lower vs. higher) and promise condition ("unknown" vs. "promise)] into a repeated-measures ANOVA. Adding Objective SES group as a between-subjects factor did not change the pattern of results described above (see *Behavioral Results*). There was a significant main effect of promise condition, $F(1,31)$ = 59.132, $p < 0.001$, η_p^2 = 0.656, with participants investing more in the "promise" condition (0.682 ± 0.027) than in the "unknown" condition (0.452 ± 0.026). There was a non-significant trend or tendency for an interaction between partner social status and promise condition, $F(1,31) = 3.863$, $p = 0.058$, $\eta^2_p = 0.111$. The pattern was the same as the pattern described above (see *Behavioral Results*). The interaction between Objective SES group and promise condition was significant, $F(1,31) = 4.784$, $p < 0.036$, $\eta^2_p = 0.134$. Further tests showed that in the low Objective SES group, the difference between investment in the "unknown" condition (0.487 ± 0.042) and the "promise" condition (0.652 ± 0.043) was smaller ($\eta^2_p = 0.562$), than in the high Objective SES group ("unknown" = 0.416 ± 0.031, "promise" condition = 0.711 ± 0.032, $\eta^2_p = 0.716$).

Importantly, and in line with the negative correlation between Objective SES and the TG behavior interaction between partner social status and promise conditions, there was a non-significant trend or tendency for an interaction between Objective SES group, partner social status, and promise condition, $F(1,31) = 3.667$, $p = 0.065$, $\eta^2_p = 0.106$ (**Figure 5D**). Further tests showed that the interaction between partner social status and promise condition was only significant in the low Objective SES group, $F(1,16) = 8.302$, $p = 0.011$, $\eta^2_p = 0.342$, such that in the "promise" condition, low Objective SES participants invested more in higher status (0.675 ± 0.041) than lower status partners (0.629 ± 0.048), $p = 0.045$; in the "unknown" condition, there was no significant difference between investment in higher status (0.470 ± 0.045) and lower status partners (0.505 ± 0.045), $p = 0.307$. The interaction between partner social status and honesty condition was not significant in the high Objective SES group, $F < 1$, $p = 0.973$.

Objective SES and MFN in the 250–310 ms Time Window Following TG Feedback

There was no correlation between Objective SES and the interaction between partner social status, promise condition, and return on the MFN in the 250–310 ms time window following TG feedback, $p = 0.208$. We do not report further analysis on the effects of Objective SES on MFN.

Objective SES and P300 Peak Amplitudes Time-Locked to the TG Feedback

There was a negative correlation between Objective SES and the interaction between partner social status, promise condition, and return on the P300 peak amplitudes in the medial central cluster time-locked to the TG feedback, $r(31) = -0.345$, $p = 0.049$ (**Figure 5A**). We chose the medial central cluster because the P300 responses were largest on these electrodes. There is also a negative correlation if we test the correlation between Objective SES and the interaction between partner social status, promise condition, and return of the P300 peak amplitudes on the CPz electrode, $r(31) = -0.358$, $p = 0.041$. Similar to the analysis of Objective SES and TG behavior, to better understand the effect of Objective SES on average P300 peak amplitudes in the medial central cluster, we conducted a median split of Objective SES and entered Objective SES group (low vs. high) as a between-subjects factor along with three within-subjects factors [partner social status (lower vs. higher) and promise condition ("unknown" vs. "promise) and return (return vs. no return)] into a repeated-measures ANOVA. The pattern of results with Objective SES included as a between-participants factor are the same as those described above (see *P300 Following TG Feedback*). There was a significant main effect of partner social status, $F(1,31) = 5.621$, $p = 0.024$, $\eta^2_p = 0.153$, with higher P300 amplitudes for higher status partners (19.475 ± 1.095) than for lower status partners (18.882 ± 1.070). There was also a significant interaction between partner social status and return $F(1,31) = 7.284$, $p = 0.011$, $\eta^2_p = 0.190$. The pattern was the same as the pattern described above (see *P300 Following TG Feedback*).

There was a significant interaction between partner social status, promise condition, return, and Objective SES, $F(1,31) = 11.086$, $p = 0.002$, $\eta^2_p = 0.263$ (**Figure 5C**). Further tests showed that the interaction between partner social status, promise condition, and return was only significant in the low Objective SES group, $F(1,16) = 9.351$, $p = 0.008$, $\eta^2_p = 0.369$. In the high Objective SES group, the interaction between partner social status, promise condition, and return was not significant, $p = 0.104$. Further tests on the interaction in the low Objective SES group showed that in the "promise" condition, the interaction between partner social status and return was significant, $F(1,16) = 13.050$, $p = 0.002$, $\eta^2_p = 0.449$: for higher status partners in the "promise" condition, P300 peak amplitudes were more positive-going in response to "return" feedback (21.984 ± 1.643) than "no return" feedback (19.049 ± 1.410), $p = 0.001$, whereas for lower status partners in the "promise" condition, there was no difference in P300 peak amplitudes in response to "return" feedback (19.559 ± 1.676) and "no return" feedback (20.881 ± 1.570), $p = 0.188$. In the "unknown" condition, the interaction between partner social status and return was not significant, $p = 0.753$. The pattern of the interaction between Objective SES group, partner social status, promise condition, and return is the same if we limit our analysis to ANOVA on the P300 peak amplitudes on the CPz electrode, $F(1,31) = 11.836$, $p = 0.002$, $\eta^2_p = 0.276$.

DISCUSSION

In the current study, we used a modified version of TG to investigate whether and how social status influences evaluation of honesty-related feedback. At the behavioral level, participants tended to be more affected by promises given by higher status Trustees than lower status Trustees, despite receiving equal feedback about lower and higher status honesty. At the neural level, when viewing TG partner feedback, MFN in the time window of 250-310 ms was more negative-going when TG partners did not return than when they did return. P300 peak amplitudes differentiated higher status return feedback (i.e., return vs. no return), but did not do so for lower status partner return feedback; moreover, P300 responses were more positive-going for higher status partner returns than for lower status partner returns. Finally, participants in low Objective SES evidenced a greater P300 effect for higher status honesty than for lower status honesty and evidenced a tendency for investing more in promises given by higher status Trustees than lower status Trustees; neither of these effects were found in participants in high Objective SES. Taken together, these findings demonstrate that social status can modulate both the behavioral responses to and the neural processing of honesty-related feedback, and suggest that higher status honesty may be perceived as more motivationally salient or rewarding than lower status honesty in individuals with low Objective SES.

Behavior

Despite the fact that participants viewed identical feedback across conditions (i.e., 50% return; 50% no return), participants invested substantially more in partners when promises were made to return at least half of the multiplied sum than when partners were not given the option to make a promise. Moreover, participants tended to be more affected by promises given by higher status than low status partners in TG. In particular, both lower and higher status promises increased the amount the participants invested in TG, in comparison with trials where promises were not available; however, higher status promises tended to increase investment to a greater extent. These behavioral findings provide support for the "social value" hypothesis, which predicts that participants would be more affected by promises given by higher status Trustees than by lower status Trustees. In contrast, the "expectation violation" hypothesis predicts that participants would be more surprised by higher status than lower status dishonesty and would thus invest less in higher status promises than lower status promises over time. No support was found for this hypothesis, and there was no evidence showing that participants' investment behavior changed over time.

While the behavioral pattern found above suggests that higher status increases the influence of promises on investment behavior, this effect is relatively weak. Similar research measuring participants' investment behavior in Trustees of lower and higher status in iterated one-shot TG found that participants invest significantly more in promises given by higher status Trustees than in promises given by lower status Trustees (Blue et al., under review). We suspect that the weakness of this effect in the current study was due to the feedback concerning the Trustee's return behavior. In Blue et al. (under review), no feedback was given concerning whether the Trustee actually kept the promise, whereas here the participants roughly knew that the Trustee broke the promise in about half of the trials. Thus, the surprising finding was that even in such harsh conditions encouraging distrust, the participants still tended to trust the high status Trustee more than the low status Trustee, a pattern replicating Blue et al. (under review).

MFN Effects on Outcome Feedback Evaluation

MFN responses were sensitive to outcome valence, as MFN responses were more negative-going for "no return" than for "return" feedback. This effect reinforces the notion that MFN encodes social expectancy violation, as not returning part of the multiplied sum is a violation of the trustworthiness norm. This trustworthiness norm in TG refers to the Investors' tendency to send around 50% of the possible amount of money to Trustees (Berg et al., 1995; Johnson and Mislin, 2011), despite the unique Nash equilibrium prediction that the Investor, as a rational and self-interested agent, should transfer no money to the Trustee, given that a rational Investor should assume that the Trustee would act in a completely self-interested way (i.e., return none of the multiplied sum to the Investor). Thus, a large portion of Investors in TG expect Trustees to reciprocate their trusting behavior by acting in a trustworthy manner, and not returning may be interpreted as a violation of the trustworthiness expectation.

Interestingly, partner social status and the promise condition did not interact to influence MFN responses to outcome feedback, especially given that previous research shows that promise information modulates responses to TG outcome feedback (Ma et al., 2015). This may have been due to the forced feedback nature of the current experiment, as participants in Ma et al. (2015) were given feedback only if they invested in the Trustee, whereas in the current study, participants were given forced feedback. Indeed, we did find that the investment decision modulated MFN responses: if the analysis of MFN in the "promise" condition is restricted to invest-only trials, the MFN effect is even more pronounced than when the analysis includes all feedback, regardless of the investment decision.

There was also a main effect of promise condition on MFN responses to TG outcome feedback, as MFN responses were more negative-going for feedback in the "unknown" condition than in the "promise" condition. This main effect is most likely due to differences in investment behavior in the two conditions. Participants invested less in the "unknown" condition (investment rate = 45%) than in the "promise" condition (investment rate = 68%). Given that past research shows that distrust decisions elicit greater MFN responses to outcome feedback, in comparison with trust decisions (Long et al., 2012), and that investment behavior modulated MFN responses to TG outcome feedback, it is most likely that this main effect is driven by the participants decreased investment in the "unknown" condition than in the "promise" condition.

Finally, "return" and "no return" decisions were more directly tied to financial payoffs than promise and partner social status information, which could make this information more salient in early MFN outcome evaluation processing. This is in line with past research showing that, in social contexts, MFN encodes stimuli that are most directly tied to financial payoff, whereas social factors are left for later processing (e.g., P300 or LPP; Leng and Zhou, 2010; Wu et al., 2011b). This could mean that the outcome evaluation system may defer to a later, more top–down stage of processing to appraise the honesty-related outcomes in the context of social status, which would suggest that the outcome evaluation in TG may be composed of earlier, semi-automatic processing which is coarse in nature and provides discrete evaluations of return feedback regardless of its relation to honesty or social status, and later top–down controlled processing of outcome evaluation, where factors such as honesty and social status can undergo higher level cognitive appraisal.

P300 Effects in Feedback Evaluation

In contrast to MFN, P300 responses to TG feedback were sensitive to the interaction between social status and return decisions. In particular, P300 responses differentiated return and no return feedback only for higher status Trustees. Moreover, higher status return feedback elicited greater P300 amplitudes than lower status return feedback. Given that P300 activity reflects affective/motivational significance (Nieuwenhuis et al., 2005; Leng and Zhou, 2010) and/or distribution of attention resources (Gray et al., 2004; Linden, 2005), the findings from the current study could suggest that higher status returns were more motivationally salient to participants than were lower status returns. Higher and lower status return likelihood and amounts were identical, which means that the increased P300 response to "return" outcomes from higher status Trustees than lower status Trustees could reflect increased perceived value or relevance of higher status "return" feedback, especially given that processing social status information is directly tied to reward-related processing in both human (Ly et al., 2011) and non-human primates (Deaner et al., 2005). Indeed, previous research using TG shows that Trustee characteristics, such as personal closeness to the Investor (i.e., "social value;" Fareri et al., 2015), modulate neural responses to Trustee feedback in brain areas related to reward processing, such as ventral striatum. Ventral striatum activity is also greater in responses to outcomes that result in reward which is shared with friends than reward which is shared with strangers (Fareri et al., 2012), suggesting that outcome evaluation is susceptible to influence of social reward. Research simultaneously measuring fMRI and EEG show that, during the anticipation of monetary gain, ventral striatum and P300 activity are positively correlated, suggesting that these two neural responses may be involved in similar motivational processing of reward-related stimuli (Pfabigan et al., 2014). Additionally, both ventral striatal and P300 activity are impaired in patients diagnosed with schizophrenia, and these impairments have both been shown to be associated with deficits in reward processing (Juckel et al., 2006; Vignapiano et al., 2016). Taken together, the P300 findings from the current study suggest that social status influences the motivational significance and/or attentional resources devoted to TG outcome feedback and that this modulation may reflect differences in perceived value of lower and higher status "return" outcomes in TG.

It is interesting that we did not find an interaction between partner social status, return, and promise information on P300 responses to TG feedback. We suspect that this may be due to individual differences in participant SES (Ly et al., 2011). Indeed, only participants in low Objective SES showed the expected interaction between partner social status, return, and promise information on P300 responses to TG feedback. In these participants, social status modulation of P300 responses to TG feedback was restricted to the "promise" condition, such that P300 responses were greater for higher status "return" than "no return" outcomes, whereas P300 responses were less sensitive to lower status promise feedback. These findings could suggest that, for participants in low Objective SES, higher status partner honesty feedback may be perceived as more motivationally salient and/or elicit greater attention allocation than lower status partner honesty feedback.

One potential explanation for these findings is that lower status individuals have the most to gain from high status cooperation in a social hierarchy (Cummins, 1996), and keeping promises is considered a sign of cooperation. This would be in line with the "social value" account and would suggest that one possible explanation for the increased P300 response to lower status than higher status honesty in participants with low Objective SES is that these individuals may value higher status honesty more than lower status honesty. Additionally, participants with low Objective SES evidenced a tendency for investing more in higher status than lower status promises, despite the equal reinforcement schedule. This behavioral finding provides further support for the "social value" account, as participants in low Objective SES may have believed they had more to gain by investing in higher status than lower status promises. Taken together, the behavioral and neural findings for low Objective SES participants could suggest that these individuals perceive higher status promises and honesty as being more valuable than that of their lower status counterparts. Given that we did not manipulate feelings of SES and given the non-significant trend or tendency of the SES behavioral interaction in the current study, future research could directly address whether changes in SES feelings (e.g., Subjective SES) could replicate the effects found in the current study and could provide more support for a causal explanation of SES in the current study.

In contrast to participants in low Objective SES, participants in high Objective SES evidenced no effects of partner social status on P300 responses or behavior. P300 responses were greater in the low Objective SES group than in the high Objective SES group, regardless of the condition, which could suggest that high Objective SES may have been associated with decreased attention to others' social information, in general. Past research shows that individuals in high SES are less attentive to others' information than individuals in low SES (Muscatell et al., 2012; Dietze and Knowles, 2016) and that high status individuals are more selective in their attention allocation (Shepherd et al., 2006). Attention-based differences between low and high SES individuals may be driven by different cognitive tendencies. Low SES individuals tend to have contextualist cognitive tendencies, whereas high SES individuals tend to have more individualistic cognitive tendencies and increased concern for goals and reward related to the self (Kraus et al., 2012). High status individuals are less reliant on others to achieve their goals, whereas low status individuals are more likely to help high status than fellow low status others, as the former is more valuable for attaining resources and protection in the future (Trivers, 1971; de Waal, 1989; Silk, 1992; Cummins, 1996, 2006; Stevens et al., 2005). Taken together, our findings regarding the effects of individual differences in Objective SES are in line with previous research, and suggest that, while high Objective SES participants are less concerned with others' social status and honesty behavior, low Objective SES participants are especially attuned to higher status others' honesty, an effect which is tied to increased investment likelihood in higher status than lower status promises.

A few points are worth mentioning. The social status manipulation (i.e., math quiz ranking) could, in fact, have influenced the way the participant viewed the lower and higher-ranking players in ways other than the prestige-based social status referred to in this study. (1) For example, despite the fact that we did not manipulate SES, participants did perceive higher ranking participants as having higher Subjective SES than that of their lower ranking counterparts. While this difference in perceived Subjective SES did not correlate with investment differences for lower and higher status partners in TG ($p = 0.969$), future research should look to manipulate partner SES while controlling for prestige-based social status to more directly address the unique effects of perceived SES on perceived trust. (2) Another possible explanation for the effect of social status on investment behavior may have been that participants may have inferred that higher status partners were happier than lower status partners after achieving their ranking (Hu et al., 2014), which could have increased perceived warmth and trustworthiness (Fiske et al., 2007). Despite the plausibility of this possibility, in the current study, participants evidenced a non-significant trend or tendency for perceiving higher status partners as *less* benevolent than their lower status counterparts, which is in line with past research (Dunn et al., 2012; Lount and Pettit, 2012), but does not support this alternative account. Moreover, status differences in perceived benevolence did not predict the TG behavior interaction between partner social status and promise conditions, and so we do not discuss it further. (3) Finally, another possible explanation is that high status partners were perceived as having put in more effort to the experiment, which could have increased their perceived trustworthiness. While we cannot rule this possibility out, it is important to note that the design of the experiment (only permitting 10 s per math question) rules out large differences in perceived effort. Taken together, the current study appears to be the start of a broader inquiry regarding the effects of social status on perceived trustworthiness.

CONCLUSION

To conclude, by manipulating prestige-based social status, this study found that participants acting as Investors in TG tended to be more affected by higher status Trustee promises than by lower status Trustee promises, despite the equal reinforcement schedule across conditions. At the neural level, in an early time window (250–310 ms), MFN responses were sensitive to return outcome, as MFN amplitudes were more negative when partners did not return than when they did return. This effect was not modulated by the Trustee's social status. In later processing, P300 responses *were* modulated by social status and return. P300 amplitudes were only sensitive to return feedback from higher status partners, and failed to distinguish lower status partner return feedback; moreover, P300 responses were more positive for higher status returns than lower status returns, which suggests that higher status positive feedback may have been perceived as more motivationally significant or rewarding than lower status positive feedback. The current study also found that the lower the participants' Objective SES, the greater their differential P300 effect for higher status over lower status honesty and the more they invested in higher status than lower status promises, suggesting that individual differences in SES affect the perceived motivational salience/reward effect of social status on honesty. Taken together, we find that social status influences the effect of promises on investment behavior in TG, and that brain responses to honesty-related feedback in social hierarchies may involve both an early MFN processing of trustworthiness outcome valence information and a later P300 cognitive appraisal process which takes into account both social status and honesty and its relation to reward.

AUTHOR CONTRIBUTIONS

PB, JH, and XZ designed the experiments and wrote the manuscript. PB and JH collected the data.

FUNDING

This work was supported by National Basic Research Program (973 Program: 2015CB856400) from the Ministry of Science and Technology of China and by grants from National Natural Science Foundation of China (91232708 and 31170972).

ACKNOWLEDGMENTS

We would like to thank Ms. Yue Li for her help during data collection.

REFERENCES

Adler, N. E., Epel, E. S., Castellazzo, G., and Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: preliminary data in healthy white women. *Health Psychol.* 19, 586–592. doi: 10.1037/0278-6133.19.6.586

Albrecht, K., von Essen, E., Fliessbach, K., and Falk, A. (2013). The influence of status on satisfaction with relative rewards. *Front. Psychol.* 4:804. doi: 10.3389/fpsyg.2013.00804

Alvarez, R. (1968). Informal reactions to deviance in simulated work organizations: a laboratory experiment. *Am. Soc. Rev.* 33, 895–912. doi: 10.2307/2092682

Aquino, K., Tripp, T. M., and Bies, R. J. (2006). Getting even or moving on? Power, procedural justice, and types of offense as predictors of revenge, forgiveness, reconciliation, and avoidance in organizations. *J. Appl. Psychol.* 91:653–668. doi: 10.1037/0021-9010.91.3.653

Bell, R., Sasse, J., Möller, M., Czernochowski, D., Mayr, S., and Buchner, A. (2016). Event-related potentials in response to cheating and cooperation in a social dilemma game. *Psychophysiology* 53, 216–228. doi: 10.1111/psyp.12561

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games Econ. Behav.* 10, 122–142. doi: 10.1006/game.1995.1027

Blader, S. L., and Chen, Y.-R. (2012). Differentiating the effects of status and power: a justice perspective. *J. Pers. Soc. Psychol.* 102, 994–1014. doi: 10.1037/a0026651

Blader, S. L., Shirako, A., and Chen, Y. R. (2016). Looking out from the top: differential effects of status and power on perspective taking. *Pers. Soc. Psychol. Bull.* 42, 723–737. doi: 10.1177/0146167216636628

Blue, P. R., Hu, J., Wang, X., van Dijk, E., and Zhou, X. (2016). When do low status individuals accept less? The interaction between self- and other-status during resource distribution. *Front. Psychol.* 7:1667. doi: 10.3389/fpsyg.2016.01667

Boksem, M. A., and De Cremer, D. (2010). Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Soc. Neurosci.* 5, 118–128. doi: 10.1080/17470910903202666

Born, A., van Eck, P., and Johannesson, M. (2017). An experimental investigation of election promises. *Polit. Psychol.* doi: 10.1111/pops.12429 [Epub ahead of print].

Boudreau, C., McCubbins, M. D., and Coulson, S. (2009). Knowing when to trust others: an ERP study of decision making after receiving information from unknown people. *Soc. Cogn. Affect. Neurosci.* 4, 23–34. doi: 10.1093/scan/nsn034

Bowles, H. R., and Gelfand, M. (2010). Status and the evaluation of workplace deviance. *Psychol. Sci.* 21, 49–54. doi: 10.1177/0956797609356509

Charness, G., and Dufwenberg, M. (2006). Promises and partnership. *Econometrica* 74, 1579–1601. doi: 10.1111/j.1468-0262.2006.00719.x

Charness, G., and Dufwenberg, M. (2010). Bare promises: an experiment. *Econ. Lett.* 107, 281–283. doi: 10.1016/j.econlet.2010.02.009

Corazzini, L., Kube, S., Maréchal, M. A., and Nicolò, A. (2014). Elections and deceptions: an experimental study on the behavioral effects of democracy. *Am. J. Polit. Sci.* 58, 579–592. doi: 10.1111/ajps.12078

Cummins, D. D. (1996). Dominance hierarchies and the evolution of human reasoning. *Minds Mach.* 6, 463–480. doi: 10.1007/BF00389654

Cummins, D. D. (1999). Cheater detection is modified by social rank. *Evol. Hum. Behav.* 20, 229–248. doi: 10.1016/S1090-5138(99)00008-2

Cummins, D. D. (2006). "Dominance, status, and social hierarchies," in *The Handbook of Evolutionary Psychology*, ed. D. M. Buss (Hoboken, NJ: Wiley), 676–697.

de Waal, F. B. (1989). Food sharing and reciprocal obligations among chimpanzees. *J. Hum. Evol.* 18, 433–459. doi: 10.1016/0047-2484(89)90074-2

Deaner, R. O., Khera, A. V., and Platt, M. L. (2005). Monkeys pay per view: adaptive valuation of social images by rhesus macaques. *Curr. Biol.* 15, 543–548. doi: 10.1016/j.cub.2005.01.044

Delgado, M. R., Frank, R. H., and Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat. Neurosci.* 8, 1611–1618. doi: 10.1038/nn1575

Dépret, E., and Fiske, S. T. (1993). "Social cognition and power: some cognitive consequences of social structure as a source of control deprivation," in *Control, Motivation, and Social Cognition*, eds G. Weary, F. Gleicher, and K. L. Marsh (New York, NY: Springer), 176–202.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Front. Psychol.* 5:781. doi: 10.3389/fpsyg.2014.00781

Dietze, P., and Knowles, E. D. (2016). Social class and the motivational relevance of other human beings: evidence from visual attention. *Psychol. Sci.* 27, 1517–1527. doi: 10.1177/0956797616667721

Donchin, E., and Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behav. Brain Sci.* 11, 357–427. doi: 10.1017/S0140525X00058027

Dubois, D., Rucker, D. D., and Galinsky, A. D. (2015). Social class, power, and selfishness: when and why upper and lower class individuals behave unethically. *J. Pers. Soc. Psychol.* 108, 436–449. doi: 10.1037/pspi0000008

Duncan-Johnson, C. C., and Donchin, E. (1977). On quantifying surprise: the variation of event-related potentials with subjective probability. *Psychophysiology* 14, 456–467. doi: 10.1111/j.1469-8986.1977.tb01312.x

Dunn, J., Ruedy, N. E., and Schweitzer, M. E. (2012). It hurts both ways: how social comparisons harm affective and cognitive trust. *Organ. Behav. Hum. Decis. Process.* 117, 2–14. doi: 10.1016/j.obhdp.2011.08.001

Fareri, D. S., Chang, L. J., and Delgado, M. R. (2015). Computational substrates of social value in interpersonal collaboration. *J. Neurosci.* 35, 8170–8180. doi: 10.1523/JNEUROSCI.4775-14.2015

Fareri, D. S., Niznikiewicz, M. A., Lee, V. K., and Delgado, M. R. (2012). Social network modulation of reward-related signals. *J. Neurosci.* 32, 9045–9052. doi: 10.1523/JNEUROSCI.0610-12.2012

Faul, F., Erdfelder, E., Lang, A.-G., and 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. doi: 10.3758/BF03193146

Fiddick, L., and Cummins, D. D. (2001). Reciprocity in ranked relationships: does social structure influence social reasoning? *J. Bioecon.* 3, 149–170. doi: 10.1023/A:1020572212265

Fiske, S. T. (2010a). Envy up, scorn down: how comparison divides us. *Am. Psychol.* 65:698. doi: 10.1037/0003-066X.65.8.698

Fiske, S. T. (2010b). "Interpersonal stratification: status, power, and subordination," in *Handbook of Social Psychology*, Vol. 2, eds S. F. Fiske, D. T. Gilbert, and G. Lindzey (Hoboken, NJ: John Wiley & Sons), 941–982. doi: 10.1002/9780470561119.socpsy002026

Fiske, S. T., Cuddy, A. J., and Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends Cogn. Sci.* 11, 77–83. doi: 10.1016/j.tics.2006.11.005

Fragale, A. R., Rosen, B., Xu, C., and Merideth, I. (2009). The higher they are, the harder they fall: the effects of wrongdoer status on observer punishment recommendations and intentionality attributions. *Organ. Behav. Hum. Decis. Process.* 108, 53–65. doi: 10.1016/j.obhdp.2008.05.002

Friedrich, D., and Southwood, N. (2011). "Promises and trust," in *Promises and Agreement: Philosophical Essays*, ed. H. Sheinman (Oxford: Oxford University Press), 277–294. doi: 10.1093/acprof:oso/9780195377958.003.0012

Galinsky, A. D., Gruenfeld, D. H., and Magee, J. C. (2003). From power to action. *J. Pers. Soc. Psychol.* 85, 453–466. doi: 10.1037/0022-3514.85.3.453

Gehring, W. J., and Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science* 295, 2279–2282. doi: 10.1126/science.1066893

Gray, H. M., Ambady, N., Lowenthal, W. T., and Deldin, P. (2004). P300 as an index of attention to self-relevant stimuli. *J. Exp. Soc. Psychol.* 40, 216–224. doi: 10.1016/S0022-1031(03)00092-1

Hajcak, G., Holroyd, C. B., Moser, J. S., and Simons, R. F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology* 42, 161–170. doi: 10.1111/j.1469-8986.2005.00278.x

Hajcak, G., Moser, J. S., Holroyd, C. B., and Simons, R. F. (2007). It's worse than you thought: the feedback negativity and violations of reward prediction in gambling tasks. *Psychophysiology* 44, 905–912. doi: 10.1111/j.1469-8986.2007.00567.x

Hamilton, V. L., and Sanders, J. (1981). The effect of roles and deeds on responsibility judgments: the normative structure of wrongdoing. *Soc. Psychol. Q.* 44, 237–254. doi: 10.2307/3033836

Haselhuhn, M. P., Kennedy, J. A., Kray, L. J., Van Zant, A. B., and Schweitzer, M. E. (2015). Gender differences in trust dynamics: women trust more than men following a trust violation. *J. Exp. Soc. Psychol.* 56, 104–109. doi: 10.1016/j.jesp.2014.09.007

Henrich, J., and Gil-White, F. J. (2001). The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evol. Hum. Behav.* 22, 165–196. doi: 10.1016/S1090-5138(00)00071-4

Hollander, E. P. (1958). Conformity, status, and idiosyncracy credit. *Psychol. Rev.* 65, 117–127. doi: 10.1037/h0042501

Holroyd, C. B., and Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* 109, 679–709. doi: 10.1037/0033-295X.109.4.679

Homans, G. C. (1961). *Social Behavior: Its Elementary Forms*. New York, NY: Harcourt Brace.

Hu, J., Blue, P. R., Yu, H., Gong, X., Xiang, Y., Jiang, C., et al. (2016). Social status modulates the neural response to unfairness. *Soc. Cogn. Affect. Neurosci.* 11, 1–10.

Hu, J., Cao, Y., Blue, P. R., and Zhou, X. (2014). Low social status decreases the neural salience of unfairness. *Front. Behav. Neurosci.* 8:402. doi: 10.3389/fnbeh.2014.00402

Jeffreys, H. (1939/1961). *The Theory of Probability*, 1st/3rd Edn. Oxford: Oxford University Press.

Jia, S., Li, H., Luo, Y., Chen, A., Wang, B., and Zhou, X. (2007). Detecting perceptual conflict by the feedback-related negativity in brain potentials. *Neuroreport* 18, 1385–1388. doi: 10.1097/WNR.0b013e3282c48a90

Johnson, G. B., and Ryu, S. R. (2010). Repudiating or rewarding neoliberalism? How broken campaign promises condition economic voting in Latin America. *Latin Am. Polit. Soc.* 52, 1–24. doi: 10.1111/j.1548-2456.2010.00096.x

Johnson, N. D., and Mislin, A. A. (2011). Trust games: a meta-analysis. *J. Econ. Psychol.* 32, 865–889. doi: 10.1016/j.joep.2011.05.007

Juckel, G., Schlagenhauf, F., Koslowski, M., Wüstenberg, T., Villringer, A., Knutson, B., et al. (2006). Dysfunction of ventral striatal reward prediction in schizophrenia. *Neuroimage* 29, 409–416. doi: 10.1016/j.neuroimage.2005.07.051

Keltner, D., Gruenfeld, D. H., and Anderson, C. (2003). Power, approach, and inhibition. *Psychol. Rev.* 110, 265–284. doi: 10.1037/0033-295X.110.2.265

Kraus, M. W., Piff, P. K., and Keltner, D. (2009). Social class, sense of control, and social explanation. *J. Pers. Soc. Psychol.* 97, 992–1004. doi: 10.1037/a0016357

Kraus, M. W., Piff, P. K., Mendoza-Denton, R., Rheinschmidt, M. L., and Keltner, D. (2012). Social class, solipsism, and contextualism: how the rich are different from the poor. *Psychol. Rev.* 119, 546–572. doi: 10.1037/a0028756

Leng, Y., and Zhou, X. (2010). Modulation of the brain activity in outcome evaluation by interpersonal relationship: an ERP study. *Neuropsychologia* 48, 448–455. doi: 10.1016/j.neuropsychologia.2009.10.002

Li, P., Jia, S., Feng, T., Liu, Q., Suo, T., and Li, H. (2010). The influence of the diffusion of responsibility effect on outcome evaluations: electrophysiological evidence from an ERP study. *Neuroimage* 52, 1727–1733. doi: 10.1016/j.neuroimage.2010.04.275

Linden, D. E. (2005). The P300: where in the brain is it produced and what does it tell us? *Neuroscientist* 11, 563–576. doi: 10.1177/1073858405280524

Long, Y., Jiang, X., and Zhou, X. (2012). To believe or not to believe: trust choice modulates brain responses in outcome evaluation. *Neuroscience* 200, 50–58. doi: 10.1016/j.neuroscience.2011.10.035

Lount, R. B., and Pettit, N. C. (2012). The social context of trust: the role of status. *Organ. Behav. Hum. Decis. Process.* 117, 15–23. doi: 10.1016/j.obhdp.2011.07.005

Ly, M., Haynes, M. R., Barter, J. W., Weinberger, D. R., and Zink, C. F. (2011). Subjective socioeconomic status predicts human ventral striatal responses to social status information. *Curr. Biol.* 21, 794–797. doi: 10.1016/j.cub.2011.03.050

Ma, Q., Meng, L., and Shen, Q. (2015). You have my word: reciprocity expectation modulates feedback-related negativity in the trust game. *PLoS One* 10:e0119129. doi: 10.1371/journal.pone.0119129

Ma, Q., Shen, Q., Xu, Q., Li, D., Shu, L., and Weber, B. (2011). Empathic responses to others' gains and losses: an electrophysiological investigation. *Neuroimage* 54, 2472–2480. doi: 10.1016/j.neuroimage.2010.10.045

Magee, J., and Galinsky, A. (2008). Social hierarchy: the self-reinforcing nature of power and status. *Acad. Manage. Ann.* 2, 351–398. doi: 10.1080/19416520802211628

Malhotra, D., and Murnighan, J. K. (2002). The effects of contracts on interpersonal trust. *Adm. Sci. Q.* 47, 534–559. doi: 10.2307/3094850

Mayer, R. C., and Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: a field quasi-experiment. *J. Appl. Psychol.* 84, 123–136. doi: 10.1037/0021-9010.84.1.123

Mayer, R. C., Davis, J. H., and Schoorman, F. D. (1995). An integrative model of organizational trust. *Acad. Manage. Rev.* 20, 709–734. doi: 10.5465/AMR.1995.9508080335

Morey, R., Rouder, J. N., and Jamil, T. (2015). *Package 'BayesFactor'*. Available at: http://bayesfactorpcl.r-forge.r-project.org/

Muscatell, K. A., Morelli, S. A., Falk, E. B., Way, B. M., Pfeifer, J. H., Galinsky, A. D., et al. (2012). Social status modulates neural activity in the mentalizing network. *Neuroimage* 60, 1771–1777. doi: 10.1016/j.neuroimage.2012.01.080

Nieuwenhuis, S., Aston-Jones, G., and Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychol. Bull.* 131, 510–532. doi: 10.1037/0033-2909.131.4.510

Oakes, J. M., and Rossi, P. H. (2003). The measurement of SES in health research: current practice and steps toward a new approach. *Soc. Sci. Med.* 56, 769–784. doi: 10.1016/S0277-9536(02)00073-4

Pfabigan, D. M., Seidel, E. M., Sladky, R., Hahn, A., Paul, K., Grahl, A., et al. (2014). P300 amplitude variation is related to ventral striatum BOLD response during gain and loss anticipation: an EEG and fMRI experiment. *Neuroimage* 96, 12–21. doi: 10.1016/j.neuroimage.2014.03.077

Phan, K. L., Sripada, C. S., Angstadt, M., and McCabe, K. (2010). Reputation for reciprocity engages the brain reward center. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13099–13104. doi: 10.1073/pnas.1008137107

Polman, E., Pettit, N. C., and Wiesenfeld, B. M. (2013). Effects of wrongdoer status on moral licensing. *J. Exp. Soc. Psychol.* 49, 614–623. doi: 10.1016/j.jesp.2013.03.012

Rosenberg, M. (1965). *Society and the Adolescent Self-Image*, Vol. 11. Princeton, NJ: Princeton University Press, 326. doi: 10.1515/9781400876136

Sato, A., Yasuda, A., Ohira, H., Miyawaki, K., Nishikawa, M., Kumano, H., et al. (2005). Effects of value and reward magnitude on feedback negativity and P300. *Neuroreport* 16, 407–411. doi: 10.1097/00001756-200503150-00020

Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science* 275, 1593–1599. doi: 10.1126/science.275.5306.1593

Shepherd, S. V., Deaner, R. O., and Platt, M. L. (2006). Social status gates social attention in monkeys. *Trends Cogn. Sci.* 4, 138–147. doi: 10.1016/j.cub.2006.02.013

Silk, J. B. (1992). The patterning of intervention among male bonnet macaques: reciprocity, revenge, and loyalty. *Curr. Anthropol.* 33, 318–325. doi: 10.1086/204073

Simpson, J. A. (2007). Psychological foundations of trust. *Curr. Dir. Psychol. Sci.* 16, 264–268. doi: 10.1111/j.1467-8721.2007.00517.x

Steckler, C. M., and Tracy, J. L. (2014). "The emotional underpinnings of social status," in *The Psychology of Social Status*, eds J. Cheng, J. Tracy, C. Anderson (New York, NY: Springer), 201–224. doi: 10.1007/978-1-4939-0867-7_10

Stevens, J. M., Vervaecke, H., de Vries, H., and Van Elsacker, L. (2005). The influence of the steepness of dominance hierarchies on reciprocity and interchange in captive groups of bonobos (*Pan paniscus*). *Behaviour* 142, 941–960. doi: 10.1163/1568539055010075

Sutton, S., Braren, M., Zubin, J., and John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science* 150, 1187–1188. doi: 10.1126/science.150.3700.1187

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Q. Rev. Biol.* 46, 35–57. doi: 10.1086/406755

Twenge, J. M., and Campbell, W. K. (2002). Self-esteem and socioeconomic status: a meta-analytic review. *Pers. Soc. Psychol. Rev.* 6, 59–71. doi: 10.1207/S15327957PSPR0601_3

Ungar, S. (1981). The effects of status and excuse on interpersonal reactions to deviant behavior. *Soc. Psychol. Q.* 44, 260–263. doi: 10.2307/3033838

Vignapiano, A., Mucci, A., Ford, J., Montefusco, V., Plescia, G. M., Bucci, P., et al. (2016). Reward anticipation and trait anhedonia: an electrophysiological investigation in subjects with schizophrenia. *Clin. Neurophysiol.* 127, 2149–2160. doi: 10.1016/j.clinph.2016.01.006

von Essen, E., and Ranehill, E. (2013). *Punishment and Status. (EFI Working Paper Series in Economics and Finance, No.* 732). Zurich: Economics.

Wahrman, R. (1970). High status, deviance and sanctions. *Sociometry* 33, 485–504. doi: 10.2307/2786321

Wahrman, R. (2010). Status, deviance, and sanctions: a critical review. *Small Group Res.* 41, 91–105. doi: 10.1177/1046496409359505

Wiggins, J. A., Dill, F., and Schwartz, R. D. (1965). On "status-liability". *Sociometry* 28, 197–209. doi: 10.2307/2785650

Wu, Y., Leliveld, M. C., and Zhou, X. (2011a). Social distance modulates recipient's fairness consideration in the dictator game: an ERP study. *Biol. Psychol.* 88, 253–262. doi: 10.1016/j.biopsycho.2011.08.009

Wu, Y., and Zhou, X. (2009). The P300 and reward valence, magnitude, and expectancy in outcome evaluation. *Brain Res.* 1286, 114–122. doi: 10.1016/j.brainres.2009.06.032

Wu, Y., Zhou, Y., van Dijk, E., Leliveld, M. C., and Zhou, X. (2011b). Social comparison affects brain responses to fairness in asset division: an ERP study with the ultimatum game. *Front. Hum. Neurosci.* 5:131. doi: 10.3389/fnhum.2011.00131

Yeung, N., Holroyd, C. B., and Cohen, J. D. (2005). ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cereb. Cortex* 15, 535–544. doi: 10.1093/cercor/bhh153

Yeung, N., and Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *J. Neurosci.* 24, 6258–6264. doi: 10.1523/JNEUROSCI.4537-03.2004

Zak, P. J., and Knack, S. (2001). Trust and growth. *Econ. J.* 111, 295–321. doi: 10.1111/1468-0297.00609

Zhou, Z., Yu, R., and Zhou, X. (2010). To do or not to do? Action enlarges the FRN and P300 effects in outcome evaluation. *Neuropsychologia* 48, 3606–3613. doi: 10.1016/j.neuropsychologia.2010.08.010

Zink, C. F., Tong, Y., Chen, Q., Bassett, D. S., Stein, J. L., and Meyer-Lindenberg, A. (2008). Know your place: neural processing of social hierarchy in humans. *Neuron* 58, 273–283. doi: 10.1016/j.neuron.2008.01.025

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 31 August 2018
doi: 10.3389/fnins.2018.00606

Transcranial Direct Current Stimulation of the Right Lateral Prefrontal Cortex Changes *a priori* Normative Beliefs in Voluntary Cooperation

Jianbiao Li[1,2], Xiaoli Liu[1]*, Xile Yin[3], Shuaiqi Li[1], Pengcheng Wang[4], Xiaofei Niu[1] and Chengkang Zhu[1]

[1] China Academy of Corporate Governance, Reinhard Selten Laboratory, Business School, Nankai University, Tianjin, China, [2] Department of Economics and Management, Nankai University Binhai College, Tianjin, China, [3] School of Business Administration, Zhejiang Gongshang University, Hangzhou, China, [4] Business School, Tianjin University of Finance and Economics, Tianjin, China

A priori normative beliefs, the precondition of social norm compliance that reflects culture and values, are considered unique to human social behavior. Previous studies related to the ultimatum game revealed that right lateral prefrontal cortex (rLPFC) has no stimulation effects on normative beliefs. However, no research has focused on the effects of *a priori* belief on the rLPFC in voluntary cooperation attached to the public good (PG) game. In this study, we used a linear asymmetric PG to confirm the influence of the rLPFC on *a priori* normative beliefs without threats of external punishment through transcranial direct current stimulation (tDCS). Participants engaged via computer terminals in groups of four (i.e., two high-endowment players with 35G$ and two low-endowment players with 23G$). They were anonymous and had no communication during the entire process. They were randomly assigned to receive 15 min of either anodal, cathodal, or sham stimulation and then asked to answer questions concerning *a priori* normative beliefs (norm.belief and pg.belief). Results suggested that anodal/cathodal tDCS significantly ($P < 0.001$) shifted the participants' *a priori* normative beliefs in opposite directions compared to the shift in the sham group. In addition, different identities exhibited varying degrees of change (28.80–54.43%). These outcomes provide neural evidence of the rLPFC mechanism's effect on the normative beliefs in voluntary cooperation based on the PG framework.

