OXFORD

Neural signatures of cooperation enforcement and violation: a coordinate-based meta-analysis

Zhong Yang,[1,2] Ya Zheng, [iD][3] Guochun Yang,[1,2] Qi Li,[1,2] and Xun Liu[1,2]

[1]CAS Key Laboratory of Behavioral Science, Institute of Psychology, 100101, Beijing, China, [2]Department of Psychology, University of the Chinese Academy of Sciences, 100049, Beijing, China and [3]Department of Psychology, Dalian Medical University, 116044, Dalian, China

Correspondence should be addressed to Qi Li, CAS Key Laboratory of Behavioral Science, Institute of Psychology, 16 Lincui Road, Chaoyang District, Beijing 100101, China. E-mail: liqi@psych.ac.cn.

Abstract

The social norms of fairness and reciprocity are fundamental to cooperation and constitute core behavioral principles. Warm glow theory suggests that cooperative behavior is driven by positive emotions, whereas inequity aversion theory proposes that cooperative behavior is necessary to avoid negative feelings. However, the precise characteristics underlying the enforcement (fairness or reciprocity) and violation (unfairness or betrayal) of cooperation remain elusive. Moreover, whether the neural mechanism of cooperation as a partner or a spectator is the same remains unclear. To resolve the above issues, we summarized the findings of human cooperation neuroimaging studies through a meta-analysis. Based on our results, cooperation enforcement activates reward-related brain areas, such as the striatum and orbitofrontal cortex, only during other-cooperation. In contrast, cooperation violation is associated with the negative emotion-related insula in both self- and other-noncooperation. Thus, people expect others to cooperate rather than themselves; however, people are disgusted when cooperation is violated by themselves or others. Taken together, cooperative behavior might be mainly driven by a process designed to avoid negative emotion, which supports the inequity aversion theory but not the warm glow theory, thereby improving our understanding of cooperation theory.

Key words: fMRI; meta-analysis; cooperation; enforcement and violation; neural signatures

Introduction

'The time is not as important as the terrain, but the terrain is not as important as unity with the people.' —Mengcius.

Cooperation constitutes the core behavioral principle of human society, being crucial for individual survival and potentially promoting alliance formation, enabling the development of science and technologies and increasing reproductive success (Hamilton, 1964; Archibugi and Pianta, 1992; Gulati, 1995). Therefore, there are various definitions of cooperation from the different perspectives of anthropologists, biologists,

psychologists and economists (Strang and Park, 2017). Reviewing the related studies, we define human cooperation as behaviors undertaken by individuals and groups to maximize long-term public benefits and to promote collective well-being (Tabibnia and Lieberman, 2007; Colman, 2009).

Various theories have attempted to explain the possibilities underlying cooperative behavior, such as warm glow and inequity aversion theories. Warm glow theory is predominant in explaining cooperation (Andreoni, 1990). This theory suggests that because cooperative behavior could be driven by positive emotions, it presumably occurs even in situations of anonymity

(Strang and Park, 2017). In contrast, the inequity aversion theory proposes an alternative explanation for cooperative human behavior in non-anonymous and anonymous settings. It hypothesizes that people feel aversion when facing unequal outcomes, regardless of whether another person is worse or better off than themselves (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). Therefore, to avoid potential negative feelings but not the induction of positive feelings, people will choose cooperation to minimize inequity (Rilling and Sanfey, 2011; Strang and Park, 2017). Given that these two theories explain cooperation from the perspective of alternative motives, the quest for a unifying theory explaining the diverse cases of cooperative behavior is still underway. Hence, an investigation of cooperative behavior in laboratory settings may serve as one step toward a better understanding of cooperation.

In terms of sharing and resource distribution as well as reciprocal exchange in social decision making, some researchers have recently further divided cooperation into fairness and reciprocity (Rilling and Sanfey, 2011). In previous studies, fairness was defined as the equitable distribution of an initial stake of money (Tabibnia and Lieberman, 2007; Tabibnia et al., 2008). In dictator game (DG) and ultimatum game (UG) tasks, the fair and unfair distributions from proposers (Weiland et al., 2012; Zaki and Mitchell, 2011) and the acceptance or rejection decisions from responders (Guroglu et al., 2010; Cheng et al., 2015) all reflected the underlying neural mechanisms of fairness enforcement or violation. Therefore, we tried to combine DG and UG tasks to reveal the underlying neural mechanisms of fairness enforcement and violation. In contrast to the UG and DG, the prisoner's dilemma game (PDG) (Poundstone, 1993), trust game (TG) and public good game (PGG) (Berg et al., 1995; Kollock, 1998) are widely used to investigate reciprocity. In the PDG, two players simultaneously determine their own reciprocal behavior in the decision stage and then see another partner's reciprocal behavior in the feedback stage. Similar to the PDG, participants in the TG may play as 'investors' or 'trustees'. As investors, they can decide their own reciprocal behavior by providing monetary endowments to the trustee and see the trustees' reciprocal or betrayal behavior based on others' return, and vice versa. Finally, in the PGG, participants can divide the monetary units between their private account and public account and see the payoff consequences of contributions to the public account. In all these games, people can choose to cooperate or not, which can be identified as self-cooperation and self-noncooperation, respectively, in addition to watching others' cooperative or noncooperative behavior, which can be identified as other-cooperation and other-noncooperation, respectively. Overall, all these games are commonly used for examining cooperation in human societies (Stanford and Bunn, 2001).

Based on these distinct economic exchange games, several researchers have proposed brain network models of cooperative decision making from a neuroscientific perspective. Stallen and Sanfey (2013) and Strang and Park (2017) summarized several separate anatomical brain areas that played important roles in cooperation. The perspective-taking network includes the dorsal medial prefrontal cortex (DMPFC). The reward-related reciprocal cooperation network involves the orbitofrontal cortex (OFC), striatum and ventromedial prefrontal cortex (VMPFC). The network involved in the anticipation of guilt or noncooperation includes the insula (IA) and anterior cingulate cortex (ACC). Rilling and Sanfey (2011) identified several models of the neural systems that mediated different types of cooperation, such as fairness and reciprocity. The fairness network, which includes the VMPFC, ventrolateral prefrontal cortex and IA, is involved in

emotional reactions, emotion regulation and aversion to offers of inequity. In addition, the reciprocity network, which includes the VMPFC, DLPFC, DMPFC, striatum and OFC, is specialized for valuing long-term benefits, expending cognitive effort for breaking reciprocal norms, building reciprocal relationships and reward processing.

Although researchers have obtained information on the above brain network models of cooperation through reviewing evidence from human neuroimaging studies and have summarized the neural mechanisms of cooperation and its subcomponents of fairness and reciprocity, the narrative review approach suffers from limitations. First, the approach of narrative reviews is subjective and lacks transparency (Borenstein et al., 2009). Different reviewers might use different inclusion criteria, weights of studies and conclusion thresholds, which makes it difficult for them to arrive at an unbiased conclusion. Second, reviews might become less useful as more information becomes available. Furthermore, the reviewers may implicitly synthesize studies according to their own opinions or assign importance to some studies, and this process might eventually become untenable as the number of studies increases (Borenstein et al., 2009). Therefore, an alternative to narrative reviews is a meta-analysis, such as a meta-analysis with a random-effect activation likelihood estimation (ALE) algorithm (Eickhoff et al., 2012), which allows statistically verifiable concurrence across functional neuroimaging studies of human cooperation and objectively reveals the regions with the highest 'likelihood' of activation.

This study combines theory-driven and data-driven approaches, and its main goal is to assess the mechanisms of cooperation enforcement and violation by performing a meta-analysis of the results reported in 85 published neuroimaging studies investigating human cooperation. On the one hand, we predict that three different patterns of results related to different cooperation models are possible. First, according to the warm glow theory, which initially insisted on positive feelings about cooperation enforcement, reward-related areas, such as the OFC and striatum, would be activated. Second, based on the inequity aversion theory, cooperation violation is associated with aversive emotion, and aversion-related brain areas, such as the IA, would be activated. Third, combining the warm glow and inequity aversion theories indicates that the neural architecture of cooperation entails both reward-related and aversion-related brain areas, reflecting cooperation enforcement and cooperation violation, respectively. On the other hand, considering that in the cooperation enforcement condition, personal interests are usually sacrificed to protect collective interests (Piliavin and Charng, 1990; Baston, 1998; Van, Sapienza, Villanueva et al., 2007). Therefore, we predict that people are more likely to expect cooperation from others than from themselves, which would activate the cooperation-related neural network.

Materials and methods

Literature search and organization

A thorough literature search was conducted using the ISI databases on 16 January 2018. The various combinations of relevant search terms used were as follows: 'cooperation/fair/ unfair/equal/equity/equivalent/inequity/ultimatum game/dictator game/prisoner's dilemma/trust game/dilemma/trust/distrust/mistrust/public good game/betray' and 'fMRI/magnetic resonance imaging/neuroimaging/functional magnetic resonance imaging' and 'social decision making'. Two researchers
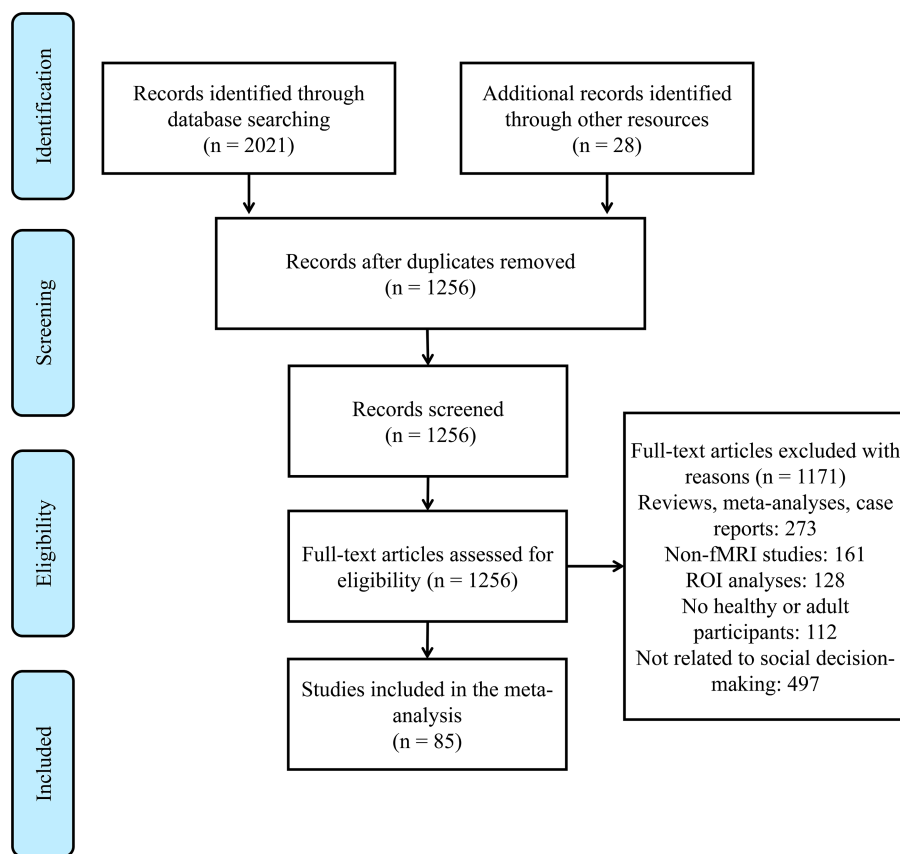
Fig. 1. Flow chart of the search strategy: the number of articles selected for the meta-analysis.

entered the search terms into databases. Then, they selected 'all fields', indicating that these relevant search items were present in any part of an article, such as the title, abstract and main text. All of the initial search results were merged to yield a total of 1256 articles after eliminating redundant entries. The exclusion criteria were (1) the published reports were not the primary empirical studies (e.g. meta-analyses, review articles, and case reports were excluded); (2) functional magnetic resonance imaging (fMRI) was not used as the imaging modality; (3) fMRI results were only based on region of interest analysis, rather than entire-brain thresholds; (4) the participants were not healthy adults (e.g. only included participants with brain damage or other mental problems, children or aged participants were excluded); (5) the results were not reported in standard stereotactic coordinate space (Talairach or Montreal Neurological Institute, MNI); and (6) the published reports were not related to social decision-making tasks. Ultimately, 85 published fMRI papers satisfied all these criteria and were used in the meta-analysis. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement guided the meta-analysis (Liberati *et al.*, 2009). The screening and exclusion procedures are shown in Figure 1.

The foci data extracted from the published reports for the ALE analysis were read as a text file, which contains the name of the first author, year and the number of subjects for this group of foci, followed by the coordinate data. The data from the studies were extracted along with the coordinates of relevant foci for ALE analysis if they met the entire-brain threshold set by the authors. In the published reports using the UG and DG, the contrasts of 'fair > unfair' and 'unfair > fair' were extracted when

participants proposed fair or unfair offers as a proposer or when they reacted to fair or unfair offers as a responder. Moreover, the contrasts of 'accept > reject' and 'reject > accept' were extracted when responders chose to accept or reject the offers in the published reports using the UG. Thus, we categorized the label of fairness with the contrasts of 'fair > unfair' and 'accept > reject' and the label of unfairness with the contrasts of 'unfair > fair' and 'reject > accept'. Moreover, the published reports using the PDG, TG and PGG were categorized with labels of 'reciprocity' (reciprocity > betrayal) and 'betrayal' (betrayal > reciprocity). Fairness and reciprocity were categorized as cooperation enforcement, while unfairness and betrayal were categorized as cooperation violation. Additionally, categorization labels of self-cooperation and self-noncooperation were applied when the participants made their own decisions to cooperate or not (propose a fair or unfair offer, reciprocate or betray), while labels of other-cooperation and other-noncooperation were applied when participants saw another person experience a fair or unfair offer, or saw others' reciprocate or betray decisions. Finally, the cooperation enforcement analysis included 534 foci from 61 published reports, whereas the cooperation violation analysis included 801 foci from 67 published reports. The term 'experiments' was adopted, as used by the BrainMap database, to refer to individual regressors or contrasts typically reported in published studies analyzing fMRI data. Each published report included only one experiment, and the numbers of experiments included in the analyses of fairness, unfairness, reciprocity, betrayal, self-cooperation, other-cooperation, self-noncooperation and other-noncooperation were 28 (175 foci), 46 (509 foci), 33 (359 foci), 21 (292 foci), 32 (214 foci), 39 (309 foci),

23 (197 foci) and 50 (576 foci), respectively. The reported coordinates in the published reports were then grouped by different spatial normalization schemes according to coordinate transformations implemented in the GingerALE toolbox (version 2.3.6, http://brainmap.org, Research Imaging Center of the University of Texas Health Science Center, San Antonio, Texas), i.e. using Brett methods to convert Talairach coordinates into MNI space (Brett, 1999).

Activation likelihood estimation

Topic-specific meta-analyses were conducted to explore activation in response to the same condition or task, and conjunction meta-analyses were conducted to explore the common regions activated across different tasks (Fox *et al., 2013*). Ten different topic-specific meta-analyses were conducted using the revised ALE algorithm (Sul *et al., 2015*) implemented in the GingerALE software, including the main analyses of all cooperation enforcement and cooperation violation and the sublists characterizing brain activation by fairness and unfairness, reciprocity and betrayal, self- and other-cooperation and self- and other-noncooperation. Moreover, four conjunction meta-analyses were conducted to identify common brain activation among fairness and reciprocity, unfairness and betrayal, self-cooperation and other-cooperation and self-noncooperation and other-noncooperation. For the topic-specific meta-analyses, ALE maps were computed with a cluster-level family-wise error (FWE)-corrected threshold (Eickhoff *et al.,* 2012; Eklund *et al.,* 2016) of $P < 0.01$, a cluster-defining threshold of $P < 0.001$ and 1000 permutations. For the conjunction meta-analyses, we used the voxel-wise minimum value of the two ALE maps calculated in the topic-specific meta-analyses. We used the number of subjects in each foci group to calculate the Full-width-half-maximum (FWHM) of the Gaussian function to blur the foci, rather than using a pre-specified FWHM for all experiments. The meta-analysis results were overlaid onto an anatomical template (Colin27_T1_seg_MNI. nii, www.brainmap.org/ale) and displayed using Mango software (http://rii.uthscsa.edu/mango/). The details (i.e. the number of included studies, subjects, peak oci, and contrasts) of the published reports included in each of the meta-analyses described above are summarized in the Supplementary Materials (Supplementary Table S6).

As the effects of the meta-analytic contrast analyses are two orders of magnitude higher than the effects of the topic-specific meta-analyses and conjunction meta-analyses, Supplementary Figure S2 ranges from 0 to 3, while the other figures range from 0 to.03.