Keywords: *a priori* normative beliefs, voluntary cooperation, identity, rLPFC, transcranial direct current stimulation

Edited by:
Yefeng Chen,
Zhejiang University, China

Reviewed by:
Walter Adriani,
Istituto Superiore di Sanità (ISS), Italy
Thomas Zoëga Ramsøy,
Neurons Inc, Denmark

****Correspondence:***
Xiaoli Liu
liuxiaoli0510@126.com

Specialty section:
This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *29 November 2017*
Accepted: *10 August 2018*
Published: *31 August 2018*

Citation:
Li J, Liu X, Yin X, Li S, Wang P, Niu X and Zhu C (2018) Transcranial Direct Current Stimulation of the Right Lateral Prefrontal Cortex Changes a priori Normative Beliefs in Voluntary Cooperation. Front. Neurosci. 12:606. doi: 10.3389/fnins.2018.00606

INTRODUCTION

Neuroscience studies on social norms prove that the human brain may have potential cognitive and neural processes that underlie the ability to learn norms, follow norms, and enforce norms by generating appropriate behavioral responses to social norm compliance and normative judgments (Güth et al., 1982; Montague and Lohrenz, 2007; Buckholtz and Marois, 2012; Liu et al., 2017). For example, Buckholtz and Marois (2012) suggested a potential neurobiological architecture that may underpin norm learning, norm compliance, and norm enforcement (social sanctions or internal sanctions). They found that a dorsal frontostriatal circuitry is essential for integrating information about sanction threats into decision-making to incentivize norm-compliant behavior. Whether or not the induction of right lateral prefrontal cortex (rLPFC) can change *a priori* normative beliefs in a controlled behavioral voluntary contribution paradigm has not been investigated in the context of social norm compliance. Therefore, changing *a priori* normative beliefs under controlled experimental conditions in healthy volunteers is necessary to clarify causally the role of rLPFC in voluntary cooperative behaviors.

Human beings are the most social creatures among all species known, because none of the other species share our capacity for stable large-scale cooperation among genetically unrelated individuals. This unique feature of human culture is made possible by cognitive capacities that permit us to establish, transmit, and enforce social norms (Fehr and Fischbacher, 2004; Buckholtz and Marois, 2012; Yin et al., 2017). A social norm is a behavioral rule that is enforced by social sanctions (Coleman, 1990) and internal sanctions (e.g., feeling of guilt) (Lindbeck, 1997). "One should not litter" is an example of a social norm. Many people do not litter even when they know that nobody is observing them because people have subjective perceptions of norms, and these subjective perceptions can guide the opinions of individuals (Buckholtz and Marois, 2012). In the context of social norms, the average person does not know the actual rates of behaviors or opinions in their community (Tankard and Paluck, 2016). As they have unreliable information about what others actually think, they need to infer what (e.g., thoughts, beliefs, desires, intentions, and motivations) is going on inside other people's heads. This subjective inference is defined as "*a priori* normative beliefs."[1] *A priori* normative beliefs are *a priori* beliefs based on perception of other people's social norms and are a reference point that guides people's behavior in social cooperation. The cooperative behaviors and actions of subjects are thought to rely strongly on the *a priori* normative beliefs in charge of regulating and coordinating thoughts and motivations under norm enforcement. Hence, extensive debates persist regarding deep neural insights into *a priori* normative beliefs and the manner of their implementation in the brain (Ruff et al., 2013; Sanfey et al., 2014).

The result of a long stream of laboratory experiments related to voluntary contributions in public good (PG) environments has already been established solidly. In the basic PGs, participants secretly decide how much of their endowment contribute into a public pool and how much remain. Contributions in the public pool, which are multiplied by a factor (greater than one and less than the number of players), are evenly divided among all participants. The actual level of contributions, which usually ranges between 40 and 60% of the total endowment (Chaudhuri et al., 2016), depends on various factors, such as the number of players and the *per capita* rate of return of the PG relative to that of the private good (Keser and Winden, 2000). Currently, although no agreement has been reached about why subjects contribute, an influential explanation is conditional cooperation. Conditional cooperation can be considered as a motivation on its own or a consequence of some fairness preferences, such as "altruism," "warm glow," "inequity aversion," or "reciprocity" (Fischbacher et al., 2001). Experiments on conditional cooperation found that subjects usually contribute similarly to their co-players (Keser and Winden, 2000; Brandts and Schram, 2004; Kocher et al., 2007; Spiller et al., 2016) and are willing to contribute to a PG when others also contribute or are expected to do so (Fischbacher and Gächter, 2010). For example, the studies of Fischbacher and Gächter (2010) on conditional cooperation indicate that individual cooperation often depends on whether a person thinks others cooperate. The existence and extent of conditional cooperation are considerably influenced by the beliefs elicited on the subjective perception of norms (e.g., people contribute nothing because they believe others will contribute nothing, Kocher et al., 2008). Two possible situations are considered before a decision is made. On the one hand, some subjects must at least know of social norms and follow them (Elster, 1989; Bicchieri, 2006). On the other hand, participants may feel that their partners may not follow a norm even if it exists (Reuben and Riedl, 2009; Spiller et al., 2016). In either case, the subject needs to infer from the belief of others. Thus, the ability to attribute thoughts to others and infer their mental states plays a crucial role in social interactions (Sellaro et al., 2015).

According to the definition of Spiller et al. (2016), belief to infer "what others do" is a kind of *a priori* normative beliefs. Previous studies provided evidence by showing that people contribute more to a PG when they expect others to contribute more as well (Kachelmeier and Shehata, 1997; Croson, 2007). On the basis of these views, we may conjecture that subjects tend to follow their *a priori* normative beliefs concerning contributions. That is, subjects consider the actions of others whom they inferred as a reference for their own behavior. In this case, the essence of *a priori* normative beliefs is a reference point that is formed in the context of common knowledge considered as a "norm." This description indicates that *a priori* normative beliefs play a key role in judging others' motives and are the basis of a subject's action in cooperation.

Human societies enforce norm by threatening norm violators with sanctions (social or internal) (Coleman, 1990; Lindbeck, 1997; Eriksson et al., 2017). Neuroscience studies on norms have mostly focused on the neural basis of sanctions (Sanfey et al., 2003, 2014; Spitzer et al., 2007; Boksem and De Cremer, 2010; Ruff et al., 2013; Xiang et al., 2013). All these studies used sanctioned cooperation based on the ultimatum game (UG). The UG consists of two players: proposer and responder. The proposer decides how much of a monetary endowment to split with the responder, while the responder could accept the offer or, if he/she deems the offer as violating a social norm, reject it (Ruff et al., 2013; Sanfey et al., 2014). These studies proved that the human brain has developed neural processes to support

[1] The normative beliefs in this paper are derived from people's perception of social norm and should be treated as *a priori* beliefs without updating or *a posteriori* beliefs. This paper treat normative beliefs as *a priori* factors and do not investigate how they evolve in the dynamic situation with feedback, as feedback (e.g., average contributions by others and corresponding payoff) on the one hand can decrease illusory ideas, on the other hand it may cause some uncontrollable noise variables (e.g., anchoring effect, Furnham and Hua, 2011).

social cooperation by punishing norm violations, which are also important in sustaining human cooperation in the PG (Fehr and Fischbacher, 2004; Reif et al., 2017).

Sanfey et al. (2003) used functional magnetic resonance imaging of UG players, who responded by complying with or violating the social norm, to investigate the neural substrates of cognitive processes involved in economic decision-making. In the study, behaviors who violated the social norm elicited activity in brain areas related to the dorsolateral prefrontal cortex (DLPFC). Spitzer et al. (2007) also found that the increase in norm compliance of individuals exhibit a strong positive correlation with activations in the right DLPFC. Similarly, a lesion of the ventromedial prefrontal cortex increases the rate of rejections of offers that violate social norms in the UG (Koenigs and Tranel, 2008). Studies on non-invasive brain stimulation [e.g., transcranial direct current stimulation (tDCS)] likewise found that interfering with the activity in the DLPFC decreases the rate of rejections (Van't et al., 2005). Mounting evidence from neuroimaging and lesion studies suggests that the DLPFC is associated with social norm violations (Aron et al., 2014; Hardung et al., 2017). Recently, the prefrontal cortex (PFC) was proven to be central to higher-level cognition (Aron et al., 2007; Azuar et al., 2014; Bahlmann et al., 2015; Nee and D'Esposito, 2016). Nee and D'Esposito [(2016), p. 17] stated that "caudal lateral prefrontal cortex (LPFC) was involved in current processing, providing selective attention to visual stimulus features, while rostral LPFC was involved in future processing, enabling the retention of information for integration into future processing. The mid LPFC appeared to synthesize both current and future processing allowing the use of current and future informed contextual information to organize behavior." In addition, an area in rLPFC is activated during a norm-compliant behavior triggered by social punishment threats (Spitzer et al., 2007), an activation that changes the social cooperation among participants (Ruff et al., 2013; Sanfey et al., 2014; Liu et al., 2017). Therefore, rLPFC, which is necessary for norm-compliant behaviors and enable humans to anticipate sanctions for norm violations and distinguish "right" from "wrong" (Ruff et al., 2013; Liu et al., 2017), is a key biological prerequisite for an evolutionarily and socially important aspect of human behavior, and its activity exerts a particularly strong effect on social cooperation.

Decision-making in social dilemmas is suggested to rely on the relative judgment of two or more alternatives and individual factors affecting judgments and decisions. (Ramsøy et al., 2015; Liu et al., 2017). Previous research proved that the tDCS of rLPFC leads to a change in the norm judgment based on voluntary cooperation (Liu et al., 2017). The results suggested that anodal/cathodal tDCS increases/decreases participants' judgment of "right contribution" (i.e., the amount individual ought to contribute) in opposite directions unlike in the sham group. Spiller et al. (2016) proved that *a priori* normative beliefs were also influenced by the "right contribution." Relying on the results and analyses presented above, we can conjecture that if *a priori* normative beliefs are influenced by the norm in other people's heads, then stimulating the same brain region (i.e., rLPFC) should also affect the *a priori* normative beliefs. Accordingly, we assume that if anodal/cathodal tDCS is applied to increase/decrease the activities of the rLPFC, the participants' *a priori* normative beliefs will be changed. Specifically, anodal tDCS will improve the *a priori* normative beliefs, whereas cathodal tDCS will deteriorate it.

Our analysis focused on two broad categories of beliefs and brain regions that are important for *a priori* normative beliefs as revealed in previous studies (Adolphs, 2009; Fishbein and Icek, 2010; Spiller et al., 2016). To provide neural evidence of *a priori* normative beliefs among different identities, we used tDCS to investigate whether the increase or decrease of rLPFC excitability among healthy participants influences *a priori* normative beliefs in voluntary cooperation. We expected that the induction of the rLPFC by applying tDCS causes a significant change in the contribution of *a priori* normative beliefs compared with that in the sham group and that treatment effects can be observed.

MATERIALS AND METHODS

Subjects

The subjects of this experiment were the same as Liu et al. (2017) and Li et al. (2018). A total of 83 healthy subjects (recruited from Nankai University students; 41 females and 42 males ranging from 20 to 30 years old) were kept in the sample. None of them had suffered from any neurological or psychiatric disorders. One participant in the anodal stimulation treatment felt discomfort, and we terminated the experiment. Participants randomly divided into three treatments, namely, cathodal ($n = 28$, 12 males), anodal ($n = 27$, 18 males), and sham ($n = 28$, 12 males) stimulation. All the participants had no ex-ante knowledge of neurological (tDCS) or PG tasks, and all voluntarily joined this study with informed consents. The experiment was performed in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of Business of Nankai University. All these 83 participants reported no adverse side effects (e.g., pain on the scalp or headaches) after the experiment.

Transcranial Direct Current Stimulation

The tDCS of the human motor cortex induces shifts in cortical excitability during and after stimulation under the electrode (Batsikadze et al., 2013; Jamil et al., 2017). These shifts are polarity-specific, with cathodal and anodal tDCS usually resulting in a decrease and an increase in cortical excitability, respectively (Iyer et al., 2005; Nitsche et al., 2008; Utz et al., 2010; Kadosh, 2013). Unilateral (Brückner and Kammer, 2017; Luo et al., 2017) and same effects exist (Marshall et al., 2005; Filmer et al., 2015) as well, although the latter is less common than the former. tDCS has become a kind of research paradigm in neural science. Thus far, brain stimulation studies in humans mostly show unidirectional maladaptive effects on decision-making, rendering participants more impulsive, selfish, or cognitively biased (Knoch et al., 2006; Chang and Sanfey, 2013; Ruff et al., 2013).

On the basis of this finding and the general role of rLPFC in behavior control (Miller and Cohen, 2001; Aron et al., 2004), we randomly sorted participants into three stimulation groups, in which the neural excitability in the rLPFC was enhanced with anodal tDCS, reduced with cathodal tDCS, or left unaltered by sham tDCS as control for possible non-neural effects of stimulation. All participants received tDCS delivered by a battery-driven stimulator (Neuro Conn, Germany) in our experiment. tDCS was applied using a set of standard 5 cm × 7 cm electrodes fixed with rubber straps, which is the most commonly used approach in tDCS (Fusco et al., 2013; Li et al., 2017). For subjects receiving tDCS, the anodal/cathodal electrode was placed over the rLPFC according to the international EEG 10–20 electrode system, and the reference electrode (cathode for anodal tDCS and anode for cathodal tDCS) was positioned over the vertex, which was consistent with the design of Ruff et al. (2013). The stimulation current was constant at 1.0 mA intensity (Ambrus et al., 2012; Meesen et al., 2014) with 15 s of ramp up and down. Participants in the anodal/cathodal group first received 15 min of stimulation. After that, the experimental task began immediately. They were requested to complete a self-report on *a priori* normative beliefs. (Schematic representation of the experimental design, see **Figure 1**) The procedures were the same for the sham group, except that the current was stopped after the first 30 s. The 30-s stimulation in the sham condition can mimic the itching sensation of real stimulation without producing any significant neural-altering effects on the cortex (Civai et al., 2015; Willis et al., 2015; Li et al., 2017). The protocol was approved by the Ethics Committee of Business of Nankai University, and all participants gave written informed consent.

Task and Procedure

The experimental task we conducted in the experiment was similar to those conducted by Spiller et al. (2016), except that tDCS was applied to the subjects before they participated in the experimental task. In the experiment, the participants engaged in anonymous social interactions with actual financial consequences via computer terminals. The unit of payoff in the experiment was game dollar (G$), and the exchange ratio was 1G$ = 1.5 Chinese Yuan (RMB). Payments were exchanged to cash after the experiment. The average duration was 60 min with payments of approximately 50RMB (7–8$).

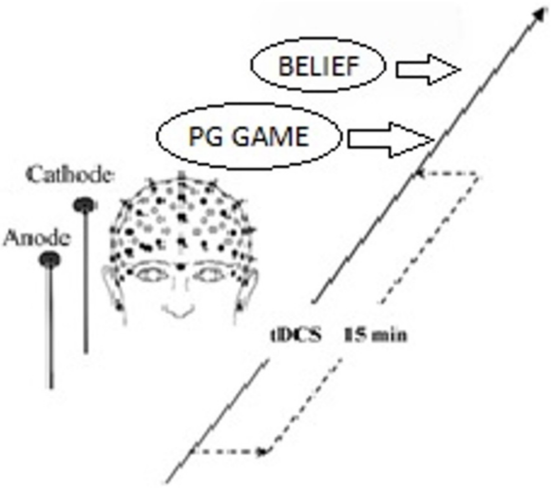

FIGURE 1 | Schematic representation of the experimental design. After 15 min of stimulation, each participant decided the amount of contribution. After that, they answered questions including two pg.belief questions and two norm.belief questions.

Subjects played a linear PG in groups of four players, two HIGH players (A1, A2) with endowments of 35G$ and two LOW players (B1, B2) with endowments of 23G$ that were asymmetric. Endowments were chosen so that 50% contributions were not an integer and not near a multiple of 5 to reduce the attraction potential of focal points (Spiller et al., 2016).

The payoff function of PG was $\pi_i = X_i - x_i + 0.6 \sum_{i=1}^{4} x_i$, where X_i was the endowment, x_i was the contribution, and $\sum_{i=1}^{4} x_i$ was the sum contributions of participants from the same group. At the beginning of each trial, the subjects were informed of their identity types (A1, A2, B1, and B2). Then they were asked to answer questions related to beliefs about themselves, voluntary cooperative level and beliefs about others. We did not focus on the beliefs about themselves and voluntary cooperative level in the current study. However, we have emphatically discussed them in Liu et al. (2017) and Li et al. (2018), respectively. In this paper, we focused on the beliefs about others which were tested by pg.belief questions and norm.belief questions:

pg.belief questions: How much do you believe your peers will contribute? If they are HIGH players (A1 or A2) and Low players (B1 or B2), respectively.

norm.belief questions: How much do you believe your peers on average think is the "right" contribution? If they are HIGH players (A1 or A2) and Low players (B1 or B2), respectively.

In each trial, the identity types of subjects were reassigned and endowments were started from the initial situation. A total of 16 trials were conducted. We assigned fixed orders (pseudorandom order) in which all identities were assigned to avoid the order effect. The subjects knew neither how many trials they would play nor any feedback about contributions and payoff.

In addition to the payoff from the contribution and non-contributed endowment, subjects were also told they could receive additional incentives, which were higher if their beliefs were closer to the actual mean of group contributions in the two pg.belief questions. For example, if the bias was less than 1G$, then they would earn 4RMB.

Statistical Analyses

The levels of beliefs were assessed using mean values (the beliefs asked during the experiment). Two types of beliefs were tested: (1) pg.belief (How much do you believe your peers will contribute?) and (2) norm.belief (How much do you believe your peers on average think is the "right" contribution?). Three treatment of tDCS stimulation groups were formed: (1) anodal, (2) sham, and (3) cathodal. The PG had two types of players, namely, (1) HIGH (35G$, A1 and A2) and (2) LOW (23G$, B1 and B2), with two types having four pairs of players: (1) HIGH for HIGH (indicates HIGH players to the question for HIGH players), (2) LOW for HIGH (indicates LOW players to the question for HIGH players), (3) HIGH for LOW (indicates HIGH players to the question for LOW players), and (4) LOW for LOW (indicates LOW players to the question for LOW players).

The levels of the two types of beliefs (norm.belief and pg.belief) were first evaluated using two-way ANOVA: 2 (types of players: HIGH and LOW) × 3 (tDCS stimulation groups: anodal, sham, and cathodal). One-way ANOVA was then performed to test the difference of norm.belief and pg.belief in three stimulation groups, respectively. Moreover, the mean levels of norm.belief and pg.belief between stimulation group and sham group were evaluated using *t*-test and rank-sum test. We also considered four pairs of players and conducted two-way ANOVA: 4 (pairs of players: HIGH for HIGH, HIGH for LOW, LOW for HIGH, LOW for LOW) × 3 (tDCS stimulation groups: anodal, sham, and cathodal).

RESULTS

Behavioral Data

We analyzed the mean values of the participants with different endowments among the three stimulation groups (**Table 1**). Results showed that the participants were sensitive to their endowment. For one thing, both HIGH and LOW players believed a higher "right" average contribution (norm.belief) relative to that of the HIGH players than to that of the LOW players. Furthermore, the players with the same initial endowment had a higher expectation of their peers (pg.belief) than those with different initial endowments, except for the pg.belief relative to LOW players in the cathodal group (8.71 < 8.89).

General Effect of tDCS Over rLPFC on *a priori* Normative Beliefs

We performed two-way ANOVA for norm.belief with the stimulation type (anodal, cathodal, and sham stimulation) as a between-subject factor and the player type (HIGH and LOW) as a within-subject factor. Significant main effects of stimulation type [$F_{(2,329)}$ = 138.38, $P < 0.001$] and player type [$F_{(1,330)}$ = 89.04, $P < 0.001$] were noted. Importantly, a significant interactive effect of stimulation type and player type was found [$F_{(2,329)}$ = 6.58, $P = 0.002$]. We also performed two-way ANOVA for pg.belief with the stimulation type (anodal, cathodal, and sham stimulation) as a between-subject factor and the player type (HIGH and LOW) as a within-subject factor. Significant main effects of stimulation type [$F_{(2,329)}$ = 114.51, $P < 0.001$] and player type [$F_{(1,330)}$ = 74.83, $P < 0.001$] were likewise observed. A significant interactive effect of stimulation type and player type [$F_{(2,329)}$ = 5.93, $P = 0.003$] was obtained (**Figure 2**).

One-way ANOVA, Kruskal–Wallis test, *t*-test, and rank-sum test were used to analyze the difference among the *a priori* normative beliefs (norm.belief and pg.belief) of the three stimulation groups. The current data show that the mean levels of norm.belief of the anodal, sham, and cathodal groups were 25.44 (SD = 7.26), 17.46 (SD = 6.13), and 12.13 (SD = 6.72), while the mean levels of pg.belief were 23.54 (SD = 8.80), 16.14 (SD = 6.15), and 10.73 (SD = 5.59), respectively. Significant differences were observed in the norm.belief and pg.belief values of the three stimulation groups [$F_{(2,329)}$ = 109.17, $P < 0.001$; Kruskal–Wallis test $P < 0.001$ and $F_{(2,329)}$ = 93.53, $P < 0.001$; Kruskal–Wallis test $P < 0.001$, respectively]. The mean levels of norm.belief and pg.belief in the anodal stimulation group were significantly higher than those in the sham stimulation group (t = 8.824, $P < 0.001$; Z = 8.031, $P < 0.001$ and t = 7.245, $P < 0.001$; Z = 7.073, $P < 0.001$, respectively, for the *t*-test and rank-sum test). The mean level of the cathodal stimulation group was significantly lower than that of the sham stimulation group (t = 6.190, $P < 0.001$; Z = 6.294, $P < 0.001$ and t = 6.888, $P < 0.001$; Z = 6.571, $P < 0.001$, respectively, for the *t*-test and rank-sum test; **Figures 3**, **4**).

Effect of tDCS Over rLPFC on *a priori* Normative Beliefs of Asymmetric Identity

We compared the level of norm.belief and pg.belief among the four pairs of players under three stimulation groups. We conducted two-way ANOVA: 4 (pairs of players: HIGH for HIGH, HIGH for LOW, LOW for HIGH, LOW for LOW) × 3 (tDCS stimulation groups: anodal, sham, and cathodal). Significant main effects of stimulation groups [$F_{(2,329)}$ = 137.64, $P < 0.001$; $F_{(2,329)}$ = 114.29, $P < 0.001$] and the pairs of players [$F_{(3,328)}$ = 29.60, $P < 0.001$; $F_{(3,328)}$ = 25.35, $P < 0.001$] to norm.belief and pg.belief were noted, respectively. Significant differences were observed, and the following results were found: norm.belief HIGH for HIGH [$F_{(2,80)}$ = 63.36, $P < 0.001$; Kruskal–Wallis test $P < 0.001$], norm.belief HIGH for LOW [$F_{(2,80)}$ = 44.28, $P < 0.001$; Kruskal–Wallis test $P < 0.001$], norm.belief LOW for HIGH [$F_{(2,80)}$ = 26.06,

TABLE 1 | Mean values of norm.belief and pg.belief.

Stimulation groups Pairs of player types	norm.belief			pg.belief		
	Anodal	Sham	Cathodal	Anodal	Sham	Cathodal
HIGH for HIGH	30.89 (6.07)	19.89 (5.05)	13.11 (6.51)	27.63 (9.00)	18.57 (5.48)	12.07 (5.58)
LOW for HIGH	29.18 (7.80)	20.96 (7.15)	14.64 (7.50)	28.74 (8.52)	18.61 (7.00)	13.25 (5.89)
HIGH for LOW	21.42 (3.08)	15.04 (4.06)	9.75 (6.08)	19.81 (5.68)	14.14 (4.37)	8.71 (5.08)
LOW for LOW	20.26 (4.42)	13.93 (4.83)	11.04 (5.97)	17.96 (6.37)	13.25 (5.67)	8.89 (4.55)

Mean values of norm.belief and pg.belief in three stimulation groups. SDs are enclosed in parentheses. Column "Pairs of player types" indicates which player type the answer was provided [e.g., norm.belief of anodal in row "HIGH for LOW" indicates the mean response of HIGH players (in the anodal stimulation group) to the question How much do you believe your peers on average think is the "right" contribution? for LOW players B1 and B2?].

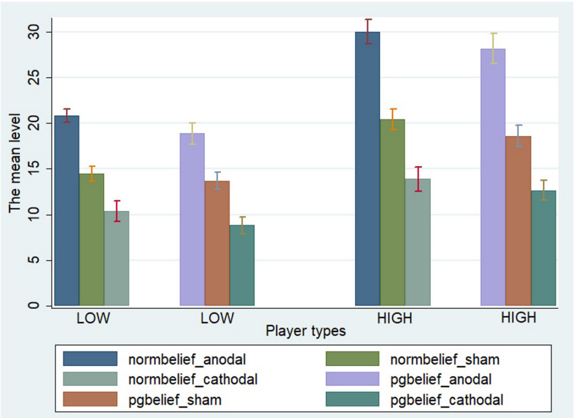

FIGURE 2 | *A priori* normative beliefs in three stimulation groups. Mean values of norm.belief and pg.belief in the three stimulation groups (anodal, sham, and cathodal) between two types of players (HIGH and LOW).

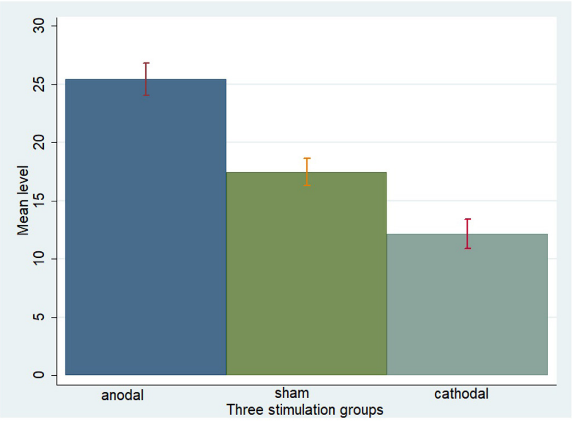

FIGURE 3 | Norm.belief in three stimulation groups. Mean values of norm.belief in the three stimulation groups (anodal, sham, and cathodal) of all players.

$P < 0.001$; Kruskal–Wallis test $P < 0.001$], norm.belief LOW for LOW [$F_{(2,80)} = 23.24$, $P < 0.001$; Kruskal–Wallis test $P < 0.001$], pg.belief HIGH for HIGH [$F_{(2,80)} = 35.66$, $P < 0.001$; Kruskal–Wallis test $P < 0.001$], pg.belief HIGH for LOW [$F_{(2,80)} = 33.03$, $P < 0.001$; Kruskal–Wallis test $P < 0.001$], pg.belief LOW for HIGH [$F_{(2,80)} = 32.70$, $P < 0.001$; Kruskal–Wallis test $P < 0.001$], and pg.belief LOW for LOW [$F_{(2,80)} = 18.22$, $P < 0.001$; Kruskal–Wallis test $P < 0.001$].

From sham to stimulation, the ratios of individual norm.belief change increased by 55.30% (HIGH to HIGH), 41.13% (HIGH to LOW), 39.27% (LOW to HIGH), and 34.71% (LOW to LOW) in the anodal group, and the matching ratios were attenuated by 34.09, 34.41, 30.15, and 35.17% in the cathodal group, respectively (**Figure 5**). The difference in improvement percentage of norm.belief among three stimulation groups with the same identities (HIGH for HIGH and LOW for LOW) is significant [$F_{(2,163)} = 74.03$, $P < 0.001$; Kruskal–Wallis test $P < 0.001$]. The difference among groups with different identities (HIGH for LOW and LOW for HIGH) is also significant [$F_{(2,163)} = 25.26$, $P < 0.001$; Kruskal–Wallis test $P < 0.001$]. This result means that the same stimulus has different effects on people of different identities.

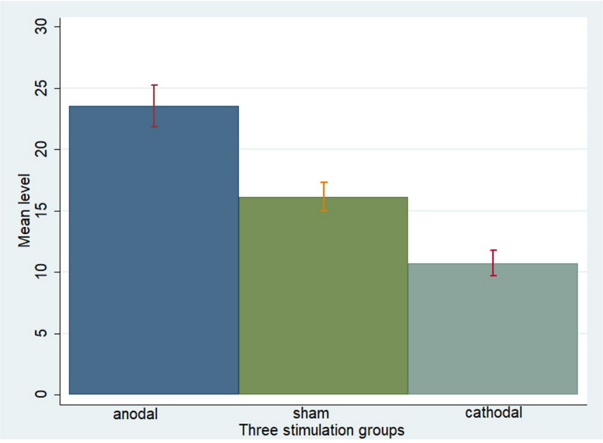

FIGURE 4 | Pg.belief in three stimulation groups. Mean values of pg.belief in the three stimulation groups (anodal, sham, and cathodal) of all players.

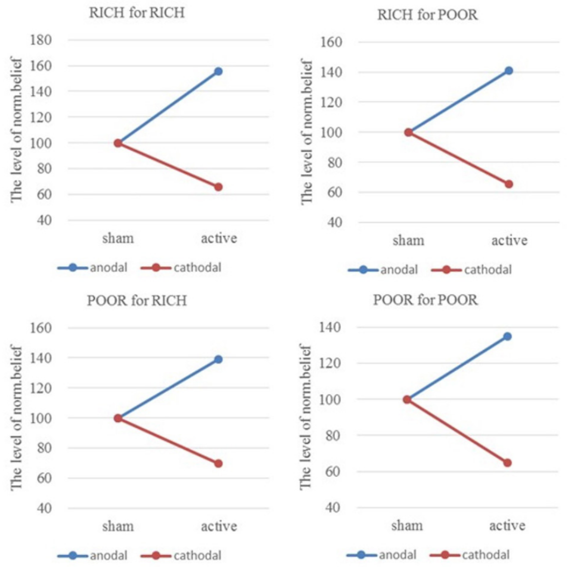

FIGURE 5 | Ratios of norm.belief of four pairs of players. From sham to stimulation, the ratios of individual norm.belief change increased by 55.30% (HIGH to HIGH), 41.13% (HIGH to LOW), 39.27% (LOW to HIGH), and 34.71% (LOW to LOW) in the anodal group, and the matching ratios were attenuated by 34.09, 34.41, 30.15, and 35.17% in the cathodal group, respectively.

Similarly, the ratios of individual pg.belief change increased by 48.79% (HIGH to HIGH), 40.10% (HIGH to LOW), 54.43% (LOW to HIGH), and 35.55% (LOW to LOW) in the anodal group, and the matching ratios were attenuated by 35.00, 38.40, 28.80, and 32.91 % in the cathodal group, respectively (**Figure 6**).

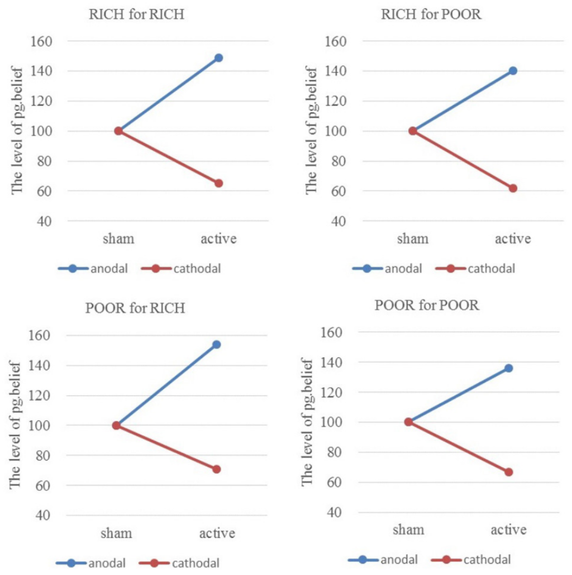

FIGURE 6 | Ratios of pg.belief of four pairs of players. From sham to stimulation, the ratios of individual pg.belief change increased by 48.79% (HIGH to HIGH), 40.10% (HIGH to LOW), 54.43% (LOW to HIGH), and 35.55% (LOW to LOW) in the anodal group, and the matching ratios were attenuated by 35.00, 38.40, 28.80, and 32.91% in the cathodal group, respectively.

The difference in improvement percentage of norm.belief among three stimulation groups with the same identities (HIGH for HIGH and LOW for LOW) is significant [$F_{(2,163)} = 50.56$, $P < 0.001$; Kruskal–Wallis test $P < 0.001$]. The difference among groups with different identities (HIGH for LOW and LOW for HIGH) is also significant [$F_{(2,163)} = 28.39$, $P < 0.001$; Kruskal–Wallis test $P < 0.001$]. The result is basically the same as that for norm.belief.

DISCUSSION

Resulting *a priori* normative beliefs in a social environment are controlled by a widespread neural network, including the rLPFC, which plays an important role in decision-making. This study investigated the influence of the neurophysiological modulation of rLPFC reactivity by means of tDCS on *a priori* normative beliefs. For this purpose, we administered anodal, cathodal, and sham stimulations on the rLPFC while subjects reported their beliefs of peers. Consistent with our hypothesis, enhancing/suppressing the activity in the rLPFC increased/decreased the level of *a priori* normative beliefs, which were tested by the self-reported contribution in the PG in contrast to the sham stimulation. Our results demonstrate that alterations of rLPFC activity can change *a priori* normative beliefs and consequently provide a causal link between rLPFC activity and *a priori* normative beliefs in voluntary cooperation.

Consistent with the results of previous research (Spitzer et al., 2007; Ruff et al., 2013; Liu et al., 2017), we also verified that rLPFC is involved in the neural mechanisms that support social cooperation. This finding is not a coincidence, as the rLPFC is a crucial brain region that is involved in the process of social norms, not only under the enforcement of sanctions based on the UG, but also under voluntary cooperation based on the PG. The former is fair norm and the latter is cooperation norm, and both belong to social norms. In addition, the present experiment sought to test the possible role of rLPFC in beliefs about voluntary cooperation norm followed by others. Ruff et al. (2013) measured some beliefs (i.e., the perceived fairness of the offer and the punishment expected) that the participants held. In their experiment, subject (Player A) was observed while he made decisions about how much of a monetary endowment to split with another participant (Player B). On the baseline condition, Player B could not punish Player A if he deemed the amount of the split to be unfair. On the punishment condition, Player B was permitted to punish Player A if he deemed the offer unfair. However, they did not measure the beliefs separately or directly assess the participants' beliefs for each treatment condition (Sanfey et al., 2014). Fortunately, our experimenters measured the *a priori* normative beliefs separately for two identities (HIGH player and LOW player) and for all colleagues in each treatment. This design enabled us to directly assess the participants' beliefs about social norms. Simultaneously, unlike our research based on the PG frame, Ruff et al. (2013) was based on the UG. UG is a kind of zero-sum game where the decision-making status of the proposer and the responder are unequal, which is not conducive to cooperation. Taken together, these differences may be the main factors that contributed to the varying results of the different research frameworks.

There is a growing interest in cognitive science and neuroscience in studying the effect of *a priori* beliefs on behavioral performance and their underlying neural mechanisms (Friston, 2010; Clark, 2013; Hohwy, 2013; Allen and Friston, 2018). What do the brain's *a priori* beliefs arise from? As Bowles (2004) suggested, there were two potential sources: one source was genes (inherited from our parents) and the other was cultural inheritance (our past experience through learning or gain). For example, a belief general prevails within certain embodied and environmental conditions in the generative sense (Allen and Friston, 2018). Heuristically, if participants were endowed with the *a priori* beliefs which could help their survival, then they will act in ways that were consistent with that *a priori* beliefs. Specifically, during minimizing prediction error which is imperative for survival, participants may necessarily incorporate self-referential information in the form of *a priori* beliefs and long-term memory to characterize their behaviors (Allen and Friston, 2018). In this process, neuromodulation of post-synaptic gain via neurotransmitters (e.g., dopamine and norepinephrine) are proved to communicate the precision of *a priori* beliefs (Feldman and Friston, 2010; Moran et al., 2013; Kanai et al., 2015).

In our experiment, the "right" contribution is self-reported rather than exogenous, that is, it is not an exact amount or proportion of the initial endowment. For example, Player A1 may think Player B1 should contribute 10G$, so he would report his belief about Player B1 on the basis of his own judgment. In the research of Ruff et al. (2013), "participants are using a fairness norm of 'equity,' whereby the optimal decision would be to split the pot of money equally between both players" (Sanfey et al., 2014, p.173). In general, the belief tested in our study based on PG was derived from the participants' own judgment about norms, whereas the belief tested in previous research based on UG was derived from external norms. Therefore, the PG without external punishment is more effective than the UG with a punishment constraint in terms of reflecting people's true beliefs in voluntary cooperation. Punishment can easily trigger negative emotions, which are associated with cognitive control. Neuroscientific findings prove that negative emotions can lead to proactive aggression (Dambacher et al., 2015) and aggressive response (Riva et al., 2015), which may interfere with the original belief. Social cooperation preferences are forced out and beliefs are changed. However, the true intentions underlying PG exert no such negative effects. To a certain extent, this outcome also shows that our research framework based on PG is more suitable than UG for cooperation norm compliance and its attached beliefs. Thus, our research provides a new paradigm for future studies on belief of social norm compliance.

In this paper, an individual think the "right" contribution is the "norm" which is based on widely shared beliefs how individual group members ought to behave in PG game. The "actual" contribution is the "compliance" that an individual truly performed in a PGs game. Participants considered the criteria for the "right" contribution believed by other subjects (norm.belief) based on the judgment that people should behave in the PGs framework. However, it is well-documented that participants might feel others not follow a norm that even if it exists (e.g., subjects contribute less than what they consider as "fair," Reuben and Riedl, 2013) and will not perform what they considered as "right" in practice. In this situation, participants believe that there is a discrepancy between "right *per se*" and "actually paid by others."

We used tDCS (Nitsche et al., 2008) in the present study to examine whether the social norm of belief and voluntary cooperation depends causally on neural processing in the previously identified rLPFC region (Spitzer et al., 2007). A methodological contribution of our study is the design that allows direct focus on the subjects' belief in voluntary cooperation. This design allows for measuring the *a priori* normative beliefs that is applicable in a specific situation and is informative of the voluntary behavior that is related to cooperation norms. For example, it could have been informative to ask participants what they believe the "right" contribution is for HIGH players A1 and A2 in each of the situations. Further analysis of the available data reveals that the same identities are more likely to behave according to the same type rather than to the different types. This phenomenon is called the identity effect, which also confirms the common saying that birds of a feather flock together. Our study is also relevant to the existing experimental economics literature (Kocher et al., 2008; Reuben and Riedl, 2009; Spiller et al., 2016), which usually identifies departures from pure self-interest payoffs by controlling other motivations. Furthermore, the valuable literature does not typically consider norm.beliefs and pg.beliefs in voluntary cooperation through tDCS stimulation. Our results offer support for this distinction with some proof. Both types of *a priori* normative beliefs can be changed by varying the neural excitability of rLPFC with tDCS and are affected in opposite manners.

However, our results only confirm the stimulation effect that tDCS anodal and cathodal stimulations of rLPFC lead to an increase and decrease in the contribution of *a priori* normative beliefs, respectively. We cannot answer why this stimulation leads to the change. Two models are actually possible: (1) tDCS anodal and cathodal stimulations of rLPFC stimulations lead to a change in the actual normative standards or (2) tDCS anodal and cathodal stimulations of rLPFC stimulations lead to no change in the normative value but rather impacts the downstream of the decision-making process, since decisions can also be influenced by other factors (e.g., cognitive ability). Both effects can also happen, and this may be a possible causal mechanism for future research. In addition, other beliefs may also matter in social decision-making (Sanfey et al., 2014). According to some scholars (Adolphs, 2009), three broad categories of beliefs exist: one's beliefs about the nonsocial environment, one's beliefs about the social environment and about what others in the group believe or do, and one's beliefs about one's self. For instance, people may have second-order beliefs, which reflect what people think their partner expects them to do with the purpose of establishing a reliable image and achieving a well-deserved social identity (Chang et al., 2011). To further examine the specificity of the present effects, other beliefs (such as second- or higher-order beliefs), may be included in future investigations into the effects of norm beliefs.