Results

Activations of cooperation enforcement and its subcomponents

Cooperation enforcement activated the right caudate and right OFC. Fairness and reciprocity resulted in consistent activation in the right caudate. Fairness additionally activated the left caudate and right OFC. Common activation of fairness and reciprocity was found in the right caudate (Figure 2 and Table 1).

Activations of cooperation violation and its subcomponents

Cooperation violation (noncooperation) activated the bilateral supplementary motor area (SMA)/bilateral dorsal ACC (DACC),

bilateral IA, left inferior parietal lobule (IPL) and left red nucleus. Unfairness caused similar brain activation patterns of the networks mentioned above, including the SMA/DACC, superior frontal cortex (SFC) and bilateral IA. Betrayal resulted in activation in the right IA and thalamus. Common activation of unfairness and betrayal occurred in the right IA (Figure 3 and Table 2).

Activations of self- and other-cooperation enforcement and violation

Other-cooperation activated the right caudate. However, self-cooperation did not show any significant activation. Additionally, other-noncooperation activated the bilateral SMA/DACC, bilateral IA and left IPL, while self-noncooperation activated the bilateral IA. Moreover, common activation of self-noncooperation and other-noncooperation occurred in the IA (Figure 4 and Table 3).

Discussion

Cooperation is crucial for survival in a society. Theories such as the warm glow and inequity aversion theories attempt to explain cooperation from different perspectives (Andreoni, 1990; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). In previous narrative reviews, fMRI studies of human cooperation have been summarized according to the opinions of the authors using experimental economics paradigms, such as the UG, DG, PDG, TG and PGG (Rilling and Sanfey, 2011; Stallen and Sanfey, 2013; Strang and Park, 2017). However, precise characterization of the neural systems underlying cooperation has been a longstanding puzzle of social neuroscience. The objective of this coordinate-based meta-analysis was to quantitatively synthesize the results of previous fMRI studies regarding cooperation, with the aim of identifying consistent activation patterns of the neural signatures underpinning this psychological phenomenon. Our results identified convergence of reported activation foci related to cooperation enforcement, fairness, reciprocity and other-cooperation, including the OFC and striatum (caudate), which are associated with reward-related brain networks (O'Doherty, 2004; Haber and Knutson, 2010; Zaki and Mitchell, 2011; Gromann *et al.,* 2014; Wang *et al.,* 2016). Furthermore, the results showed that cooperation violation, unfairness, betrayal and self- and other-noncooperation were correlated with activity in the DACC, IA, IPL and red nucleus. These regions are thought to convey conflict (Harle *et al.,* 2012) and aversive affective experiences (Corradi-Dell'Acqua *et al.,* 2016). In addition, the striatum, which is a reward-related area, was activated only under other-cooperation conditions, and the IA, which is associated with inequity aversion (Hsu *et al.,* 2008; Rilling and Sanfey, 2011), was activated in response to all noncooperation conditions. Based on these findings, cooperative behavior might be mainly driven by a process designed to avoid negative emotion, which supports the inequity aversion theory but is not captured by the warm glow theory.

Differing from previous narrative reviews of the neural basis of cooperation behavior, the current study used a meta-analysis to summarize the neural basis of cooperation behavior. Our meta-analytical study deviates from these review studies in some aspects. First, the three reviews focused on different perspectives of the neural mechanism of cooperation. Stallen and Sanfey (2013) provided an overview that external incentives motivate cooperation and examined the role of reward, learning
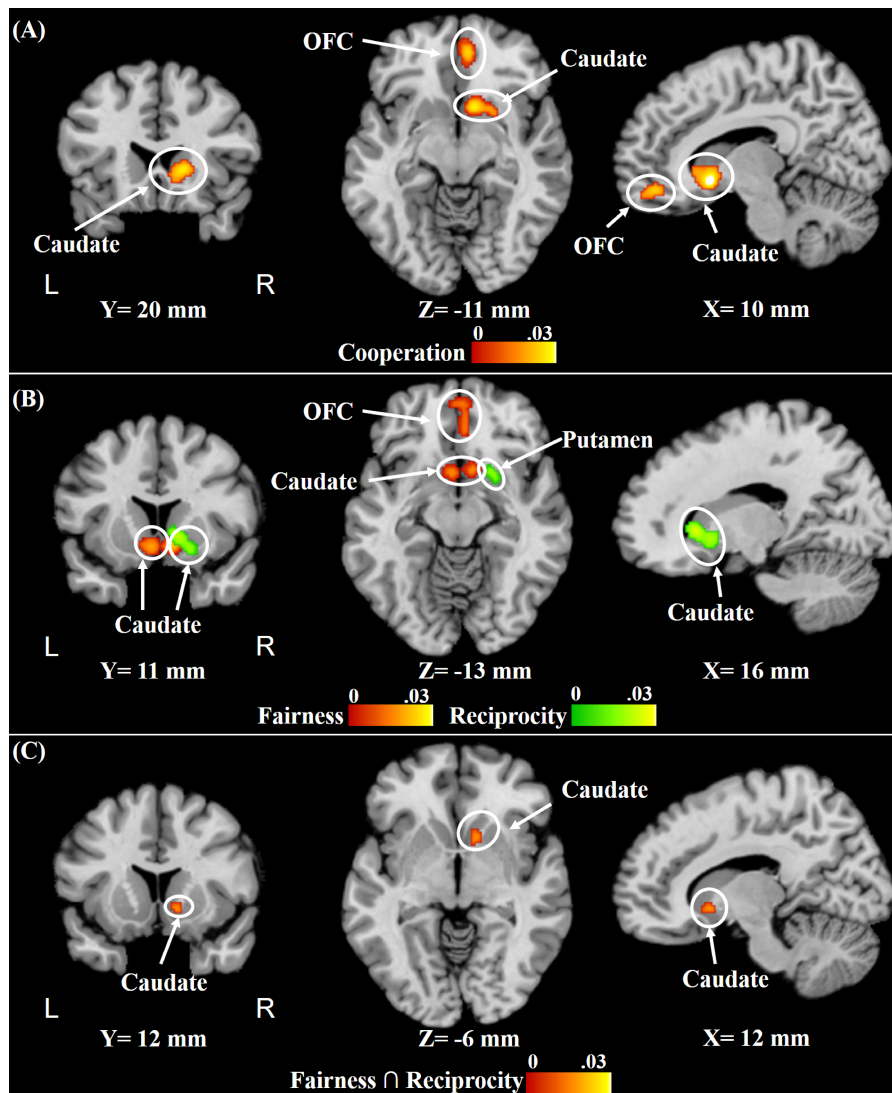
Fig. 2. Results of the topic-specific meta-analyses of **(A)** cooperation enforcement and **(B)** fairness and reciprocity (for A and B, cFWE, P < 0.01, and uncorrected cluster-defining, P < 0.001) and conjunction meta-analysis of **(C)** common regions activated in response to fairness and reciprocity. The bars represent the ALE values.

social context and social emotions on the neural mechanisms underlying cooperation. Strang and Park (2017) stated that cooperative behavior has often been associated with positive emotions and the activation of the striatum, while noncooperative behavior was linked to negative emotions and activation of the IA. Our study explored the neural mechanism of cooperation from the perspectives of self-other. The findings of our meta-analysis suggest the involvement of different brain networks in self- and other-enforcement and violation of cooperation, indicating that distinct norms might be used when people consider their own and others' cooperative behaviors. Second, the three reviews tried to answer different theoretical questions. The first review summarized the important factors influencing the neural mechanisms underlying cooperation. The second review proposed that cooperative behavior is accompanied by emotional reactions. Our meta-analysis supported the inequity aversion theory from the perspective of the neural mechanism of cooperation and solved the controversy regarding the theories of different model of cooperation. Finally, the research method differs from the previous reviews on coopera-

tion. Previously published reports are narrative reviews (Stallen and Sanfey, 2013; Strang and Park, 2017). These reviews tend to collect relevant evidence based on the authors' personal perspectives. Therefore, their conclusions are often subjective and the information collected is incomplete (Borenstein *et al.*, 2009).

Moreover, a related meta-analysis has been performed to investigate the neural mechanisms of trust behavior (Bellucci *et al.*, 2017). This study focused on examining brain activation in three investment game stages, including the trust, reciprocity and feedback stages. However, these researchers were not concerned with the two opposite aspects of these three stages, such as trust–distrust, reciprocity–betrayal and positive feedback–negative feedback. Differing from this study, our study explored the common and distinct brain networks underlying human cooperation from the perspective of fairness and reciprocity as well as the self and others. Our study investigated not only 'cooperation', 'fairness' and 'reciprocity' but also their opposites, i.e. 'non-cooperation', 'unfairness' and 'betrayal', from the perspectives of the self and others.

Table 1. Significant clusters that were activated in response to cooperation enforcement, fairness and reciprocity (cFWE, $P < 0.01$, and uncorrected cluster-defining, $P < 0.001$), and common regions activated in response to fairness and reciprocity

Brain regions	BA	MNI			Cluster size (mm³)	Extrema value ($\times 10^{-2}$)
		x	y	z		
Cooperation						
Caudate	–	12	12	−6	2784	3.55
Caudate	–	16	20	0		2.70
OFC	32	6	48	−14	1352	3.10
Fairness						
Caudate	–	10	12	−10	1520	2.54
Caudate	–	−4	10	−10		1.98
OFC	10	4	54	−16	1104	1.75
OFC	32	4	40	−16		1.40
Reciprocity						
Caudate	–	16	20	0	2432	2.47
Caudate	–	14	10	−4		2.24
LN/putamen	–	22	10	−12		1.74
Fairness ∩ reciprocity						
Caudate	–	12	12	−6	200	1.71

Abbreviations: LN, lentiform nucleus.

Table 2. Significant clusters that were activated in response to cooperation violation, unfairness and betrayal (cFWE, $P < 0.01$, and uncorrected cluster-defining, $P < 0.001$) and common regions activated in response to unfairness and betrayal

Brain regions	BA	MNI			Cluster size (mm³)	Extrema Value ($\times 10^{-2}$)
		x	y	z		
Noncooperation						
SMA/DACC	32	−6	16	46	8376	5.31
SMA/DACC	32	8	26	34		4.42
SMA/DACC	6	6	22	44		2.65
IA	–	34	24	−2	7840	7.38
IA	–	−34	18	−10	6744	5.73
IPL	7	−32	−54	48	1872	3.52
RN	–	−4	−24	−4	1408	4.26
Unfairness						
SMA/DACC	32	−4	16	48	9840	4.64
SMA/DACC	32	8	26	34		4.41
SMA/DACC	6	6	22	44		2.62
IA	–	34	24	−2	7112	5.97
IA	–	−30	22	4	6784	4.91
Betrayal						
IA	–	36	24	−12	1816	2.89
Thalamus	–	0	−22	0	1184	2.43
Unfairness ∩ betrayal						
IA	–	36	22	−12	1720	2.86

Abbreviations: RN, red nucleus.

Brain networks of cooperation

Core brain networks of cooperation enforcement. The core areas involved in cooperative enforcement, fairness and reciprocity are centered on the right striatum (caudate-putamen) and right OFC, which are correlated with reward processing. The striatum is involved in representing positive emotions, which have been regarded as a major factor driving cooperative behavior (Moll *et al.,* 2006; Harbaugh *et al.,* 2007; Hare *et al.,* 2010; Strang and Park, 2017). Although the right striatum was activated in response to all cooperative conditions in the current study, we inferred that this brain area might represent general social norm enforcement, rather than specific cooperative behavior. Moreover, the striatum also plays a role in coding outcome values of social interactions (Rilling *et al.,* 2002; Phan *et al.,* 2010; Baez-Mendoza and Schultz, 2013; Fareri and Delgado, 2014;

Hughes *et al.,* 2016) or is reliably linked to learning signals that help inform about another's reputation in cooperative behavior (King-Casas *et al.,* 2005; Fouragnan *et al.,* 2013; Fareri *et al.,* 2015). Thus, the activation of this brain area may represent the benefit of cooperation or learning behavior to predict another's behavior and make decisions accordingly. Furthermore, the OFC has been shown to be involved in increasing sensitivity to distant rewards (Schultz *et al.,* 2000; O'Doherty *et al.,* 2001; Rilling *et al.,* 2002). Based on other findings from neuroimaging studies, the OFC appears to code stimulus-reward value (O'Doherty, 2004), including the relative values of different rewarding stimuli (Tremblay and Schultz, 1999; Watanabe, 1999) and the values of offered and chosen goods (Padoa-Schioppa and Assad, 2006). Thus, the activation of the OFC may play roles in integrating reward values and learning the opponent's reputation to guide decisions. This area seems to be required for the establishment
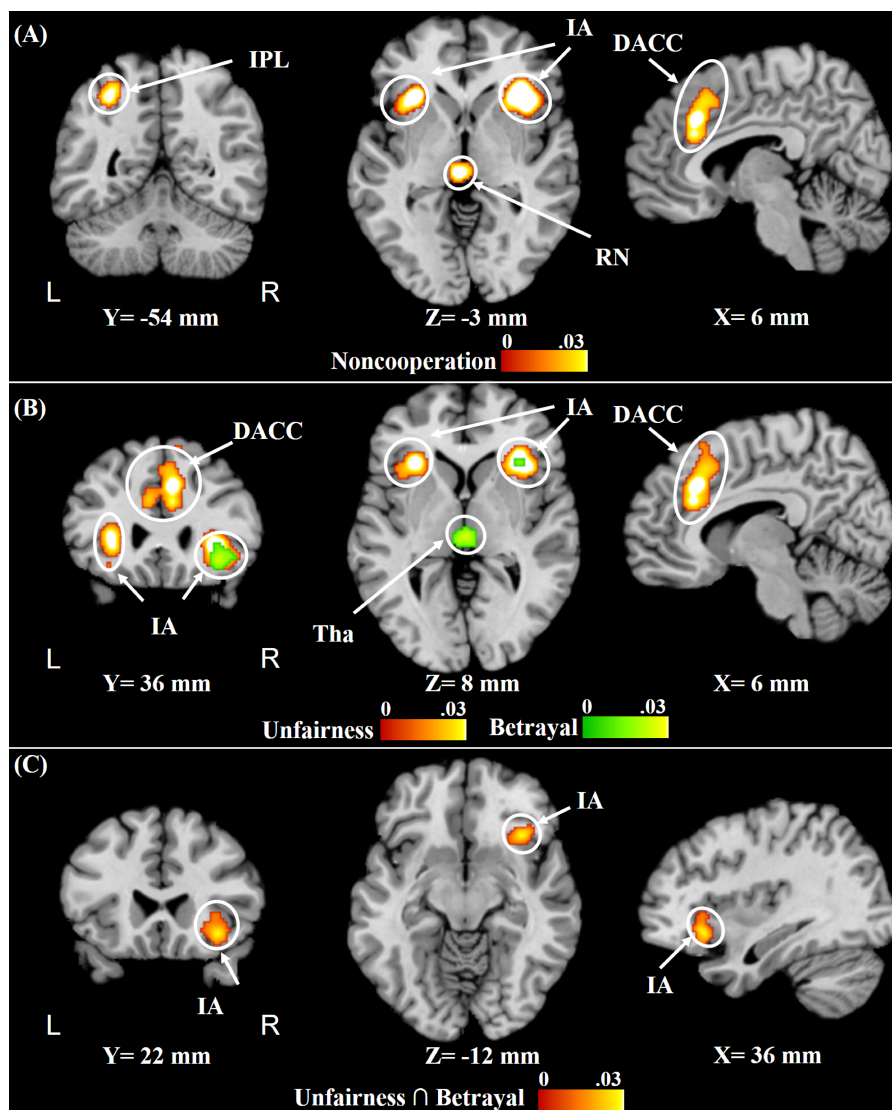
Fig. 3. Results of the topic-specific meta-analyses of **(A)** cooperation violation and **(B)** unfairness and betrayal (for A and B, cFWE, P < 0.01, and uncorrected cluster defining, P < 0.001) and conjunction meta-analysis of **(C)** common regions activated in response to unfairness and betrayal. The bars represent the ALE values. Abbreviation: Tha, thalamus.

of cooperation and for the maintenance of cooperative behavior (Decety *et al.*, 2004; Hare *et al.*, 2010; Zaki and Mitchell, 2011).