CONCLUSION

Our finding reveals that rLPFC stimulation affects beliefs in the cooperation norm. Anodal tDCS on the rLPFC can improve the contribution of *a priori* normative belief, whereas cathodal tDCS on the rLPFC can deteriorate it. This research is a promising step toward understanding how neurobiological mechanisms are connected to beliefs in cooperation norms.

AUTHOR CONTRIBUTIONS

JL and XL designed the experiment. XL, XN, and CZ performed the experiment. XL and XY analyzed the data. XL and SL drew the figures. XL wrote the manuscript. JL, XL, and XY revised the manuscript. All authors approved the final version of the manuscript to be published.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant Nos. 71673152, 71602051, 71602138, 71372094, and 71533002) and the National Social Science Foundation of China (16BJY035).

REFERENCES

Adolphs, R. (2009). The social brain: neural basis of social knowledge. *Psychology* 60, 693–716. doi: 10.1146/annurev.psych.60.110707.163514

Allen, M., and Friston, K. J. (2018). From cognitivism to autopoiesis: towards a computational framework for the embodied mind. *Synthese* 195, 2459–2482. doi: 10.1007/s11229-016-1288-5

Ambrus, G. G., Almoyed, H., Chaieb, L., Sarp, L., Antal, A., and Paulus, W. (2012). The fade-in–short stimulation–fade out approach to sham tDCS–reliable at 1 mA for naive and experienced subjects, but not investigators. *Brain Stimul.* 5, 499–504. doi: 10.1016/j.brs.2011.12.001

Aron, A. R., Behrens, T. E., Smith, S., Frank, M. J., and Poldrack, R. A. (2007). Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *J. Neurosci.* 27, 3743–3752. doi: 10.1523/JNEUROSCI.0519-07.2007

Aron, A. R., Robbins, T. W., and Poldrack, R. A. (2004). Right inferior frontal cortex: addressing the rebuttals. *Front. Hum. Neurosci.* 8:905. doi: 10.3389/fnhum.2014.00905

Aron, A. R., Robbins, T. W., and Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends Cogn. Sci.* 18:177. doi: 10.1016/j.tics.2013.12.003

Azuar, C., Reyes, P., Slachevsky, A., Volle, E., Kinkingnehun, S., Kouneiher, F., et al. (2014). Testing the model of caudo-rostral organization of cognitive control in the human with frontal lesions. *Neuroimage* 84:1053. doi: 10.1016/j.neuroimage.2013.09.031

Bahlmann, J., Aarts, E., and D'Esposito, M. (2015). Influence of motivation on control hierarchy in the human frontal cortex. *J. Neurosci.* 35, 3207–3217. doi: 10.1523/JNEUROSCI.2389-14.2015

Batsikadze, G., Moliadze, V., Paulus, W., Kuo, M. F., and Nitsche, M. A. (2013). Partially non-linear stimulation intensity-dependent effects of direct current stimulation on motor cortex. *J. Physiol.* 591, 1987–2000. doi: 10.1113/jphysiol.2012.249730

Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

Boksem, M. A., and De Cremer, D. (2010). Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Soc. Neurosci.* 5, 118–128. doi: 10.1080/17470910903202666

Bowles, S. (2004). *Microeconomics: Behavior, Institutions and Evolution. Microeconomics: Behavior, Institutions, and Evolution*. Princeton, NJ: Princeton University Press.

Brandts, J., and Schram, A. (2004). Cooperation and noise in public goods experiments: applying the contribution function approach. *J. Public Econ.* 79, 399–427. doi: 10.1016/S0047-2727(99)00120-6

Brückner, S., and Kammer, T. (2017). Both anodal and cathodal transcranial direct current stimulation improves semantic processing. *Neuroscience* 343:269. doi: 10.1016/j.neuroscience.2016.12.015

Buckholtz, J. W., and Marois, R. (2012). The roots of modern justice: cognitive and neural foundations of social norms and their enforcement. *Nat. Neurosci.* 15:655. doi: 10.1038/nn.3087

Chang, L. J., and Sanfey, A. G. (2013). Great expectations: neural computations underlying the use of social norms in decision-making. *Soc. Cogn. Affect. Neurosci.* 8, 277–284. doi: 10.1093/scan/nsr094

Chang, L. J., Smith, A., Dufwenberg, M., and Sanfey, A. G. (2011). Triangulating the neural, psychological, and economic bases of guilt aversion. *Neuron* 70, 560–572. doi: 10.1016/j.neuron.2011.02.056

Chaudhuri, A., Paichayontvijit, T., and Smith, A. (2016). Belief heterogeneity and contributions decay among conditional cooperators in public goods games. *J. Econ. Psychol.* 58, 15–30. doi: 10.1016/j.joep.2016.11.004

Civai, C., Miniussi, C., and Rumiati, R. I. (2015). Medial prefrontal cortex reacts to unfairness if this damages the self: a tDCS study. *Soc. Cogn. Affect. Neurosci.* 10:, 1054–1060. doi: 10.1093/scan/nsu154

Clark, A. (2013). Are we predictive engines? perils, prospects, and the puzzle of the porous perceiver. *Behav. Brain Sci.* 36, 233–253. doi: 10.1017/S0140525X12002440

Coleman, J. (1990). "Foundations of Social Theory," in *Proceedings of the International Symposium on Mobile Agents*, Cambridge, MA.

Croson, R. T. (2007). Theories of commitment, altruism and reciprocity: evidence from linear public goods games. *Econ. Inquiry* 45, 199–216. doi: 10.1111/j.1465-7295.2006.00006.x

Dambacher, F., Schuhmann, T., Lobbestael, J., Arntz, A., Brugman, S., and Sack, A. T. (2015). Reducing proactive aggression through non-invasive brain stimulation. *Soc. Cogn. Affect. Neurosci.* 10, 1303–1309. doi: 10.1093/scan/nsv018

Elster, J. (1989). *The Cement of Society: A Survey of Social Order*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511624995

Eriksson, K., Andersson, P. A., and Strimling, P. (2017). When is it appropriate to reprimand a norm violation? the roles of anger, behavioral consequences, violation severity, and social distance. *Judgm. Decis. Mak.* 12, 396–407.

Fehr, E., and Fischbacher, U. (2004). Social norms and human cooperation. *Trends Cogn. Sci.* 8, 185–190. doi: 10.1016/j.tics.2004.02.007

Feldman, H., and Friston, K. (2010). Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* 4:215. doi: 10.3389/fnhum.2010.00215

Filmer, H. L., Dux, P. E., and Mattingley, J. B. (2015). Dissociable effects of anodal and cathodal tDCS reveal distinct functional roles for right parietal cortex in the detection of single and competing stimuli. *Neuropsychologia* 74, 120–126. doi: 10.1016/j.neuropsychologia.2015.01.038

Fischbacher, U., and Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Econ. Rev.* 100, 541–556. doi: 10.1257/aer.100.1.541

Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Econ. Lett.* 71, 397–404. doi: 10.1016/S0165-1765(01)00394-9

Fishbein, M., and Icek, A. (2010). *Predicting and Changing Behavior: The Reasoned Action Approach*. New York, NY: Psychology Press.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Furnham, A., and Hua, C. B. (2011). A literature review of the anchoring effect. *J. Soc.Econ.* 40, 35–42. doi: 10.1016/j.socec.2010.10.008

Fusco, A., De, A. D., Morone, G., Maglione, L., Paolucci, T., Bragoni, M., et al. (2013). The ABC of tDCS: effects of anodal, bilateral and cathodal montages of transcranial direct current stimulation in patients with stroke-a pilot study. *Stroke Res. Treatment* 2013:837595. doi: 10.1155/2013/837595

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum game bargaining. *Econ. Behav. Organ.* 3, 367–388. doi: 10.1016/0167-2681(82)90011-7

Hardung, S., Epple, R., Jäckel, Z., Eriksson, D., Uran, C., Senn, V., et al. (2017). A functional gradient in the rodent prefrontal cortex supports behavioral inhibition. *Curr. Biol.* 27, 549–555. doi: 10.1016/j.cub.2016.12.052

Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199682737.001.0001

Iyer, M. B., Mattu, U., Grafman, J., Lomarev, M., Sato, S., and Wassermann, E. M. (2005). Safety and cognitive effect of frontal DC brain polarization in healthy individuals. *Neurology* 64, 872–875. doi: 10.1212/01.WNL.0000152986.07469.E9

Jamil, A., Batsikadze, G., Kuo, H. I., Labruna, L., Hasan, A., Paulus, W., et al. (2017). Systematic evaluation of the impact of stimulation intensity on neuroplastic after-effects induced by transcranial direct current stimulation. *J. Physiol.* 595, 1273–1288. doi: 10.1113/JP272738

Kachelmeier, S. J., and Shehata, M. (1997). Internal auditing and voluntary cooperation in firms: a cross-cultural experiment. *Account. Rev.* 72, 407–431.

Kadosh, R. C. (2013). Using transcranial electrical stimulation to enhance cognitive functions in the typical and atypical brain. *Transl. Neurosci.* 4, 20–33. doi: 10.1523/JNEUROSCI.4927-12.2013

Kanai, R., Komura, Y., Shipp, S., and Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140169. doi: 10.1098/rstb.2014.0169

Keser, C., and Winden, F. V. (2000). Conditional cooperation and voluntary contributions to public goods. *Scand. J. Econ.* 102, 23–39. doi: 10.1111/1467-9442.00182

Knoch, D., Pascualleone, A., Meyer, K., Treyer, V., and Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832. doi: 10.1126/science.1129156

Kocher, M. G., Cherry, T., Kroll, S., Netzer, R. J., and Sutter, M. (2007). *Conditional Cooperation on Three Continents. Faculty of Economics and Statistics*. Innrain: University of Innsbruck.

Kocher, M. G., Cherry, T., Kroll, S., Netzer, R. J., and Sutter, M. (2008). Conditional cooperation on three continents. *Econ. Lett.* 101, 175–178. doi: 10.1016/j.econlet.2008.07.015

Koenigs, M., and Tranel, D. (2008). Prefrontal cortex damage abolishes brand-cued changes in cola preference. *Soci. Cogn. Affect. Neurosci.* 3, 1–6. doi: 10.1093/scan/nsm032

Li, J., Liu, X., Yin, X., Wang, G., Niu, X., and Zhu, C. (2018). Transcranial direct current stimulation altered voluntary cooperative norms compliance under equal decision-making power. *Front. Hum. Neurosci.* 12:265. doi: 10.3389/fnhum.2018.00265

Li, J., Yin, X., Li, D., Liu, X., Wang, G., and Qu, L. (2017). Controlling the anchoring effect through transcranial direct current stimulation (tDCS) to the right dorsolateral prefrontal cortex. *Front. Psychol.* 8:1079. doi: 10.3389/fpsyg.2017.01079

Lindbeck, A. (1997). Incentives and social norms in household behavior. *Am. Econ. Rev.* 87, 370–377.

Liu, X., Li, J., Wang, G., Yin, X., Li, S., and Fu, X. (2017). Transcranial direct current stimulation of the rLPFC shifts normative judgments in voluntary cooperation. *Neuroscience* doi: 10.1016/j.neulet.2017.10.020 [Epub ahead of print].

Luo, J., Ye, H., Zheng, H., Chen, S., and Huang, D. (2017). Modulating the activity of the dorsolateral prefrontal cortex by tDCS alters distributive decisions behind the veil of ignorance via risk preference. *Behav. Brain Res.* 328, 70–80. doi: 10.1016/j.bbr.2017.03.045

Marshall, L., Mölle, M., Siebner, H. R., and Born, J. (2005). Bifrontal transcranial direct current stimulation slows reaction time in a working memory task. *BMC Neurosci.* 6:23. doi: 10.1186/1471-2202-6-23

Meesen, R. L., Thijs, H., Leenus, D. J., and Cuypers, K. (2014). A single session of 1 mA anodal tDCS-supported motor training does not improve motor performance in patients with multiple sclerosis. *Restor. Neurol. Neurosci.* 2, 293–300. doi: 10.3233/RNN-130348

Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167

Montague, P. R., and Lohrenz, T. (2007). To detect and correct: norm violations and their enforcement. *Neuron* 56, 14–18. doi: 10.1016/j.neuron.2007.09.020

Moran, R. J., Campo, P., Symmonds, M., Stephan, K. E., Dolan, R. J., and Friston, K. J. (2013). Free energy, precision and learning: the role of cholinergic neuromodulation. *J. Neurosci.* 33, 8227–8236. doi: 10.1523/JNEUROSCI.4255-12.2013

Nee, D. E., and D'Esposito, M. (2016). The hierarchical organization of the lateral prefrontal cortex. *eLife* 5:e12112. doi: 10.7554/eLife.12112

Nitsche, M. A., Cohen, L. G., Wassermann, E. M., Priori, A., Lang, N., Antal, A., et al. (2008). Transcranial direct current stimulation: State of the art 2008. *Brain Stimul.* 1, 206–223. doi: 10.1016/j.brs.2008.06.004

Ramsøy, T. Z., Skov, M., Macoveanu, J., Siebner, H. R., and Fosgaard, T. R. (2015). Empathy as a neuropsychological heuristic in social decision-making. *Soc. Neurosci.* 10, 179–191. doi: 10.1080/17470919.2014.965341

Reif, C., Rübbelke, D., and Löschel, A. (2017). Improving voluntary public good provision through a non-governmental, endogenous matching mechanism: experimental evidence. *Environ. Resour. Economics* 67, 559–589. doi: 10.1007/s10640-017-0126-7

Reuben, E., and Riedl, A. (2009). Enforcement of contribution norms in public good games with heterogeneous populations. *Games Econ. Behav.* 77, 122–137. doi: 10.1016/j.geb.2012.10.001

Reuben, E., and Riedl, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations ☆. *Games Econ. Behav.* 77, 122–137. doi: 10.1016/j.geb.2012.10.001

Riva, P., Romero Lauro, L. J., Dewall, C. N., Chester, D. S., and Bushman, B. J. (2015). Reducing aggressive responses to social exclusion using transcranial direct current stimulation. *Soc. Cogn. Affect. Neurosci.* 10, 352–356. doi: 10.1093/scan/nsu053

Ruff, C. C., Ugazio, G., and Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science* 342, 482–484. doi: 10.1126/science.1241399

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755–1758. doi: 10.1126/science.1082976

Sanfey, A. G., Stallen, M., and Chang, L. J. (2014). Norms and expectations in social decision-making. *Trends Cogn. Sci.* 18, 172–174. doi: 10.1016/j.tics.2014.01.011

Sellaro, R., Güroğlu, B., Nitsche, M. A., Wildenberg, W. P., Massaro, V., Durieux, J., et al. (2015). Increasing the role of belief information in moral judgments by stimulating the right temporoparietal junction. *Neuropsychologia* 77, 400–408. doi: 10.1016/j.neuropsychologia.2015.09.016

Spiller, J., Ufert, A., Vetter, P., and Ulrike, W. (2016). Norms in an asymmetric Public Good experiment. *Econ. Lett.* 142, 35–44. doi: 10.1016/j.econlet.2016.01.014

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., and Fehr, E. (2007). The Neural Signature of Social Norm Compliance. *Neuron* 56, 185–196. doi: 10.1016/j.neuron.2007.09.011

Tankard, M. E., and Paluck, E. L. (2016). Norm perception as a vehicle for social change. *Soc. Issues Policy Rev.* 10, 181–211. doi: 10.1111/sipr.12022

Utz, K. S., Dimova, V., Oppenlander, K., and Kerkhoff, G. (2010). Electrified minds: transcranial direct current stimulation (tDCS) and galvanic vestibular stimulation (GVS) as methods of non-invasive brain stimulation in neuropsychology—a review of current data and future implications. *Neuropsychologia* 48, 2789–2810. doi: 10.1016/j.neuropsychologia.2010.06.002

Van't, W. M., Kahn, R. S., Sanfey, A. G., and Aleman, A. (2005). Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making. *Neuroreport* 16, 1849–1852. doi: 10.1097/01.wnr.0000183907.08149.14

Willis, M. L., Murphy, J. M., Ridley, N. J., and Vercammen, A. (2015). Anodal tDCS targeting the right orbitofrontal cortex enhances facial expression recognition. *Soc. Cogn. Affect. Neurosci.* 10, 1677–1683. doi: 10.1093/scan/nsv057

Xiang, T., Lohrenz, T., and Montague, P. R. (2013). Computational substrates of norms and their violations during social exchange. *J. Neurosci.* 33, 1099–1108a. doi: 10.1523/JNEUROSCI.1642-12.2013

Yin, Y., Yu, H., Su, Z., Zhang, Y., and Zhou, X. (2017). Lateral prefrontal/orbitofrontal cortex has different roles in norm compliance in gain and loss domains: a transcranial current stimulation (tDCS) study. *Eur. J. Neurosci.* 46, 2088–2095. doi: 10.1111/ejn.13653

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Neuroscience

CORRECTION
published: 16 August 2019
doi: 10.3389/fnins.2019.00849

Corrigendum: Transcranial Direct Current Stimulation of the Right Lateral Prefrontal Cortex Changes *a priori* Normative Beliefs in Voluntary Cooperation

Jianbiao Li[1,2], Xiaoli Liu[1], Xile Yin[3], Shuaiqi Li[1], Pengcheng Wang[4], Xiaofei Niu[1] and Chengkang Zhu[1]*

[1] *China Academy of Corporate Governance, Reinhard Selten Laboratory, Business School, Nankai University, Tianjin, China,* [2] *Department of Economics and Management, Nankai University Binhai College, Tianjin, China,* [3] *School of Business Administration, Zhejiang Gongshang University, Hangzhou, China,* [4] *Business School, Tianjin University of Finance and Economics, Tianjin, China*

Keywords: *a priori* normative beliefs, voluntary cooperation, identity, rLPFC, transcranial direct current stimulation

A Corrigendum on

Transcranial Direct Current Stimulation of the Right Lateral Prefrontal Cortex Changes *a priori* Normative Beliefs in Voluntary Cooperation
by Li, J., Liu, X., Yin, X., Li, S., Wang, P., Niu, X., et al. (2018). Front. Neurosci. 12:606. doi: 10.3389/fnins.2018.00606

In the original article, there was an error. The participants in the experiment were insufficiently described.

A correction has been made to the ***Materials and methods***, subsection ***Subjects***:

"The subjects of this experiment were the same as Liu et al. (2017) and Li et al. (2018). A total of 83 healthy subjects (recruited from Nankai University students; 41 females and 42 males ranging from 20 to 30 years old) were kept in the sample. None of them had suffered from any neurological or psychiatric disorders. One participant in the anodal stimulation treatment felt discomfort, and we terminated the experiment. Participants randomly divided into three treatments, namely, cathodal ($n = 28$, 12 males), anodal ($n = 27$, 18 males), and sham ($n = 28$, 12 males) stimulation. All the participants had no ex-ante knowledge of neurological (tDCS) or PG tasks, and all voluntarily joined this study with informed consents. The experiment was performed in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of Business of Nankai University. All these 83 participants reported no adverse side effects (e.g., pain on the scalp or headaches) after the experiment."

In addition, the experiment procedure was insufficiently described.

A correction has been made to the ***Materials and methods***, subsection ***Tasks and Procedure***, *paragraph three*:

"The payoff function of PG was $\pi_i = X_i - x_i + 0.6\sum_{i=1}^{4} x_i$, where X_i was the endowment, x_i was the contribution, and $\sum_{i=1}^{4} x_i$ was the sum contributions of participants from the same group. At the beginning of each trial, the subjects were informed of their identity types (A1, A2, B1, and B2). Then they were asked to answer questions related to beliefs about themselves, voluntary cooperative level and beliefs about others. We did not focus on the beliefs about themselves and voluntary cooperative level in the current study. However, we have emphatically discussed them

OPEN ACCESS

Approved by:
Frontiers Editorial Office,
Frontiers Media SA, Switzerland

****Correspondence:***
Xiaoli Liu
liuxiaoli0510@126.com

Specialty section:
This article was submitted to
Decision Neuroscience,
a section of the journal
Frontiers in Neuroscience

Received: *17 July 2019*
Accepted: *30 July 2019*
Published: *16 August 2019*

Citation:
Li J, Liu X, Yin X, Li S, Wang P, Niu X and Zhu C (2019) Corrigendum: Transcranial Direct Current Stimulation of the Right Lateral Prefrontal Cortex Changes a priori Normative Beliefs in Voluntary Cooperation.
Front. Neurosci. 13:849.
doi: 10.3389/fnins.2019.00849

in Liu et al. (2017) and Li et al. (2018), respectively. In this paper, we focused on the beliefs about others which were tested by pg.belief questions and norm.belief questions:"

The authors apologize for these errors and state that they do not change the scientific conclusions of the article in any way. The original article has been updated.

REFERENCES

Li, J., Liu, X., Yin, X., Wang, G., Niu, X., and Zhu, C. (2018). Transcranial direct current stimulation altered voluntary cooperative norms compliance under equal decision-making power. *Front. Hum. Neurosci.* 12:265. doi: 10.3389/fnhum.2018.00265

Liu, X., Li, J., Wang, G., Yin, X., Li, S., and Fu, X. (2017). Transcranial direct current stimulation of the rLPFC shifts normative judgments in voluntary cooperation. *Neuroscience.* doi: 10.1016/j.neulet.2017.10.020

ORIGINAL RESEARCH
published: 12 June 2018
doi: 10.3389/fpsyg.2018.00932

Fourth-Party Evaluation of Third-Party Pro-social Help and Punishment: An ERP Study

Jianbiao Li[1,2], Shuaiqi Li[1]*, Pengcheng Wang[3], Xiaoli Liu[1], Chengkang Zhu[1], Xiaofei Niu[1], Guangrong Wang[4] and Xile Yin[1]

[1] Reinhard Selten Laboratory, China Academy of Corporate Governance, Business School, Nankai University, Tianjin, China, [2] Nankai University Binhai College, Tianjin, China, [3] International Business School, Tianjin University of Finance and Economics, Tianjin, China, [4] Neural Decision Science Laboratory, Weifang University, Weifang, China

Pro-social behaviors have been adequately studied by neuroscientists. However, few neural studies have focused on the social evaluation of pro-social behaviors, and none has compared the neural correlates of different pro-social decision evaluations. By fourth-party evaluation of third-party punishment/help dictator game paradigm, we explored the third-party pro-social behaviors and derived feedback-related negativity (FRN) from the electroencephalogram. Different from previous event-related potentials (ERP) studies, we simultaneously focused on two different third-party pro-social behaviors, which were called third-party help and third-party punishment. For the first time, we compared the different neural processes of fourth-party evaluation on third-party help and punishment. Behavioral results showed that fourth-party bystanders appreciated the help behavior of the third party even more than the punishment behavior. ERP results revealed that fourth-party bystanders' FRN amplitudes were modulated by the third-party behaviors. Under the assignment condition (70:30) with help/punishment magnitude 45 and (90:10) with magnitude 80, the third-party help elicited a larger FRN than third-party punishment; whereas under the condition (90:10) with help/punishment magnitude 45, the difference between FRN amplitudes disappeared. These results indicated that fourth-party bystanders ultimately agreed more with helpful third parties; however, after they witnessed the norm violation, they expected the third parties to punish the norm violators immediately. This phenomenon appears only when the third-party actors can achieve justice between norm violators and victims.

Keywords: pro-social behaviors, fourth-party evaluation, feedback-related negativity, third-party help, third-party punishment

OPEN ACCESS

Edited by:
Yefeng Chen,
Zhejiang University, China

Reviewed by:
Francesca Marina Bosco,
Università degli Studi di Torino, Italy
Jan Antfolk,
Åbo Akademi University, Finland
Yi-Wen Wang,
Tianjin Normal University, China

****Correspondence:***
Shuaiqi Li
clodelee@126.com

Specialty section:
This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Psychology

Received: *28 September 2017*
Accepted: *22 May 2018*
Published: *12 June 2018*

Citation:
Li J, Li S, Wang P, Liu X, Zhu C, Niu X, Wang G and Yin X (2018) Fourth-Party Evaluation of Third-Party Pro-social Help and Punishment: An ERP Study. Front. Psychol. 9:932. doi: 10.3389/fpsyg.2018.00932

INTRODUCTION

The evaluation of pro-social behaviors accurately reflects the ethical standards of a society. Behavioral and experimental economics have achieved some insights on the social evaluation of pro-social behaviors using the experimental paradigm, in which the fourth-party bystanders may evaluate third-party help or third-party punishment behaviors (Raihani and Bshary, 2015). Subsequently, they found that, on one hand, third parties who took punishment action on selfish dictators or helped victims were rewarded by bystanders more frequently than third parties who did not respond to a selfish dictator or a victim. On the other hand, third-party helpers were more likely to be rewarded than third-party punishers.

Neuroscience studies have not explored the neuronal mechanisms underlying such behavioral outcomes. Existing neuroscience studies have examined the motivations, brain processes, and even genetic factors of third parties who punished norm violators (Strobel et al., 2011; Qu et al., 2014) and helped the victims (Hu et al., 2015). These studies also examined the effects of situations and individual differences or individual heterogeneity on these brain processes (Knoch et al., 2010; Sun et al., 2015; Morese et al., 2016; Mothes et al., 2016). However, none have focused on the behavioral and neurophysiological foundations of how fourth-party bystanders, also called social publics, perceived and evaluated these third-party pro-social behaviors. The current study aims to classify the brain processes of fourth-party bystanders when evaluating third-party pro-social behaviors by assessing neuronal markers (electroencephalogram: EEG). It also investigates the neural differences between the evaluations of pro-social help and punishment.

In the situation of norm violation, pro-social behaviors are actions that are executed by third parties and driven by their other regarding preferences (Buchan et al., 2006). Third parties may be concerned with two potential justice targets when thinking about achieving justice and taking pro-social actions (Gromet and Darley, 2009). According to the targets of other-regarding, third-party pro-social behaviors can be divided into two kinds: helping victims when they demonstrate compensatory concerns or punishing norm violators when they demonstrate punitive concerns (Leliveld et al., 2012; Gummerum et al., 2016).

Psychological and behavioral studies have investigated the motivational structure of third-party behaviors using several empirical and ingenious experimental paradigms (Fehr and Gächter, 2002; Boyd et al., 2003; Fehr and Fischbacher, 2003, 2004; Leliveld et al., 2012). In these studies, the dominant motive of third-party punishment is to maintain the social norm and benefit all the members of our human society. Third-party pro-social behaviors are best accounted for by the hypothesis that people promote the welfare of others as an ultimate end and not by alternative hypotheses that treat these behaviors as instrumental toward ulterior benefits, such as future reciprocation or gaining social approval (Fehr and Fischbacher, 2004).

However, from an evolutionary perspective, reciprocity and reputation cannot be excluded from third-party behaviors (Raihani and Bshary, 2015). Numerous theorists have shown that pro-social help, which comprises actions that benefit others at one's own expense, can be sustained if help behaviors are made visible and the helper will receive helping in return (Milinski et al., 2001; Seinen and Schram, 2006; Tomasello and Vaish, 2013). As for third-party punishment, punitive reputation may play a crucial role in motivating third parties to take punishment actions. Moreover, individuals cooperate because the threat of punishment makes it beneficial for them to do so (dos Santos et al., 2011, 2013). Punishment also plays the role of a signal that shows that the punisher cares about others, is trustworthy, and shows sympathy (Ye et al., 2011; Jordan and Rand, 2017). Punishment can lead to long-term benefits if it influences the punisher's reputation, thereby making the punisher more likely to receive help in future interactions (Jordan et al., 2016).

Neuroscience studies recently started to investigate the brain processes of third parties involved in pro-social behaviors. The pro-social decision-making process is associated with activity in the large-scale nervous system, which includes multiple prefrontal, limbic, and subcortical regions (Sanfey et al., 2003; Wang et al., 2016b; Li et al., 2017). Strobel et al. (2011) showed that third-party punishment elicited stronger activation in the ventral striatum compared with that when no punishment is implemented. They also found that when punishment occurs, the activity of the left dorsal lateral prefrontal cortex is weaker than when no punishment occurs. Hu et al. (2015) revealed that both third-party help and punishment activates the bilateral striatum. They also found that third-party help and punishment involves two different networks; specifically, third-party help involves the bilateral striatum and the right lateral prefrontal cortex, and third-party punishment involves the bilateral striatum and the left lateral prefrontal cortex as well as ventral medial prefrontal cortex. Recently, David et al. (2017) further investigated the different neural mechanisms underlying third-party help and punishment. They revealed that the dorsal anterior cingulate cortex showed higher response during the help (vs. punishment) choice when the (un-)fairness of the proposer's offer was considered by the participants (i.e., offender-focused).

Several studies have used event-related potential (ERP) technique on assessing the neural processes of third-party behaviors because the use of EEG provides high temporal resolution, which is useful for further investigation on the neural processes of punishment decisions especially over the time course (Mothes et al., 2016). Qu et al. (2014) examined the effect of unfairness degrees and punishment decisions on Ne/ERN amplitudes. They found that the Ne/ERN amplitudes were more negative for not punishment decisions than for punishment decisions. Sun et al. (2015) used a similar experimental paradigm and found a medial frontal negativity (MFN) effect, and this effect was modulated by unfairness levels. Mothes et al. (2016) suggested that the amplitudes of feedback-related negativity (FRN) were more pronounced when participants witnessed unfair offers. Hence, MFN (including ERN and FRN) amplitudes, which were related to the activation of anterior cingulate cortex (ACC) (Nieuwenhuis et al., 2004), were sensitive to fairness norm violations, and participants elicited larger MFN effects when they did not take punishment actions.

All these neuroscience studies focused on the motivations and brain processes of third-party pro-social punishers (Qu et al., 2014) and the effects of individual differences, such as altruistic tendency and empathy (Sun et al., 2015; Mothes et al., 2016). These studies, however, did not investigate the third-party pro-social help behaviors or the social evaluations of third-party pro-social behaviors. Loke et al. (2011) partly discussed the evaluation of pro-social help. They found that neural correlates of bystanders' evaluation about pro-social helping behaviors exist. However, the authors mainly focused on the comparison between evaluations of assistance or not when someone obviously needed help or not. Their study did not investigate the differences between pro-social help and pro-social punishment, and their experimental paradigm was not a norm violation paradigm.

We aim to explore the brain processes of the fourth-party evaluation of third-party pro-social behaviors under the situation of norm violation. Specifically, we attempt to investigate the neural differences between the evaluations of pro-social help and punishment by answering the following question: Do bystanders always consider helping victims (or punishing norm violators) a better choice than punishing (or helping)?

FRN and Forth-Party Expectation on Third-Party Pro-social Behaviors

To this end, we used the ERP technology and an adopted third-party punishment/help dictator game paradigm, in which a fourth-party evaluator is added. The high temporal resolution of ERP allowed us to catch the initial psychological processes of fourth-party bystanders after witnessing the third-party behaviors. In the ERP analyses, we focused on the FRN, which is referred to as a negative-going ERP peak between 200 and 350 ms (Miltner et al., 1997; Sun et al., 2015) at the front to central recording sites in the vicinity of ACC. The ACC is considered to be sensitive to detecting cognitive conflicts (Liu et al., 2004; Nieuwenhuis et al., 2004). Studies showed that modulation in ACC, as well as DLPFC and lPFC activities following fair and unfair offers of proposers, plays an important role in pro-social help or punishment decisions (Strobel et al., 2011; Hu et al., 2015; David et al., 2017).

Recent EEG studies examined the role of ACC-related FRN or ERN component in pro-social behavior scenarios (Qu et al., 2014; Sun et al., 2015; Mothes et al., 2016). They assumed that the FRN component is an indicator that reflects whether outcomes matched expectations (Oliveira et al., 2007; Sun et al., 2015; Mothes et al., 2016). When outcomes were unexpected, a larger FRN was elicited compared with those in the expected outcomes (Sun et al., 2015). Moreover, some studies demonstrated that FRN was elicited even when participants witnessed other individual's behaviors (Yeung et al., 2005; Koban and Pourtois, 2014). Thus, FRN is a reliable indicator even in the perspective of fourth-party bystanders. We hypothesize that if the fourth party expect third-party actors to punish the norm violators, then third-party punishment will elicit a smaller FRN than third-party help, which goes against the expectation of the fourth party. Conversely, if the fourth party expect third-party help more, the help will elicit a smaller FRN than the punishment.

FRN and Fourth-Party Evaluation Scores

The second point we are interested in is that whether the FRN amplitude characteristics of the fourth party following the third-party actions will predict fourth-party evaluation scores. Given that the ERP has high temporal resolution, studies mostly focus on the characteristic of EEG within 1 s or even 800 ms following the epochs. In such a limited time, few cognitive resources were used in the brain processes of individuals, and thus, they are cognitively constrained (Cappelletti et al., 2011). Therefore, in a dual-system perspective, the individual's deliberative capacity was limited in a short time and their expected actions might be different from the situation when time is sufficient (Blechert et al., 2012). As emotion was considered to be a determining factor of the automatic processes, in a short time, individual's expected behaviors were more possibly modulated by emotional reactions spontaneously (Qu et al., 2014; Yang et al., 2017). Bystanders would rapidly elicit empathic anger at witnessing injustice or harm to someone else, and the empathic anger is considered as a motivation underlying third-party punishment and the expectation of third-party punishment (Fehr and Gächter, 2002; Fehr and Fischbacher, 2004; Batson et al., 2007; Van Doorn et al., 2014; FeldmanHall et al., 2015). However, the evaluation of third-party help requires more cognitive resources (Loke et al., 2011; Erlandsson et al., 2014). Thus, the ERP characteristics, which were extracted in a time shorter than 1 s mostly reflected the brain processes involving third-party punishment evaluation compared with help evaluation. We expect that smaller FRN amplitudes following third-party actions may not always predict higher fourth-party evaluation scores.

We addressed the above issues in two studies. In Study 1, participants witnessed the third party punish the unfair dictator or help the victim receiver, when the third party can reach a fair between the dictator and the victim under somewhat unfair offer condition and cannot reach a fair under extremely unfair offer condition. In Study 2, participants can turn the unfair offer into a fair one under extremely unfair offer condition.

MATERIALS AND METHODS

Study 1: Third Party Can Reach a Fair Under Somewhat Unfair Condition and Cannot Do That Under Extremely Unfair Condition

Participants

A total of 24 healthy volunteers from Nankai University participated in this study for monetary compensation. Three subjects were excluded due to technical problems and severe artifacts in the EEG data. The brain activities of 21 subjects (13 women, 8 men; mean age = 23.3 years; range = 21–25 years) were fully analyzed. All participants were right-handed and native Chinese speakers. They had normal or corrected-to-normal vision and had no history of psychiatric or neurological disorders. Written informed consent was obtained before we conducted the experiment. The study protocol was approved by the Ethics Committee of Business School of Nankai University.

We used E-Prime experimental program to present the game. Color bars were applied to present the assignments of the dictator and the final payoffs of the dictator and the receiver after the third party took actions. The horizontal viewing angle of each target picture was 3°, and the vertical viewing angle was 1.5°.

Stimuli and Task

We introduced fourth-party bystanders in a third-party punishment of dictator game to adopt a modified paradigm of this game (Raihani and Bshary, 2015; Mothes et al., 2016). In the experiment, the main unit of analysis was defined as a "trial," where four persons were referred to as "dictator," "receiver," "third party," and "bystander". However, neutral terms (P1, P2, P3, and P4) were used in the experimental instructions. Participants were assigned the role of fourth-party bystander. In each trial, they first witnessed the decision of the dictator who distributed 100 Yuan between himself and the receiver. Then they witnessed the decisions of third parties that can turn the unfair offer into a fair one. Thus, third parties were given a starting endowment of 50 Yuan. Third parties were given the opportunity to adjust the initial distribution. They can pay 15 Yuan to reduce the dictator's bonus by 45 (i.e., TPP) or to increase the receiver's bonus by 45 (i.e., TPH).

Subsequently, the participants can rate the third-party actions using a five-point Likert scale. The score determined the magnitude that the participants agreed with the third-party actions. A score of "1" indicated that the participant strongly disagreed with the third-party's decision, "5" indicated that the participant strongly agreed with the third-party's decision, and "3" was a neutral score.

We presented two predetermined assignments (70:30) and (90:10). In the (70:30) situation, the TPH or TPP actions can achieve almost absolute fairness between the dictator and the receiver. On the one hand, if the third party punished the dictator, the payoffs of the dictator and the receiver were 25 and 30, respectively. On the other hand, if the third party helped the receiver, the payoffs of the dictator and the receiver were 70 and 75, respectively. In the (90:10) situation, third-party behaviors could hardly achieve fairness between the dictator and the receiver. The TPH action resulted in payoffs (90:55) and the TPP in (45:10). The 2 × 2 conditions were fulfilled to compare the ERP responses with the fourth-party evaluation of TPH and TPP, which realized or at least attempted to realize the fairness between the dictator and the receiver.

With each condition containing 40 trials, a total of 160 experimental trials were performed. We randomly interspersed 40 control trials between these 160 trials to prevent anticipation effect of the participants. In these control trials, the decision of the third party was neither to punish the dictator nor to help the receiver. We did not analyze the EEG of these control trials.

Procedures

Electroencephalogram recording was conducted in a small, sound-attenuated, and electrically shielded chamber. After EEG electrodes were attached, participants sat in a comfortable chair approximately 100 cm in front of a 23-inch computer monitor. Before the tasks began, all participants read the instructions carefully and were asked to take one or more 5-trial practice until the tasks were understood. **Figure 1** shows the time course of a single trial. Each trial began with the presentation of a single centrally located white fixation cross for 500 ms. Next, a blank screen was presented for 400–800 ms. Afterwards, the decision of the dictator, that is, to distribute 100 Yuan between himself and the receiver, was presented. Subsequently, decisions of the third party and the payoffs of the dictator and the receiver were presented at the center of the screen for 2000 ms. After the ERP, an evaluation display with five options (1, 2, 3, 4, and 5) was shown until the participants pressed the button of a five-key response pad.

The entire experiment comprised 160 test trials, 40 control trials, and 5 practice trials. Only the test trials were used for ERP analysis. Trials appeared in five blocks of 40 trials. Each block was separated by a break, the duration of which was determined by the participant. All 200 trials were performed within 15–25 min, during which these trials were randomly presented. E-Prime software was used to control the display of the stimuli and the acquisition of behavioral data (Version 2.0, Psychology Software Tools, Inc.).

EEG Acquisition

The EEG was recorded continuously with a 40-channel NuAmps DC amplifier (Compumedics Neuroscan, Inc., Charlotte, NC, United States). According to the International 10–20 System, 32 active Ag/AgCl electrodes were used. The impedances of all electrodes were kept below 10 kΩ. The reference electrode and the ground electrode were positioned at AFz. Electrodes below and above the left eye, as well as those located on the outer canthi of each eye, measured the bipolar vertical and horizontal electro-oculogram activities. Meanwhile, online EEG was digitized at a sampling rate of 1000 Hz using a 22-bit A/D converter.

Further offline processing was performed with Neuroscan Curry Software (Version 7.0.11, Compumedics Neuroscan, Inc., Charlotte, NC, United States). While offline, the reference of EGG signals was reset to the average of the left and right mastoids. Eye-blink artifacts were corrected, and the artifact rejection method excluded epochs with the EEG amplitude of any channel exceeding ±100 μV. The EEG data were band-pass filtered between 0.1 and 30 Hz. Subjects had no fewer than 40 artifact-free epochs in each condition, and the accepted epochs were baseline corrected. For each stimulus, we extracted 1000 ms epochs, with a 200-ms pre-stimulus period used as baseline.

EEG Analysis

The 1000-ms epochs were extracted in the markers "P3 decides to: Punish" and "P3 decides to: Help" starting at 200 ms before presentation of the third-party decisions. Mean amplitudes were then used for the FRN analysis. We found that maximum amplitudes of FRN were obtained at approximately 300 ms after participants witnessed third-party decisions over multiple frontal electrodes by visual inspection of grand averaged waveforms under TPH and TPP conditions. We then selected three electrodes in the midline area (Fz, FCz, and Cz) for statistical analysis (Mothes et al., 2016; Navarro-Cebrian et al., 2016). Previous studies suggested that maximum FRN amplitudes were often observed at mediofrontal electrodes, which corresponded to our observation (Gehring and Willoughby, 2002; Mothes et al., 2016). To further investigate the ERP characteristics, data from these three electrodes in a 270–330-ms time window were used.