In contrast to reciprocity, fairness was additionally distinguished by modulating activity in the left striatum and right OFC. Previous studies have shown that the left striatum is activated when participants win more than their counterpart (Bault *et al.*, 2011) and that it. Additionally, the left striatum encodes social ranking and positive social comparison (Hsu *et al.*, 2008; Wang *et al.*, 2013) and shows a striking hemodynamic response correlated with income distribution depending on work effort (Cappelen *et al.*, 2014). Fairness also resulted in extra activation of the right OFC. The OFC and caudate are considered the major players in reward processing because they are the main projection areas of two distinct dopaminergic pathways: the mesocortical and mesolimbic pathways. Previous studies have attempted to differentiate the roles of these two structures in terms of reward stages, relating the striatum to reward anticipation and associating the medial OFC to the outcome of a reward (Suzuki *et al.*, 2011). In situations of fairness deci-

sion making, the OFC not only processes the absolute reward value (Kringelbach, 2005; O'Doherty, 2007) but also codes relative values of different rewarding stimuli (Elliott *et al.*, 2003), including the valuation of equality in the outcomes between the players (Aoki *et al.*, 2014), the value of offered goods (Padoa-Schioppa and Assad, 2006), the reduction of inequality (Vostroknutov *et al.*, 2012) and the assessment of equitable resource distribution (Rilling *et al.*, 2002; Tricomi *et al.*, 2010). In light of these recent findings, the activations of the left caudate and right OFC might represent the specific encoding of subjective feelings about equitableness. Additionally, two meta-analytic contrast analyses between fairness and reciprocity were conducted, and the results were similar to the results of the topic-specific meta-analyses (Supplementary Table S3), supporting the inferred roles of the activation of these brain regions.

Core brain networks of cooperation violation. A neural network including the IA, DACC, IPL and red nucleus, which are specific to
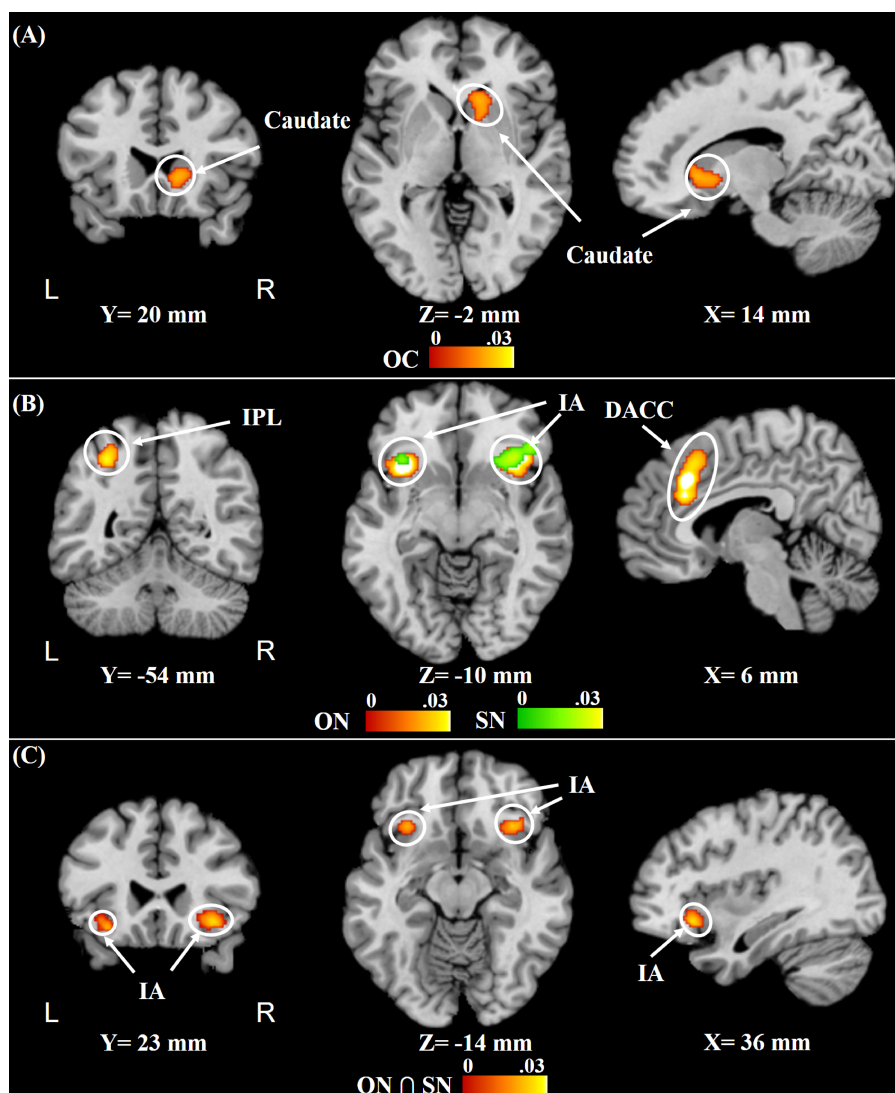
Fig. 4. Results of the topic-specific meta-analyses of **(A)** other-cooperation, **(B)** self- and other-noncooperation (for A and B, cFWE, $P < 0.01$ and uncorrected cluster-defining $P < 0.001$) and conjunction meta-analysis of **(C)** common regions activated in response to self- and other-noncooperation. The bars represent the ALE values. Abbreviations: OC, other-cooperation; SN, self-noncooperation; ON, other-noncooperation.

human cooperation violation, was identified. The IA is related to a negative affective processing area and is involved in the neural instantiation of these processes. It frequently codes physical pain (Critchley *et al.*, 2000), social pain (Goldin *et al.*, 2008; Hein *et al.*, 2010), the extent of cooperation (Sanfey *et al.*, 2003) and the negative emotions aroused during noncooperation (Sanfey *et al.*, 2003; Tabibnia *et al.*, 2008). In addition, activation of the IA was observed under all noncooperation conditions, including unfairness and betrayal, suggesting that when faced with a violation of cooperative behavior, people feel guilty and negative emotions will be elicited (Krajbich *et al.*, 2009). Although previous studies suggested that the IA was associated with negative emotions, recent neuroimaging studies indicated that IA may signal norm violations when results deviated from an expected outcome (Civai *et al.*, 2012). As cooperation is a default social norm and constitutes the core behavioral principle of human society, unfairness and betrayal reflect a violation of social norms. Thus, the activation of the IA under all noncooperation conditions might indicate the underlying process of detecting social norm deviations.

We also observed activation of the DACC in noncooperation and unfairness. Based on the involvement of the DACC in conflict processing (Botvinick *et al.*, 2001; Pochon *et al.*, 2008), the activation of the DACC may reflect the conflict between self-interest and social norms when facing unfairness and noncooperative behaviors and the promotion of norm enforcement by sacrificing self-interest (Rilling and Sanfey, 2011). Moreover, the DACC is associated with tracking expectation violations, such as the social norm deviation (Klucharev *et al.*, 2009) and the detection of social prediction errors (Chang and Sanfey, 2013). The DACC indicates a specific neural signal when people's behaviors break their expectations, which in turn may serve as an emotional indicator motivating people to avoid the violation of social norms (Stallen and Sanfey, 2013).

Activation of the IPL and red nucleus was found in noncooperation. A previous study suggests that the IPL plays a critical role in social interactions (Iacoboni and Dapretto, 2006) and exhibits greater activation during noncooperation than cooperation, which may be due to deeper mind-set synchronization to mentalize and predict an opponent's actions and intentions

Table 3. Significant clusters that were activated in response to self- and other-cooperation enforcement and violation (cFWE, P < 0.01, and uncorrected cluster-defining, P < 0.001), and common regions activated in response to self-noncooperation and other noncooperation

Brain regions	BA	MNI			Cluster size (mm³)	Extrema Value (× 10⁻²)
		x	y	z		
OC						
Caudate	–	14	20	−2	1664	2.07
ON						
SMA/DACC	32	−4	16	48	6968	4.41
SMA/DACC	32	8	26	34		4.41
SMA/DACC	6	6	22	44		2.62
SMA/DACC	32	−6	32	26	1304	3.07
IA	–	34	24	−4	5776	6.22
IA	–	−34	18	−10	5520	4.86
IPL	40	−32	−52	48	1432	2.99
SN						
IA	–	36	24	−12	1896	2.58
IA	–	−32	22	−16	904	2.25
SN ∩ ON						
IA	–	36	24	−12	1208	2.58
IA	–	−32	22	−16	496	2.25

Abbreviations: OC, other-cooperation; SN, self-noncooperation; ON, other-noncooperation.

when they compete with others (Decety et al., 2004; Liu et al., 2015). Furthermore, the red nucleus is situated in the brain stem and is implicated in motoric coordination and movement control. Some studies have suggested that the red nucleus may play an important role in emotion processing (Günther et al., 2017), such as subjective experience of affect, emotional salience detection (Nioche et al., 2009) and anticipatory pleasure (Yan et al., 2015). The activation of the red nucleus may be linked to unpleasant feelings under noncooperation conditions.

The current meta-analysis identified a specific betrayal-related brain area in the thalamus. The thalamus plays an important role in regulating arousal and awareness levels (Van der Werf et al., 2002). In current reciprocal exchange, people may experience a higher level of awareness due to the fear of retaliation when they choose to betray another or may be vigilant to others' harmful behavior when they view others' betrayal. Alternatively, the thalamus is involved in risk assessment during economic exchange (Miedl et al., 2010) and in feedback processing for uncertain rewards (Winkler et al., 2013). Betrayal is a high-risk behavior that violates social reciprocity to maximize the self-benefit. We inferred that the activation of the thalamus observed in this study may reflect the assessment of this high-risk behavior.

Core brain networks of self- and other-cooperation and noncooperation

The results of our meta-analysis showed different brain networks of self- and other-enforcement and violation of cooperation, which indicated that there might be distinct norms when people consider their own and others' behaviors. We identified involvement of the reward-related striatum in response to other-cooperation. Therefore, we postulated that the evaluation and expectation of others' behaviors might occur in strict accordance with cooperation norms, and the norm-abiding behavior of cooperation is encoded as a social reward (Kringelbach and Rolls, 2004). The appreciation of other people's cooperative behaviors also promotes people to standardize their behavior in accordance with the cooperation norms. Interestingly, we did not observe any significant activation in regions associated

with self-cooperation, which might be due to the consideration of multiple competing motivations, such as a balance of self-interest and social norms (Civai et al., 2012), decision strategies (Hampton et al., 2008), a fear of being punished (Cheng et al., 2015; Fehr and Gachter, 2002) and reputation concerns (Nowak and Sigmund, 1998), rather than simple and automatic reward-seeking behaviors when people decide to reciprocate in economic decision making processes. Decisions regarding a person's own reciprocal behaviors might be determined through the integration of multiple factors (Nowak and Sigmund, 1998; Hampton et al., 2008; Civai et al., 2012). Another alternative explanation is that self-fairness and self-reciprocity may involve distinct neural networks. As the number of published reports of self-fairness and self-reciprocity is insufficient to conduct separate ALE analyses, the neural mechanisms of self-cooperation under different conditions remain to be clarified. Thus, the common and distinct neural mechanisms between the self-fairness and self-reciprocating conditions must be clarified in the future to verify this proposed explanation. Overall, based on the results from the present study, cooperative behavior is not necessarily driven by positive emotions, particularly during self-cooperation, which did not support the warm glow theory.

Moreover, we observed common activation of the IA in self- and other-noncooperation. This result suggested that the aversion aroused by noncooperation might be a kind of automatic response, as it occurred not only when people saw others' noncooperative behaviors but also when they made decisions to maximize their self-interest. Extra activation of the DACC and IPL during other-noncooperation was also revealed. As mentioned previously, the DACC might reflect social prediction errors as others' noncooperation behaviors violating their expectations as well as the conflict of self-interest and social norms. Since we did not find the activation of the DACC in self-noncooperation, we inferred that conflict is feeble when people make decisions by themselves, as self-interest with strong motivation. The IPL belongs to the mirror neuron system and plays a critical role in interpersonal interactions (Iacoboni and Dapretto, 2006; Iacoboni and Mazziotta, 2007). The IPL is involved in the ability to identify with others, self-other distinction and perspective taking (Lamm et al., 2007;

Schulte-Rüther *et al.,* 2007). When others do not cooperate, people must explain the noncooperative behavior of others from their own perspective (Farrer and Frith, 2002; Meltzoff and Decety, 2003).

Based on these findings, we postulated that people make cooperative decisions by themselves in social interactions through a complex process to balance multiple motivations but mainly to avoid the negative emotion aroused by the noncooperative behavior.

A summary of cooperation processing

Based on the identification of networks involved in processing cooperation/noncooperation, we summarized the cooperation model from the economic field. First, the IA may be a core brain area involved in human cooperative behaviors. Considering that activation of the IA was found in all noncooperation conditions, disgust with uncooperative behavior might be one of the most important motivations for self-motivated cooperative behavior and could represent a universal psychological mechanism of cooperative behavior. Second, the striatum and OFC might be the driving force promoting cooperative behavior. Although reward areas, such as the striatum and OFC, are activated when seeing others' cooperation, it is not necessary to activate these areas, especially under self-cooperation. Thus, the social approval of others' cooperation comes from the heart, which pushes human behavior to obey cooperation norms. In addition, the frontoparietal network may play an important role in evaluating and monitoring others' cooperative behaviors. People may not only feel aversion when facing others' noncooperation but also monitor deviation from the cooperation norm and evaluate these behaviors from the perspective of others. We also conducted meta-analytic contrast analyses between fairness and reciprocity, unfairness and betrayal, and self-noncooperation and other-noncooperation. The results (Figure S2 and Supplementary Tables S3–S5) were similar to the results of the topic-specific meta-analyses, which supported the proposed brain networks underlying cooperation and our conclusions.

Caveats

Several methodological caveats should be noted. First, the coordinate-based ALE method is based on reported peak activations, not on the effect size of activation at the peak. Future meta-analyses would benefit from an image-based approach employing statistical maps without a threshold to calculate activation likelihood (Maumet and Nichols, 2015). Second, because a limited number of published reports were included in this meta-analysis (particularly in the meta-analytic contrast analyses between the perspectives of the self and others), the statistical power of the results could be increased in future meta-analyses by including the rapidly accumulating published neuroimaging studies investigating cooperation. Third, the potential functions of the brain regions involved in cooperation enforcement and violation were evaluated in the context of evidence from neuroscientific studies from previous decades, but the specific functions of many brain areas are still not fully understood. For example, although activations of the OFC and striatum in fairness and reciprocity, respectively, were identified in this meta-analysis, the specific dissociated reward function of the two areas in cooperation has scarcely been discussed in previous studies. As mentioned in the introduction, the combination of DG and UG tasks is mainly

based on the observation that the decision-making behaviors of both the proposers and the respondents in these tasks reflect fairness enforcement and violation (Guroglu *et al.,* 2010; Zaki and Mitchell, 2011; Weiland *et al.,* 2012; Cheng *et al.,* 2015). Additionally, the combination likely enhances the reliability of the results by increasing the numbers of the published reports available for ALE analyses (Eickhoff and Etkin, 2016; Eickhoff *et al.,* 2017). Notably, proposers who make an unfair offer might utilize different cognitive and neural processes than responders who accept or reject offers. A further analysis of the underlying neural mechanism of fairness enforcement or the violation for proposers and responders, respectively, is needed to exclude the possibility that the proposers and responders have different positions in the economic distribution. Because we were unable to analyze the underlying neural mechanism in proposers due to the insufficient sample size, we excluded the published reports examining the proposers and reanalyzed the 17 remaining published reports to explore the brain networks of fairness enforcement and violation in the responders. The results consistently showed that fairness enforcement activated the striatum (caudate) and OFC, while the fairness violation activated the IA and DACC, supporting the current conclusion (Supplementary Figure S1 and Supplementary Tables S1 and S2). Finally, the meta-analysis of imaging results allowed us to test some hypotheses proposed in previous studies. Gender-specific activation of the neural reward system, including the striatum, was reported during decision-making processes in the DG (Soutschek et al., 2017). When a sufficient number of studies specifically involving female or male participants are available, an informative assessment would be to test whether this gender difference in fairness behavior is supported by a meta-analysis of imaging results. Similarly, other interesting topics, such as gender difference, personality characteristics, development and others, would also be explored using a meta-analysis to improve our understanding of the neural mechanisms of cooperative behavior from multiple perspectives.

Conclusions

Our meta-analysis directly profits from the rich scientific legacy of theoretical and experimental disciplines to comprehensively capture the neural processes underlying cooperation. Based on our results, cooperation enforcement activated reward-related brain areas, such as the striatum and OFC, only during other-cooperation. In contrast, cooperation violation evoked the activation of a brain network related to aversive feelings and cognitive conflict between self-interest and social norms *both from the perspective of self and others*. Moreover, cooperation is likely mainly driven by avoiding the negative emotions aroused by noncooperation, rather than pursuing the positive emotions aroused by cooperation. Therefore, the results of the current meta-analysis were not consistent with the warm glow theory but support the inequity aversion theory, namely, inequity aversion is the main driver of cooperative behavior.

Author Contributions

Q.L. designed and organized the study. X.L. supervised the study. Z.Y. performed the literature search, data extraction and manuscript preparation. Y.Z. participated in performing the literature search and data extraction. G.C.Y. revised the manuscript and provided technical support for the data analysis.

Supplementary data

Supplementary data are available at *SCAN* online.