For all analyses of variance (ANOVA), p-values were corrected using the Greenhouse–Geisser correction whenever the sphericity assumption has been violated. $p < 0.05$ was considered significant. Significant interaction was analyzed by the simple-effect model. Bonferroni correction was implemented to adjust

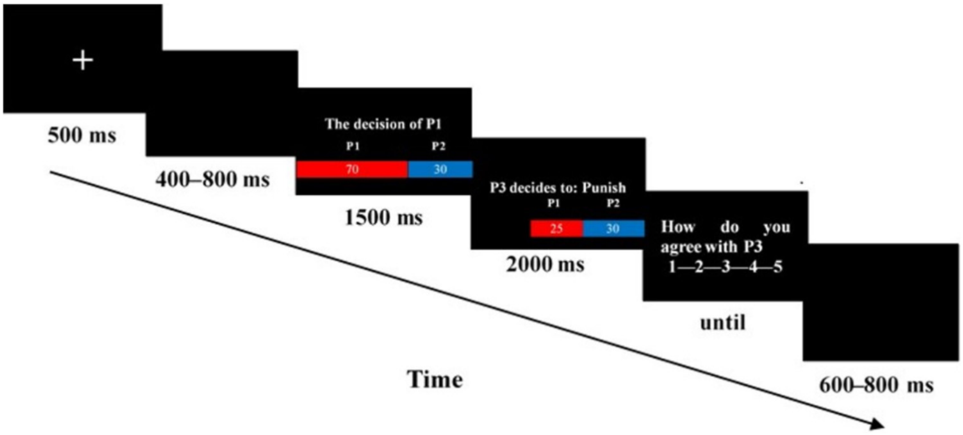

FIGURE 1 | Time course of a single trial in Study 1. Each trial began with a 500-ms fixation point followed by the blank screen, which were randomized between 400 and 800 ms. A screen displaying the decision of P1 was shown for 1500 ms. Then, the stimulus presentation was shown for 2000 ms. Afterwards, an evaluation screen appeared until the participants responded. The intertrial interval was randomized to last between 600 and 800 ms.

for multiple comparisons. Statistics were analyzed with the IBM SPSS 19.0 software.

Study 2: Third Party Can Reach a Fair Under Extremely Unfair Condition

Participants

A total of 19 healthy volunteers (10 women, 9 men; mean age = 22.8 years; range = 19–24 years) from Nankai University participated in Study 2. Their brain activities were all fully analyzed. In contrast to Study 1, numbers, not color bars, were applied to present the assignments of the dictator. The horizontal viewing angle of each target picture was 3°, and the vertical viewing angle was 1.5°.

Task and Procedure

The same game was used as in Study 1, except for the magnitude of TPH and TPP, which changed. In Study 2, third parties were given a starting endowment of 50 Yuan. They can pay 20 Yuan to reduce the dictator's bonus by 80 (i.e., TPP) or to increase the receiver's bonus by 80 (i.e., TPH). The payoffs of dictators can be cut down to 0 but can never be below 0. Thus, third-party actors can achieve fairness between the dictator and the receiver in the (90:10) situation. The TPH action resulted in payoffs (90:90) and the TPP in (10:10).

With each condition (TPH and TPP) containing 50 trials, a total of 100 experimental trials were performed. We randomly interspersed 40 control trials among these 100 experiment trials. In these control trials, allocations (50:50), (65:45), (70:30), (95:5), and (100:0) were included. All trials added up to 140.

The procedure for each trial in Study 2 was the same as that in Study 1. The differences were that we replaced the color bars with numbers and removed the presence of final payoffs. **Figure 2** shows the time course of a single trial in Study 2.

The entire experiment comprised 100 test trials, 40 control trials, and 5 practice trials. Only the test trials were used for ERP analysis. Trials appeared in three blocks of 40 trials and one block of 20 trials. Each block was separated by a break, the duration of which was determined by the participant. All 140 trials were performed within 15–20 min, during which these trials were randomly presented.

EEG Acquisition and Analysis

The EEG acquisition and offline processing were the same as that in Study 1. The epoch selection was also similar. We found that the maximum amplitudes of FRN were obtained at approximately 300 ms after participants witnessed third-party decisions over multiple frontal electrodes by visual inspection of grand averaged waveforms under TPH and TPP conditions. We also selected three electrodes in the midline area (Fz, FCz, and Cz) for statistical analysis. To further investigate the ERP characteristics, data from these three electrodes in a 270–350-ms time window were used, which was different from the procedure in Study 1.

RESULTS

Study 1: Achieving Fairness Under (70:30) and Not Achieving Fairness Under (90:10)

Behavior Results

For the assignment (70:30), 85.71% (18/21) of the fourth-party bystanders evaluated the TPH better than TPP, 14.29% (3/21) of the fourth party considered that TPH was nearly the same as TPP, and none of the bystanders preferred TPP. For the assignment (90:10), 66.67% (14/21) of the fourth-party bystanders rated the TPH higher, 4.76% (1/21) of the fourth party considered that TPH was nearly the same as TPP, and 28.57% (6/21) of the bystanders preferred TPP.

The fourth-party evaluation was performed using 2 × 2 repeated measures ANOVA with factors assignments (70:30, 90:10) and third-party behaviors (TPH vs. TPP). Significant effects [$F(1,832) = 41.672$, $p < 0.001$] and [$F(1,832) = 736.341$,

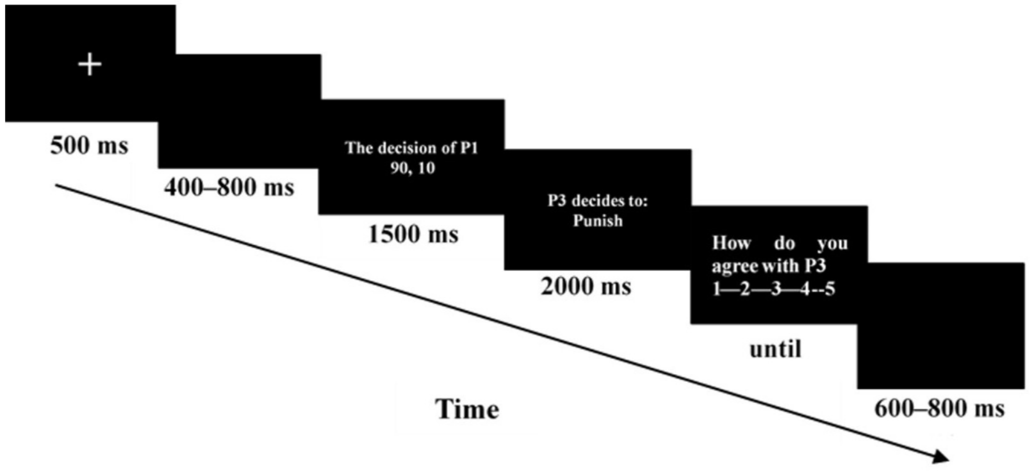

FIGURE 2 | Time course of a single trial in Study 2. Each trial began with a 500-ms fixation point followed by blank screen, which were randomized between 400 and 800 ms. A screen displaying the decision of P1 was shown for 1500 ms. Then, the stimulus presentation was shown for 2000 ms. Afterwards, an evaluation screen appeared until the participants responded. The intertrial interval was randomized to last between 600 and 800 ms.

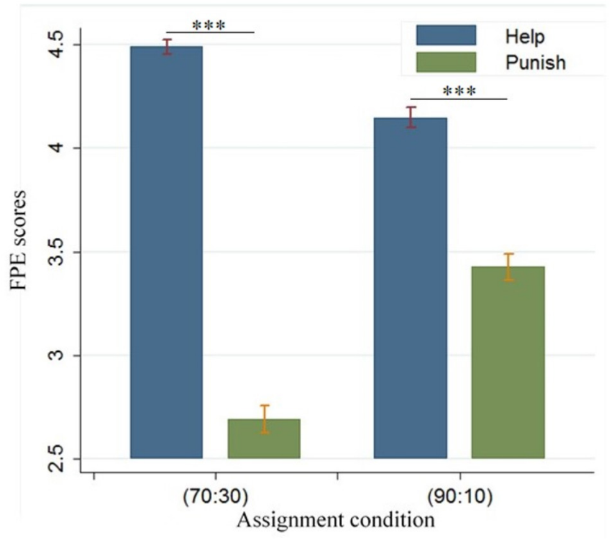

FIGURE 3 | Behavioral results of fourth-party evaluation. Fourth-party evaluation scores are differentiated among (70:30, 90:10) × (TPH, TPP) conditions. "*" represents that $p < 0.1$, "**" represents that $p < 0.05$, and "***" represents that $p < 0.01$.

$p < 0.001$] were yielded for factor condition (70:30, 90:10) and (TPP, TPH) (see **Figure 3**). A significant interaction effect occurred between first-party assignments and third-party behaviors [$F(1,832) = 312.219$, $p < 0.001$]. The fourth-party evaluation of TPH (mean = 4.49, $sd = 0.725$) was higher compared with the fourth-party evaluation of TPP (mean = 2.69, $sd = 1.389$) under the condition of (70:30) [$t(832) = 39.534$, $p < 0.001$]. We also found that the fourth-party evaluation of TPH (mean = 4.15, $sd = 1.050$) was higher than that of TPP (mean = 3.43, $sd = 1.329$) under the condition of (90:10) [$t(839) = 10.301$, $p < 0.001$].

TABLE 1 | Regression results of fourth-party evaluation.

Condition Variables	A		B
	(70:30)	(90:10)	(90:10)
	Forth-party evaluation		Forth-party evaluation
Third-party behaviors	−1.809***	−0.722*	−1.675***
	(0.255)	(0.382)	(0.313)
Constant	4.499***	4.148***	4.652***
	(0.135)	(0.197)	(0.0924)
Observations	1,656	1,680	1,883
R-squared	0.375	0.056	0.370

A is the cluster regression result of study 1. B is the cluster regression result of study 2. "Third-party behaviors" is a dumb variable in which "0" represents TPH and "1" represents TPH. "" represents that p < 0.1, "**" represents that p < 0.05, and "***" represents that p < 0.01.*

We performed cluster regressions under condition (70:30) and (90:10) separately (see **Table 1**, A). In these regressions, we used the third-party behaviors as independent variables, the forth-party evaluation as dependent variables and participants as cluster indicators. We found the results were similar with those of ANOVA, the forth-party evaluation was more positive when third-party behavior was TPH compared to TPP under the condition of (70:30) (coef. = −1.809, $p < 0.001$), and the difference was also found under condition of (90:10) (coef. = −0.721, $p = 0.073$). The fourth-party bystanders agreed to the help behavior of the third party even more than punishment.

ERP Results

FRN: 270–330 ms

We assessed the ERPs evoked by TPH and TPP under the assignment conditions of (70:30) and (90:10). We submitted stimulus-induced activity in the FRN time range to 2 × 2 × 3 repeated measures ANOVA with the factors of first-party assignments (70:30, 90:10), third-party behaviors (TPH vs. TPP), and sites (Fz, FCz, and Cz). However, no significant differences were found between the (70:30) and (90:10) conditions [$p > 0.05$] and among electrodes [$p > 0.05$]. A significant difference was found between the TPH and TPP conditions [$F(1,20) = 16.652$, $p = 0.001$]. A significant interaction effect occurred between first-party assignments and third-party behaviors [$F(1,20) = 7.794$, $p = 0.011$]. We divided the data into two parts based on the first-party assignments and examined the difference between TPH and TPP.

FRN: (70:30)

Under the condition of (70:30), we conducted 2 × 3 repeated measures ANOVA with factors of third-party behaviors (TPH vs. TPP) and sites (Fz, FCz, and Cz). The result showed no significant effect of electrodes and no significant interaction effect between third-party behaviors and electrodes (all $p > 0.05$). Activity in the FRN time range was significantly more negative when the fourth-party bystanders witnessed TPH than when they witnessed TPP, as indicated by a main effect of third-party behaviors [$F(1,20) = 36.571$, $p < 0.001$]. Thus, the typical FRN of fourth-party bystanders was observed, and its topography is illustrated in **Figure 4**.

FRN: (90:10)

Under the condition of (90:10), we conducted 2 × 3 repeated measures ANOVA with factors of third-party behaviors (TPH vs. TPP) and sites (Fz, FCz, and Cz). The result showed no significant effect of electrodes and no significant interaction effect between third-party behaviors and electrodes (all $p > 0.05$). No significant difference between TPH and TPP was also observed [$p > 0.05$]. Statistical *post hoc* tests showed that the different waves were not significantly different from zero on all three electrodes (all $p > 0.05$).

Study 2: Achieving Fairness Under (90:10)

Behavior Results

The fourth-party evaluation was analyzed using ANOVA with third-party behaviors (TPH vs. TPP) under allocation (90:10). Significant effects [$F(1,1881) = 1107.06$, $p < 0.001$] were obtained for factor condition (TPP and TPH). The fourth-party evaluation of TPH (mean = 4.651, *sd* = 0.679) was higher compared with the fourth-party evaluation of TPP (mean = 2.977, *sd* = 1.389) [$t(1881) = 33.273$, $p < 0.001$]. The fourth-party bystanders agreed to the help behavior of the third party even more than punishment. We also performed a cluster regression in which the third-party behaviors were independent variables, the forth-party evaluation were dependent variables and the identities of participants were cluster indicators (see **Table 1**, B). We found that the forth-party evaluation decreased when the third-party behaviors changed from TPH to TPP (coef. = −1.675, $p < 0.001$).

ERP Results

FRN: 270–350 ms

We assessed the ERPs evoked by TPH and TPP under the assignment conditions of (90:10) and the punishment or help magnitude 80 condition. We submitted stimulus-induced activity in the FRN time range to 2 × 3 repeated measures ANOVA with third-party behaviors (TPH vs. TPP) and sites (Fz, FCz, and Cz). A significant difference was found between the TPH and TPP conditions [$F(1,18) = 26.134$, $p < 0.001$]. However, no significant differences were found among electrodes, and no significant interaction effect occurred between third-party behaviors and electrodes (all $p > 0.05$). The typical FRN of fourth-party bystanders in Study 2 was observed and its topography is illustrated in **Figure 5**.

DISCUSSION

Interest in behavioral and neurophysiological research on pro-social behaviors has been growing in recent years. However, very few related studies focused on the issue of how social publics perceived and evaluated pro-social behaviors and the neural correlates. To understand the possible explanatory and modulatory factors of fourth-party evaluation of pro-social behaviors, our study examines the interplay between pro-social behavior types (i.e., third-party help and punishment), fourth-party evaluations, and the FRN component. Our study is the first to investigate the ERP correlates of social public evaluations on different kinds of pro-social behaviors.

First Expectation: Agreement/Disagreement of Fourth-Party With the Third-Party Help/Punishment

In line with our first expectations, the behavioral data demonstrated that the fourth-party participants showed different feelings regarding third-party help and punishment. A comparison of the evaluation scores indicated that fourth parties agreed to third-party help more than punishment regardless of the first-party assignment decisions and the punishment or help magnitudes. The results corresponded to the concept of Raihani and Bshary (2015), who first discussed the fourth-party evaluation on third-party behaviors.

We examined the relation between the fourth-party evaluation, pro-social behavior types, and the FRN component associated with ACC-dependent responses toward unexpected outcomes (Hauser et al., 2014). The ERP data illustrated that a more negative FRN was exhibited by third-party help compared with punishment between 270 and 330 ms under assignment condition (70:30) with punishment/help magnitude of 45 and assignment condition (90:10) with punishment/help magnitude 80. Given that previous studies found that larger FRN amplitudes were observed for unexpected or surprise events (Oliveira et al., 2007; Sun et al., 2015; Mothes et al., 2016), we can deduce that

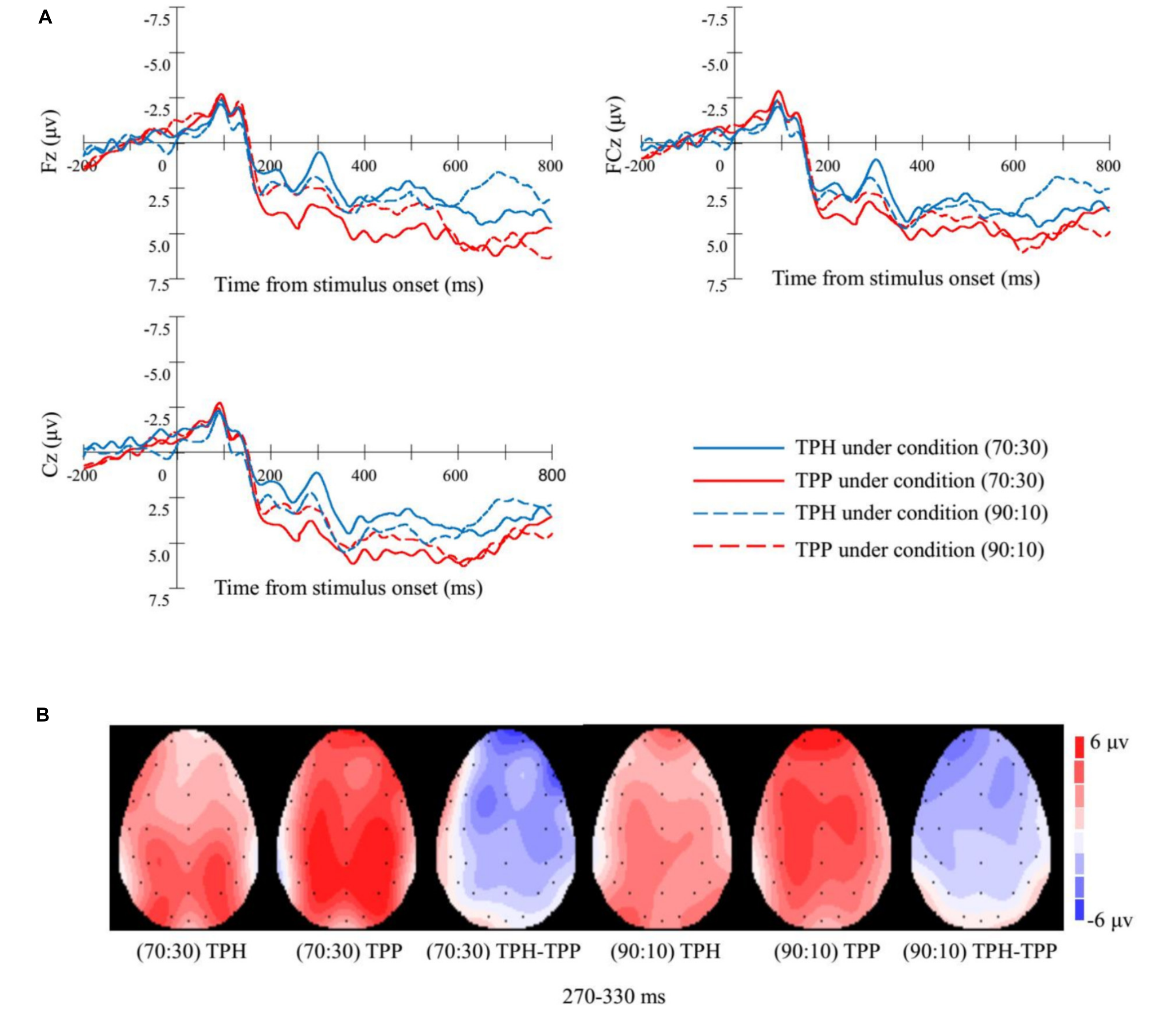

FIGURE 4 | Event-related potentials (ERP) waveforms and topographic voltage maps of study 1. **(A)** Grand average ERP waveforms at selected electrodes of Study 1. ERP differentiated between TPP and TPH conditions: FRN between 270 and 330 ms at central sites (Fz, FCz, and Cz). **(B)** Topographic voltage maps of mean amplitude of Study 1. Voltage maps showing the scalp distributions with significant effects of (70:30, 90:10) × (TPH, TPP) obtained in the 270–330 ms epochs, and the difference between TPH and TPP (i.e., TPH-TPP).

third-party punishment is more likely to be expected by the fourth-party bystanders than third-party help.

Feedback-related negativity has been substantially investigated in the third-party punishment of dictator game and similar paradigm, such as ultimatum game (Boksem and De Cremer, 2010; Wu et al., 2011; Qu et al., 2013; Sun et al., 2015; Mothes et al., 2016). The FRN was extracted immediately after the receiver realized the fair or unfair offer in the ultimatum game paradigm, or the third-party actor witnessed the assignment of the dictator in the third-party punishment of dictator game. These studies concluded that unfair offers or assignments elicited more pronounced FRN amplitude compared with fair offers. However, we cannot use this idea to interpret our results because in our experimental paradigm, which was adopted from the third-party punishment of dictator game, the FRN was evoked by the third-party behaviors instead of the dictator's offers. Moreover, when the fourth-party bystanders evaluated third-party behaviors, they faced the same assignment from the dictator. Even now, the analysis of FRN composition was still useful in our study. Because except for unfairness, previous studies also found that FRN is sensitive to negative outcomes (Boksem and De Cremer, 2010), others' negative situations (Yeung et al., 2005; Koban and Pourtois, 2014; Wei et al., 2015), or unexpected events (Oliveira et al., 2007; Sun et al., 2015; Mothes et al., 2016). Thus, in the present study, larger FRN values reflected that

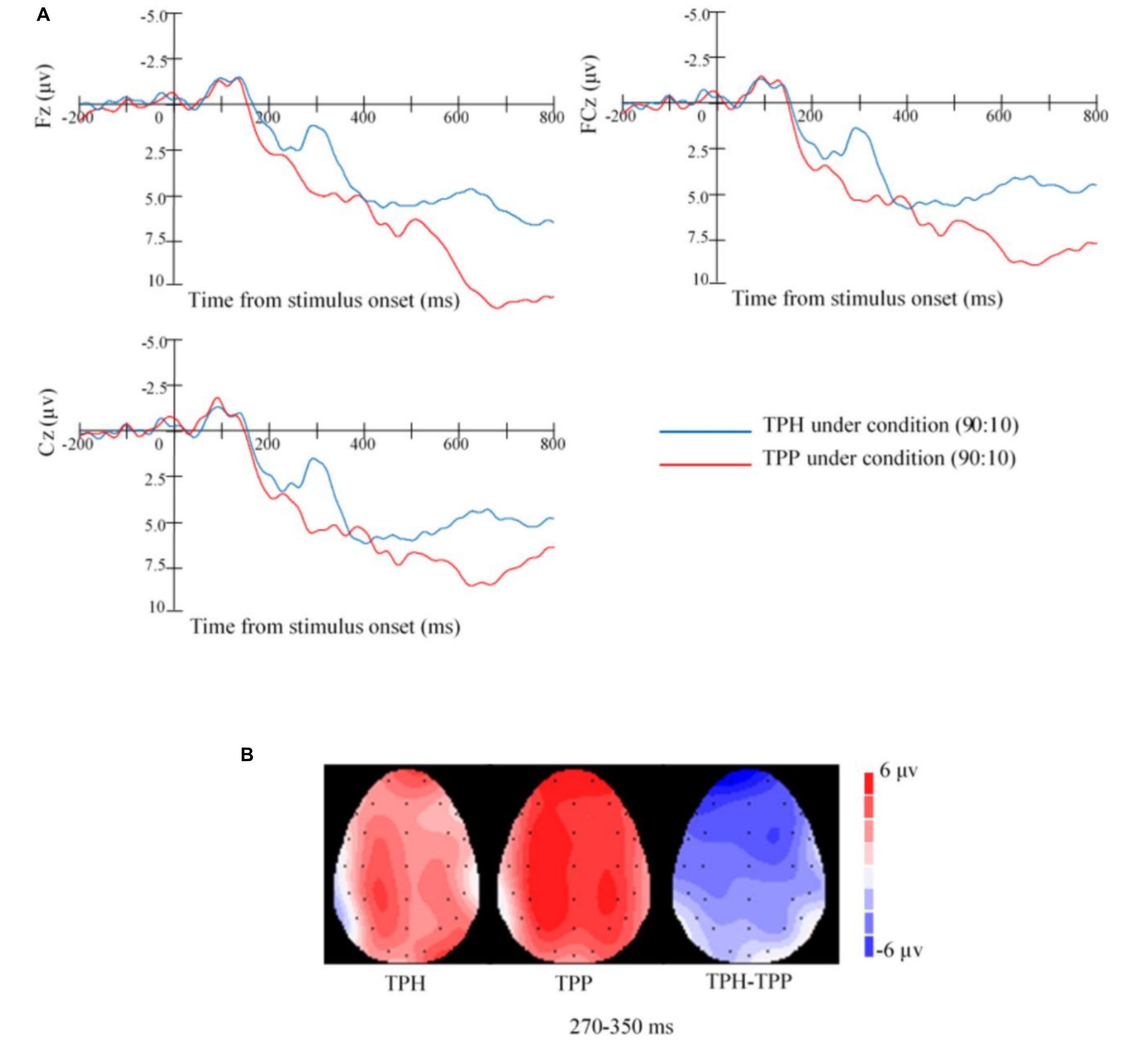

FIGURE 5 | Event-related potentials waveforms and topographic voltage maps of study 2. **(A)** Grand average ERP waveforms at selected electrodes of Study 2. ERP waveforms that are differentiated between TPP and TPH conditions: FRN between 270 and 350 ms at central sites (Fz, FCz, and Cz). **(B)** Topographic voltage maps of mean amplitude of Study 2. Voltage maps showing the scalp distributions of significant effects of (TPH, TPP) obtained in the 270–350 ms epochs and the difference between TPH and TPP (i.e., TPH-TPP).

TPH somewhat violated the expectancy of the fourth-party bystanders.

Second Expectation: FRN Characteristics May Not Predict Fourth-Party Evaluation Scores

We found that the ERP result showed that third-party punishment was more likely to be expected by the fourth-party bystanders than third-party help. This result was somewhat not in accordance with the behavioral result, which showed that fourth-party bystanders agreed to third-party help more than to punishment. Accordingly, the final behavioral results showed that fourth-party bystanders agreed to third-party help more, whereas the temporary ERP responses reflected that the bystanders did not expect third-party actors to help the victims at first, which appeared to agree with our second expectation.

The case wherein FRN amplitudes did not predict final behaviors was also observed in some other studies. For example, Mothes et al. (2016) found that larger FRN amplitudes were not associated with third-party punishment. Larger FRN were elicited by unfairness, and those of third parties that would not make pro-social punishment because their levels of involvement were low. Boksem and De Cremer (2010) made another interpretation to this phenomenon; they believed that the ERP technology had high temporal resolution, which can be used to evaluate the processes immediately after the event (fair and unfair assignment). Thus, the FRN, which was locked to the witness of the dictator's decisions, was elicited by the evaluation of the fair or unfair assignment and would not be influenced by the response preparation, which would take place at least several seconds later.

We partly followed the ideas of Boksem and De Cremer (2010) and introduced dual-process system theory to understand why third-party help elicited larger FRN, which indicated that third-party help was against with the expectation of the fourth party but acquired more agreement finally. We suspected that something happened during the process of the subconscious evaluation of third-party pro-social behaviors turning into evaluation behaviors. From a perspective of dual-process system to decision making, the evaluation decision of the fourth party was made by the interaction of two different processing systems, which were called the automatic and the controlled systems (Lieberman et al., 2007; Adolphs, 2009; Qu et al., 2014; Yang et al., 2017). Automatic system was considered to be a fast, spontaneous, short-sighted, and intuitive system, which seldom require cognitive resources. The automatic system was also called the affective system because it was mostly driven by affections or emotions (Cappelletti et al., 2011). By contrast, controlled system requires quantities of cognitive resources; thus it is deliberate, effortful, and slow. Previous studies also called the dual-process system as two-step process system (Cappelletti et al., 2011; Rand et al., 2012; Sun et al., 2015; Yang et al., 2017). Intuitive proposals were made automatically during the first step. In the second step, actors made tradeoffs between the proposals from the first step. In this deliberative step, motivational consideration, social contextual consideration, and quantities of cognitive resources were injected.

The third-party punishment and expectation of third-party punishment might be related to the automatic system or proposals generated in the first step, as they were considered to be driven by emotional factors (Qu et al., 2014). Crockett et al. (2010) believed that impulsive emotional responses induced by unfairness may play an important role in driving third-party punishment (Crockett et al., 2010). Qu et al. (2014) suggested that emotional factors were uncontrolled and automatic when participants made decisions. Emotional factors can provide great power to punish and make third-party punishment decisions spontaneously and unconsciously (Olatunji and Puncochar, 2014). On the contrary, not all punishment behaviors were cogitative and conscious. Punishment tended to be automatically taken when norm violations were witnessed. Sun et al. (2015) found that when participants yearned for fairness during an unfairness experience, a greater MFN was elicited, which was in accordance to the results of other studies (Mothes et al., 2016; Wang et al., 2016a). They also supported the idea of Qu et al. (2014) that third-party punishment, which was an automatic intuitive proposal, occurred in the early stage of the outcome evaluation. Empathic anger, which would rapidly arise upon witnessing injustice or harm to someone else, was considered a key factor for third-party punishment and third-party punishment expectation (Fehr and Gächter, 2002; Fehr and Fischbacher, 2004; Batson et al., 2007; Van Doorn et al., 2014; FeldmanHall et al., 2015). The expectation of third-party punishment was an automatic intuitive proposal and occurred in the early stage after social publics witnessed the unfair assignment decision of the first party (Qu et al., 2014; Sun et al., 2015; Mothes et al., 2016).

Different from the third-party punishment expectations, which were intuitively or subconsciously driven, the decisions and expectations of third-party help were more complex. Except for probably emotional reactions, third-party help was affected by some other factors, such as moral judgment, perceived responsibility or duty to help, or even the perceived utility of helping (Erlandsson et al., 2014). Carlo (2014) believed that pro-social help occurs in social contexts. He suggested that except for the intrinsic processes, such as sympathy, internalized values or principles, or a strong pro-social or moral identity, pro-social help may also be motivated by external or social context concerns (e.g., social approval, social power, and money). Loke et al. (2011) found that reasoning about pro-social help, which was a kind of moral judgment, played an important role in pro-social help evaluation.

Compared with the evaluations of third-party punishment, more cognitive resources were inputted by the bystanders when they evaluated third-party help decisions. Thus, the fourth-party evaluation of third-party help tended to form in the second step of decision-making or social information processes, which involved a more deliberative and controlled phase and were sensitive to social context. Therefore, immediately after witnessing the unfairness decision, fourth-party bystanders may subconsciously expect third-party actors to punish the norm violators. With more cognitive resources and moral or social context concerns related to the third-party help evaluation introduced into the evaluating process, the bystanders' expectations or evaluations tended to change. Social public expected others to punish the norm violators immediately after witnessing a norm violation, but ultimately agreed to helpful third parties more.

Fairness, Help, and Punishment

The present study also found that the effect of third-party behaviors on the FRN of fourth-party bystanders was modulated by first-party assignments. Based on ERP results, we found that the FRN amplitude of third-party help was significantly more negative than that of third-party punishment only when the third-party behaviors can achieve fairness between the dictator and the receiver regardless of the first-party assignment. No significant difference in the FRN amplitude was found between third-party help and punishment when third-party actors cannot safeguard fairness under condition (90:10) with help/punishment magnitude 45. We speculated that under condition (90:10) with help/punishment magnitude 45, the third-party behaviors could hardly achieve justice between norm violators and victims. As a result, bystanders tended to regard third-party help and punishment as same behaviors, and they subconsciously believed that the two behaviors were all bad choices. Only after they realized that the third-party actors can afford to safeguard fairness, they would focus on the difference between third-party help and punishment.

The present study also found that the effect of third-party behaviors on the FRN of fourth-party bystanders was modulated by first-party assignments. Based on ERP results, we found that the FRN amplitude of third-party help was significantly more negative than that of third-party punishment under condition (70:30) with help/punishment magnitude 45 and (90:10) with help/punishment magnitude 80. No significant difference in the FRN amplitude was found between third-party help and punishment under condition (90:10) with help/punishment magnitude 45. These results indicated that first-party assignments and third-party behaviors could not determine the forth-party evaluation separately. Under condition (70:30) with help/punishment magnitude 45 and (90:10) with help/punishment magnitude 80, the third-party actors could achieve fairness between the dictator and the receiver. However, third-party actors cannot safeguard fairness under condition (90:10) with help/punishment magnitude 45. Whether the third-party behaviors could achieve fairness between the dictator and the receiver is a precondition of the neural difference between third-party help evaluation and punishment evaluation.

If the third-party behaviors could hardly achieve justice between norm violators and victims, bystanders tended to regard third-party help and punishment as same behaviors and consider that the two behaviors were all bad choices. Only after they realized that the third-party actors can afford to safeguard fairness, they would focus on the difference between third-party help and punishment. The occurrence of this situation needed to be traced back to the neural processing in our brains. Kahneman (2011) suggested that our brains were tended to be lazy in our daily life to save cognition resources. Our brains preferred to encode the things we perceived into binary categories when cognition resources were limited. This phenomenon was significant in the neural processing reflected by FRN. Holroyd et al. (2006) found that FRN appeared to reflect a binary categorization of the outcomes as either good or not good. The TPH and TPP under condition (90:10) with help/punishment magnitude 45 were all "no good" behaviors for bystanders, because the third-party actors didn't achieve justice between norm violators and victims. The forth-party bystanders would not distinguish between the two bad choices. As a result, the FRN amplitudes of TPP and TPH appeared to be the same under condition (90:10) with help/punishment magnitude 45.

CONCLUSION

This study is the first to examine the neural correlations of the fourth-party evaluation of third-party pro-social behaviors using ERP technology. Behavioral results showed that fourth-party bystanders agreed to third-party help more than to third-party punishment. However, the tendency was decreased with the increase in the unfairness of the first-party assignment. Third-party help elicited more negative FRN amplitude at least under the assignment (70:30) with help/punishment magnitude of 45 and (90:10) condition with help/punishment magnitude 80. Specifically, no difference in the FRN amplitudes was observed under (90:10) with help/punishment magnitude of 45. These results indicated that, although bystanders finally agreed that third party should help the victims more, they expected third-party actors to punish the norm violators immediately after they witnessed the norm violation. However, this phenomenon appeared only when the fourth-party bystanders believed that the third-party actors can safeguard fairness.

Limitations

Potential limitations of the studies reported here must be emphasized. In the present study, we used dual-process system theory to explain our finding that third-party help evoked larger FRN but obtained more behavioral agreement compared with punishment. We suspected that the FRN reflected a relatively automatic process during which third-party punishment was expected. However, the high evaluation score of third-party help mainly resulted from a more deliberate process. Though previous studies somewhat supported our interpretation (Boksem and De Cremer, 2010; Sun et al., 2015), direct experimental evidence that can distinguish between the proposed automatic and deliberate evaluations was still needed. Although the automatic process was difficult to orient, weakening the controlled system and the deliberative process by time pressure, cognitive load, or some other methods was possible (Cappelletti et al., 2011). Future studies can perform some of these methods to distinguish between automatic and deliberate processes and produce more persuasive results.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Ethics Committee of Business School of Nankai University with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Ethics Committee of Business School of Nankai University.

AUTHOR CONTRIBUTIONS

JL and SL conceived and designed the study. JL, SL, PW, GW, and XY designed the experimental stimuli and procedure. SL, CZ, and XL implemented experimental protocols and collected data. SL, XN, and GW analyzed the data. SL, JL, and XY wrote and revised the paper.

FUNDING

This research was supported by the National Natural Science Foundation of China (Grants Nos. 71673152, 71372094, 71533002, and 71602138).

REFERENCES

Adolphs, R. (2009). The social brain: neural basis of social knowledge. *Annu. Rev. Psychol.* 60, 693–716. doi: 10.1146/annurev.psych.60.110707.163514

Batson, C. D., Eklund, J. H., Chermok, V. L., Hoyt, J. L., and Ortiz, B. G. (2007). An additional antecedent of empathic concern: valuing the welfare of the person in need. *J. Pers. Soc. Psychol.* 93, 65–74. doi: 10.1037/0022-3514.93.1.65

Blechert, J., Sheppes, G., Di Tella, C., Williams, H., and Gross, J. J. (2012). See what you think: reappraisal modulates behavioral and neural responses to social stimuli. *Psychol. Sci.* 23, 346–353. doi: 10.1177/0956797612438559

Boksem, M. A., and De Cremer, D. (2010). Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Soc. Neurosci.* 5, 118–128. doi: 10.1080/17470910903202666

Boyd, R., Gintis, H., Bowles, S., and Richerson, P. J. (2003). The evolution of altruistic punishment. *Proc. Natl. Acad. Sci. U.S.A.* 100, 3531–3535. doi: 10.1073/pnas.0630443100

Buchan, N. R., Johnson, E. J., and Croson, R. T. (2006). Let's get personal: an international examination of the influence of communication, culture and social distance on other regarding preferences. *J. Econ. Behav. Organ.* 60, 373–398. doi: 10.1016/j.jebo.2004.03.017

Cappelletti, D., Güth, W., and Ploner, M. (2011). Being of two minds: ultimatum offers under cognitive constraints. *J. Econ. Psychol.* 32, 940–950. doi: 10.1016/j.joep.2011.08.001

Carlo, G. (2014). The development and correlates of prosocial moral behaviors. *Handb. Moral Dev.* 2, 208–234. doi: 10.4324/9780203581957.ch10

Crockett, M. J., Clark, L., Lieberman, M. D., Tabibnia, G., and Robbins, T. W. (2010). Impulsive choice and altruistic punishment are correlated and increase in tandem with serotonin depletion. *Emotion* 10, 855–862. doi: 10.1037/a0019861

David, B., Hu, Y., Krüger, F., and Weber, B. (2017). Other-regarding attention focus modulates third-party altruistic choice: An fMRI study. *Sci. Rep.* 7:43024. doi: 10.1038/srep43024

dos Santos, M., Rankin, D. J., and Wedekind, C. (2011). The evolution of punishment through reputation. *Proc. R. Soc. Lond. B Biol. Sci.* 278, 371–377. doi: 10.1098/rspb.2010.1275

dos Santos, M., Rankin, D. J., and Wedekind, C. (2013). Human cooperation based on punishment reputation. *Evolution* 67, 2446–2450. doi: 10.1111/evo.12108

Erlandsson, A., Björklund, F., and Bäckström, M. (2014). Perceived utility (not sympathy) mediates the proportion dominance effect in helping decisions. *J. Behav. Decis. Making* 27, 37–47. doi: 10.1002/bdm.1789

Fehr, E., and Fischbacher, U. (2003). The nature of human altruism. *Nature* 425, 785–791. doi: 10.1038/nature02043

Fehr, E., and Fischbacher, U. (2004). Third-party punishment and social norms. *Evol. Hum. Behav.* 25, 63–87. doi: 10.1016/S1090-5138(04)00005-4

Fehr, E., and Gächter, S. (2002). Altruistic punishment in humans. *Nature* 415, 137–140. doi: 10.1038/415137a

FeldmanHall, O., Dalgleish, T., Evans, D., and Mobbs, D. (2015). Empathic concern drives costly altruism. *Neuroimage* 105, 347–356. doi: 10.1016/j.neuroimage.2014.10.043

Gehring, W. J., and Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science* 295, 2279–2282. doi: 10.1126/science.1066893

Gromet, D. M., and Darley, J. M. (2009). Punishment and beyond: achieving justice through the satisfaction of multiple goals. *Law Soc. Rev.* 43, 1–38. doi: 10.1111/j.1540-5893.2009.00365.x

Gummerum, M., Van Dillen, L. F., Van Dijk, E., and López-Pérez, B. (2016). Costly third-party interventions: the role of incidental anger and attention focus in punishment of the perpetrator and compensation of the victim. *J. Exp. Soc. Psychol.* 65, 94–104. doi: 10.1016/j.jesp.2016.04.004

Hauser, T. U., Iannaccone, R., Stämpfli, P., Drechsler, R., Brandeis, D., Walitza, S., et al. (2014). The feedback-related negativity (FRN) revisited: new insights into the localization, meaning and network organization. *Neuroimage* 84, 159–168. doi: 10.1016/j.neuroimage.2013.08.028

Holroyd, C. B., Hajcak, G., and Larsen, J. T. (2006). The good, the bad and the neutral: electrophysiological responses to feedback stimuli. *Brain Res.* 1105, 93–101. doi: 10.1016/j.brainres.2005.12.015

Hu, Y., Strang, S., and Weber, B. (2015). Helping or punishing strangers: neural correlates of altruistic decisions as third-party and of its relation to empathic concern. *Front. Behav. Neurosci.* 9:24. doi: 10.3389/fnbeh.2015.00024

Jordan, J. J., Hoffman, M., Nowak, M. A., and Rand, D. G. (2016). Uncalculating cooperation is used to signal trustworthiness. *Proc. Natl. Acad. Sci. U.S.A.* 113, 8658–8663. doi: 10.1073/pnas.1601280113

Jordan, J. J., and Rand, D. G. (2017). *The Drive to Appear Trustworthy Shapes Punishment and Moral Outrage in One-Shot Anonymous Interactions*. New York, NY: Social Science Research Network.

Kahneman, D. (2011). *Thinking Fast and Slow*. New York, NY: Farrar, Straus, and Giroux.