Conflict of interest

None declared.

References

Andreoni, J. (1990). Impure altruism and donations to public goods: a theory of warm-glow giving. *Economic Journal*, **100**(401), 464–77.

Aoki, R., Matsumoto, M., Yomogida, Y., *et al.* (2014). Social equality in the number of choice options is represented in the ventromedial prefrontal cortex. *Journal of Neuroscience*, **34**(18), 6413–21. doi: 10.1523/jneurosci.4427-13.2014.

Archibugi, D., Pianta, M. (1992). *The Technological Specialization of Advanced Countries: A Report to the EEC on International Science and Technology Activities*, Brussels-Luxembourg: Springer Science & Business Media.

Baez-Mendoza, R., Schultz, W. (2013). The role of the striatum in social behavior. *Frontiers in Neuroscience*, **7**, 14.

Baston, C. (1998). *Altruism and Prosocial Behavior G*, 4th edn, New York: McGraw-Hill.

Bault, N., Joffily, M., Rustichini, A., Coricelli, G. (2011). Medial prefrontal cortex and striatum mediate the influence of social comparison on the decision process. *Proceedings of the National Academy of Sciences*, **108**(38), 16044–9.

Bellucci, G., Chernyak, S.V., Goodyear, K., Eickhoff, S.B., Krueger, F. (2017). Neural signatures of trust in reciprocity: a coordinate-based meta-analysis. *Human Brain Mapping*, **38**(3), 1233–1248. doi: 10.1002/hbm.23451.

Berg, J., Dickhaut, J., McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, **10**(1), 122–42.

Bolton, G.E., Ockenfels, A. (2000). ERC: a theory of equity, reciprocity, and competition. *American Economic Review*, **90**(1), 166–93.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R. (2009). Why perform a meta-analysis. In: *Introduction to Meta-Analysis*, Chichester: John Wiley & Sons, Ltd, pp. 9–14.

Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., Cohen, J.D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, **108**(3), 624–52.

Brett, M. (1999). The MNI brain and the Talairach atlas, Cambridge Imagers. Available: http://www.mrc-cbu.cam.ac.uk/Imaging/mnispace.html.

Cappelen, A.W., Eichele, T., Hugdahl, K., Specht, K., Sorensen, E.O., Tungodden, B. (2014). Equity theory and fair inequality: a neuroeconomic study. *Proceedings of the National Academy of Sciences of the United States of America*, **111**(43), 15368–72.

Chang, L.J., Sanfey, A.G. (2013). Great expectations: neural computations underlying the use of social norms in decision-making. *Social Cognitive and Affective Neuroscience*, **8**(3), 277–84.

Cheng, X., Zheng, L., Li, L., *et al.* (2015). Power to punish norm violations affects the neural processes of fairness-related decision making. *Frontiers in Behavioral Neuroscience*, **9**, 344.

Civai, C., Crescentini, C., Rustichini, A., Rumiati, R.I. (2012). Equality versus self-interest in the brain: differential roles of anterior insula and medial prefrontal cortex. *NeuroImage*, **62**(1), 102–12.

Colman, A.M. (2009). Cooperation: the political psychology of effective human interaction. *Applied Cognitive Psychology*, **23**(4), 599–600.

Corradi-Dell'Acqua, C., Tusche, A., Vuilleumier, P., Singer, T. (2016). Cross-modal representations of first-hand and vicarious pain, disgust and fairness in insular and cingulate cortex. *Nature Communications*, **7**, 10904.

Critchley, H.D., Elliott, R., Mathias, C.J., Dolan, R.J. (2000). Neural activity relating to generation and representation of galvanic skin conductance responses: a functional magnetic resonance imaging study. *Journal of Neuroscience*, **20**(8), 3033–40.

Decety, J., Jackson, P.L., Sommerville, J.A., Chaminade, T., Meltzoff, A.N. (2004). The neural bases of cooperation and competition: an fMRI investigation. *NeuroImage*, **23**(2), 744–51.

Eickhoff, S.B., Etkin, A. (2016). Going beyond finding the "lesion": a path for maturation of neuroimaging. *American Journal of Psychiatry*, **173**(3), 302–3.

Eickhoff, S.B., Bzdok, D., Laird, A.R., Kurth, F., Fox, P.T. (2012). Activation likelihood estimation meta-analysis revisited. *NeuroImage*, **59**(3), 2349–61.

Eickhoff, S.B., Laird, A.R., Fox, P.M., Lancaster, J.L., Fox, P.T. (2017). Implementation errors in the GingerALE software: description and recommendations. *Human Brain Mapping*, **38**(1), 7–11.

Eklund, A., Nichols, T.E., Knutsson, H. (2016). Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America*, **113**(28), 7900–5.

Elliott, R., Newman, J.L., Longe, O.A., Deakin, J.F.W. (2003). Differential response patterns in the striatum and orbitofrontal cortex to financial reward in humans: a parametric functional magnetic resonance imaging study. *Journal of Neuroscience*, **23**(1), 303–7.

Fareri, D.S., Delgado, M.R. (2014). Social rewards and social networks in the human brain. *The Neuroscientist*, **20**(4), 387–402.

Fareri, D.S., Chang, L.J., Delgado, M.R. (2015). Computational substrates of social value in interpersonal collaboration. *Journal of Neuroscience*, **35**(21), 8170–80.

Farrer, C., Frith, C.D. (2002). Experiencing oneself vs another person as being the cause of an action: the neural correlates of the experience of agency. *NeuroImage*, **15**(3), 596–603.

Fehr, E., Gachter, S. (2002). Altruistic punishment in humans. *Nature*, **415**(6868), 137–40.

Fehr, E., Schmidt, K.M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, **114**(3), 817–68.

Fouragnan, E., Chierchia, G., Greiner, S., Neveu, R., Avesani, P., Coricelli, G. (2013). Reputational priors magnify striatal responses to violations of trust. *Journal of Neuroscience*, **33**(8), 3602–11.

Fox, P.T., Laird, A.R., Eickhoff, S.B., *et al.* (2013). *User manual for GingerALE 2.3*, San Antonio, TX: UT Health Science Center San Antonio.

Goldin, P.R., McRae, K., Ramel, W., Gross, J.J. (2008). The neural bases of emotion regulation: reappraisal and suppression of negative emotion. *Biological Psychiatry*, **63**(6), 577–86.

Gromann, P.M., Shergill, S.S., de Haan, L. et al.(2014). Reduced brain reward response during cooperation in first-degree

relatives of patients with psychosis: an fMRI study. *Psychological Medicine*, **44**(16), 3445–54.

Gulati, R. (1995). Social structure and Alliance formation patterns: a longitudinal analysis. *Administrative Science Quarterly*, **40**(4), 619–52.

Günther, V., Zimmer, J., Kersting, A., Hoffmann, K.-T., Lobsien, D., Suslow, T. (2017). Automatic processing of emotional facial expressions as a function of social anhedonia. *Psychiatry Research: Neuroimaging*, **270**, 46–53.

Guroglu, B., van den Bos, W., Rombouts, S.A., Crone, E.A. (2010). Unfair? It depends: neural correlates of fairness in social context. *Social Cognitive and Affective Neuroscience*, **5**(4), 414–23.

Haber, S.N., Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology*, **35**, 4–26.

Hamilton, W.D. (1964). The genetical evolution of social behaviour. II. *Journal of Theoretical Biology*, **7**(1), 17–52.

Hampton, A.N., Bossaerts, P., O'Doherty, J.P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences of the United States of America*, **105**(18), 6741–6.

Harbaugh, W.T., Mayr, U., Burghart, D.R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science*, **316**(5831), 1622–5.

Hare, T.A., Camerer, C.F., Knoepfle, D.T., O'Doherty, J.P., Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *Journal of Neuroscience*, **30**(2), 583–90.

Harle, K.M., Chang, L.J., vant Wout, M., Sanfey, A.G. (2012). The neural mechanisms of affect infusion in social economic decision-making: a mediating role of the anterior insula. *NeuroImage*, **61**(1), 32–40.

Hein, G., Silani, G., Preuschoff, K., Batson, C.D., Singer, T. (2010). Neural responses to Ingroup and Outgroup Members' suffering predict individual differences in costly helping. *Neuron*, **68**(1), 149–60.

Hsu, M., Anen, C., Quartz, S.R. (2008). The right and the good: distributive justice and neural encoding of equity and efficiency. *Science*, **320**(5879), 1092–5.