Knoch, D., Gianotti, L. R., Baumgartner, T., and Fehr, E. (2010). A neural marker of costly punishment behavior. *Psychol. Sci.* 21, 337–342. doi: 10.1177/0956797609360750

Koban, L., and Pourtois, G. (2014). Brain systems underlying the affective and social monitoring of actions: an integrative review. *Neurosci. Biobehav. Rev.* 46, 71–84. doi: 10.1016/j.neubiorev.2014.02.014

Leliveld, M. C., Dijk, E., and Beest, I. (2012). Punishing and compensating others at your own expense: the role of empathic concern on reactions to distributive injustice. *Eur. J. Soc. Psychol.* 42, 135–140. doi: 10.1002/ejsp.872

Li, J., Yin, X., Li, D., Liu, X., Wang, G., and Qu, L. (2017). Controlling the anchoring effect through transcranial direct current stimulation (tDCS) to the right dorsolateral prefrontal cortex. *Front. Psychol.* 8:1079. doi: 10.3389/fpsyg.2017.01079

Lieberman, D., Tooby, J., and Cosmides, L. (2007). The architecture of human kin detection. *Nature* 445, 727–731. doi: 10.1038/nature05510

Liu, X., Banich, M. T., Jacobson, B. L., and Tanabe, J. L. (2004). Common and distinct neural substrates of attentional control in an integrated Simon and spatial Stroop task as assessed by event-related fMRI. *Neuroimage* 22, 1097–1106. doi: 10.1016/j.neuroimage.2004.02.033

Loke, I. C., Evans, A. D., and Lee, K. (2011). The neural correlates of reasoning about prosocial–helping decisions: an event-related brain potentials study. *Brain Res.* 1369, 140–148. doi: 10.1016/j.brainres.2010.10.109

Milinski, M., Semmann, D., Bakker, T. C., and Krambeck, H. J. (2001). Cooperation through indirect reciprocity: image scoring or standing strategy? *Proc. R. Soc. Lond. B Biol. Sci.* 268, 2495–2501. doi: 10.1098/rspb.2001.1809

Miltner, W. H., Braun, C. H., and Coles, M. G. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a "generic" neural system for error detection. *J. Cogn. Neurosci.* 9, 788–798. doi: 10.1162/jocn.1997.9.6.788

Morese, R., Rabellino, D., Sambataro, F., Perussia, F., Valentini, M. C., Bara, B. G., et al. (2016). Group membership modulates the neural circuitry underlying third party punishment. *PLoS One* 11:e0166357. doi: 10.1371/journal.pone.0166357

Mothes, H., Enge, S., and Strobel, A. (2016). The interplay between feedback-related negativity and individual differences in altruistic punishment: An EEG study. *Cogn. Affect. Behav. Neurosci.* 16, 276–288. doi: 10.3758/s13415-015-0388-x

Navarro-Cebrian, A., Knight, R. T., and Kayser, A. S. (2016). Frontal monitoring and parietal evidence: mechanisms of error correction. *J. Cogn. Neurosci.* 28, 1166–1177. doi: 10.1162/jocn_a_00962

Nieuwenhuis, S., Holroyd, C. B., Mol, N., and Coles, M. G. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neurosci. Biobehav. Rev.* 28, 441–448. doi: 10.1016/j.neubiorev.2004.05.003

Olatunji, B. O., and Puncochar, B. D. (2014). Delineating the influence of emotion and reason on morality and punishment. *Rev. Gen. Psychol.* 18, 186–207. doi: 10.1037/gpr0000010

Oliveira, F. T., McDonald, J. J., and Goodman, D. (2007). Performance monitoring in the anterior cingulate is not all error related: expectancy deviation and the representation of action-outcome associations. *J. Cogn. Neurosci.* 19, 1994–2004. doi: 10.1162/jocn.2007.19.12.1994

Qu, C., Wang, Y., and Huang, Y. (2013). Social exclusion modulates fairness consideration in the ultimatum game: an ERP study. *Front. Hum. Neurosci.* 7:505. doi: 10.3389/fnhum.2013.00505

Qu, L., Dou, W., You, C., and Qu, C. (2014). The processing course of conflicts in third-party punishment: an event-related potential study. *Psych J.* 3, 214–221. doi: 10.1002/pchj.59

Raihani, N. J., and Bshary, R. (2015). Third-party punishers are rewarded, but third-party helpers even more so. *Evolution* 69, 993–1003. doi: 10.1111/evo.12637

Rand, D. G., Greene, J. D., and Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature* 489, 427–430. doi: 10.1038/nature11467

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755–1758. doi: 10.1126/science.1082976

Seinen, I., and Schram, A. (2006). Social status and group norms: indirect reciprocity in a repeated helping experiment. *Eur. Econ. Rev.* 50, 581–602. doi: 10.1016/j.euroecorev.2004.10.005

Strobel, A., Zimmermann, J., Schmitz, A., Reuter, M., Lis, S., Windmann, S., et al. (2011). Beyond revenge: neural and genetic bases of altruistic punishment. *Neuroimage* 54, 671–680. doi: 10.1016/j.neuroimage.2010.07.051

Sun, L., Tan, P., Cheng, Y., Chen, J., and Qu, C. (2015). The effect of altruistic tendency on fairness in third-party punishment. *Front. Psychol.* 6:820. doi: 10.3389/fpsyg.2015.00820

Tomasello, M., and Vaish, A. (2013). Origins of human cooperation and morality. *Annu. Rev. Psychol.* 64, 231–255. doi: 10.1146/annurev-psych-113011-143812

Van Doorn, J., Zeelenberg, M., and Breugelmans, S. M. (2014). Anger and prosocial behavior. *Emot. Rev.* 6, 261–268. doi: 10.1177/1754073914523794

Wang, G., Li, J., Li, Z., Wei, M., and Li, S. (2016a). Medial frontal negativity reflects advantageous inequality aversion of proposers in the ultimatum game: an ERP study. *Brain Res.* 1639, 38–46. doi: 10.1016/j.brainres.2016.02.040

Wang, G., Li, J., Yin, X., Li, S., and Wei, M. (2016b). Modulating activity in the orbitofrontal cortex changes trustees' cooperation: a transcranial direct current stimulation study. *Behav. Brain Res.* 303, 71–75. doi: 10.1016/j.bbr.2016.01.047

Wei, W., Wang, L., Shang, Z., and Li, J. C. (2015). Non-sympathetic FRN responses to drops in others' stocks. *Soc. Neurosci.* 10, 616–623. doi: 10.1080/17470919.2015.1013222

Wu, Y., Zhou, Y., van Dijk, E., Leliveld, M. C., and Zhou, X. (2011). Social comparison affects brain responses to fairness in asset division: an ERP study with the ultimatum game. *Front. Hum. Neurosci.* 5:131. doi: 10.3389/fnhum.2011.00131

Yang, Z., Jin, C., Qi, Y., Zheng, Y., and Liu, X. (2017). The influence of emotion on fairness-related decision making: a critical review of theories and evidence. *Front. Psychol.* 8:1592. doi: 10.3389/fpsyg.2017.01592

Ye, H., Tan, F., Ding, M., Jia, Y., and Chen, Y. (2011). Sympathy and punishment: evolution of cooperation in public goods game. *J. Artif. Soc. Soc. Simul.* 14:20. doi: 10.18564/jasss.1805

Yeung, N., Holroyd, C. B., and Cohen, J. D. (2005). ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cereb. Cortex* 15, 535–544. doi: 10.1093/cercor/bhh153

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 20 August 2018
doi: 10.3389/fpsyg.2018.01321

Does Gender Make a Difference in Deception? The Effect of Transcranial Direct Current Stimulation Over Dorsolateral Prefrontal Cortex

Mei Gao[1], *Xiaolan Yang*[2,3]*, *Jinchuan Shi*[3], *Yiyang Lin*[1] *and Shu Chen*[1,4]

[1] *College of Economics, Zhejiang University, Hangzhou, China,* [2] *School of Business and Management, Shanghai International Studies University, Shanghai, China,* [3] *Academy of Financial Research, Zhejiang University, Hangzhou, China,* [4] *Interdisciplinary Center for Social Sciences, Zhejiang University, Hangzhou, China*

Neuroimaging studies have indicated a correlation between dorsolateral prefrontal cortex (DLPFC) activity and deceptive behavior. We applied a transcranial direct current stimulation (tDCS) device to modulate the activity of subjects' DLPFCs. Causal evidence of the neural mechanism of deception was obtained. We used a between-subject design in a signaling framework of deception, in which only the sender knew the associated payoffs of two options. The sender could freely choose to convey the truth or not, knowing that the receiver would never know the actual payment information. We found that males were more honest than females in the sham stimulation treatment, while such gender difference disappeared in the right anodal/left cathodal stimulation treatment, because modulating the activity of the DLPFC using right anodal/left cathodal tDCS only significantly decreased female subjects' deception.

Keywords: deception, dorsolateral prefrontal cortex, transcranial direct current stimulation, gender difference, cheap talk

OPEN ACCESS

Edited by:
Qi Li,
University of Chinese Academy of Sciences (UCAS), China

Reviewed by:
Haiyan Wu,
University of Chinese Academy of Sciences (UCAS) and Institute of Psychology (CAS), China
Jianbiao Li,
Nankai University, China
Yiping Zhong,
Hunan Normal University, China

***Correspondence:**
Xiaolan Yang
yangxiaolan1996@126.com

Specialty section:
This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Psychology

Received: *03 December 2017*
Accepted: *09 July 2018*
Published: *20 August 2018*

Citation:
Gao M, Yang X, Shi J, Lin Y and Chen S (2018) Does Gender Make a Difference in Deception? The Effect of Transcranial Direct Current Stimulation Over Dorsolateral Prefrontal Cortex. Front. Psychol. 9:1321. doi: 10.3389/fpsyg.2018.01321

INTRODUCTION

Deception is a complex human behavior that is prevalent in finance, politics and interpersonal relationships. It is widespread in various sectors of society and has important economic consequences (Gachter and Schulz, 2016). Numerous fraud scandals in recent years have greatly damaged the economy and the stability of financial markets (Abrantes-Metz et al., 2012; Sapienza and Zingales, 2012). In a situation involving asymmetric information, businessmen, politicians and others may deliberately take advantage of private information to deceitfully improve their self-earnings (Gneezy, 2005; Clotsfigueras et al., 2015). Therefore, determining what maintains human honesty and how to prevent deceptive behavior, especially in the economy, is a fundamental problem.

In reality, most fraudsters in social economies expect to increase their profits by a series of lies that lead to the decreasing earnings of others. It must be emphasized that honest choices are always associated with conflicts between self-interest and others' interests. It's obvious that financial honesty concerns moral norms that help us to resist the temptation of making more money by behaving dishonestly (Villeval, 2014). Why people sometime could sacrifice monetary payoffs and be truthful? The moral conflicts elicited by dishonest gain play a significant role in human deceptive behavior (Mead et al., 2009), while little is known about the neural process of human when resolving the conflict between honesty and monetary gain.

Some studies relied on instructed-lying paradigms show that deception requires the host of executive functions as people need to inhibit the disclosure of the truth to make deceptive responses (Hu et al., 2011). Conflict related deception involves executive function in dorsalateral prefrontal cortex (DLPFC), ventralateral prefrontal cortex (VLPFC), medial frontal cortex (MFC) and anterior cingulate cortex (ACC) (Ganis et al., 2003; Abe et al., 2006). As studies using instructed-lying paradigms typically examine deception ability rather than deceptive behavior (Sip et al., 2008), other studies pay attention to the neural mechanisms underlying spontaneous deception (Greene and Paxton, 2009; Ding et al., 2013). Wu et al. (2009) found a more positive P300 amplitude triggered by self-determined response than that triggered by forced responses. What's more, the N2, which indicates subjects' conflict detection, was more negative elicited by deceptive response than that elicited by honest response. It seems that the brain response of both instructed deception and spontaneous deception is conflict related. However, in the most studies of spontaneous deception, subjects' gains were not directly associated with their dishonest or honest decisions. That is, they didn't face the moral trade-off between deceptive behavior and self-interest. Only one study has investigated spontaneous deception considering the moral conflict between honesty and self-interest (Greene and Paxton, 2009). In a simple game asking subjects undergoing functional magnetic resonance imaging (fMRI) to self-report the accuracy of coin-flip predictions, they found that increased activity in the DLPFC was closely associated with dishonest subjects' decisions compared to subjects behaving honestly, both when telling lies and occasionally telling truth. As neuroimaging studies can only demonstrate a correlation between the activity of certain cortex areas and deceptive behavior, the causal effect remains unknown. Thus, the neural basis of deceptive behavior in DLPFC remains unexplored especially in the setting involving moral conflict between honesty and personal gain.

Increasingly, brain stimulation techniques are being used in research (Li et al., 2017; Maréchal et al., 2017; Wang et al., 2017). Such techniques can enable direct observations of how modulating the activity of the DLPFC affects subjects' deceptive behavior. Maréchal et al. (2017) tested subjects' honest behavior using a die-rolling task with transcranial direct current stimulation (tDCS) over the right dorsolateral prefrontal cortex (rDLPFC). They showed that honesty was enhanced after anodal stimulation of the rDLPFC. To the best of our knowledge, this was the first paper to demonstrate that honesty can be strengthened through non-invasive stimulation of the DLPFC. Further research using different experimental paradigms is needed to excavate the neural mechanism of honest behavior robustly.

Some studies of cheating behavior have adopted a "cheap talk sender–receiver" game (e.g., Gneezy, 2005; Zhu et al., 2014), in which only the sender knew the associated payoffs of two options and freely chose to convey the truth or not, knowing that the receiver would never know the actual payment information. This game contains the conflict between self-interest and honesty.

Unlike the die-rolling task adopted by Maréchal et al. (2017), measuring aggregate-level of honesty, the cheap talk sender–receiver game enables us to utilize individual-level data to analyze the deceptive behavior. Except that we collected individual-level data of deception, we could clearly justify whether subjects made an honest decision or not in our study, while in the die-rolling task, there were some probability that subjects reported the profit-maximizing outcome because that was the actual outcome. Obviously, it was not the case that we were intend to investigate, where subjects needed to decide whether to behave dishonestly for self-interest or not. What's more, though subjects were anonymous in both experiments, subjects knew that their deceptive behavior could be observed by experimenters only in the cheap-talk sender-receiver game. Therefore, the psychic cost of deception is different in the two experiments. Although the classical die-rolling experimental study conducted by Fischbacher and Follmiheusi (2013) showed that the reported distribution was not significantly changed when the remainder was given to another subject instead of being kept by the experimenter, the two experiments are still different in paradigm itself.

To investigate the effect of tDCS on individuals' deceptive behavior related to the moral conflict between self-interest and others' interests, our experiment used a cheap talk sender–receiver game in which senders had private information about the real allocation of money between themselves and their paired receivers and then decided to send an honest/dishonest message about the allocation to receivers. We adopted a between-subject design to test whether various tDCS treatments changed subjects' honesty by comparing the subjects' deceptive behavior among different stimulation treatments. Our goal was to find a causal relationship between DLPFC and deceptive behavior, and to compare the exact effect of different stimulations on the honesty of subjects when there are interest conflicts between senders and receivers and when there are no interest conflicts.

Since a sender might choose to tell the truth strategically if he/she expected the receiver not to follow his/her message, the cheap talk sender–receiver game in our experiment might confound honesty with strategic motives. An additional questionnaire was also conducted to directly verify the senders' strategic consideration.

MATERIALS AND METHODS

Subjects

Hundred and eighty subjects were recruited from different majors at Zhejiang University via an advertisement posted on the school bulletin board system. Subjects were grouped in pairs and randomly assigned the role of sender or receiver. Ninety subjects (46 females, mean age = 21.4 ± 2.07 years, all right-handed) who acted as senders, were randomly assigned to three treatment groups: right anodal/left cathodal stimulation (n = 30), left anodal/right cathodal stimulation (n = 30) or sham (n = 30) treatment. The experiment lasted around 40 min, and the average payment of the subjects was CNY 22.88 (approximately 3.46 dollars). To learn the senders' beliefs regarding the reaction of the receivers, 57 senders (29 females, mean age = 21.02 ± 2.2 years, all right-handed) were asked

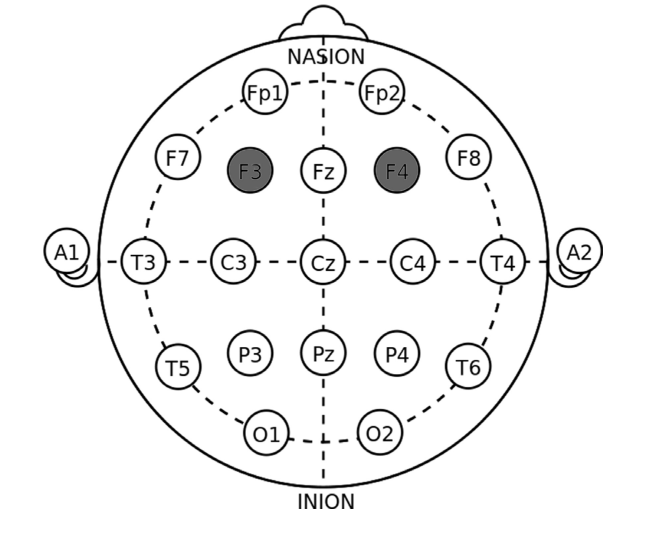

FIGURE 1 | Electrode placements in DLPFC stimulations.

to sequentially complete a questionnaire. All of the subjects gave written informed consent, and the study was approved by the Zhejiang University ethics committee before the start of the experiment. No subjects reported any adverse side effects regarding pain on the scalp or headaches after the experiment.

Transcranial Direct Current Stimulation

Transcranial direct current stimulation (tDCS), a non-invasive brain stimulation technique, was delivered by a battery-driven multichannel non-invasive wireless neurostimulator (Starlab, Spain). A constant 2-mA current flow lasting for 20 min with 30 s of ramp up and down was applied via a pair of saline-soaked sponge electrodes (5 cm × 7 cm) fixed on the scalp of the participant with a rubber belt. tDCS facilitates neural excitability depending on electrode polarity. The anodal electrode enhances cortical excitability while the cathodal electrode weakens it (Nitsche and Paulus, 2000). As in the study by Gandiga et al. (2006), the current delivered in the sham stimulation treatment only lasted for 30 s once it reached 2 mA. This short-lived but perceptible stimulation was designed to make the subjects feel as if they had received the true stimulation treatment.

Electrodes placed over F3 and F4 can effectively influence the DLPFC area (Fecteau et al., 2007a,b; Boggio et al., 2009). As shown in **Figure 1**, the anodal (cathodal) electrode was placed over the right F4 and the cathodal (anodal) electrode was placed over the left F3 in the right anodal/left cathodal (left anodal/right cathodal) treatment based on the International 10-20 System for electrode placement.

Experimental Design

The cheap talk sender–receiver game is a two-player communication game in which one player (the sender) has private information and the other (the receiver) makes the final allocation decision (Gneezy, 2005; Zhu et al., 2014). In the experiment, the subjects were grouped in pairs and randomly assigned the role of sender or receiver. A screen was set up to separate senders and receivers, so that the sender and the receiver in a pair never meet. The game was composed of 12 trials. For each group, only the subject who played the role of the sender was informed about the monetary payoffs of two options, A and B, in each trial. The sender had to send one of two messages to the other subject in the role of receiver:

> Message 1: "Option A will earn you more money than option B."
> Message 2: "Option B will earn you more money than option A."

After receiving the message sent by the sender, the receiver chose the option to be carried out. Crucially, all senders knew that receivers would never be informed of the payoffs associated with each option. Therefore, they could choose either the honest or dishonest message. At the end of the game, we randomly chose one of the 12 trials to determine the real payoff for the subjects.

The monetary consequences varying across trials are displayed in **Table 1**, following Zhu et al. (2014). For instance, option A corresponds to CNY15 to the sender and CNY5 to the receiver, and option B corresponds to CNY5 to the sender and CNY15 to the receiver in Trial 1. It is obvious that the sender's honest choice, that is, sending message 2, "Option B will earn you more money than option A," to the receiver, will damage his/her own payoff. Thus, there is a conflict between self-interest and others' interests, as an honest message will result in the sender allocating less money to himself/herself but more to the receiver. These trials are referred to as "conflict trials" (C). There are also "no-conflict trials" (NC), in which the sender's interest is aligned with the receiver's interest (Trials 5 and 9).

Experimental Procedure

At the beginning of the experiment, subjects were randomly assigned to a sender or a receiver and asked to sign the written informed consent form. Then, the researchers placed tDCS devices on the sender's head for a 20-min stimulation and told them to seat themselves comfortably and relax. The devices were taken away when the stimulation ended. After a public reading of experimental instructions, the experiment was conducted by the software z-Tree (Fischbacher, 2007). Trials were presented one by one. At the end of the experiment, the computer randomly selected one trial as the payoff. The final payments were the combination of a show-up fee and the payoffs in the selected trials according to receivers' decisions.

In the experiment, we collected each subject's percentage of honest choices and "amount given" in conflict trials and in no-conflict trials. Following Zhu et al. (2014) we adopted a measure of honesty called "amount given," which was defined as the amount that senders were willing to allocate to receivers according to the message sent by senders. Taking Trial 1 as an example, if the sender tells the truth (message 2), then the amount given is CNY15 in Option B; while if the sender lies (message 1), it is CNY5 in Option A.

TABLE 1 | The cheap talk sender–receiver game.

Trial number	Option A		Option B		Interest conflict	Honest message
	Self	Other	Self	Other		
1	15	5	5	15	C	2
2	10	5	5	20	C	2
3	6	5	10	4.99	C	1
4	5	10	10	5	C	1
5	8	10	10	12	NC	2
6	6	5	5	6	C	2
7	5	20	20	5	C	1
8	6	5	5	15	C	2
9	10	6	10	5	NC	1
10	10	12	12	10	C	1
11	5	10	6	5	C	1
12	10	4.99	4	5	C	2

RESULTS

Effect of tDCS on Deceptive Behavior

To test whether different tDCS treatments changed subjects' deceptive behavior, we compared the percentage of honest and dishonest choices after the treatment. Deception was substantial in the sham stimulation treatment group, in which senders cheated in half of the trials (**Figure 2A**). However, the deceptive behavior was concentrated in the conflict trials, and in no-conflict trials, the cheating proportion was only 8.3%.

A 2 conflict condition (conflict trials vs. no-conflict trials) × 3 stimulation type (right anodal/left cathodal stimulation vs. left anodal/right cathodal stimulation vs. sham stimulation) ANOVA on the average percentage of honest choices revealed a significant main effect of conflict condition, $F_{1,87} = 114.675$, $p < 0.000$, with subjects choosing less honest choices in conflict trials (mean = 49.4%) compared to no-conflict trials (mean = 87.2%). Though the interaction of conflict condition and stimulation type was not significant, $F_{1,87} = 2.102$, $p = 0.128$, a main effect of stimulation type was found, $F_{1,87} = 3.389$, $p = 0.038$. The average percentage of honest choices was higher after the right anodal/left cathodal stimulation in conflict trials (R+/L− mean = 59%, sham mean = 44.7%, $p = 0.021$), but there was no significant difference after the left anodal/right cathodal stimulation in conflict trials (L+/R− mean = 44.7%, sham mean = 44.7%, $p > 0.1$). However, tDCS had little effect on senders' deceptive behavior in no-conflict trials no matter what type of stimulation was used. In other words, right anodal/left cathodal stimulation made senders more honest only when they had to resolve the trade-off between self-interest and honesty.

A 2 conflict condition (conflict trials vs. no-conflict trials) × 3 stimulation type (right anodal/left cathodal stimulation vs. left anodal/right cathodal stimulation vs. sham stimulation) ANOVA on the amount given revealed a significant main effect of the interaction of conflict condition and stimulation type was found, $F_{2,87} = 3.361$, $p = 0.039$. Similarly, **Figure 2B** shows that senders were willing to give more to receivers after right anodal/left cathodal tDCS only in the conflict trials (R+/L− mean = 9.649, sham mean = 8.622, $p = 0.05$), whereas the left anodal/right cathodal tDCS had no influence on the amounts given to receivers (L+/R− mean = 8.725, sham mean = 8.622, $p > 0.1$). Senders were more honest after the right anodal/left cathodal tDCS of the DLPFC, especially in the conflict trials.

Figure 2C shows the response times for senders' honest and dishonest decisions, revealing the different behavioral patterns between conflict trials and no-conflict trials. According to the 2 conflict condition (conflict trials vs. no-conflict trials) × 3 stimulation type (right anodal/left cathodal stimulation vs. left anodal/right cathodal stimulation vs. sham stimulation) ANOVA on the response times, senders spent more time choosing the message to send in the conflict trials than in the no-conflict trials regardless of the stimulation type (conflict trials = 9.061, no-conflict trials = 5.928, $p < 0.000$). In light of the main effect of stimulation type, there was no significant difference in response time among three stimulation types ($F_{2,87} = 0.563$, $p > 0.1$).

Gender Difference

As **Figure 3** shows, for females, the average percentage of honest choices was higher after the right anodal/left cathodal stimulation in the conflict trials (R+/L− mean = 57.5%, sham mean = 35%, $t_{1,28} = 3.38$, $p = 0.002$), while for males, the effect of tDCS was not significant (R+/L− mean = 60.7%, sham mean = 53.1%, $t_{1,28} = 0.97$, $p = 0.341$). To further determine the gender difference in deceptive behavior, we applied a two-way ANOVA with the percentage of honest choices in the conflict trials as the dependent variable, while gender and stimulation type served as independent variables. We found that males were more likely to make honest choices than females in the conflict trials in the sham stimulation treatment (males: mean = 53.1%, females: mean = 35%, $p = 0.014$). However, no significant difference in the percentage of honest choices between males and females was observed after the right anodal/left cathodal stimulation in the conflict trials (males: mean = 60.7%, females: mean = 57.5%, $p = 0.656$). Only females seemed to be altered by tDCS and became more honest after the right anodal/left cathodal stimulation (females: R+/L− mean = 57.5%, sham mean = 35%,

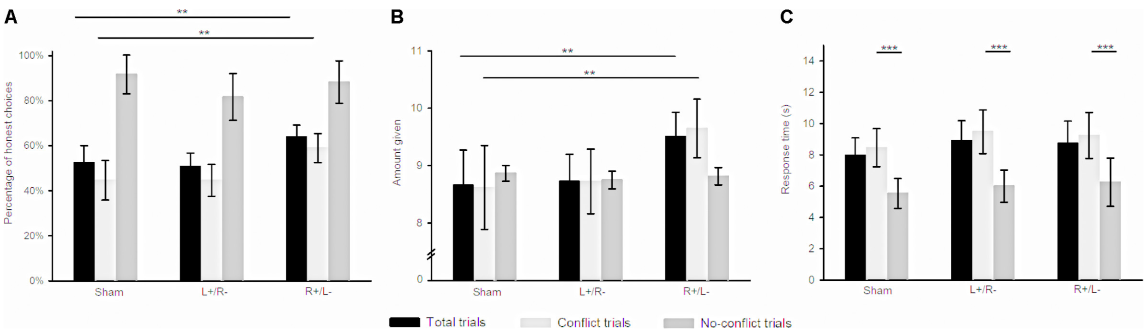

FIGURE 2 | Deceptive behavior in different trials and treatments after the stimulations. Error bars indicate 95% confidence intervals. **(A)** Average percentage of honest choices. **(B)** Average amount given. **(C)** Average response time. Significance level: *** 1 percent, ** 5 percent, * 10 percent.

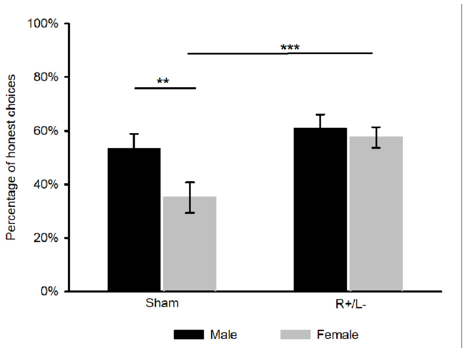

FIGURE 3 | Average percentage of honest choices in conflict trials after the sham and right anodal/left cathodal stimulations over DLPFC for males and females, respectively. Error bars indicate 95% confidence intervals. Significance level: *** 1 percent, ** 5 percent, * 10 percent.

TABLE 2 | Regression results for deceptive behavior.

	Full sample
Left	−0.04 (−0.25)
Right	0.52*** (2.98)
Male	0.45*** (3.18)
Sender-interest gap	−0.16*** (−7.62)
Receiver-interest gap	0.12*** (7.05)
Trial	−0.01 (−0.28)
Pseudo R^2	0.0788
P-value	0.0000
Observation	900

*Significance level: *** 1 percent, ** 5 percent, * 10 percent; t-values in parentheses.*

p = 0.007; males: R+/L− mean = 60.7%, sham mean = 53.1%, p = 0.883).

Empirical Analysis

We employed a logit model to examine the effect of tDCS on deceptive behavior in conflict trials. The dependent variable was *Honesty*, which was a dummy variable and equaled one if the sender sent the honest message to receiver, and otherwise zero. As there were three stimulation types, we set a dummy variable *Left* to be one for left anodal/right cathodal tDCS and otherwise zero, and another dummy variable *Right* to be one for right anodal/left cathodal tDCS and otherwise zero. *Sender-interest gap* and *receiver-interest gap* were two variables representing the absolute difference between the payoff of the two options for senders in each trial and the absolute difference between the payoff of the two options for receivers in each trial. We also included *Trial* as a control variable. **Table 2** provided the results of the logit models.

According to the regression results of the full sample, *Right* was significant and its coefficient was positive, while the coefficient of *Left* was not significant. It meant that senders were more likely to send the honest message after the right anodal/left cathodal tDCS over DLPFC which was consistent with the test results in Section "Effect of tDCS on Deceptive Behavior." The estimated coefficient of *Male* was significant and positive. That is, males were more honest than females. In addition, the significantly negative coefficient of *Sender-interest gap* indicated that the higher the absolute difference between the payoff in the two options for senders, the less likelihood that senders sent the honest message which could decrease self-interest. However, the significant positive coefficient of *Receiver-interest gap* indicated that the higher the absolute difference between the payoff in the two options for receivers, the more likelihood that senders sent the honest message which could increase the interests of others. These findings might provide evidence that deception related to the trade-off between self-interest and others' interests.

Questionnaire

To directly learn the sender's belief, in the questionnaire, we asked senders whether they believed that the receiver would follow their messages. If the answer was no, we further asked them whether they would deliberately send the honest message to mislead the receiver. In fact, only 5.26% (3 of 57) of the senders admitted that they would choose to tell the truth because they expected the receivers not to follow their messages. Therefore, according to the supplementary questionnaire, we can conclude that the senders' strategic considerations were nearly non-existent in our experiment, and indeed, the transcranial direct current stimulation influenced their deceptive behavior.

DISCUSSION

The objective of this study was to investigate the effect of modulating the activity of the DLPFC on deception. In the experiment, we used a between-subject design and a cheap talk sender–receiver task from which we were able to measure the honest/dishonest decisions of subjects and uncover the effect of tDCS on deception by comparing different treatments. Direct evidence of a causal relationship between DLPFC and deceptive behavior was provided. We found that modulating the activity of the DLPFC using right anodal/left cathodal tDCS significantly decreased subjects' deception; they became more honest after right anodal/left cathodal stimulation of the DLPFC. A gender difference in deceptive behavior was also observed. To better learn the sender's beliefs regarding the receiver's reaction to messages, we used an additional questionnaire. Only 5.26% senders in the questionnaire would deliberately choose to be honest because they believed receivers would not follow their messages. The results implied that most of the senders did not have strategic considerations.

Conflict between self-interest and others' interest is of great importance in subjects' deceptive behavior. Gneezy (2005) defines four categories of lie: (1) white lies, which may be helpful, or at least do no harm to anyone; (2) lies that help others but harm the liar; (3) lies that may not help the liar but harm others or harm both sides; and (4) lies that increase the liar's payoffs and decrease others' payoffs. His study showed that people may dislike cheating but will lie for considerable benefits when there is interest conflict. The focus of our paper is the third and fourth categories, especially the fourth, which is relevant to many economic events. In our experiment, deception was ubiquitous in the sham stimulation treatment. Senders are expected to manipulate receivers to improve their own interests and damage receivers' interests in conflict trials. We found that subjects were significantly more honest in no-conflict trials than in conflict trials. To some extent, it suggests that deception is self-interest driven (Mead et al., 2009).

Our study also demonstrated the important role of DLPFC in modulating self-interested driven deceptive behavior. Importantly, we found that deception could be significantly decreased with right anodal/left cathodal stimulation of the DLPFC, which may help to detecting deception using neurotechnologies. People became more honest after such stimulation in terms of both the percentage of honest choices made and the amount given. Moreover, the effect of tDCS on deceptive behavior was only significant when senders' own interests were in conflict with receivers' interests, which was partly in support of the results of Zhu et al. (2014) showing that DLPFC patients behaved differently in conflict trials and no-conflict trials. It is reasonable that modulating the activity in DLPFC will affect subjects' deceptive behavior, because conflict related deception in our experiment needs the executive control which is an important function of DLPFC (Macdonald et al., 2000). Our research extended the effect of tDCS on deceptive behavior when honest choices were associated with conflict between subjects' self-interest and others' interests, which was different from the study by Maréchal et al. (2017), which only considered self-interest and honesty, regardless of others' benefit.

The gender did make differences in the effect of transcranial direct current stimulation over dorsolateral prefrontal cortex on deception. Experimental economic studies in the literature have shown that gender differences are substantial concerning risk aversion, corruption, competitiveness, as well as deception (Gneezy et al., 2003; Dreber and Johannesson, 2008; Croson and Gneezy, 2009; Frank et al., 2011; Marchewka et al., 2012). In the sham stimulation, we found that gender differences were significant in deceptive behavior and males were more honest than females in conflict trials. This was not consistent with the previous study by Dreber and Johannesson (2008), which replicated the task used by Gneezy (2005) and showed that men were more likely than women to lie for higher amounts of money. One possible explanation for the different findings is that the different monetary allocation between senders and receivers in our experiment compared with Gneezy (2005) might affect senders' deceptive behavior[1]. We also found that the deceptive behavior of females was significantly decreased after the right anodal/left cathodal stimulation of the DLPFC but the effect was not significant for males. Because the percentage of honest choices was already high for males in the sham stimulation, the small amount of room for improvement in the honesty of males may have resulted in the insignificant effect of tDCS on males' deceptive behavior.

Two limitations of our study should be noted. One is the problem of the focality of tDCS. Specifically, it is hard to determine whether the observed effects of tDCS were due to selective modulation of the target area or due to the inevitable widespread and non-selective modulation over the cortex (Sellaro et al., 2016). Second, there is a remained question that whether the neural process of conflict related deception in DLPFC is specialized for resolving moral conflicts between self-interest and others' interest or it is just a general brain response to conflict resolution.

In sum, our results suggest that the neural basis of deception is mainly managed by the activity of the DLPFC. Modulating the activity of the DLPFC using right anodal/left cathodal tDCS significantly decreased subjects' deception. Honesty is very important in economic and social relationships, so it is meaningful to explore its neural process to have a better understanding of the basis of people's deceptive behavior. We may design other deception games including four kinds of lies defined by Gneezy (2005) to investigate the effects of tDCS over DLPFC on deception with different interest conflicts, and we should also add moral attitude measurement and other conflict

[1] There might also be other factors making the result uncertain. Gylfason et al. (2013) used the Gneezy (2005)'s design but found no significant gender differences.

related task to the experimental design to better understand the neural mechanism of deception in further study.

AUTHOR CONTRIBUTIONS

XY, MG, and YL designed the experiments. MG and YL performed the experiments. SC and MG analyzed the data. SC and YL drew the figures. XY, MG, JS, YL, and SC wrote the manuscript, revised the manuscript, and finally approved the version to be published.

FUNDING

This work was supported by the National Natural Science Foundation of China (Project 71303210) and the Fundamental Research Funds for the Central Universities.

REFERENCES

Abe, N., Suzuki, M., Tsukiura, T., Mori, E., Yamaguchi, K., Itoh, M., et al. (2006). Dissociable roles of prefrontal and anterior cingulate cortices in deception. *Cereb. Cortex* 16, 192–199. doi: 10.1093/cercor/bhi097

Abrantes-Metz, R. M., Kraten, M., Metz, A. D., and Seow, G. S. (2012). Libor manipulation? *J. Bank. Financ.* 36, 136–150. doi: 10.1016/j.jbankfin.2011.06.014

Boggio, P., Zaghi, S., and Fregni, F. (2009). Modulation of emotions associated with images of human pain using anodal transcranial direct current stimulation (tDCS). *Neuropsychologia* 47, 212–217. doi: 10.1016/j.neuropsychologia.2008.07.022

Clotsfigueras, I., Hernangonzalez, R., and Kujal, P. (2015). Information asymmetry and deception. *Front. Behav. Neurosci.* 9:109. doi: 10.3389/fnbeh.2015.00109

Croson, R., and Gneezy, U. (2009). Gender differences in preferences. *J. Econ. Lit.* 47, 448–474. doi: 10.1257/jel.47.2.448

Ding, X. P., Gao, X., Fu, G., and Lee, K. (2013). Neural correlates of spontaneous deception: a functional near-infrared spectroscopy (fNIRS) study. *Neuropsychologia* 51, 704–712. doi: 10.1016/j.neuropsychologia.2012.12.018

Dreber, A., and Johannesson, M. (2008). Gender differences in deception. *Econ. Lett.* 99, 197–199. doi: 10.1016/j.econlet.2007.06.027

Fecteau, S., Knoch, D., Fregni, F., Sultani, N., Boggio, P. S., and Pascualleone, A. (2007a). Diminishing risk-taking behavior by modulating activity in the prefrontal cortex: a direct current stimulation study. *J. Neurosci.* 27, 12500–12505. doi: 10.1523/JNEUROSCI.3283-07.2007

Fecteau, S., Pascualleone, A., Zald, D. H., Liguori, P., Theoret, H., Boggio, P. S., et al. (2007b). Activation of prefrontal cortex by transcranial direct current stimulation reduces appetite for risk during ambiguous decision making. *J. Neurosci.* 27, 6212–6218. doi: 10.1523/JNEUROSCI.0314-07.2007

Fischbacher, U. (2007). Z-Tree. Zurich toolbox for readymade economic experiments. *Exp. Econ.* 10, 171–178. doi: 10.1007/s10683-006-9159-4

Fischbacher, U., and Follmiheusi, F. (2013). Lies in disguise: an experimental study on cheating. *J. Eur. Econ. Assoc.* 11, 525–547. doi: 10.1111/jeea.12014

Frank, B., Lambsdorff, J. G., and Boehm, F. (2011). Gender and corruption: lessons from laboratory corruption experiments. *Eur. J. Dev. Res.* 23, 59–71. doi: 10.1057/ejdr.2010.47

Gachter, S., and Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature* 531, 496–499. doi: 10.1038/nature17160

Gandiga, P. C., Hummel, F. C., and Cohen, L. G. (2006). Transcranial DC stimulation (tDCS): a tool for double-blind sham-controlled clinical studies in brain stimulation. *Clin. Neurophysiol.* 117, 845–850. doi: 10.1016/j.clinph.2005.12.003

Ganis, G., Kosslyn, S. M., Stose, S., Thompson, W. L., and Yurgeluntodd, D. A. (2003). Neural correlates of different types of deception: an fMRI investigation. *Cereb. Cortex* 13, 830–836. doi: 10.1093/cercor/13.8.830

Gneezy, U. (2005). Deception: the role of consequences. *Am. Econ. Rev.* 95, 384–394. doi: 10.1257/0002828053828662

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in competitive environments: gender differences. *Q. J. Econ.* 118, 1049–1074. doi: 10.1162/00335530360698496

Greene, J. D., and Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12506–12511. doi: 10.1073/pnas.0900152106

Gylfason, H. F., Arnardottir, A. A., and Kristinsson, K. (2013). More on gender differences in lying. *Econ. Lett.* 119, 94–96. doi: 10.1016/j.econlet.2013.01.027

Hu, X., Wu, H., and Fu, G. (2011). Temporal course of executive control when lying about self- and other-referential information: an ERP study. *Brain Res.* 1369, 149–157. doi: 10.1016/j.brainres.2010.10.106

Li, J., Yin, X., Li, D., Liu, X., Wang, G., and Qu, L. (2017). Controlling the anchoring effect through transcranial direct current stimulation (tDCS) to the right dorsolateral prefrontal cortex. *Front. Psychol.* 8:1079. doi: 10.3389/fpsyg.2017.01079

Macdonald, A. W., Cohen, J. D., Stenger, V. A., and Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science* 288, 1835–1838. doi: 10.1126/science.288.5472.1835

Marchewka, A., Jednorog, K., Falkiewicz, M., Szeszkowski, W., Grabowska, A., and Szatkowska, I. (2012). Sex, lies and fMRI—gender differences in neural basis of deception. *PLoS One* 7:e43076. doi: 10.1371/journal.pone.0043076

Maréchal, M. A., Cohn, A., Ugazio, G., and Ruff, C. C. (2017). Increasing honesty in humans with noninvasive brain stimulation. *Proc. Natl. Acad. Sci. U.S.A.* 114, 4360–4364. doi: 10.1073/pnas.1614912114

Mead, N. L., Baumeister, R. F., Gino, F., Schweitzer, M. E., and Ariely, D. (2009). Too tired to tell the truth: self-control resource depletion and dishonesty. *J. Exp. Soc. Psychol.* 45, 594–597. doi: 10.1016/j.jesp.2009.02.004

Nitsche, M. A., and Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *J. Physiol.* 527, 633–639. doi: 10.1212/WNL.57.10.1899

Sapienza, P., and Zingales, L. (2012). A trust crisis. *Int. Rev. Financ.* 12, 123–131. doi: 10.1111/j.1468-2443.2012.01152.x

Sellaro, R., Nitsche, M. A., and Colzato, L. S. (2016). The stimulated social brain: effects of transcranial direct current stimulation on social cognition. *Ann. N. Y. Acad. Sci.* 1369, 218–239. doi: 10.1111/nyas.13098

Sip, K. E., Roepstorff, A., Mcgregor, W. B., and Frith, C. D. (2008). Detecting deception: the scope and limits. *Trends Cogn. Sci.* 12, 48–53. doi: 10.1016/j.tics.2007.11.008

Villeval, M. C. (2014). Behavioural economics: professional identity can increase dishonesty. *Nature* 516, 48–49. doi: 10.1038/nature14068

Wang, P., Wang, G., Niu, X., Shang, H., and Li, J. (2017). Effect of transcranial direct current stimulation of the medial prefrontal cortex on the gratitude of individuals with heterogeneous ability in an experimental labor market. *Front. Behav. Neurosci.* 11:217. doi: 10.3389/fnbeh.2017.00217

Wu, H., Hu, X., and Fu, G. (2009). Does willingness affect the N2-P3 effect of deceptive and honest responses? *Neurosci. Lett.* 467, 63–66. doi: 10.1016/j.neulet.2009.10.002

Zhu, L., Jenkins, A. C., Set, E., Scabini, D., Knight, R. T., Chiu, P. H., et al. (2014). Damage to dorsolateral prefrontal cortex affects tradeoffs between honesty and self-interest. *Nat. Neurosci.* 17, 1319–1321. doi: 10.1038/nn.3798

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer HW and handling Editor declared their shared affiliation.

ORIGINAL RESEARCH
published: 23 November 2017
doi: 10.3389/fpsyg.2017.02033

Stimulating the Right Temporoparietal Junction with tDCS Decreases Deception in Moral Hypocrisy and Unfairness

Honghong Tang[1,2,3], *Peixia Ye*[2,3,4], *Shun Wang*[2,3,4], *Ruida Zhu*[2,3,4], *Song Su*[1]*, *Luqiong Tong*[1] *and Chao Liu*[1,2,3,4]*

[1] Business School, Beijing Normal University, Beijing, China, [2] State Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China, [3] Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, Beijing, China, [4] Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing, China

Self-centered and other-regarding concerns play important roles in decisions of deception. To investigate how these two motivations affect deception in fairness related moral hypocrisy, we modulated the brain activity in the right temporoparietal junction (rTPJ), the key region for decision making involved in self-centered and other-regarding concerns. After receiving brain stimulation with transcranial direct current stimulation (tDCS), participants finished a modified dictator game. In the game, they played as proposers to make allocations between themselves and recipients and had a chance to deceive by misreporting their totals for allocations. Results show that deception in moral hypocrisy was decreased after anodal stimulation than sham and cathodal stimulation, only when participants know that their reported totals (appearing fair) would be revealed to recipients rather than being unrevealed. Anodal stimulation also increased offers to recipients than cathodal stimulation regardless of the revelation of reported totals. These findings suggest that enhancing the activity of rTPJ decreased deception caused by impression management rather than self-deception in moral hypocrisy and unfairness through facilitating other-regarding concerns and weakening non-material self-centered motivations. They provide causal evidence for the role of rTPJ in both other-regarding concerns and non-material self-centered motivations, shedding light on the way to decrease moral hypocrisy.

Keywords: deception, fairness, moral hypocrisy, impression management, self-deception, transcranial direct current stimulation (tDCS), right temporoparietal junction (rTPJ)

OPEN ACCESS

Edited by:
*Paul E. M. Phillips,
University of Washington,
United States*

Reviewed by:
*Alexander Soutschek,
University of Zurich, Switzerland
Xiaolin Zhou,
Peking University, China*

Correspondence:
*Song Su
sus@bnu.edu.cn
Chao Liu
liuchao@bnu.edu.cn*

Specialty section:
*This article was submitted to
Decision Neuroscience,
a section of the journal
Frontiers in Psychology*

Received: *27 July 2017*
Accepted: *06 November 2017*
Published: *23 November 2017*

Citation:
Tang H, Ye P, Wang S, Zhu R, Su S, Tong L and Liu C (2017) Stimulating the Right Temporoparietal Junction with tDCS Decreases Deception in Moral Hypocrisy and Unfairness. Front. Psychol. 8:2033. doi: 10.3389/fpsyg.2017.02033

INTRODUCTION

Deception is commonly used in social interaction, in which liars often intentionally and strategically give false statements to mislead others. Motivations that affect deception have attracted researchers' attention for years. Although people lie mostly for material benefits for themselves, they also lie for non-material self-centered factors, such as regulating feelings or improving self-presentation (DePaulo et al., 1996; Toma et al., 2008). Those non-material self-centered motivations in deception, which aim at making people appear kinder, fairer, smarter or more attractive instead of being truly so, are consistent with motivations in moral hypocrisy that has been commonly defined as the phenomenon to appear moral instead of being truly moral (Batson et al., 1997, 1999).

Moral hypocrisy is closely linked with deliberate or unconscious deception. It has been proposed to be caused by impression management which aims to protect one's social image in other's eyes through deception and self-deception that targets on protecting one's self-concept of morality when people transgress moral principles (Batson et al., 1999, 2002; Valdesolo and DeSteno, 2008). These non-material self-centered motivations make moral hypocrisy sensitive to both social contexts and threats of moral self. Considering the different directions of them, they might lead people to behave differently when appearing moral would be perceived by others than not.

Moral hypocrisy could be classified into different forms based on the existence of public claims (Graham et al., 2015). Moral deception or moral duplicity that observed when people appear fair through flipping a coin but misreporting the results of the coin (Batson et al., 1999, 2002; Lönnqvist et al., 2014) and moral double standards that used in moral judgment when people evaluate their own moral transgressions less harshly than others (Valdesolo and DeSteno, 2007, 2008) have been treated as interpersonal moral hypocrisy. Moral weakness which describes the conflicts between moral values and behaviors, can exist without public claims, is classified as intrapersonal moral hypocrisy. Although interpersonal moral hypocrisy could engage self-deception to make it more successful through dealing threat of moral self (Batson et al., 1999), it essentially relies on social context and might be more sensitive to changes driven by impression management than self-deception.

Researchers also try to reduce moral hypocrisy and most of them focus on changing the processing of self-concept. For example, some of them found that increasing concerns of self-concept can reduce moral hypocrisy by increasing the self-awareness with a mirror (Batson et al., 1999) or priming religious motivations through religious concepts (Carpenter and Marshall, 2009). Others show that increasing cognitive load to limit cognitive processing of protecting self-concept can also decrease moral hypocrisy (Valdesolo and DeSteno, 2008). However, how concerns of others affect interpersonal moral hypocrisy is still ambiguous. Studies have found that people show other-regarding concerns when they decide whether to deceive or not. They care about the harms, losses or feelings of others in deception (Biziou-van-Pol et al., 2015). Half of honest people are led by other-regarding preferences to be honest (Sheremeta and Shields, 2013), and people decrease deception and lower perceived fairness of deception when they consider the loss of others (Gneezy, 2005). Another study also shows that imaging others' thoughts and feelings in the same situation reduce moral hypocrisy (Batson et al., 2003), indicating the role of other-regarding concerns in moral hypocrisy. However, other-regarding concerns might either decrease interpersonal moral hypocrisy through leading people to be actually prosocial and care others' feelings and payoffs, or increase interpersonal moral hypocrisy by enhancing the self-centered motivation to endorse other-regarding moral principles to protect ones' social image (Szabados and Soifer, 2004). Although both these two accesses require the perspective-taking mechanism, they have opposite effects on interpersonal moral hypocrisy. Thus, in the current study, we modulated the other-regarding concerns through brain stimulation techniques to investigate how it would affect moral hypocrisy.

Neural imaging studies show that the right temporoparietal junction (rTPJ) is a key brain region for social cognition and decision making involved in self and other presentations (Decety and Lamm, 2007; Murray et al., 2012). On the one hand, activity in rTPJ is engaged in understanding other's mental states in theory of mind (Saxe et al., 2006; Van der Meer et al., 2011). It contributes to successful strategic deception in social interaction through inferring other's beliefs and intentions (Bhatt et al., 2010; Tang et al., 2015). On the other hand, rTPJ is active in decisions involved self-centered and other-regarding concerns. When facing the choices between selfish and generous alternatives, TPJ inhibits selfish motivation then facilitates generosity (Strombach et al., 2015). The activity in rTPJ is also associated with altruistic allocations in dictator game (Morishima et al., 2012), and altruistic third-party punishment for unfair behaviors (David et al., 2017).

Recent studies also show causal links between the function of the rTPJ and self-centered and other-regarding concerns in behaviors with non-invasive brain stimulation techniques. For example, increasing excitability of rTPJ with anodal stimulation of transcranial direct current stimulation (tDCS) enhances performances in perspective-taking task (Santiesteban et al., 2012); decreasing excitability of rTPJ with cathodal stimulation of tDCS weakens cognitive empathy in theory of mind (Mai et al., 2016). Moreover, strengthening TPJ with tDCS increases inequality aversion in advantageous situations (Luo et al., 2017), and disrupting rTPJ with disruptive transcranial magnetic stimulation (TMS) decreases the ability to overcome egocentricity, suppressing pro-social choices (Soutschek et al., 2016). These results indicate that modulating the activity in rTPJ could change both self-centered and other-regarding concerns in behaviors.

In this study, we stimulated the rTPJ with tDCS techniques to explore how non-material self-centered motivations and other-regarding concerns affect fairness related moral hypocrisy. We used a revised version of dictator game, in which participants played as the proposer and had a chance to deceive about the total amount of money units (MUs) for allocation, then made a division between self and the recipient. The recipient cannot reject the allocation, providing the opportunity for participants to act unfairly instead of being unfair through appearing fair and excluding the effects of materialistic self-interest on moral hypocrisy. To investigate the tDCS effect on impression management and self-deception in moral hypocrisy, we manipulated whether participants' reported totals of allocation would be revealed to recipients or not. We predicted participants to deceive more when the reported totals would be revealed than unrevealed for they concern social image. And this discrepancy would be changed by increasing other-regarding concerns through tDCS stimulation on rTPJ.

MATERIALS AND METHODS

Participants

Ninety-six participants [58 females, age (mean ± SD): 22.36 ± 2.37] were recruited as proposers in two waves (72 participants in the first wave and 24 participants in the second wave). They were randomly assigned into the anodal [$n = 32$ (7 in the second wave)), cathodal ($n = 30$ (5 in the second wave)] or sham group [$n = 34$ (12 in the second wave)]. One participant in the sham group who was skeptic about the tDCS stimulation in the first wave, and three participants in the second wave who said that they thought the recipients were not real humans (one in the anodal group and two in the sham group) was excluded in the analysis (final $N = 92$). All participants were healthy students and paid according to their performances in the experiment (about 40–50 RMB). This study was approved by the Institutional Review Board of the State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University.

Procedure and Design

A 3 (tDCS: Anodal vs. Cathodal vs. Sham) × 2 (Revelation: Reported Revealed vs. Unrevealed) mixed design was run, in which the tDCS was a between-subject factor and the revelation of reported totals (whether the recipient would know the reported totals) was a within-subject factor. Firstly, participants filled the Interpersonal Reactivity Index scale (IRI) (Davis, 1980) which measures the tendency of empathy and then randomly received either anodal, cathodal or sham stimulation over the rTPJ with a constant-current stimulator (DC-Stimulator Plus, NeuroConn GmbH, Germany). A saline-soaked pair of surface sponge electrodes (in 35 cm^2 size) was used, in which the anodal or cathodal one was placed over P6 and CP6 in the international 10–20 EEG system in the brain (Jurcak et al., 2007; Santiesteban et al., 2012), and the reference one was placed over the left cheek. With a current of 1.5 mA, 15 s fade in and fade out, participants in the anodal and cathodal groups received 20 min stimulation, and participants in the sham group received anodal stimulation for 15 s (Keeser et al., 2011; Santiesteban et al., 2012).

Next, all participants were instructed to play a dictator game (Forsythe et al., 1994), in which participants played as the proposer and made a division between themselves and different recipients in 32 trials (photos of confederate recipients were shown). Participants were instructed to play with different real recipients whose photos were collected before the experiment and would be shown in each trial. They were told that they would randomly gain a total amount of money for allocation from the computer and the amount would be only known by themselves (four monetary units [MUs] (8, 10, 12, or 14) were randomly extracted in each trial and the range was not told to participants). Next, they needed to report an amount of the money for allocation (providing a chance to tell a lie) and made a division between themselves and recipients. In half of the trails, both their reported totals and offers would be revealed to recipients (Reported Revealed); in another half of trials only offers and nothing about the totals would be revealed to recipients (Unrevealed). Their divisions would determine the payoffs between themselves and the recipients, and recipients would not know the true totals in both conditions. After the instructions, participants answered checking questions including "Will your divisions affect recipients' payoffs?" "Will your true totals for allocation would be known by others?" "What will the recipient would know in the reported revealed and unrevealed condition?" and practiced to ensure that they understand the game. In each trial, participants would see a screen about pairing recipients for them, then know whether their reported totals would be revealed not before they saw the photos of a recipient. After that, they gained the total for allocation, reported the total and made the offer to the recipient. Finally, the gains would be revealed, in which participants were told that recipients would see both gains of them and the offers based on their reported totals in the reported revealed condition or recipients would only see the offers in the unrevealed condition (**Figure 1**).

After they finished the game, they rated how much different they perceived between the reported revealed and unrevealed conditions when they made decisions from 1 (Not different at all) to 7 (Strongly different). They also rated how fair of being the proposer in this game, and how fair the offers 5:5, 7:3, 8:2, 9:1 were from 1 (Not fair at all) to 7 (Strongly fair). Note that the offer 9:1 in this question means when the true total was 10, the proposer kept 9 and offered 1 to the recipient. Then they filled the Positive and Negative Affect Schedule scale (PANAS) (Watson et al., 1988) to measure their emotional states in the experiment. Finally, all of them were debriefed with questions including "What the purpose of this experiment in your opinion?" "How will these recipients feel after the experiment?" They were told the objective of this experiment and were required not to talk this study with others. To check whether they really believed that they played against real humans, participants in the second wave were also required to write down their strategies in the reported revealed and unrevealed conditions, their thoughts about the recipients and who they thought the recipients were. After that, they were also directly asked about whether they regarded recipients as real humans and knew that their divisions would take effect on recipients when they made decisions in the experiment. Only 3 in 24 participants (12.5%) reported that they didn't believe these recipients were real humans and didn't consider the recipients' payoffs would be affected by their divisions. Their data has been excluded in the analysis.

We compared the percentage of participants who actually deceived in each group, analyzed the deception rate [percentage of deceptive trials to all trials (%)], mean magnitude of dishonesty (the true total minus the reported total), and offer proportion (proportion of offers to the true total amount) with 3 (tDCS: Anodal vs. Cathodal vs. Sham) × 2 (Revelation: Reported Revealed vs. Unrevealed) mixed ANOVA.

RESULTS

Percentage of participants who actually showed deception after receiving anodal stimulation (74%) was less than cathodal (97%) and sham (94%) stimulation in the reported revealed condition

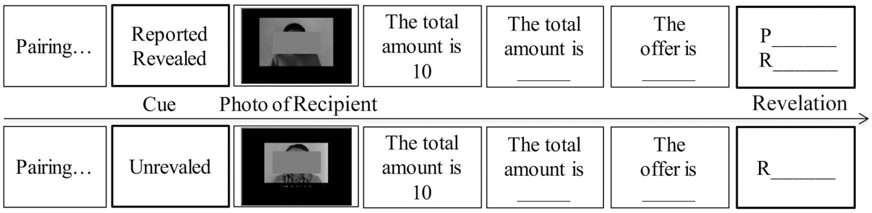

FIGURE 1 | The procedure of the experiment with two example trials that proposers were told that both gains of P (proposer) and R (recipient) would be revealed (Reported Revealed) or only the offer to R would be revealed (Unrevealed).

[$\chi^2(2) = 8.66$, $p = 0.01$]. No such difference was found in the unrevealed condition [anodal: 77%, cathodal: 87%, sham: 94%, $\chi^2(2) = 3.35$, $p = 0.19$]. On the deception rate, main effect of Revelation was found [$F(1,89) = 12.51$, $p = 0.001$, $\eta_p^2 = 0.12$], and no main effect of tDCS or interaction of tDCS × Revelation was significant ($Fs < 2.05$, $ps > 0.14$) (**Table 1**). Analysis on the magnitude of dishonesty showed significant main effect of Revelation [$F(1,89) = 8.05$, $p = 0.006$, $\eta_p^2 = 0.08$] and significant interaction of tDCS × Revelation [$F(2,89) = 3.37$, $p = 0.039$, $\eta_p^2 = 0.07$] (**Figure 2A**). Participants had greater dishonesty in the reported revealed condition than in the unrevealed condition only in the cathodal [$t(29) = 2.17$, $p = 0.038$] and sham [$t(30) = 2.53$, $p = 0.02$] groups but not in the anodal group. The tDCS effect was significant in the reported revealed [$F(2,89) = 4.07$, $p = 0.02$, $\eta_p^2 = 0.08$] but not in the unrevealed condition [$F(2,89) = 0.54$, $p = 0.58$]. That is, anodal stimulation on rTPJ reduced dishonesty than cathodal [$t(59) = -2.68$, $p = 0.01$, Cohen's $d = 0.68$] and sham [$t(60) = -2.19$, $p = 0.03$, Cohen's $d = 0.60$] stimulation in the reported revealed but not in the unrevealed condition ($ts < 0.97$, $ps > 0.34$).

Analysis on offer proportion showed significant main effect of Revelation [$F(1,89) = 15.50$, $p < 0.001$, $\eta_p^2 = 0.15$] and marginally significant main effects of tDCS [$F(1,89) = 3.06$, $p = 0.052$, $\eta_p^2 = 0.06$] (**Figure 2B**). No significant interaction of tDCS × Revelation [$F(2,89) = 0.38$, $p = 0.69$, $\eta_p^2 = 0.008$] was found. Anodal stimulation significantly increased their offers than cathodal stimulation in both reported revealed [$t(59) = 2.26$, $p = 0.02$, Cohen's $d = 0.60$] and unrevealed conditions [$t(59) = 2.23$, $p = 0.03$, Cohen's $d = 0.57$]. The main effect of tDCS on fairness was also marginally significant in the rating of four offers (5:5, 7:3, 8:2 and 9:1) [$F(1,89) = 2.47$, $p = 0.09$, $\eta_p^2 = 0.05$], in which anodal stimulation significantly and sham stimulation marginally decreased the fair ratings of the offers 8:2 [anodal: $t(59) = -2.01$, $p = 0.049$, Cohen's $d = 0.51$; sham: $t(59) = -1.77$, $p = 0.08$, Cohen's $d = 0.46$] and 9:1 [anodal: $t(59) = -2.12$, $p = 0.038$, Cohen's $d = 0.54$; sham: $t(59) = -2.00$, $p = 0.05$, Cohen's $d = 0.51$] (**Figure 2C**).

No significant difference was found for the response time either when participants reported the total or when participants made the offer (see response time in two conditions in **Table 1**) ($Fs < 2.07$, $ps > 0.13$), indicating that our results were not caused by tDCS changed participants' cognitive ability in this game. Participants' perceived difference between the two conditions in decisions (anodal: 3.48 ± 2.06; sham: 3.58 ± 1.98; cathodal: 3.17 ± 1.97) and perceived fairness of being the proposer in this game (anodal: 3.35 ± 1.70; sham: 2.81 ± 1.49; cathodal: 3.50 ± 1.96) were not affected by tDCS stimulation on rTPJ ($Fs < 1.38$, $ps > 0.26$). In addition, IRI scores (including perspective taking, fantasy, and empathic concern) before the brain stimulation ($Fs < 1.76$, $ps > 0.18$), and PANAS scores at the end of the experiment were not different among three groups ($Fs < 2.58$, $ps > 0.08$). These results excluded the possibilities that difference of participants' behaviors was caused by their essential perception of the conditions or being the proposers *per se*, or they were different in empathy or emotional state.

DISCUSSION

The present study examined the role of self-centered and other-regarding concerns in deception in fairness related moral hypocrisy through stimulating rTPJ by tDCS. We found that deception in moral hypocrisy was increased by revealing appearing fair without true fairness to recipients than not and this effect was decreased by anodal stimulation on rTPJ rather than cathodal and sham stimulation. Anodal stimulation on rTPJ increased truly fairness than cathodal stimulation regardless of the revelation of appearing fair and led participant to rate extremely unfair offers less fair. These findings suggest that exciting activity in rTPJ increases other-regarding concerns

TABLE 1 | Deception rate (%), response time (RT: ms) when participants reported the total and made the offer (mean).

	Deception rate (%)		Reporting totals (RT: ms)		Making offers (RT: ms)	
	Reported Revealed	Unrevealed	Reported Revealed	Unrevealed	Reported Revealed	Unrevealed
Anodal	54	52	1265	1382	1198	1211
Sham	71	57	1166	1293	1168	1209
Cathodal	70	57	1273	1276	1087	1223

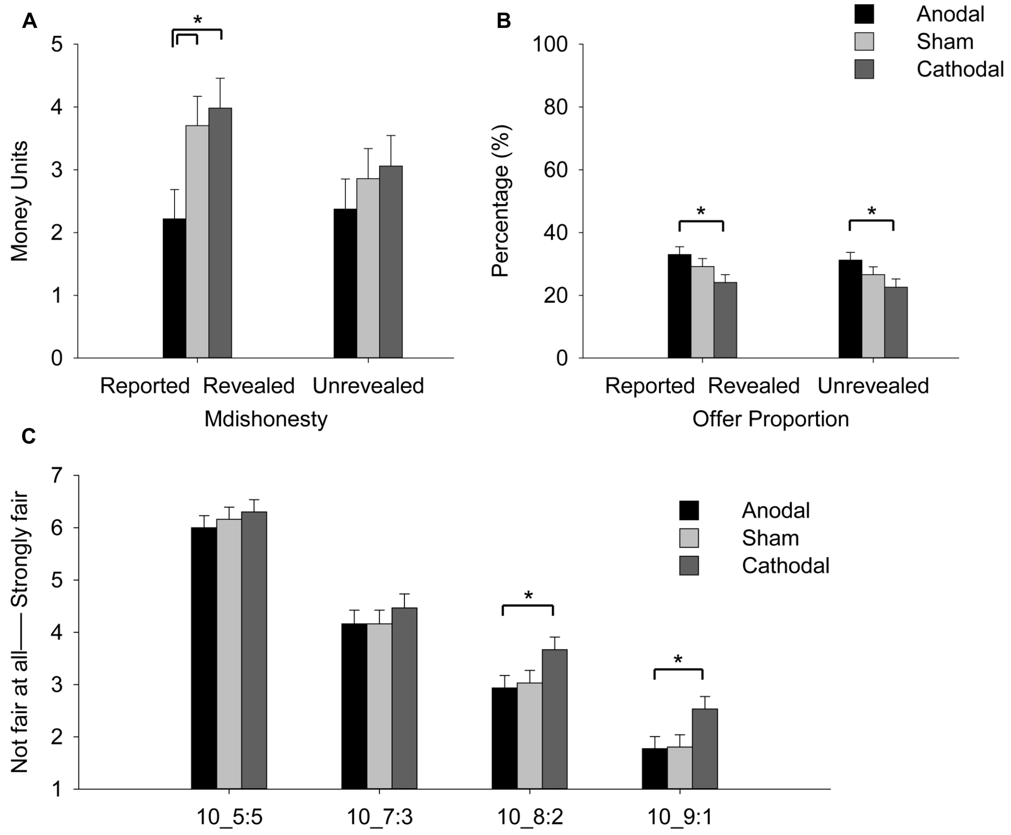

FIGURE 2 | (A) Mean magnitude of dishonesty after receiving anodal, cathodal and sham stimulation with tDCS over rTPJ. **(B)** Mean offer proportion after tDCS stimulation. **(C)** Fairness rating of 5:5, 7:3, 8:2, and 9:1 offers based on the true total as 10 after the task (*$p < 0.05$). Error bars indicate standard errors.

then increases truly fair behaviors. Specifically, it decreases non-material self-centered deception in moral hypocrisy when social image concerns exist but not when social image concerns are lacking.

Previous studies have discussed how rTPJ contributes to deception through understanding other's minds (Bhatt et al., 2010; Tang et al., 2015). In those cases, rTPJ processes beliefs or intentions of others, and helps to build one's reputation in social interaction then assists deception. However, our findings confirmed the causal role of rTPJ in deception with a different access. In the current study, it is unnecessary for participants to mentalize how recipients' responses would affect their own gains in the current trial, or to build the reputation for future materialistic reward. The repeated one-shot dictator game in which the recipients cannot reject allocations and recipients were different in each trial removed effects of both current and long-term social interaction and material reward on deception.

Results that enhancing rTPJ decreased the deception in moral hypocrisy provided more information for this access. When the reported total would be revealed, it is hard to separate the effects of self-deception and impression management motivations in moral hypocrisy. In contrast, when the reported total is not revealed, it lacks social image concerns then leads self-deception motivation to be more prominent (Batson et al., 2002; von Hippel and Trivers, 2011). Our findings that participants deceived a lot in the unrevealed condition confirmed the existing of self-deception in moral hypocrisy, and the cathodal and sham group deceived more in the reported revealed than in the unrevealed condition support that participants concerned social image in other's eyes. Moreover, anodal stimulation decreased the difference of deception between these two conditions through decreasing deception in the reported revealed condition, suggesting that rTPJ is only involved in moral hypocrisy driven by impression management but not by self-deception.

Exciting rTPJ increased truly fair behaviors provided further explanation for these results. In line with previous findings that rTPJ inhibits selfish motivations to maximize materialistic benefit and facilities other-regarding behaviors in allocation (Morishima et al., 2012; Strombach et al., 2015; Luo et al., 2017), our results show that exciting rTPJ increased other-regarding concerns regardless of whether fairness would be perceived or not. Moreover, the enhancement of other-regarding concerns decreases deception in moral hypocrisy driven by concerns of social image rather than increasing the moral hypocrisy by endorsing other-regarding moral principles to protect one's social image (Szabados and Soifer, 2004). These findings provide causal evidence for the role of other-regarding concerns in reducing moral hypocrisy (Batson et al., 2003; Graham et al., 2015), and indicate that this effect might be caused by TPJ constructs social contexts through integrating social information and reorients people's attention to social stimuli (Carter and Huettel, 2013), then exciting rTPJ prompts people to pay more attention to interpersonal processes involved in impression management (Schlenker and Weigold, 1992). That is, increasing other-regarding concerns facilitates considering other's evaluations and expectations, therefore, decreases deception in moral hypocrisy driven by impression management rather than self-deception.

Another possibility is that rTPJ is not involved in self-deception processing. Recently, researchers investigated the neural correlates of self-deception and impression management with the Balanced Inventory of Desirable Responding (BIDR) scale through functional magnetic resonance imaging (fMRI) technique (Farrow et al., 2015; Paulhus, unpublished). They found that impression management is correlated with activity in the left TPJ, whereas self-deception is not correlated with activity in bilateral TPJ. As the authors noted, the reason why the fMRI study did not find the relationship between impression management and rTPJ might be they did not directly measure participants' hypocritical behaviors based on impression management and self-deception (Farrow et al., 2015). In line with this study, we found that the rTPJ is engaged in processing one's public image but not in promoting self-concept. One potential explanation is self-deception involves the mechanisms for action selection and interpretation to justify self-serving unethical behaviors and diminish the threat to moral self (Mijović-Prelec and Prelec, 2010; Shaul et al., 2015). These mechanisms might be more closely related to cognitive control system since increasing cognitive load disrupts rationalization and justification of one's own moral transgressions (Valdesolo and DeSteno, 2008) and repeatedly exposing the truth decreases self-deception (Chance et al., 2015).

One limitation of this study is it is hard to obtain the baseline of self-deception in the interpersonal moral hypocrisy which essentially involves concerns of both self and others. Future studies using intrapersonal moral hypocrisy paradigms would provide more evidence for how self-centered and other-regarding concerns affect self-deception. Another limitation is we used photos of real humans to construct the social context between participants and recipients rather real participants. However, only very few participants suspected that they played against real recipients, indicating that most participants decreased moral hypocrisy for protecting the social image.

Taken together, our study is the first one to investigate how activity in rTPJ affects deception in fairness related moral hypocrisy. The results support that rTPJ is involved in other-regarding behaviors and contributes to decreasing deception in moral hypocrisy through facilitating interpersonal processes. Future studies about how cognition for deception and fairness is processed in moral hypocrisy would be helpful to understand the role of rTPJ in decisions for non-material reward.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Institutional Review Board of the State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University with written informed consent from all subjects. All subjects gave written informed consent. The protocol was approved by the Institutional Review Board of the State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University.

AUTHOR CONTRIBUTIONS

HT, SW, and CL designed experiment. HT, PY, SW, and LT performed experiment. RZ and LT analyzed data and drew figures. HT, PY, SW, RZ, SS, and CL wrote the manuscript. HT, RZ, SS, LT, and CL revised the manuscript and HT, PY, SW, RZ, SS, LT, and CL finally approved the version to be published.

FUNDING

This work was supported by the National Key R&D Program of China (2017YFC0803402), the National Basic Research Program of China (2013CB837300), the National Natural Science Foundation of China (NSFC) (31170971, 71572018, 71302024), the Major Project of National Social Science Foundation (12&ZD228), the Beijing Municipal Science and Technology Commission (Z151100003915122), the National Program for Support of Top-notch Young Professionals.

ACKNOWLEDGMENT

Thanks for the help of Wenli Zhang and Rui Su in the initial data collection for this study.

REFERENCES

Batson, C. D., Kobrynowicz, D., Dinnerstein, J. L., Kampf, H. C., and Wilson, A. D. (1997). In a very different voice: unmasking moral hypocrisy. *J. Pers. Soc. Psychol.* 72, 1335–1348. doi: 10.1037/0022-3514.72.6.1335

Batson, C. D., Lishner, D. A., Carpenter, A., Dulin, L., Harjusola-Webb, S., Stocks, E. L., et al. (2003). "As you would have them do unto you": Does imagining yourself in the other's place stimulate moral action? *Pers. Soc. Psychol. Bull.* 29, 1190–1201. doi: 10.1177/0146167203254600

Batson, C. D., Thompson, E. R., and Chen, H. (2002). Moral hypocrisy: addressing some alternatives. *J. Pers. Soc. Psychol.* 83, 330–339. doi: 10.1037//0022-3514.83.2.330

Batson, C. D., Thompson, E. R., Seuferling, G., Whitney, H., and Strongman, J. A. (1999). Moral hypocrisy: appearing moral to oneself without being so. *J. Pers. Soc. Psychol.* 77, 525–537. doi: 10.1037//0022-3514.77.3.525

Bhatt, M. A., Lohrenz, T., Camerer, C. F., and Montague, P. R. (2010). Neural signatures of strategic types in a two-person bargaining game. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19720–19725. doi: 10.1073/pnas.1009625107

Biziou-van-Pol, L., Haenen, J., Novaro, A., Liberman, A. O., and Capraro, V. (2015). Does telling white lies signal pro-social preferences? *Judgm. Decis. Mak.* 10, 538–548. doi: 10.2139/ssrn.2617668

Carpenter, T. P., and Marshall, M. A. (2009). An examination of religious priming and intrinsic religious motivation in the moral hypocrisy paradigm. *J. Sci. Study Relig.* 48, 386–393. doi: 10.1111/j.1468-5906.2009.01454.x

Carter, R. M., and Huettel, S. A. (2013). A nexus model of the temporal–parietal junction. *Trends Cogn. Sci.* 17, 328–336. doi: 10.1016/j.tics.2013.05.007

Chance, Z., Gino, F., Norton, M. I., and Ariely, D. (2015). The slow decay and quick revival of self-deception. *Front. Psychol.* 6:1075. doi: 10.3389/fpsyg.2015.01075

David, B., Hu, Y., Krüger, F., and Weber, B. (2017). Other-regarding attention focus modulates third-party altruistic choice: an fMRI study. *Sci. Rep.* 7:43024. doi: 10.1038/srep43024

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog Sel. Doc. Psychol.* 10, 85.

Decety, J., and Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *Neuroscientist* 13, 580–593. doi: 10.1177/1073858407304654

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., and Epstein, J. A. (1996). Lying in everyday life. *J. Pers. Soc. Psychol.* 70, 979–995. doi: 10.1037/0022-3514.70.5.979

Farrow, T. F., Burgess, J., Wilkinson, I. D., and Hunter, M. D. (2015). Neural correlates of self-deception and impression-management. *Neuropsychologia* 67, 159–174. doi: 10.1016/j.neuropsychologia.2014.12.016

Forsythe, R., Horowitz, J. L., Savin, N. E., and Sefton, M. (1994). Fairness in simple bargaining experiments. *Games Econ. Behav.* 6, 347–369. doi: 10.1006/game.1994.1021

Gneezy, U. (2005). Deception: the role of consequences. *Am. Econ. Rev.* 95, 384–394. doi: 10.1257/0002828053828662

Graham, J., Meindl, P., Koleva, S., Iyer, R., and Johnson, K. M. (2015). When values and behavior conflict: moral pluralism and intrapersonal moral hypocrisy. *Soc. Personal. Psychol. Compass* 9, 158–170. doi: 10.1111/spc3.12158

Jurcak, V., Tsuzuki, D., and Dan, I. (2007). 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage* 34, 1600–1611. doi: 10.1016/j.neuroimage.2006.09.024

Keeser, D., Meindl, T., Bor, J., Palm, U., Pogarell, O., Mulert, C., et al. (2011). Prefrontal transcranial direct current stimulation changes connectivity of resting-state networks during fMRI. *J. Neurosci.* 31, 15284–15293. doi: 10.1523/JNEUROSCI.0542-11.2011

Lönnqvist, J.-E., Irlenbusch, B., and Walkowitz, G. (2014). Moral hypocrisy: impression management or self-deception? *J. Exp. Soc. Psychol.* 55, 53–62. doi: 10.1016/j.jesp.2014.06.004

Luo, J., Chen, S., Huang, D., Ye, H., and Zheng, H. (2017). Whether modulating the activity of the temporalparietal junction alters distribution decisions within different contexts: evidence from a tDCS study. *Front. Psychol.* 8:224. doi: 10.3389/fpsyg.2017.00224

Mai, X., Zhang, W., Hu, X., Zhen, Z., Xu, Z., Zhang, J., et al. (2016). Using tDCS to explore the role of the right temporo-parietal junction in theory of mind and cognitive empathy. *Front. Psychol.* 7:380. doi: 10.3389/fpsyg.2016.00380

Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. C., and Fehr, E. (2012). Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron* 75, 73–79. doi: 10.1016/j.neuron.2012.05.021

Murray, R. J., Schaer, M., and Debbané, M. (2012). Degrees of separation: a quantitative neuroimaging meta-analysis investigating self-specificity and shared neural activation between self-and other-reflection. *Neurosci. Biobehav. Rev.* 36, 1043–1059. doi: 10.1016/j.neubiorev.2011.12.01

Mijović-Prelec, D., and Prelec, D. (2010). Self-deception as self-signalling: a model and experimental evidence. *Philos. Trans. R. Soc. B* 365, 227–240. doi: 10.1098/rstb.2009.0218

Santiesteban, I., Banissy, M. J., Catmur, C., and Bird, G. (2012). Enhancing social ability by stimulating right temporoparietal junction. *Curr. Biol.* 22, 2274–2277. doi: 10.1016/j.cub.2012.10.018

Saxe, R., Moran, J. M., Scholz, J., and Gabrieli, J. (2006). Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. *Soc. Cogn. Affect. Neurosci.* 1, 229–234. doi: 10.1093/scan/nsl034

Schlenker, B. R., and Weigold, M. F. (1992). Interpersonal processes involving impression regulation and management. *Annu. Rev. Psychol.* 43, 133–168. doi: 10.1146/annurev.ps.43.020192.001025

Shaul, S., Francesca, G., Rachel, B., and Shahar, A. (2015). Self-serving justifications: doing wrong and feeling moral. *Curr. Dir. Psychol. Sci.* 24, 125–130. doi: 10.1177/0963721414553264

Sheremeta, R. M., and Shields, T. W. (2013). Do liars believe? Beliefs and other-regarding preferences in sender–receiver games. *J. Econ. Behav. Organ.* 94, 268–277. doi: 10.1016/j.jebo.2012.09.023

Soutschek, A., Ruff, C. C., Strombach, T., Kalenscher, T., and Tobler, P. N. (2016). Brain stimulation reveals crucial role of overcoming self-centeredness in self-control. *Sci. Adv.* 2:e1600992. doi: 10.1126/sciadv.1600992

Strombach, T., Weber, B., Hangebrauk, Z., Kenning, P., Karipidis, I. I., Tobler, P. N., et al. (2015). Social discounting involves modulation of neural value signals by temporoparietal junction. *Proc. Natl. Acad. Sci. U.S.A.* 112, 1619–1624. doi: 10.1073/pnas.1414715112

Szabados, B., and Soifer, E. (2004). Hypocrisy: ethical investigations. *Dialogue* 45, 395.

Tang, H., Mai, X., Wang, S., Zhu, C., Krueger, F., and Liu, C. (2015). Interpersonal brain synchronization in the right temporo-parietal junction during face-to-face economic exchange. *Soc. Cogn. Affect. Neurosci.* 11, 23–32. doi: 10.1093/scan/nsv092

Toma, C. L., Hancock, J. T., and Ellison, N. B. (2008). Separating fact from fiction: an examination of deceptive self-presentation in online dating profiles. *Pers. Soc. Psychol. Bull.* 34, 1023–1036. doi: 10.1177/0146167208318067

Valdesolo, P., and DeSteno, D. (2007). Moral hypocrisy social groups and the flexibility of virtue. *Psychol. Sci.* 18, 689–690. doi: 10.1111/j.1467-9280.2007.01961.x

Valdesolo, P., and DeSteno, D. (2008). The duality of virtue: deconstructing the moral hypocrite. *J. Exp. Soc. Psychol.* 44, 1334–1338. doi: 10.1016/j.jesp.2008.03.010

Van der Meer, L., Groenewold, N. A., Nolen, W. A., Pijnenborg, M., and Aleman, A. (2011). Inhibit yourself and understand the other: neural basis of distinct processes underlying Theory of Mind. *Neuroimage* 56, 2364–2374. doi: 10.1016/j.neuroimage.2011.03.053

von Hippel, W., and Trivers, R. (2011). The evolution and psychology of self-deception. *Behav. Brain Sci.* 34, 1–16. doi: 10.1017/S0140525X10001354

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070. doi: 10.1037/0022-3514.54.6.1063

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Psychology

ORIGINAL RESEARCH
published: 01 June 2018
doi: 10.3389/fpsyg.2018.00860

Functional Dissociation of the Posterior and Anterior Insula in Moral Disgust

Xiaoping Ying[1,2], Jing Luo[3], Chi-yue Chiu[4], Yanhong Wu[5], Yan Xu[1]* and Jin Fan[6,7,8,*]

[1] Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education, Faculty of Psychology, Beijing Normal University, Beijing, China, [2] Institute of Sociology, Chinese Academy of Social Sciences, Beijing, China, [3] School of Psychology, Capital Normal University, Beijing, China, [4] Department of Psychology, The Chinese University of Hong Kong, Shatin, Hong Kong, [5] School of Psychological and Cognitive Sciences, Peking University, Beijing, China, [6] Department of Psychology, Queens College, The City University of New York, Flushing, NY, United States, [7] Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, United States, [8] Fishberg Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, United States, [9] The Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, United States

The insula is thought to be involved in disgust. However, the roles of the posterior insula (PI) and anterior insula (AI) in moral disgust have not been clearly dissociated in previous studies. In this functional magnetic resonance imaging study, the participants evaluated the degree of disgust using sentences related to mild moral violations with different types of behavioral agents (mother and stranger). The activation of the PI in response to the stranger agent was significantly higher than that in response to the mother agent. In contrast, the activation of the AI in response to the mother agent was significantly higher than that in response to the stranger agent. These data suggest a clear functional dissociation between the PI and AI in which the PI is more involved in the primary level of moral disgust than is the AI, and the AI is more involved in the secondary level of moral disgust than is the PI. Our results provide key evidence for understanding the principle of embodied cognition and particularly demonstrate that high-level moral disgust is built on more basic disgust via a mental construction approach through a process of embodied schemata.