Hughes, B.L., Ambady, N., Zaki, J. (2016). Trusting outgroup, but not ingroup members, requires control: neural and behavioral evidence. *Social Cognitive and Affective Neuroscience*, **12**(3), 372–81.

Iacoboni, M., Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience*, **7**(12), 942–51.

Iacoboni, M., Mazziotta, J.C. (2007). Mirror neuron system: basic findings and clinical applications. *Annals of Neurology*, **62**(3), 213–8.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C.F., Quartz, S.R., Montague, P.R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science*, **308**(5718), 78–83.

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., Fernández, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron*, **61**(1), 140–51.

Kollock, P. (1998). Social dilemmas: the anatomy of cooperation. *Annual Review of Sociology*, **24**(1), 183–214.

Krajbich, I., Adolphs, R., Tranel, D., Denburg, N.L., Camerer, C.F. (2009). Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *Journal of Neuroscience*, **29**(7), 2188–92.

Kringelbach, M.L. (2005). The human orbitofrontal cortex: linking reward to hedonic experience. *Nature Reviews Neuroscience*, **6**(9), 691–702.

Kringelbach, M.L., Rolls, E.T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Progress in Neurobiology*, **72**(5), 341–72.

Lamm, C., Batson, C.D., Decety, J. (2007). The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience*, **19**(1), 42–58.

Liberati, A., Altman, D.G., Tetzlaff, J., *et al.* (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine*, **6**(7), e1000100.

Liu, T., Saito, H., Oi, M. (2015). Role of the right inferior frontal gyrus in turn-based cooperation and competition: a near-infrared spectroscopy study. *Brain and Cognition*, **99**, 17–23.

Maumet, C., Nichols, T., E. (2015). Do the units matter? Validity of intensity based meta-analysis in the presence of unit mismatches. In: 21st annual meeting of the Organization for Human Brain Mapping (OHBM), Honolulu, United States. http://www.hal.inserm.fr/inserm-01149475.

Meltzoff, A.N., Decety, J. (2003). What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **358**(1431), 491–500.

Miedl, S.F., Fehr, T., Meyer, G., Herrmann, M. (2010). Neurobiological correlates of problem gambling in a quasi-realistic blackjack scenario as revealed by fMRI. *Psychiatry Research: Neuroimaging*, **181**(3), 165–73.

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., Grafman, J. (2006). Human fronto–mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(42), 15623–8.

Nioche, C., Cabanis, E.A., Habas, C. (2009). Functional connectivity of the human red nucleus in the brain resting state at 3T. *American Journal of Neuroradiology*, **30**(2), 396–403.

Nowak, M.A., Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature*, **393**(6685), 573–7.

O'Doherty, J.P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Current Opinion in Neurobiology*, **14**(6), 769–76.

O'Doherty, J.P. (2007). Lights, camembert, action! The role of human orbitofrontal cortex in encoding stimuli, rewards, and choices. *Annals of the New York Academy of Sciences*, **1121**(1), 254–72.

O'Doherty, J.P., Kringelbach, M.L., Rolls, E.T., Hornak, J., Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience*, **4**(1), 95–102.

Padoa-Schioppa, C., Assad, J.A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, **441**(7090), 223–6.

Phan, K.L., Sripada, C.S., Angstadt, M., McCabe, K. (2010). Reputation for reciprocity engages the brain reward center. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(29), 13099–104.

Piliavin, J.A., Charng, H.-W. (1990). Altruism: a review of recent theory and research. *Annual Review of Sociology*, **16**(1), 27–65.

Pochon, J.-B., Riis, J., Sanfey, A.G., Nystrom, L.E., Cohen, J.D. (2008). Functional imaging of decision conflict. *Journal of Neuroscience*, **28**(13), 3468–73.

Poundstone, W. (1993). *Prisoner's Dilemma/John von Neumann, Game Theory and the Puzzle of the*, Anchor: Bomb.

Rilling, J.K., Sanfey, A.G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology*, **62**, 23–48.

Rilling, J.K., Gutman, D.A., Zeh, T.R., Pagnoni, G., Berns, G.S., Kilts, C.D. (2002). A neural basis for social cooperation. *Neuron*, **35**(2), 395–405.

Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., Cohen, J.D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, **300**(5626), 1755–8.

Schulte-Rüther, M., Markowitsch, H.J., Fink, G.R., Piefke, M. (2007). Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: a functional magnetic resonance imaging approach to empathy. *Journal of Cognitive Neuroscience*, **19**(8), 1354–72.

Schultz, W., Tremblay, L., Hollerman, J.R. (2000). Reward processing in primate orbitofrontal cortex and basal ganglia. *Cerebral Cortex*, **10**(3), 272–83.

Soutschek, A., Burke, C.J., Raja Beharelle, A., *et al.* (2017). The dopaminergic reward system underpins gender differences in social preferences. *Nature Human Behaviour*, **1**(11), 819–27.

Stallen, M., Sanfey, A.G. (2013). The cooperative brain. *The Neuroscientist*, **19**(3), 292–303.

Stanford, C.B., Bunn, H.T. (2001). *Meat-Eating and Human Evolution*, Oxford: Oxford University Press.

Strang, S., Park, S.Q. (2017). Human Cooperation and Its Underlying Mechanisms. In M. Wühr S. Krach (Eds.) *Social Behavior from Rodents to Humans: Neural Foundations and Clinical Implications (pp. 223−239)*. Cham: Springer International Publishing.

Sul, S., Tobler, P.N., Hein, G. et al.(2015). Spatial gradient in value representation along the medial prefrontal cortex reflects individual differences in prosociality. *Proceedings of the National Academy of Sciences of the United States of America*, **112**(25), 7851–6.

Suzuki, S., Niki, K., Fujisaki, S., Akiyama, E. (2011). Neural basis of conditional cooperation. *Social Cognitive and Affective Neuroscience*, **6**(3), 338–47.

Tabibnia, G., Lieberman, M.D. (2007). Fairness and cooperation are rewarding: evidence from social cognitive neuroscience. *Annals of the New York Academy of Sciences*, **1118**, 90–101.

Tabibnia, G., Satpute, A.B., Lieberman, M.D. (2008). The sunny side of fairness - preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science*, **19**(4), 339–47.

Tremblay, L., Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. *Nature*, **398**(6729), 704–8.

Tricomi, E., Rangel, A., Camerer, C.F., O'Doherty, J.P. (2010). Neural evidence for inequality-averse social preferences. *Nature*, **463**(7284), 1089–91.

Van der Werf, Y.D., Witter, M.P., Groenewegen, H.J. (2002). The intralaminar and midline nuclei of the thalamus. Anatomical and functional evidence for participation in processes of arousal and awareness. *Brain Research Reviews*, **39**(2–3), 107–40.

Ven, A.H.V.D., Sapienza, H.J., Villanueva, J. (2007). Entrepreneurial pursuits of self- and collective interests. *Strategic Entrepreneurship Journal*, **1**(3–4), 353–70. doi: 10.1002/sej.34

Vostroknutov, A., Tobler, P.N., Rustichini, A. (2012). Causes of social reward differences encoded in human brain. *Journal of Neurophysiology*, **107**(5), 1403–12.

Wang, K.S., Smith, D.V., Delgado, M.R. (2016). Using fMRI to study reward processing in humans: past, present, and future. *Journal of Neurophysiology*, **115**(3), 1664–78.

Wang, K.C., Yang, J., Qiu, J. (2013). Spatiotemporal cortical activation underlying reward processing under "worse off than some, better off than many" comparison: an ERP study. *International Journal of Psychology and Behavioral Sciences*, **3**(5), 123–30.

Watanabe, M. (1999). Attraction is relative not absolute. *Nature*, **398**(6729), 661–3.

Weiland, S., Hewig, J., Hecht, H., Mussel, P., Miltner, W.H. (2012). Neural correlates of fair behavior in interpersonal bargaining. *Social Neuroscience*, **7**(5), 537–51.

Winkler, A.D., Hu, S., Li, C.-s. R. (2013). The influence of risky and conservative mental sets on cerebral activations of cognitive control. *International Journal of Psychophysiology*, **87**(3), 254–61.

Yan, C., Yang, T., Yu, Q.-j., Jin, Z., Cheung, E. F. C., Liu, X., & Chan, R. C. K. (2015). Rostral medial prefrontal dysfunctions and consummatory pleasure in schizophrenia: a meta-analysis of functional imaging studies. *Psychiatry Research: Neuroimaging*, **231**(3), 187–96.

Zaki, J., Mitchell, J.P. (2011). Equitable decision making is associated with neural markers of intrinsic value. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(49), 19761–6.