Keywords: moral disgust, agent, posterior insula, anterior insula, fMRI

OPEN ACCESS

Edited by:
Qi Li,
University of Chinese Academy of Sciences (UCAS), China

Reviewed by:
Fei Luo,
Institute of Psychology (CAS), China
Pei Sun,
Tsinghua University, China

Correspondence:
Yan Xu
xuyan@bnu.edu.cn
Jin Fan
jin.fan@qc.cuny.edu

Specialty section:
This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Psychology

Received: *31 December 2017*
Accepted: *14 May 2018*
Published: *01 June 2018*

Citation:
Ying X, Luo J, Chiu C-y, Wu Y, Xu Y and Fan J (2018) Functional Dissociation of the Posterior and Anterior Insula in Moral Disgust. Front. Psychol. 9:860. doi: 10.3389/fpsyg.2018.00860

INTRODUCTION

Morality is the center of our attitudes and behaviors in daily social life and beyond (Haidt and Kesebir, 2010). Moral judgment has been generally recognized to encompass not only reasoning but also emotion and affection (Greene and Haidt, 2002), and disgust has a strong impact on moral judgment and is rudimentary to moral emotion (Miller, 2008; Rozin et al., 2008; Giubilini, 2016). Neuroimaging studies have shown that the insula is involved in physical disgust, moral judgment (Moll et al., 2005), and detecting norm violations (Xiang et al., 2013; Cheng et al., 2017). However, how moral disgust is encoded and represented in the insula remains unclear.

The insular cortex, which is a key region responsible for encoding and re-encoding feelings, consists of regions with variable cell structures or cytoarchitectures ranging from granular in the posterior portion to agranular in the anterior portion (Flynn, 1999; Varnavas and Grand, 1999). The posterior-to-anterior progression, which includes increasingly complex representations in the human insula, indicates that the posterior insula (PI) plays a role in encoding more primary emotions, the mid-insula plays a role in encoding contextual integration (Craig, 2002, 2009), and the anterior insula (AI) plays a role in encoding introspective awareness of emotion and bodily states (Critchley et al., 2004; Paulus and Stein, 2006). This hypothesis provides a new perspective for understanding how a complicated, high-level mentality or emotionality is built or developed from more basic feelings.

A neuroimaging study investigating the relationship between love and sexual desire revealed that the anterior part of the insula was significantly activated by feelings of love, whereas the posterior part of the left insula was significantly activated by primary feelings, such as sexual desire (Cacioppo et al., 2012). A study investigating the neurodevelopmental changes in the circuits underlying empathy and sympathy from childhood to adulthood found a significant negative correlation between age and the degree of activation in the PI and a positive correlation in the anterior portion of the insula (Decety and Michalska, 2010), suggesting that a higher level of frontalization of inhibitory capacity and a greater top–down modulation of activity occur in primitive emotion-processing regions during individual development (Yurgelun-Todd, 2007). In a study investigating fairness in relation to moral judgments, the PI was selectively associated with the processing of the objective aspects of fairness, whereas the more anterior part, i.e., the mid-insula, was involved in the processing of the contextual aspects of fairness, suggesting that the mid-insula performs a re-encoding function for the integration of context with inequality (Haidt and Kesebir, 2010; Wright et al., 2011).

However, studies investigating the involvement of the anterior and posterior insula in moral judgment and disgust have been inconsistent. Most studies report that the anterior part of the insula was activated in moral indignation/disgust relative to pure disgust (Moll et al., 2005), while passively viewing pictures depicting social moral violations relative to viewing these pictures with an endeavor to decrease emotional reactions (Harenski and Hamann, 2006), in deontological guilt relative to altruistic guilt (Basile et al., 2011), while retrieving personal guilt or shameful memories (Wagner et al., 2011), and in guilt associated with prejudice (Fourie et al., 2014). In addition, compared with the processing of easy moral dilemmas, the anterior part of the insula was involved in the processing of difficult personal moral dilemmas (Greene et al., 2004) and difficult dilemmas where the to-be-sacrificed person was humanized as a full-blown individual with mental states (Majdandžić et al., 2012). In contrast to the involvement of the anterior part of the insula, the involvement of the PI in moral processing, such as in a comparison between moral indignation and a neutral condition (Moll et al., 2005) or between sociomoral violation actions and physically repulsive actions (Schaich Borg et al., 2008), has only been occasionally reported.

To date, no study has doubly dissociated the function of the PI and AI in moral disgust. Given that functional segregation has been generally established in this extensive and cytoarchitectonically diverse cortical region (Flynn, 1999; Varnavas and Grand, 1999), the double dissociation of the PI and AI in moral disgust could have important theoretical implications for moral cognition and emotion, particularly for the theory that disgust in response to moral violations is built on more basic types of disgust (such as that associated with distaste for food and body waste products) through a process of embodied schemata, which refers to patterns of experience that are based on bodily knowledge or sensation (Haidt et al., 1997).

In this study, we separated the following two components involved in the representation of moral disgust: the primary moral disgust component represented in the PI and the secondary component represented in the AI. The dissociation of these two components was achieved by requiring participants to process moderate moral transgression behaviors (e.g., speaking loudly on the telephone in a public place or saying dirty words in a public place) with different behavioral agents (stranger or mother). Moral indignation toward a stranger who behaves immorally was relatively primary and featured feelings of anger and hate; thus, the PI was challenged. However, moral indignation toward the mother was relatively secondary, required relatively high levels of integration and regulation and featured feelings of shame or guilt; thus, the AI was involved.

MATERIALS AND METHODS

Participants

Thirty-six healthy, right-handed students with normal or corrected-to-normal vision participated in this study. No participants had a history of neurological or psychiatric disorders or head injury. Of these participants, six participants were excluded from fMRI analysis due to device or technical errors, and one participant was excluded due to excessive head movement (>3 mm). The final sample included 29 participants (14 females; mean age 22.4 ± 2.40; range 19–28 years). The participants were compensated for their time. Before the fMRI scan, written informed consent approved by the local Ethics Committee at Beijing Normal University was obtained from each participant. Before the fMRI experiment, we asked the participants to answer a list of self-developed questions regarding their relationship with their mothers, and only individuals who indicated a close relationship with their mothers [quantified by answering a 4 or 5 on a five-point Likert scale from 1 ("very bad") to 5 ("very good")] were included in the sample. The participants also completed the Chinese version of the Yale-Brown Obsessive Compulsive Scale (Goodman et al., 1989) and Toronto Alexithymia Scale (Parker et al., 2003).

Materials

The experimental stimuli included 60 sentences describing situations of moral transgressions (e.g., speaking loudly on the telephone in a public place) that frequently occur in daily life and reliably evoke moral disgust. The severity of the moral transgressions was controlled at a moderate level because a serious transgression (such as "killing someone") was inapplicable to the participants' mothers. The length and complexity of the sentences were carefully controlled. Another sample of participants (N = 100) who did not participate in the formal fMRI experiment rated the severity of the moral transgressions and the emotional arousal associated with each moral transgression event. The severity of the moral transgressions was rated using a six-point Likert scale from 1 ("not serious at all") to 6 ("extremely serious"), and five emotions, i.e., disgust, anger, surprise, sadness and disappointment, were rated using a seven-point Likert scales from 0 ("no feeling at all") to 6 ("extremely strong"). The 60 sentences had a similar length and complexity and were divided into three equal groups according to the severity and emotional arousal ratings. No statistically significant differences were observed in the severity of the moral transgressions, moral disgust or other types of emotional arousal among the three groups of materials (**Supplementary Table S1**). In the fMRI experiment, each group of materials was assigned to only one of the three experimental conditions that used the "stranger," "mother," or "best friend" as the behavioral agent (**Table 1**). The assignment of a given agent to a given group of materials were counter-balanced across the participants.

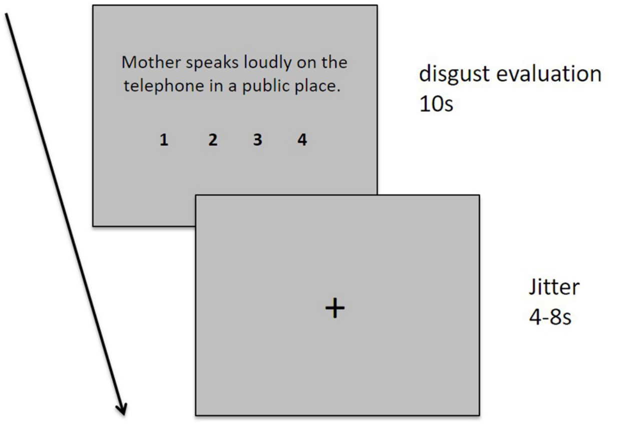

FIGURE 1 | Task design. Each sentence was displayed for 10 s, followed by the presentation of a fixation cross for 4–8 s. A fixation cross was presented for 30 s at the start and end of the task and for 16 s during each trial.

Task Design and Procedures

During the experimental fMRI scan session, the participants were asked to read and evaluate 60 sentences describing different moral transgression events with different behavioral agents (mother, stranger, and best friend) one-by-one. Each sentence was presented for 10 s, followed by a cross fixation phase of a varied duration ranging from 4 to 8 s. During the 10-s sentence presentation stage, the participants were instructed to read and comprehend the situation described by the sentence and evaluate their degree of disgust using a four-point Likert scale from 1 ("not disgusting at all") to 4 ("extremely disgusting"). The participants were required to indicate their evaluation by pressing one of four buttons using their index, middle, fourth, or little finger of their right hand. The degree of disgust and the numbers 1, 2, 3, and 4 were presented below the sentence (see **Figure 1** for a detailed description of each sentence).

To prevent the participants from frequently switching between the behavioral agent of the moral transgressions, 20 sentences in each condition were separated into two sub-groups with 10 sentences in each sub-group, and the 10 sentences in each sub-group were presented successively in one block. Therefore, two blocks of each of the three experimental conditions involved the mother, stranger and best friend as the behavioral agent, and the participants completed a total of six blocks. The sequences of the block presentations were counter-balanced across the participants with the restriction that two blocks in the same condition could never be presented successively. During the periods between the blocks, a fixation (cross-viewing) was presented for 16 s. In addition, 30-s fixation periods were presented at the beginning and end of the session.

TABLE 1 | Sample sentences from the experimental materials.

	Stranger	Best friend	Mother
Moral disgust	Stranger says dirty words in a public place	Best friend chats at a concert	Mother speaks on the telephone loudly in a public place

Image Acquisition

All MRI scans were acquired using a Siemens MAGNETOM Trio 3T MR scanner at the Imaging Center for Brain Research at Beijing Normal University. Foam padding and a plastic brace were used to minimize head movement. For the functional imaging, the whole-brain coverage of 33 axial slices was acquired using a T2-weighted echo-planar imaging sequence based on the blood oxygenation level-dependent (BOLD) contrast with the following parameters: 2000 ms repetition time (TR), 30 ms echo time (TE), 90° flip angle, 4.0-mm slice thickness, 0.6-mm gap, 64 × 64 data matrix, 200-mm field of view (FOV), and 3.1 × 3.1 × 4.0-mm voxel size. In addition, 3D structural brain scans were also acquired for each participant using a T1-weighted anatomical scan with the following parameters: 2530-ms TR, 3.39-ms TE, 7° flip angle, 256 × 256 data matrix, 256-mm FOV, 1.3 × 1.0 × 1.3-mm voxel size, and Bandwidth (BW) = 190 Hz/pixel.

Image Data Analysis

The event-related analyses of the fMRI data from the moral disgust task were conducted using a statistical parametric mapping package (SPM8; Wellcome Trust Centre for Neuroimaging, London, United Kingdom). In the preprocessing of the data, each image volume was slice-time corrected, realigned, unwarped to the first volume, co-registered to the structural scan images, spatially normalized to the Montreal Neurological Institute (MNI) ICBM152 space based on the normalization parameters of the T1 image, subsampled to a voxel size of 2 × 2 × 2 mm, and finally spatially smoothed using a Gaussian kernel of 8 mm full-width half-maximum.

For statistical analysis, a general linear model (GLM) was constructed to analyze the functional scans from each participant with a duration of 10 s by regressing the observed event-related

BOLD signals on the regressors to identify the relationship between the hemodynamic responses and task events. Low-frequency drifts in the signal were removed using a high-pass filter with a 128-s cutoff. Regressors were created by convolving a train of delta functions representing the sequence of individual events using the default SPM basis function, which consists of a synthetic hemodynamic response function (HRF) composed of two gamma functions (Friston et al., 1998). Three regressors were used for the three conditions (mother, best friend and stranger). The 6 parameters generated during the motion correction were also entered as covariates. In addition, HRF related to trials in which the participants failed to respond was also modeled separately and explicitly to partial out error-related activity. Linear contrasts of the parameter estimates were performed to identify the effects of the three conditions and the difference between every two conditions in each session. Then, the first level contrasts were aggregated into a second level, and one-sample *t*-tests were performed to compute the group-level statistics using a random-effects model.

Regions of Interest (ROIs) and Psychophysiological Interaction (PPI) Analysis

To define the regions of interest (ROIs), we first conducted contrasts between the stranger condition and the mother condition. To test our hypotheses regarding the role of the AI and PI in moral disgust, ROI analyses were performed based on the templates developed by Lin and colleagues (Lin et al., 2013), which consisted of six insula regions, including the left and right AI (LAI and RAI), left and right PI (LPI and RPI), and left and right middle insula (LMI and RMI). The ROIs were defined by masking the six abovementioned insula regions on the whole brain results of a given contrast (e.g., the contrast of "mother condition minus stranger condition" and the contrast of "stranger condition minus mother condition"). The significance level was set at an uncorrected threshold of $p < 0.05$ with a cluster extent of at least 5 contiguous voxels. The LAI was activated in the mother condition minus the stranger condition. The LPI was activated in the stranger condition minus the mother condition (see **Table 2** for details). ROIs as clusters were created for the LAI and LPI. The BOLD signal changes were extracted from each ROI for the contrast between the stranger condition and the mother condition. Separate psychophysiological interaction (PPI) analyses were also performed using the LAI or LPI ROIs as seeds.

Psychophysiological interaction analyses provide a measure of functional connectivity change among different brain regions depending on a specific psychological context (Friston et al., 1997). This analysis was achieved using a moderator derived from the product of the activity of a source region and the psychological context. The LAI and LPI were derived from the ROI analysis and identified in moral disgust by the saliency level of the contrast between the stranger condition and the mother condition (see the Results). We aimed to determine whether the AI and PI functionally interact with regions involved in secondary and primary moral disgust processing, respectively. PPI analysis was performed to identify the region(s) that had differential connectivity with the AI and PI modulated by the difference between the stranger agent and the mother agent in moral disgust.

RESULTS

Before presenting the results, the following two points should be noted. First, in this paper, we focused on the results of the mother condition and stranger condition, and the results of the best friend condition are not reported in the present paper. This condition was omitted because the main goal of this study is to dissociate the function of the PI and AI in moral disgust. The ideal way to achieve this goal is to perform a direct contrast between a very intimate relationship, i.e., the mother condition, and a very distant relationship, i.e., the stranger condition. Furthermore, the best friend condition complicates the situation because the nature of friendship is unclear and could vary from person to person (**Supplementary Figure S1**). Second, regarding the brain imaging results, we only focused on the ROIs in the anterior, middle and PI and brain regions found to be functionally connected to these ROIs. We conducted and inspected the results of the whole-brain analysis and confirmed the general validity of the results. For example, we confirmed that the brain activation we observed during the visual, linguistic, and cognitive control processing in our moral judgment task was similar to that reported in other related studies. However, these results are not the focus of this study and are provided in the Supplementary Materials.

Behavioral Results

The behavioral data analyzed included the disgust ratings (DR), response times (RT), and severity ratings (SR). The online recording of the participants' behavioral responses during the MRI scanning showed that the participants required a significantly longer duration to complete the DRs in the stranger condition [Mean = 4.47, standard deviation (SD) = 0.824] than in the mother condition (Mean = 4.26, *SD* = 0.854) [$t_{(28)} = 2.55$, $p < 0.05$] (**Figure 2A**), and the participants rated the strangers performing moral transgressions as significantly more disgusting (Mean = 2.87, SD = 0.539) than those of the mothers (Mean = 2.64, *SD* = 0.494) [$t_{(28)} = 2.76$, $p < 0.01$] (**Figure 2B**).

TABLE 2 | Brain activation of the insula in a contrast between the stranger and mother conditions.

Insula region	Side	x	y	Z	T	Z	P	K
Moral disgust: Mother > Stranger								
Anterior	Left	−34	18	−16	1.83	1.77	<0.05	43
Moral disgust: Stranger < Mother								
Posterior	Left	−38	−10	22	2.28	2.17	<0.05	66
		−42	−8	12	2.22	2.12	<0.05	
middle	Right	42	2	0	2.02	1.94	<0.05	18

P < 0.05, K > 5 of 2 mm × 2 mm × 2 mm voxels, LAI (−34 18 −16), LPI (−38 −10 22), RMI (42 2 0).

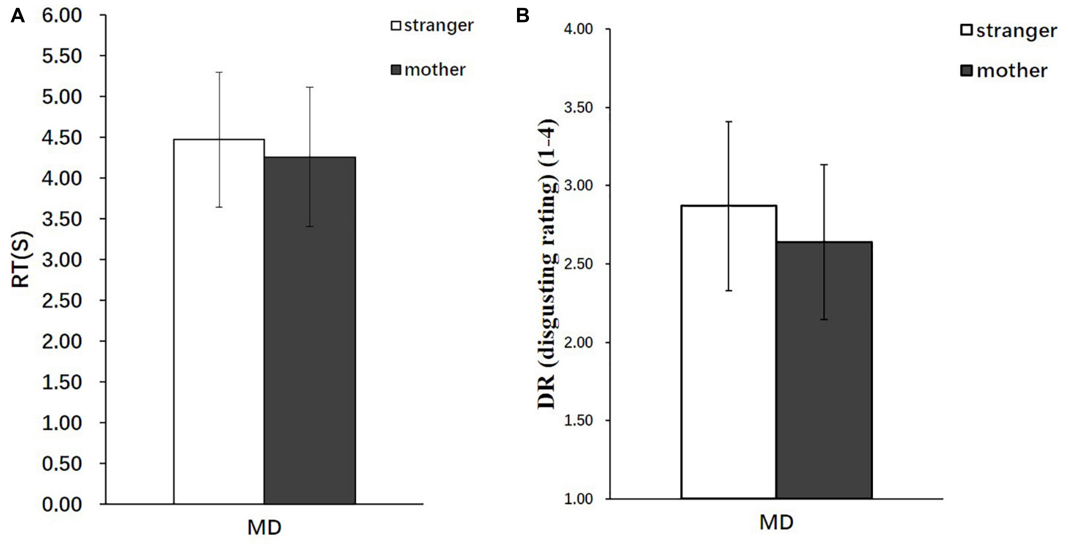

FIGURE 2 | Behavioral results (*n* = 29). **(A)** RT in the stranger condition and mother condition. **(B)** DR in the stranger condition and mother condition. Error bars represent the SD.

fMRI Results

ROI Analysis

We performed an ROI analysis of the clusters of the AI and PI based on the activation of these two regions in the stranger condition and the mother condition of moral disgust (**Figure 3** and **Table 2**). The coordinates of the ROIs were as follows: AI, the center of the cluster at [−34 18 −16], and PI, the center of the cluster at [−38 −10 22].

The β value of the mother and stranger conditions were extracted from the AI and PI ROIs. For the AI, the β value in the stranger condition (Mean = 0.16, *SD* = 0.292) was lower than that in the mother condition (Mean = 0.29, *SD* = 0.298). For the PI, the β value in the stranger condition (Mean = 0.20, *SD* = 0.164) was greater than that in the mother condition (Mean = 0.13, *SD* = 0.155) (**Figure 3**).

PPI Analysis

The PPI of the AI and PI seeds represents how the agent (mother/stranger) modulates the change of the connectivity between the seeds regions and other brain regions. In the mother condition, the AI was more functionally connected with bilateral prefrontal cortex (PFC) relative to the stranger condition (**Figure 4**), whereas in the stranger condition, the PI was more functionally connected with thalamus and amygdala, the AI was more functionally connected with anterior cingulate cortex (ACC), and both PI and AI were more functionally connected with temporo-parietal junction (TPJ), relative to mother condition (**Table 3**).

DISCUSSION

The behavioral results indicated that the RTs in the mother condition were quicker than those in the stranger condition, which could be due to people devoting less time to thinking negative thoughts about their mother because these thoughts may evoke strong unpleasant feelings and the desire for avoidance (Li et al., 2011). Unsurprisingly, the moral transgressions performed by the mother were rated as less disgusting and less severe than those performed by the stranger. This bias could be related to the participants' internal tendency to favor their mothers in moral judgments. In particular, in our Chinese participants who may mentally represent themselves and their mothers by the same cognitive-brain mechanism (Zhu et al., 2007), this bias could be more obvious (Hwang, 2006).

Critically, the brain imaging results exhibited a double dissociation between the AI and PI in which an AI activation was found in the mother condition minus stranger condition contrast, while a PI activation was found in the stranger condition minus mother condition contrast. Furthermore, PPI analysis indicated that these PI and AI areas were functionally connected to widely distributed areas, including areas that are necessary for the representation of sensations and feelings, emotion regulation, and theory of mind (ToM).

The Role of the AI in Moral Disgust

Relative to the stranger condition, the mother condition was associated with AI activation. Similar activation was reported among people who were required to recall personal guilt experiences (Shin et al., 2000) or internally generate deontological guilt (Basile et al., 2011). Generally, the role of the ventral AI observed in this study has been proposed to mediate the core affect representing broadly tuned motivational states (e.g., excitement) with associated subjective feelings (Wager and Barrett, 2004).

Several hypotheses could be applied to explain why the mother condition was associated with more AI activation than was the stranger condition. For instance, making a moral judgment in

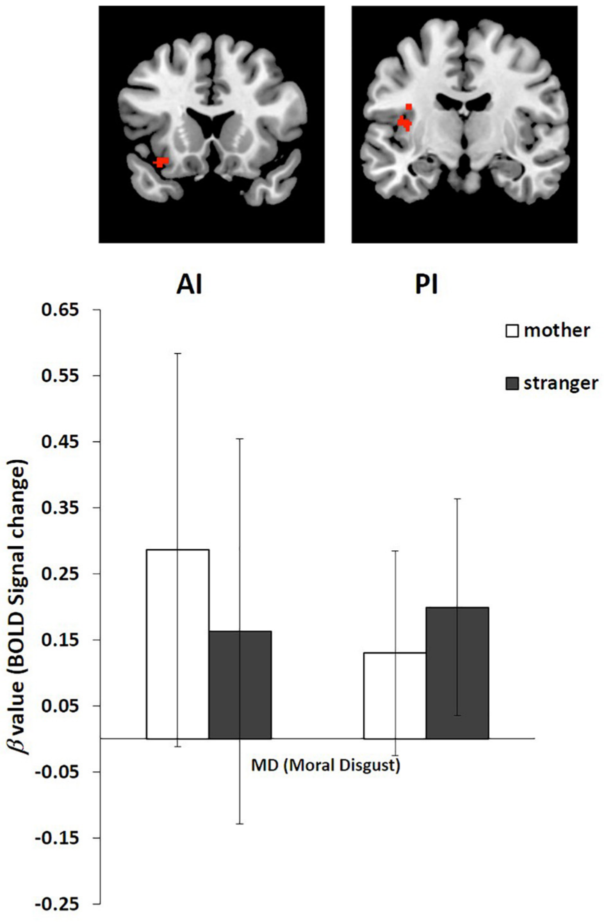

FIGURE 3 | Regions of interest (ROI) analysis of BOLD signal change in the AI and PI in the stranger condition and mother condition. Error bars represent the SD.

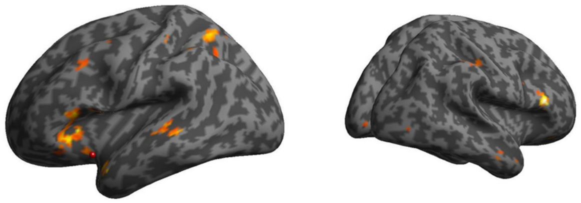

FIGURE 4 | Psychophysiological interactions (PPIs) of the AI in the mother condition minus stranger condition. The bilateral prefrontal cortex was functionally connected with the AI in the mother condition minus stranger condition.

TABLE 3 | Significant PPIs of the AI and PI seeds.

Region	x	y	z	T	Z	K
AI: Positive PPI						
R insula	36	−16	22	2.93	2.71	34
	46	4	−10	2.36	2.24	56
L TPJ	−34	−38	30	3.61	3.25	109
R TPJ	64	−30	32	2.99	2.76	600
R ACC	18	32	20	3.81	3.39	68
	2	28	−2	2.31	2.19	111
L ACC	−2	32	20	2.17	2.07	71
AI: Negative PPI						
L insula	−26	24	4	2.8	2.61	104
	−34	24	−4	2.48	2.34	
PI: Positive PPIs						
R thalamus	16	−16	18	3.14	2.88	99
R TPJ	40	−32	40	2.66	2.49	191
R amygdala	22	−2	−14	22	−2	69
L caudate	−8	20	4	2.89	2.68	160
R caudate	10	16	12	2.42	2.29	125
L ACC	−20	32	24	2.57	2.42	48

$P < 0.05$, uncorrected, 2 mm × 2 mm × 2 mm voxels. TPJ, temporo-parietal junction. ACC, anterior cingulate cortex.

the mother condition could require more integrative processing. AI activation could be related to the feelings that are represented on a more integrative level relative to the less integrative level, such as the feeling of love relative to sexual desire (Cacioppo et al., 2012) or the processing of the contextual aspects of fairness relative to the processing of objective aspects (Wright et al., 2011). Compared with making a bad moral judgment against a stranger, making a judgment against one's mother could involve more conflict of self-interest. Learning to make moral judgments based on considerations beyond self-interest is a fundamental aspect of moral development that can be achieved by communicating and interacting with many more people than one's parents or the process of deliberate (moral) persuasion intentionally generated by certain people to vividly demonstrate their value judgment (Bloom, 2010). Therefore, more integrative processing could be required by this type of moral judgment. A second possibility regarding the involvement of the AI in the mother condition could be the requirement for more processes of emotion regulation. A previous study found that the development of emotion regulation capacity with age could be accompanied by a posterior-to-anterior progression in insula activity in response to empathy- or sympathy-eliciting stimuli (Decety and Michalska, 2010). In this study, more emotion regulatory processes could be required in people evaluating their mothers' immoral behaviors than thinking of a stranger performing the same behaviors. Third, according to a previous study that found that the AI could function together with other prefrontal and temporal-parietal areas as a mechanism of "guilt aversion" that motivates people to choose to cooperate if they can better serve their interests by acting selfishly (Chang et al., 2011), the mother condition could contain more components of "guilt aversion" because of the close relationship with the mother. This closeness might result in stronger AI activation. Finally, in the moral context, the AI was found to be selectively activated in negative moral verdicts that identified an act as morally wrong regardless of whether the acts transgressed against moral principles more or less or required more or less moral deliberation (Schaich Borg et al., 2011). This finding is generally consistent with previous studies indicating that the activity in the AI correlated with the rejection of unfair offers (Sanfey et al., 2003), rejection of inequitable allocations (Hsu et al., 2008), decisions not to donate to charity (Moll et al., 2006), decisions not to purchase in a shopping task (Knutson et al., 2007), and verdicts of disbelief (Harris et al., 2008). Although the participants in our study made comparable negative moral verdicts in both conditions, making a fair judgment in the mother condition might require more resolution and extensive processing of negative moral verdicts and, thus, evoke more AI activation.

The Role of PI in Moral Disgust

Relative to the mother condition, the stranger condition was associated with PI activation. The involvement of the PI in moral-related tasks has been much more rarely reported than that of the AI. The PI is known to function in primary representations of emotionally relevant somato-sensory signals (Craig, 2002), such as primary pain, temperature, and touch perception, including facilitative touch (Björnsdotter et al., 2009; Löken et al., 2009; Lamm et al., 2011). Studies have reported PI activation in participants reading phrases that elicited moral indignation compared to that in participants reading neutral phrases (Moll et al., 2005) and in participants processing sociomoral acts (the immoral ones) compared to participants processing pathogenic acts (physically repulsive ones) (Schaich Borg et al., 2008). Notably, the examples of moral violations used in the Borg and colleagues' study were more serious than the mild violations used in the present study. In the previous study, the materials that the participants read included statements, such as "You watching your sister masturbate" or "You killing your sister's child." A study investigating major depressive disorder (MDD), including excessive proneness to self-blaming emotions, such as guilt and shame, exhibited an increasing PI activation in response to shame relative to that in response to guilt, implying that the specific function of PI in generating moral disgust related feeling (Pulcu et al., 2014). Additionally, PI activation was observed in social rejection, particularly when the rejection is powerfully elicited (Kross et al., 2011). An intracranial electroencephalography study found that, in contrast to the AI, which showed an initial fast response to social exclusion with a rapidly fading signal, the PI showed a more persistent activation pattern, implying that the PI represents a more primary aspect of disgust that does not decay over time (Cristofori et al., 2013). In addition to social rejection, PI activation was also observed in the processing of unfair offers in the ultimatum game, and its activation level could be modulated by emotion regulation strategies (Kirk et al., 2011), although the AI is much more frequently reported to be involved in the ultimatum game (Rilling et al., 2002; Sanfey et al., 2003; King-Casas et al., 2008). Both the social rejection in the Cyberball task and the unfair offers in the ultimatum game could be types of moral violations due to their nature, and the participants are the victims of these immoralities. This type of deep and painful feeling could eventually evokes a body sensation-like PI activation. In the present study, the participants likely perceived themselves as the victims of moral violations more in the stranger condition than in the mother condition, and this type of sympathy and empathy with the victims could contribute to significantly challenging the PI.

Functional Connectivity

In the stranger condition, several areas exhibit activation with stronger functional connectivity with the seed regions of the AI or PI. For example, the connectivity between the bilateral PFC and AI was stronger in the mother condition, which was consistent with our speculation that the mother condition required higher levels of integration (Wright et al., 2011) and modulation (Decety and Michalska, 2010), including the ones for disgust aversion (Chang et al., 2011) and related negative moral verdicts (Schaich Borg et al., 2011).

However, in the stranger condition, more areas showed stronger functional connections with the PI or AI seed regions. First, stronger connectivity was observed between the thalamus and the PI, which was consistent with the observation that the PI receives input from the thalamus and implies that the moral violation of the stranger evoked a more basic form of disgust. Second, the stronger connectivity between the amygdala and the PI seed region in the stranger condition was also consistent with the higher level of disgust reported by the participants in the stranger condition. In contrast to the connectivity with the thalamus and amygdala, the seed region showing stronger connectivity with the ACC was located in the AI rather than in the PI. This finding was consistent with the hypothesis that the co-activation of the ACC, amygdala, caudate, ventral striatum and the AI represented emotion and motivation values (Craig, 2002). However, the present study only found enhanced connectivity between the amygdala and PI but not with the AI, and both the PI and AI seed regions were increasingly connected to the caudate in the stranger condition. Therefore, our results were only partially consistent with the abovementioned hypothesis (Craig, 2002). Finally, the TPJ, which is among the most important areas for ToM and moral judgment, was more functionally connected with both the PI and AI seed regions in the stranger condition. As previously mentioned, in the stranger condition, the participants could be more likely to perceive themselves as the victims of the moral violations, which may lead to more sympathy and empathy processes that not only evoke activation in the PI but also result in enhanced connectivity between the TPJ and posterior and anterior portions of the insula.

The results of the insula's functional connectivity with other brain regions were also consistent with the behavioral results in which the RTs in the mother condition were shorter than those in the stranger condition. First, the insular seed region had more functional connectivity with other brain areas in the stranger condition than in the mother condition, suggesting that wider and more extensive information processing (and maybe longer RTs) occurred in the former condition. Second, the stranger condition exhibited more functional connectivity between the insula (including both the AI and PI) and TPJ than did the mother condition, implying that the stranger condition relied more heavily on reasoning processes based on ToM that might have taken longer time. Third, in the mother condition, the AI had stronger connectivity with the PFC, whereas in the stranger condition, the PI had stronger connectivity with the amygdala and thalamus. One possible explanation for this difference is that the enhanced connectivity between the insula and PFC in the mother condition could be related to inhibitory processes caused by the individuals' reluctance to think negative thoughts about their mother, and this inhibitory process might prevent individuals from further processing the sentences about their mothers.

In this study, although the processing of the materials evoked complicated feelings, emotions and cognitive processes, the feeling of moral disgust could be essentially involved in this complicated processing. Due to its well-established role in disgust, the insula could play a key role in representing moral disgust. However, in the present study, we could not completely justify that the observed insular activation did represent moral disgust rather than other feelings or thoughts. We did not find a significant correlation between the insular activation and individual subjective evaluations of disgust toward the immoral events. A possible interpretation is that the subjective evaluation of moral disgust is a holistic impression consisting of complicated cognitions, emotions, experiences, and social attitudes toward the transgression event. The element of disgust represented by the insula was not sufficiently strong to be reflected by this subjective evaluation. Further studies should adopt specific judgments that are more sensitive to detect the disgust element in moral judgment and verify the role of the insula in disgust representation.

In summary, in this study, we doubly dissociated two insular components in the processing of moral transgression events, and the component located in the posterior region was more activated in the stranger condition, while the other component located in the anterior region was more activated in the mother condition. Given that both the PI and AI were positively activated in the mother and stranger conditions (the signal change in the AI and PI regions was positive in both conditions), we propose that these two components may have been generally involved in both conditions regardless of the behavioral agent of the moral transgression (mother or stranger), and the double dissociation between the AI and PI implies that the stranger and mother conditions could rely on one of the two components more than the other. Based on the already known function of the PI and AI in emotion representations and re-representation and the consideration of the distinctive moral emotions involved in stranger and mother conditions, we hypothesize that the PI and AI might represent primary and secondary levels of moral disgust, respectively. Specifically, the PI component represents people's basic moral disgust that is directly embodied by the sensory components of physical disgust, whereas the AI components represent a secondary level of moral disgust that is related to the affective components of physical disgust. This result demonstrated the mechanism of embodied schemata from a cognitive neuroscience perspective and showed how disgust in response to moral violations is built on more basic types of disgust (such as the disgust associated with distaste for food and body waste products) and how it develops a more integrative and abstract form of mental representation.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Institutional Review Board of the State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University with written informed consent from all subjects. All subjects gave written informed consent. The protocol was approved by the Institutional Review Board of the State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was supported by funding programs from the National Social Science Foundation of China (12AZD116) and the National Natural Science Foundation of China (31671124, 31671160).

ACKNOWLEDGMENTS

We thank Alex Dufford for the valuable comments regarding manuscript. We thank Dr. Qiong Wu, Dr. Tingting Wu, Tingting Guo, Dr. Thomas P. Naidich, Dr. Shaozheng Qin, and Dr. Pengfei Xu for the comments regarding image data analysis, Dr. Furong Huang for assistance with the figures, Yan Zhang and Teng Rong for assistance with the data collection, and Yufan Zhang and Wenjin Hong for assistance with preparing the experimental materials.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00860/full#supplementary-material

FIGURE S1 | Reaction time, disgust rating, and severity rating in all three conditions: Stranger, Best Friend, and Mother.

TABLE S1 | Statistical analysis of moral disgust materials in the three groups.

REFERENCES

Basile, B., Mancini, F., Macaluso, E., Caltagirone, C., Frackowiak, R. S., and Bozzali, M. (2011). Deontological and altruistic guilt: evidence for distinct neurobiological substrates. *Hum. Brain Mapp.* 32, 229–239. doi: 10.1002/hbm.21009

Björnsdotter, M., Löken, L., Olausson, H., Vallbo, Å., and Wessberg, J. (2009). Somatotopic organization of gentle touch processing in the posterior insular cortex. *J. Neurosci.* 29, 9314–9320. doi: 10.1523/JNEUROSCI.0400-09.2009

Bloom, P. (2010). How do morals change? *Nature* 464, 490–490. doi: 10.1038/464490a

Cacioppo, S., Bianchi-Demicheli, F., Frum, C., Pfaus, J. G., and Lewis, J. W. (2012). The common neural bases between sexual desire and love: a multilevel kernel density fMRI analysis. *J. Sex. Med.* 9, 1048–1054. doi: 10.1111/j.1743-6109.2012.02651.x

Chang, L. J., Smith, A., Dufwenberg, M., and Sanfey, A. G. (2011). Triangulating the neural, psychological, and economic bases of guilt aversion. *Neuron* 70, 560–572. doi: 10.1016/j.neuron.2011.02.056

Cheng, X., Zheng, L., Li, L., Zheng, Y., Guo, X., and Yang, G. (2017). Anterior insula signals inequalities in a modified ultimatum game. *Neuroscience* 348, 126–134. doi: 10.1016/j.neuroscience.2017.02.023

Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nat. Rev. Neurosci.* 3, 655–666. doi: 10.1038/nrn894

Craig, A. D. (2009). How do you feel — now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70. doi: 10.1038/nrn2555

Cristofori, I., Moretti, L., Harquel, S., Posada, A., Deiana, G., Isnard, J., et al. (2013). Theta signal as the neural signature of social exclusion. *Cereb. Cortex* 23, 2437–2447. doi: 10.1093/cercor/bhs236

Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A., and Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nat. Neurosci.* 7, 189–195. doi: 10.1038/nn1176

Decety, J., and Michalska, K. J. (2010). Neurodevelopmental changes in the circuits underlying empathy and sympathy from childhood to adulthood. *Dev. Sci.* 13, 886–899. doi: 10.1111/j.1467-7687.2009.00940.x

Flynn, F. G. (1999). Anatomy of the insula functional and clinical correlates. *Aphasiology* 13, 55–78. doi: 10.1080/026870399402325

Fourie, M. M., Thomas, K. G., Amodio, D. M., Warton, C. M., and Meintjes, E. M. (2014). Neural correlates of experienced moral emotion: an fMRI investigation of emotion in response to prejudice feedback. *Soc. Neurosci.* 9, 203–218. doi: 10.1080/17470919.2013.878750

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., and Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* 6, 218–229. doi: 10.1006/nimg.1997.0291

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., and Turner, R. (1998). Event-related fMRI: characterizing differential responses. *Neuroimage* 7, 30–40. doi: 10.1006/nimg.1997.0306

Giubilini, A. (2016). What in the world is moral disgust? *Aust. J. Philos.* 94, 227–242. doi: 10.1080/00048402.2015.1070887

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., et al. (1989). The yale-brown obsessive compulsive scale: I. development, use, and reliability. *Arch. Gen. Psychiatry* 46, 1006–1011. doi: 10.1001/archpsyc.1989.01810110048007

Greene, J., and Haidt, J. (2002). How (and where) does moral judgment work? *Trends Cogn. Sci.* 6, 517–523. doi: 10.1016/S1364-6613(02)02011-9

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., and Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44, 389–400. doi: 10.1016/j.neuron.2004.09.027

Haidt, J., and Kesebir, S. (2010). "Morality," in *Handbook of Social Psychology*, eds S. Fiske, D. Gilbert, and G. Lindzey (Hobeken, NJ: John Wiley & Sons).

Haidt, J., Rozin, P., Mccauley, C., and Imada, S. (1997). Body, psyche, and culture: the relationship between disgust and morality. *Psychol. Dev. Soc.* 9, 107–131. doi: 10.1177/097133369700900105

Harenski, C. L., and Hamann, S. (2006). Neural correlates of regulating negative emotions related to moral violations. *Neuroimage* 30, 313–324. doi: 10.1016/j.neuroimage.2005.09.034

Harris, S., Sheth, S. A., and Cohen, M. S. (2008). Functional neuroimaging of belief, disbelief, and uncertainty. *Ann. Neurol.* 63, 141–147. doi: 10.1002/ana.21301

Hsu, M., Anen, C., and Quartz, S. R. (2008). The right and the good: distributive justice and neural encoding of equity and efficiency. *Science* 320, 1092–1095. doi: 10.1126/science.1153651

Hwang, K.-K. (2006). Moral face and social face: contingent self-esteem in Confucian society. *Int. J. Psychol.* 41, 276–281. doi: 10.1080/00207590544000040

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., and Montague, P. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science* 321, 806–810. doi: 10.1126/science.1155236902

Kirk, U., Harvey, A., and Montague, P. R. (2011). Domain expertise insulates against judgment bias by monetary favors through a modulation of ventromedial prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10332–10336. doi: 10.1073/pnas.1019332108

Knutson, B., Rick, S., Wimmer, G. E., Prelec, D., and Loewenstein, G. (2007). Neural predictors of purchases. *Neuron* 53, 147–156. doi: 10.1016/j.neuron.2006.11.010

Kross, E., Berman, M. G., Mischel, W., Smith, E. E., and Wager, T. D. (2011). Social rejection shares somatosensory representations with physical pain. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6270–6275. doi: 10.1073/pnas.1102693108

Lamm, C., Decety, J., and Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* 54, 2492–2502. doi: 10.1016/j.neuroimage.2010.10.014

Li, Q., Qin, S., Rao, L.-L., Zhang, W., Ying, X., Guo, X., et al. (2011). Can Sophie's choice be adequately captured by cold computation of minimizing losses? An fMRI study of vital loss decisions. *PLoS One* 6:e17544. doi: 10.1371/journal.pone.0017544

Lin, C.-S., Hsieh, J.-C., Yeh, T.-C., Lee, S.-Y., and Niddam, D. M. (2013). Functional dissociation within insular cortex: the effect of pre-stimulus anxiety on pain. *Brain Res.* 1493, 40–47. doi: 10.1016/j.brainres.2012.11.035

Löken, L. S., Wessberg, J., Morrison, I., McGlone, F., and Olausson, H. (2009). Coding of pleasant touch by unmyelinated afferents in humans. *Nat. Neurosci.* 12, 547–548. doi: 10.1038/nn.2312

Majdandžić, J., Bauer, H., Windischberger, C., Moser, E., Engl, E., and Lamm, C. (2012). The human factor: behavioral and neural correlates of humanized perception in moral decision making. *PLoS One* 7:e47698. doi: 10.1371/journal.pone.0047698

Miller, G. (2008). The roots of morality. *Science* 320, 734–737. doi: 10.1126/science.320.5877.734

Moll, J., de Oliveira-Souza, R., Moll, F. T., Ignácio, F. A., Bramati, I. E., Caparelli-Dáquer, E. M., et al. (2005). The moral affiliations of disgust: a functional MRI study. *Cogn. Behav. Neurol.* 18, 68–78. doi: 10.1097/01.wnn.0000152236.46475.a7

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., and Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15623–15628. doi: 10.1073/pnas.0604475103

Parker, J. D., Taylor, G. J., and Bagby, R. M. (2003). The 20-Item Toronto Alexithymia scale: III. Reliability and factorial validity in a community population. *J. Psychosom. Res.* 55, 269–275. doi: 10.1016/S0022-3999(02)00578-0

Paulus, M. P., and Stein, M. B. (2006). An insular view of anxiety. *Biol. Psychiatry* 60, 383–387. doi: 10.1016/j.biopsych.2006.03.042

Pulcu, E., Lythe, K., Elliott, R., Green, S., Moll, J., Deakin, J. F., et al. (2014). Increased amygdala response to shame in remitted major depressive disorder. *PLoS One* 9:e86900. doi: 10.1371/journal.pone.0086900

Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., and Kilts, C. (2002). A neural basis for social cooperation. *Neuron* 35, 395–405. doi: 10.1016/S0896-6273(02)00755-9

Rozin, P., Haidt, J., and McCauley, C. R. (2008). "Disgust," in *Handbook of Emotions*, 3rd Edn, eds M. Lewis, J. M. Haviland-Jones, and L. F. Barrett (New York, NY: Guilford Press), 757–776.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755–1758. doi: 10.1126/science.1082976

Schaich Borg, J., Lieberman, D., and Kiehl, K. A. (2008). Infection, incest, and iniquity: investigating the neural correlates of disgust and morality. *J. Cogn. Neurosci.* 20, 1529–1546. doi: 10.1162/jocn.2008.20109

Schaich Borg, J., Sinnott-Armstrong, W., Calhoun, V. D., and Kiehl, K. A. (2011). Neural basis of moral verdict and moral deliberation. *Soc. Neurosci.* 6, 398–413. doi: 10.1080/17470919.2011.559363

Shin, L. M., Dougherty, D. D., Orr, S. P., Pitman, R. K., Lasko, M., Macklin, M. L., et al. (2000). Activation of anterior paralimbic structures during guilt-related script-driven imagery. *Biol. Psychiatry* 48, 43–50. doi: 10.1016/S0006-3223(00)00251-1

Varnavas, G. G., and Grand, W. (1999). The insular cortex: morphological and vascular anatomic characteristics. *Neurosurgery* 44, 127–136. doi: 10.1097/00006123-199901000-00079

Wager, T. D., and Barrett, L. F. (2004). *From Affect to Control: Functional Specialization of the Insula in Motivation and Regulation*. Available at: www.columbia.edu/cu/psychology.tor/

Wagner, U., N'Diaye, K., Ethofer, T., and Vuilleumier, P. (2011). Guilt-specific processing in the prefrontal cortex. *Cereb. Cortex* 21, 2461–2470. doi: 10.1093/cercor/bhr016

Wright, N. D., Symmonds, M., Fleming, S. M., and Dolan, R. J. (2011). Neural segregation of objective and contextual aspects of fairness. *J. Neurosci.* 31, 5244–5252. doi: 10.1523/JNEUROSCI.3138-10.2011

Xiang, T., Lohrenz, T., and Montague, P. R. (2013). Computational substrates of norms and their violations during social exchange. *J. Neurosci.* 33, 1099–1108. doi: 10.1523/JNEUROSCI.1642-12.2013

Yurgelun-Todd, D. (2007). Emotional and cognitive changes during adolescence. *Curr. Opin. Neurobiol.* 17, 251–257. doi: 10.1016/j.conb.2007.03.009

Zhu, Y., Zhang, L., Fan, J., and Han, S. (2007). Neural basis of cultural influence on self-representation. *Neuroimage* 34, 1310–1316. doi: 10.1016/j.neuroimage.2006.08.047

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 26 March 2018
doi: 10.3389/fnins.2018.00193

tDCS Over DLPFC Leads to Less Utilitarian Response in Moral-Personal Judgment

Haoli Zheng[1,2,3], *Xinbo Lu*[1,3] *and Daqiang Huang*[2,3]*

[1] Center for Economic Behavior and Decision-Making, Neuro & Behavior EconLab (NBEL), Zhejiang University of Finance and Economics, Hangzhou, China, [2] Interdisciplinary Center for Social Sciences, Zhejiang University, Hangzhou, China, [3] School of Economics, Zhejiang University of Finance and Economics, Hangzhou, China

The profound nature of moral judgment has been discussed and debated for centuries. When facing the trade-off between pursuing moral rights and seeking better consequences, most people make different moral choices between two kinds of dilemmas. Such differences were explained by the dual-process theory involving an automatic emotional response and a controlled application of utilitarian decision-rules. In neurocognitive studies, the bilateral dorsolateral prefrontal cortex (DLPFC) has been demonstrated to play an important role in cognitive "rational" control processes in moral dilemmas. However, the profile of results across studies is not entirely consistent. Although one transcranial magnetic stimulation (TMS) study revealed that disrupting the right DLPFC led to less utilitarian responses, other TMS studies indicated that inhibition of the right DLPFC led to more utilitarian choices. Moreover, the right temporoparietal junction (TPJ) is essential for its function of integrating belief and intention in moral judgment, which is related to the emotional process according to the dual-process theory. Relatively few studies have reported the causal relationship between TPJ and participants' moral responses, especially in moral dilemmas. In the present study, we aimed to demonstrate a direct link between the neural and behavioral results by application of transcranial direct current stimulation (tDCS) in the bilateral DLPFC or TPJ of our participants. We observed that activating the right DLPFC as well as inhibiting the left DLPFC led to less utilitarian judgments, especially in moral-personal conditions, indicating that the right DLPFC plays an essential role, not only through its function of moral reasoning but also through its information integrating process in moral judgments. It was also revealed that altering the excitability of the bilateral TPJ using tDCS negligibly altered the moral response in non-moral, moral-impersonal and moral-personal dilemmas, indicating that bilateral TPJ may have little influence over moral judgments in moral dilemmas.

Keywords: moral dilemma, dorsolateral prefrontal cortex, temporoparietal junction, transcranial direct current stimulation, dual-process theory, theory of mind

OPEN ACCESS

Edited by:
Chao Liu,
Beijing Normal University, China

Reviewed by:
Sara Palumbo,
Università degli Studi di Pisa, Italy
Yasuki Noguchi,
Kobe University, Japan

****Correspondence:***
Daqiang Huang
hdqjerry@zju.edu.cn

Specialty section:
This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *30 December 2017*
Accepted: *09 March 2018*
Published: *26 March 2018*

Citation:
Zheng H, Lu X and Huang D (2018) tDCS Over DLPFC Leads to Less Utilitarian Response in Moral-Personal Judgment. Front. Neurosci. 12:193. doi: 10.3389/fnins.2018.00193

INTRODUCTION

The nature of moral judgment has been debated for centuries. To analyze the moral brain of humans, a valid measurement is by observing participants' responses to moral dilemmas, which present a story involving a trade-off between pursuing moral rights and seeking better consequences (Borg et al., 2006). When people make moral judgments in conflicts between harm and moral rights, both reason and emotion are considered important forces driving moral judgments. Greene et al. (2001) classified moral dilemmas into two categories: moral-impersonal dilemmas (e.g., switch dilemma) and moral-personal dilemmas (e.g., footbridge dilemma). Most people may find it appropriate to save five lives at the expense of one by turning a switch in a classic switch dilemma (Thomson, 1986), whereas in a footbridge dilemma, they may consider it inappropriate to push a stranger off the footbridge in order to stop the train, which may also save the lives of five people (Greene et al., 2001). By considering both reason and emotion as essential forces in moral decisions, such differences in moral responses are explained by the dual-process theory (Greene et al., 2001, 2004). According to the dual-process theory, moral decisions are made involving an automatic emotional response and a controlled application of rational utilitarian decision-rules. The moral emotional response is considered too strong to be overwhelmed by the cognitive reasoning process in moral-personal dilemmas while in contrast, participants may favor the utilitarian choice in moral-impersonal dilemmas because the weaker emotional response is manipulated by rational cognitive control (Greene, 2007).

The cognitive reasoning process has been directly related to the involvement of the dorsolateral prefrontal cortex (DLPFC) in moral decisions (Greene et al., 2001, 2004). According to the dual-process theory, the right DLPFC may lead to utilitarian choices through its influence over the cognitive rational control process. However, the profile of results across studies is not entirely consistent. It has been revealed that damage to the frontal cortex leads to utilitarian moral judgments that rely solely on best results. Recent transcranial magnetic stimulation (TMS) studies have also raised questions regarding the role of the right DLPFC restricted to rational cognitive control. Using low-frequency repetitive transcranial magnetic stimulation (rTMS), Knoch et al. (2006) revealed that disrupting the function of participants' right DLPFC reduced the rejection rates of their partners' intentionally unfair offers, leading to a more utilitarian judgment in an economic interaction. Moreover, Tassy et al. (2011) applied rTMS over participants' right DLPFC while subjecting them to moral tasks and demonstrated that disrupting the right DLPFC alters moral judgment, increasing the probability of utilitarian responses. These TMS studies indicated that suppressing the right DLPFC may result in more utilitarian judgments, suggesting that the right DLPFC function not only participates in a rational cognitive control process but also integrates emotions in moral judgments, especially in high-conflict moral dilemmas (Tassy et al., 2011). In contrast, Jeurissen et al. (2014) revealed that TMS-induced disruption of the DLPFC in moral-personal decisions leads to less utilitarian decisions, which supported the dual-process theory. The contradiction of the observations in these two studies may due to their relatively small sample sizes which may affect the robustness of the results.

Emotional response is associated with the bilateral temporoparietal junction (TPJ), which plays a significant role in the process of belief attribution in moral judgments (Ruby and Decety, 2001; Vogeley et al., 2001; Gallagher and Frith, 2003; Schleim et al., 2010; Mai et al., 2016). When individuals make moral decisions, the bilateral TPJ is centrally involved in understanding others by reasoning about the content of mental states (Saxe and Kanwisher, 2003; Jeurissen et al., 2014). Previous studies have demonstrated people making moral judgments depending more substantially on beliefs and intentions rather than on results and consequences (Surber, 1977; Shultz et al., 1986; Baird and Moses, 2001; Baird and Astington, 2004). This type of behavior may be interpreted by the theory of mind: the ability to attribute mental states, such as beliefs and intentions, to moral agents, which also play a crucial role in the process of moral judgment (Borg et al., 2006; Cushman et al., 2006; Young et al., 2007). The right TPJ is associated with beliefs because its activity was observed to be significantly higher when participants read false belief stories (Sommer et al., 2007; Aichhorn et al., 2009; Young and Dodell, 2010). Using TMS, Young et al. (2010) demonstrated a direct causal link between the disruption of the right TPJ and the decreasing influence of beliefs in moral judgment. More recently, Sellaro et al. (2015) demonstrated that anodal stimulation of the right TPJ enhanced the role of belief in moral judgment, suggesting that the right TPJ integrates beliefs and intentions into participants' moral judgments. Using transcranial direct current stimulation (tDCS), Ye et al. (2015) revealed that the bilateral TPJ is indispensable for integrating intentions in moral judgment. Leloup et al. (2016) also indicated that the right TPJ may play multiple roles in moral cognition, in relation to the methodological aspects of the use of tDCS. However, the moral tasks in these studies were moral judgments involving both intentions and consequences rather than moral dilemmas. Using moral dilemma tasks, Jeurissen et al. (2014) revealed that disrupting the function of TPJ affects only moral-impersonal conditions in moral dilemmas.

Although cognitive reasoning and emotional processes, as identified in the dual-process theory, have been associated with the activity of the DLPFC and TPJ (Greene et al., 2001, 2004), no conclusive results have been demonstrated in previous neural imaging and stimulation studies. In the current study, using tDCS which allows cortical excitability to be directly manipulated, we aimed to investigate whether modulating the excitability of the bilateral DLPFC (or TPJ) can directly influence our participants' moral judgments by affecting their cognitive reasoning or emotional processes. Furthermore, we enlarged the sample size to 20 participants in each group with total of 100 valid subjects to examined the robustness of the double-dissociation effect between DLPFC and TPJ on the outcome of a moral decision. The casual relationship between the activity of bilateral DLPFC (or TPJ) and individuals' moral judgments may be revealed by comparing their judgments among different types of stimulations of the bilateral DLPFC (or TPJ).

MATERIALS AND METHODS

Subjects

One hundred right-handed healthy subjects (mean age 21.4 years, ranging from 17 to 30 years; 52 females) with no history of neurological or psychiatric problems participated in the study for payment. All the participants were naïve to tDCS and moral judgment tasks, had normal or corrected-to-normal vision, and provided their written informed consent, which was approved by the Zhejiang University ethics committee. The entire experiment lasted approximately 30 min, and each participant received a payment of 50 RMB Yuan (approximately 7.576 US dollars) upon completion of their tasks. None of the participants reported any adverse side effects concerning pain on the scalp or headaches after the experiment.

tDCS

For tDCS, a weak direct current was applied to the scalp via two saline-soaked surface sponge electrodes (35 cm^2). The current was constant and was delivered by a battery-driven stimulator (Multichannel noninvasive wireless tDCS neurostimulator, Starlab, Barcelona, Spain). It was adjusted to induce cortical excitability of the target area without any physiological damage to the participants. Various configurations of the current had various effects on cortical excitability; anodal stimulation enhanced cortical excitability, whereas cathodal stimulation suppressed it (Nitsche and Paulus, 2000).

The participants were randomly assigned to receive right anodal/left cathodal tDCS over DLPFC (n = 20, 12 females), left anodal/right cathodal tDCS over DLPFC (n = 20, 10 females), right anodal/left cathodal tDCS over TPJ (n = 20, 11 females), left anodal/right cathodal tDCS over TPJ (n = 20, 9 females) or sham stimulation (n = 20, 10 females). For right anodal/left cathodal stimulation over DLPFC, the anodal electrode was placed over the right DLPFC at the F4 position according to the international EEG 10/20 system, whereas the cathodal electrode was placed over the left DLPFC at the F3 position. For left anodal/right cathodal stimulation, the placement was reversed. For right anodal/left cathodal and left anodal/right cathodal tDCS stimulations over TPJ, the placement of electrodes was identical to those over DLPFC (**Figures 1**, **2**). For sham stimulation, the procedures were the same (the placement of electrodes was either over the bilateral DLPFC or over the bilateral TPJ), but the current lasted for only the first 30 s. The participants may have felt the initial itching, but there was actually no current for the rest of the stimulation. This method of sham stimulation has been shown to be reliable (Gandiga et al., 2006). The current was constant and of 2 mA in intensity, with a 30 s ramp up and down; the safety and efficiency of this stimulation has been demonstrated in previous studies. Before the moral judgment task, the laboratory assistant put a tDCS device on the participant's head for stimulation. After 20 min of stimulation, the participant was then asked to complete a moral judgment task.

Task and Procedure

After the participants received tDCS stimulation for 20 min (single-blinded, sham-controlled), they completed a moral judgment task (the computer program for the task was written in visual C#), which was similar to Greene's design (Greene et al., 2001). The moral dilemma task involved 12 stories, including 4 non-moral dilemmas, 4 moral-impersonal dilemmas and 4 moral-personal dilemmas (Supplementary Material). Moral dilemmas were presented in a pseudorandom order, the order of stories counterbalanced across runs, ensuring that same type of moral dilemma was never immediately repeated. Each participant read the 12 stories as text, then rated the degree of appropriateness of the protagonists' actions on a 10-point scale (1 = completely appropriate; 10 = completely inappropriate). Upon completing the moral task, participants had to complete a questionnaire before receiving their payments.

RESULTS

The reaction time and the response rating data were statistically evaluated using the SPSS software (version 22, SPSS Inc., Chicago, IL, USA). The significance level was set at 0.05 for all analyses.

Response

Response ratings from the right anodal/left cathodal tDCS over TPJ, left anodal/right cathodal tDCS over TPJ, right anodal/left cathodal tDCS over DLPFC, left anodal/right cathodal tDCS over DLPFC and sham groups were analyzed by repeated measures analyses of variance (ANOVAs) with dilemma type as a within-subject factor and tDCS stimulation type as a between-subject factor. A significant influence of dilemma type was observed [$F_{(2,\ 790)} = 673.587, p < 0.001$, partial $\eta^2 = 0.630$]. The utilitarian responses of protagonists were considered more inappropriate in moral-personal dilemmas (average rating of 7.54) than those in moral-impersonal dilemmas (average rating of 4.08, $p < 0.001$) or in non-moral dilemmas (average rating of 1.81, $p < 0.001$).

Notably, there was a significant interaction effect involving the dilemma type and stimulation type [$F_{(8,\ 790)} = 2.633$, $p = 0.008$, partial $\eta^2 = 0.026$]. *Post hoc* analyses (Bonferroni) revealed that in the personal dilemma tasks, the response ratings obtained in the right anodal/left cathodal DLPFC group (average rating of 8.475) were significantly higher than those obtained in the sham group (average rating of 6.863, $p = 0.002$). No other significant effects were observed in the impersonal dilemma or non-moral dilemma tasks (**Figure 3**).

Reaction Times

All trials in which reaction times were too long (>30 s) were excluded from data analysis (Jeurissen et al., 2014). Reaction times obtained following right anodal/left cathodal tDCS over TPJ, left anodal/right cathodal tDCS over TPJ, right anodal/left cathodal tDCS over DLPFC, left anodal/right cathodal tDCS over DLPFC and from the sham groups were analyzed by repeated measures analyses of variance (ANOVAs) with the dilemma type as a within-subject factor and tDCS stimulation type as a between-subject factor. No significant influence of

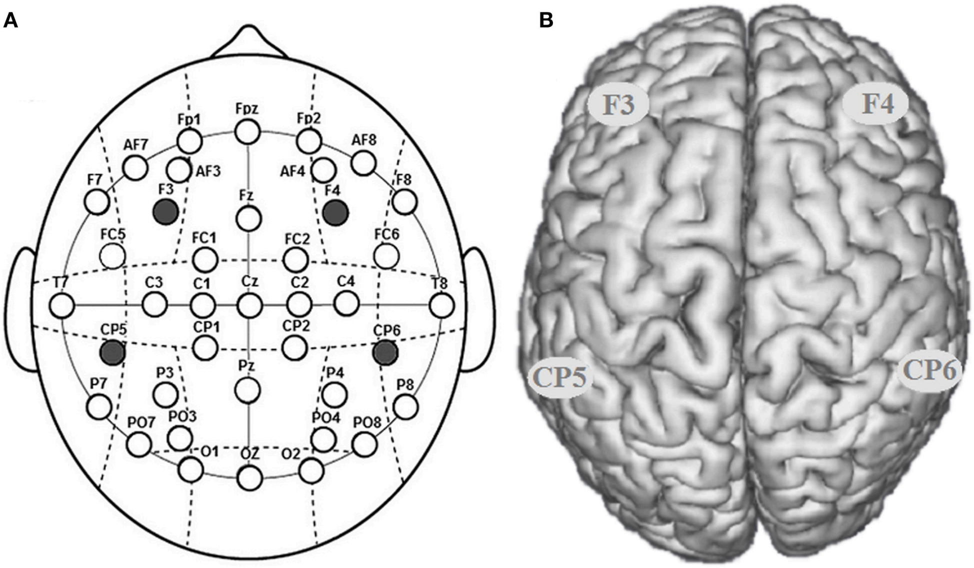

FIGURE 1 | Locations of the electrode positions. **(A)** Schematic of the electrode positions DLPFC (F3, F4) and TPJ (CP5, CP6) based on the international EEG 10-20 system. **(B)** Locations of the DLPFC (F3, F4) and the TPJ (CP5, CP6) of the human brain.

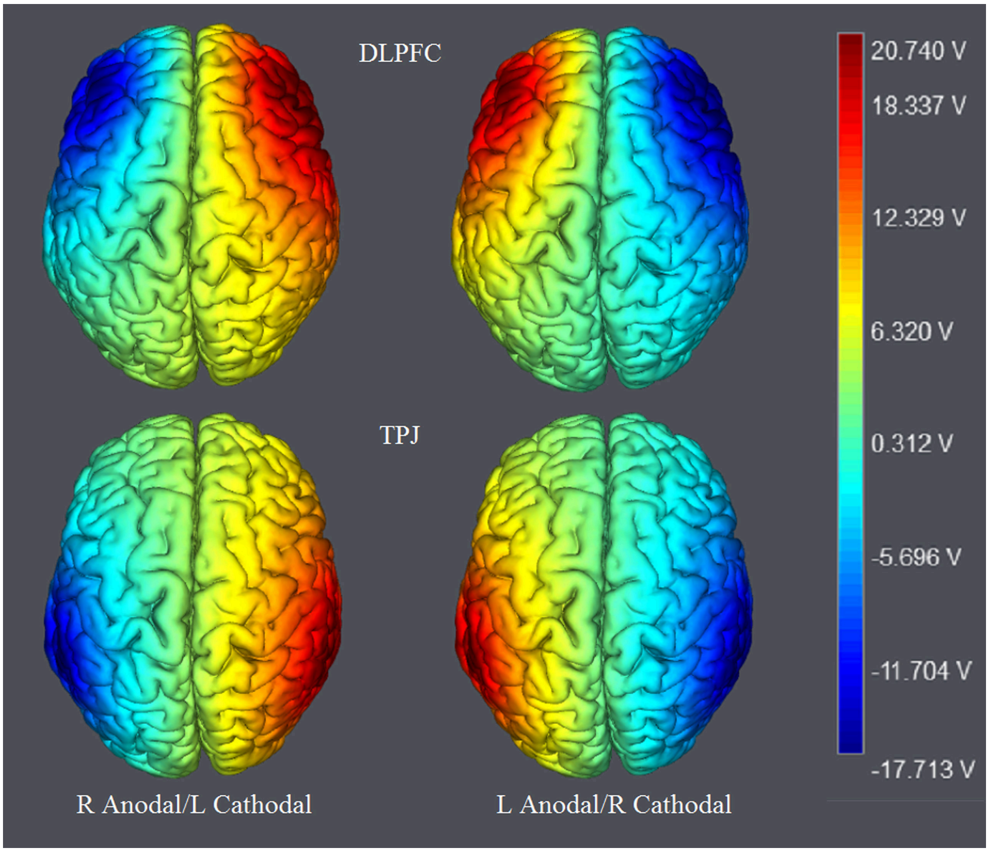

FIGURE 2 | The stimulation modes of tDCS treatments. The axis represents the range of input voltage from −17.713 to 20.740 V.

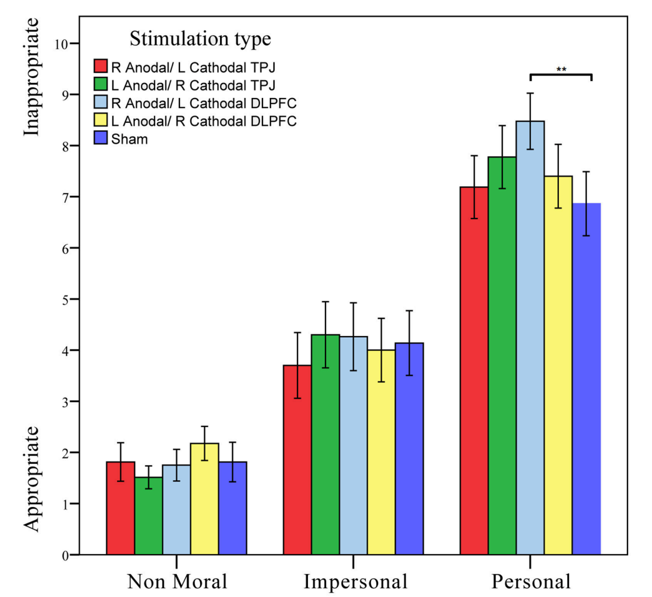

FIGURE 3 | Data of moral response ratings. The moral response ratings of participants with moral-personal, moral-impersonal and non-moral conditions across stimulations. Error bars indicate 95% confidence intervals. Asterisks indicate statistical significance of difference between treatments.

tDCS stimulation type was observed [$F_{(4,\,369)} = 0.710$, $p = 0.585$, partial $\eta^2 = 0.008$]. No significant interaction effect involving dilemma type and stimulation type was observed [$F_{(8,\,738)} = 0.585$, $p = 0.790$, partial $\eta^2 = 0.006$]. The reaction times in moral-impersonal dilemmas (mean = 8,357ms) were significantly higher than that in moral-personal dilemmas (mean = 7,098, $p < 0.007$) or in non-moral dilemmas (mean = 7,072, $p < 0.006$). Crucially, there was a significant negative correlation between reaction times and response ratings within the moral-personal condition (coefficient = −0.218, $p < 0.001$, Pearson correlation). There was also a significant positive correlation between reaction times and response ratings within the non-moral condition (coefficient = 0.189, $p < 0.001$, Pearson correlation). No significant correlation between reaction times and response ratings within the moral-impersonal condition was observed (coefficient = −0.218, $p = 0.424$, Pearson correlation) (**Table 1**).

DISCUSSION

DLPFC and Moral Dilemma

The dual-process theory hypothesizes that in high-conflict moral-personal dilemmas, stronger rational cognitive control is required to overrule the initial emotional impulse. Using TMS, Jeurissen et al. (2014) supported the dual-process theory by revealing that disruption of the right DLPFC leads to less utilitarian choices. It was explained by the dual-process theory that the DLPFC is majorly involved in "rational cognitive" control superseding emotional impulse, which is not strong enough. In contrast, Tassy et al. (2011) observed that disrupting the function of the right DLPFC leads to more utilitarian choices in moral dilemmas. Moreover, it has also been claimed that the dual-process theory may not be sufficient to explain various aspects of moral cognitions (Buckholtz and Marois, 2012; Van Bavel et al., 2015) suggested that the role of DLPFC in prosocial behaviors may be not solely restricted to its rational cognitive control process. In the current study, we observed that activation of the right DLPFC and inhibition of the left DLPFC by tDCS led to less utilitarian choices, especially in moral-personal conditions, supporting the claim that apart from its function in rational cognitive control process, the right DLPFC also plays an essential role in integrating emotional information in moral judgments. Such an emotional integration process was only observed in high-conflict dilemmas, such as moral-personal dilemmas. When confronting moral-personal dilemmas, the conflict of pursuing moral rights and seeking better consequences was stronger than moral-impersonal and non-moral dilemmas. In moral-personal dilemmas, the strengthened excitability of the right DLPFC weighed more on the initial emotional impulse through its emotion integrating process, resulting in less utilitarian moral response. In contrast, in moral-impersonal dilemmas, when the conflict of moral rights and better results was much weaker than in moral-personal condition, the enhancement of right DLPFC negligibly altered moral decisions.

Observations of the current study may be explained by the hypothesis provided by Buckholtz and Marois (2012). Buckholtz and Marois (2012) revealed that the dual-process theory could not completely explain the role of DLPFC in altruistic punishment games. According to the dual-process theory, if the role of the right DLPFC is solely in rational cognitive control process, its inhibition may result in less utilitarian choices in punishment games, which is not consistent with the finding that this brain region is activated to a greater extent when participants decide to punish protagonists in third-party interactions (Buckholtz et al., 2008). The role of the right DLPFC may be that it selects a specific response from among possible response options by integrating information about harm and blame with context-specific rules. In the current case of moral decisions, the more appropriate explanation regarding the role of the right DLPFC may be that it selects a specific moral response from these possible options by integrating information about moral rights and utilitarian consequences with dilemma-specific contents following moral rules.

fMRI studies revealed that the excitability of the frontal cortex was higher in moral-personal situations than that in non-moral and moral-impersonal situations (Greene et al., 2001, 2004). More recently, Jeurissen et al. (2014) revealed that TMS-induced disruption of the DLPFC only affects moral-personal decisions, leading to less utilitarian moral choices. However, several TMS studies, using moral decision tasks or prosocial economic games measuring fairness of participants (e.g., Ultimatum game), indicated that disruption of the DLPFC may result in more utilitarian choices (Knoch et al., 2006; Tassy et al., 2011). Moreover, whether enhancing the activity of bilateral DLPFC alters participants' non-moral, moral-impersonal and moral-personal decisions remains unknown. In the current study, we observed that modulating the excitability of the bilateral

TABLE 1 | The mean and SD of reaction time across moral contents and stimulations.

Moral Content	Stimulation type	R Anodal/L Cathodal over DLPFC	L Anodal/R Cathodal over DLPFC	R Anodal/L Cathodal over TPJ	L Anodal/R Cathodal over TPJ	Sham
Non-moral	Mean (ms)	6994.535	7092.519	6309.587	7433.849	7528.311
	SD	577.772	589.065	609.454	589.065	605.206
Moral impersonal	Mean (ms)	7619.179	8485.320	8869.920	8966.988	7843.793
	SD	699.502	713.174	737.859	713.174	732.717
Moral personal	Mean (ms)	6700.983	7595.328	6997.196	7532.256	6663.313
	SD	607.357	619.229	640.662	619.229	636.197

DLPFC altered the participants' moral judgments, especially in moral-personal situations. Our results confirmed the casual relationship between the activity of DLPFC and the moral decisions of participants in moral-personal conflicts. No such causal relationship between the activity of DLPFC and the moral choices in moral-impersonal or non-moral conflicts was observed.

Moreover, the left DLPFC has also been revealed that enhancing the activity of this brain region may induce a shift in moral judgment toward more non-utilitarian actions (Kuehne et al., 2015). No such effect was observed in our study. The difference between the findings of our study and the observations of the previous study may due to the variety in experimental designs and the stimulation locations. Kuehne et al. (2015) placed the active electrode over the left DLPFC with the reference electrode over the right parietal cortex, while in the current study the target and reference electrodes were placed over the bilateral DLPFC. Kuehne et al. (2015) performed a within-subject study and we performed a between-subject study which may also lead to inconsistent findings. In addition, our finding may also be the result of a combination stimulation effect over the bilateral DLPFC. The function of the left DLPFC may be justified through unilateral stimulation in further studies.

TPJ and Moral Dilemma

According to the dual-process theory, the emotional response may have been influenced by the TPJ through its function described in the theory of mind. TMS and tDCS studies have demonstrated that altering the activity of TPJ may change moral decisions, especially in conditions involving beliefs and intentions (Young et al., 2007; Young and Saxe, 2008; Sellaro et al., 2015; Ye et al., 2015). Jeurissen et al. (2014) revealed that disruption of TPJ leads to less utilitarian choices, especially in moral-impersonal dilemmas. In the current study, we observed that neither enhancing nor reducing the excitability of bilateral TPJ altered moral decisions, regardless of the dilemmas being moral-personal, moral-impersonal or non-moral conditions. Since the responses of participants confronting moral dilemmas were not identical to the moral judgments involving beliefs and intentions, the mechanism of altering moral decisions through the theory of mind in other moral judgments may not be available for the moral response in moral dilemmas. The observations in our current study do not support the findings of previous studies. However, the findings in the current study may be due to the combination of bilateral anodal and cathodal tDCS stimulations of TPJ, and further study is needed focusing on separating the influences of the left and right TPJ to discuss the functions of these brain regions, respectively.

To sum up, we conclude that in moral dilemmas, altering the activation of the bilateral DLPFC may change moral responses by altering its information integrating process in moral decisions, especially in high conflict moral-personal dilemmas, while modulating the excitability of TPJ has no significant effect over moral responses through its function, as described in the theory of mind.

Reaction Times and Moral Responses

The observation that those saying more "appropriate" to moral-personal dilemmas exhibit longer reaction times also indicates that they experienced greater emotional interference in high-conflict moral dilemmas (Greene et al., 2001). On contrast, no such correlation between reaction times and response ratings within the moral-impersonal condition was observed and the data within the non-moral condition exhibit a significant opposite direction. The relationship between behavioral moral responses and respective reaction times further proved that the emotional interference may play an essential role in moral dilemmas, especially in moral-personal dilemmas.

Limitations

One limitation of the current study is that although our findings in the DLPFC confirmed that modulating the excitability of the right DLPFC altered participants' moral judgments through its function of moral information integrating process, the mechanism underlying the bilateral DLPFC altering the moral response in moral-personal dilemmas remains to be revealed and discussed. Another deficiency of our study is that the results of our experiment were based on stimulation of the bilateral DLPFC or TPJ, and may reflect the combination of bilateral anodal and cathodal tDCS stimulations. Future studies focused on separating the influences of the left and right DLPFC or TPJ and discussing the functions of these brain regions, are required.

CONCLUSION

In summary, our findings provide important information regarding the impact of tDCS on the DLPFC of healthy participants, especially with respect to moral-personal dilemmas.

Activating the right DLPFC while inhibiting the left DLPFC by tDCS may lead to less utilitarian responses in moral judgment, especially in moral-personal dilemmas, supporting the claim that the right DLPFC plays an essential role, not only through its function of moral reasoning but also through its emotional information integrating process in moral judgments. Moreover, neither enhancing nor reducing the excitability of the bilateral TPJ altered participants' moral decisions, regardless of the dilemmas being moral-personal, moral-impersonal or non-moral conditions, indicating that the bilateral TPJ may have little influence over moral judgments in moral dilemmas.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the guideline of tDCS experiment, Zhejiang University ethics committee with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Zhejiang University ethics committee.

AUTHOR CONTRIBUTIONS

HZ, XL, and DH designed experiment; HZ and DH performed experiment; HZ, XL, and DH analyzed data; HZ drew figures; HZ, XL, and DH wrote the manuscript; HZ, XL, and DH revised the manuscript and HZ, XL, and DH finally approved the version to be published.

ACKNOWLEDGMENTS

This work was supported by the Zhejiang Philosophy and Social Sciences Planning Project (17NDJC166YB), the National Social Science Fund of China (Grant number: 13AZD061, 15ZDB134) and the National Nature Science Foundation of China (Grant number: 71403242).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2018.00193/full#supplementary-material

REFERENCES

Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., and Ladurner, G. (2009). Temporo-parietal junction activity in theory-of-mind tasks: falseness, beliefs, or attention. *J. Cogn. Neurosci.* 21, 1179–1192. doi: 10.1162/jocn.2009.21082

Baird, J. A., and Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Dir. Child Adolesc. Dev.* 2004, 37–49. doi: 10.1002/cd.96

Baird, J. A., and Moses, L. J. (2001). Do preschoolers appreciate that identical actions may be motivated by different intentions? *J. Cogn. Dev.* 2, 413–448. doi: 10.1207/s15327647jcd0204_4

Borg, J. S., Hynes, C., Van Horn, J., Grafton, S., and Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: an fMRI investigation. *J. Cogn. Neurosci.* 18, 803–817. doi: 10.1162/jocn.2006.18.5.803

Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., et al. (2008). The neural correlates of third-party punishment. *Neuron* 60, 930–940. doi: 10.1016/j.neuron.2008.10.016

Buckholtz, J. W., and Marois, R. (2012). The roots of modern justice: cognitive and neural foundations of social norms and their enforcement. *Nat. Neurosci.* 15, 655–661. doi: 10.1038/nn.3087

Cushman, F., Young, L., and Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: testing three principles of harm. *Psychol. Sci.* 17, 1082–1089. doi: 10.1111/j.1467-9280.2006.01834.x

Gallagher, H. L., and Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends Cogn. Sci.* 7, 77–83. doi: 10.1016/s1364-6613(02)00025-6

Gandiga, P. C., Hummel, F. C., and Cohen, L. G. (2006). Transcranial DC stimulation (tDCS): a tool for double-blind sham-controlled clinical studies in brain stimulation. *Clin. Neurophysiol.* 117, 845–850. doi: 10.1016/j.clinph.2005.12.003

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends Cogn. Sci.* 11, 322–323. doi: 10.1016/j.tics.2007.06.004

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., and Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44, 389–400. doi: 10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., and Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science* 293, 2105–2108. doi: 10.1126/science.1062872

Jeurissen, D., Sack, A. T., Roebroeck, A., Russ, B. E., and Pascualleone, A. (2014). TMS affects moral judgment, showing the role of DLPFC and TPJ in cognitive and emotional processing. *Front. Neurosci.* 8:8. doi: 10.3389/fnins.2014. 00018

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., and Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832. doi: 10.1126/science.1129156

Kuehne, M., Heimrath, K., Heinze, H. J., and Zaehle, T. (2015). Transcranial direct current stimulation of the left dorsolateral prefrontal cortex shifts preference of moral judgments. *PLoS ONE* 10:e127061. doi: 10.1371/journal.pone.0127061

Leloup, L., Miletich, D. D., Andriet, G., Vandermeeren, Y., and Samson, D. (2016). Cathodal transcranial direct current stimulation on the right temporo-parietal junction modulates the use of mitigating circumstances during moral judgments. *Front. Hum. Neurosci.* 10:355. doi: 10.3389/fnhum.2016.00355

Mai, X., Zhang, W., Hu, X., Zhen, Z., Xu, Z., Zhang, J., et al. (2016). Using tDCS to explore the role of the right temporo-parietal junction in theory of mind and cognitive empathy. *Front. Psychol.* 7:380. doi: 10.3389/fpsyg.2016.00380

Nitsche, M. A., and Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *J. Physiol.* 527, 633–639. doi: 10.1111/j.1469-7793.2000.t01-1-00633.x

Ruby, P., and Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nat. Neurosci.* 4, 546–550. doi: 10.1038/87510

Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind." *Neuroimage* 19, 1835–1842. doi: 10.1016/s1053-8119(03)00230-1

Schleim, S., Spranger, T. M., Erk, S., and Walter, H. (2010). From moral to legal judgment: the influence of normative context in lawyers and other academics. *Soc. Cogn. Affect. Neurosci.* 6, 48–57. doi: 10.1093/scan/nsq010

Sellaro, R., Güroglu, B., Nitsche, M. A., Wildenberg, W. P. M. V.D., Massaro, V., Durieux, J., et al. (2015). Increasing the role of belief information in moral judgments by stimulating the right temporoparietal junction. *Neuropsychologia* 77, 400–408. doi: 10.1016/j.neuropsychologia.2015.09.016

Shultz, T. R., Wright, K., and Schleifer, M. (1986). Assignment of moral responsibility and punishment. *Child Dev.* 57, 177–184. doi: 10.2307/1130649

Sommer, M., Döhnel, K., Sodian, B., Meinhardt, J., Thoermer, C., and Hajak, G. (2007). Neural correlates of true and false belief reasoning. *Neuroimage* 35, 1378–1384. doi: 10.1016/j.neuroimage.2007.01.042

Surber, C. F. (1977). Development processes in social inference: averaging of intentions and consequences in moral judgment. *Dev. Psychol.* 13, 654–665. doi: 10.1037/0012-1649.13.6.654

Tassy, S., Oullier, O., Duclos, Y., Coulon, O., Mancini, J., Deruelle, C., et al. (2011). Disrupting the right prefrontal cortex alters moral judgement. *Soc. Cogn. Affect. Neurosci.* 7, 282–288. doi: 10.1093/scan/nsr008

Thomson, J. J. (1986). *Rights, Restitution, and Risk: Essays in Moral Theory*. Cambridge, MA: Harvard University Press.

Van Bavel, J. J., Feldmanhall, O., and Mende-Siedlecki, P. (2015). The neuroscience of moral cognition: from dual processes to dynamic systems. *Curr. Opin. Psychol.* 6, 167–172. doi: 10.1016/j.copsyc.2015.08.009

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happ,é, F., Falkai, P., et al. (2001). Mind reading: neural mechanisms of theory of mind and self-perspective. *Neuroimage* 14, 170–181. doi: 10.1006/nimg.2001.0789

Ye, H., Chen, S., Huang, D., Zheng, H., Jia, Y., and Luo, J. (2015). Modulation of neural activity in the temporoparietal junction with transcranial direct current stimulation changes the role of beliefs in moral judgment. *Front. Hum. Neurosci.* 9:659. doi: 10.3389/fnhum.2015.00659

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., and Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial stimulation reduces the role of beliefs in moral judgments. *Proc. Natl. Acad. Sci. U.S.A.* 107, 6753–6758. doi: 10.1073/pnas.0914826107

Young, L., Cushman, F., Hauser, M., and Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8235–8240. doi: 10.1073/pnas.0701408104

Young, L., and Dodell, F. D. R. (2010). What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia* 48, 2658–2664. doi: 10.1016/j.neuropsychologia.2010. 05.012

Young, L., and Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *Neuroimage* 40, 1912–1920. doi: 10.1016/j.neuroimage.2008.01.057

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership