Distinguishing neural correlates of context-dependent advantageous- and disadvantageous-inequity aversion

Xiaoxue Gao (高晓雪)[a,b,c,d,1], Hongbo Yu (于宏波)[a,b,e,f,1], Ignacio Sáez[g,h], Philip R. Blue[a,b,c,d], Lusha Zhu (朱露莎)[b,c,d,i,j], Ming Hsu (许明)[g,h], and Xiaolin Zhou (周晓林)[a,b,c,d,j,k,2]

[a]Center for Brain and Cognitive Sciences, Peking University, Beijing 100871, China; [b]School of Psychological and Cognitive Sciences, Peking University, Beijing 100871, China; [c]Key Laboratory of Machine Perception, Ministry of Education, Peking University, Beijing 100871, China; [d]Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, China; [e]Department of Psychology, Yale University, New Haven, CT 06520; [f]Department of Experimental Psychology, University of Oxford, Oxford OX2 6GG, United Kingdom; [g]Haas School of Business, University of California, Berkeley, CA 94720; [h]Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720; [i]Peking–Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China; [j]Peking University–IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, China; and [k]Institute of Psychological and Brain Sciences, Zhejiang Normal University, Zhejiang 321004, China

Edited by Ernst Fehr, University of Zurich, Zurich, Switzerland, and accepted by Editorial Board Member Michael S. Gazzaniga July 9, 2018 (received for review February 10, 2018)

Humans can integrate social contextual information into decision-making processes to adjust their responses toward inequity. This context dependency emerges when individuals receive more (i.e., advantageous inequity) or less (i.e., disadvantageous inequity) than others. However, it is not clear whether context-dependent processing of advantageous and disadvantageous inequity involves differential neurocognitive mechanisms. Here, we used fMRI to address this question by combining an interactive game that modulates social contexts (e.g., interpersonal guilt) with computational models that enable us to characterize individual weights on inequity aversion. In each round, the participant played a dot estimation task with an anonymous coplayer. The coplayer would receive pain stimulation with 50% probability when either of them responded incorrectly. At the end of each round, the participant completed a variant of dictator game, which determined payoffs for him/herself and the coplayer. Computational modeling demonstrated the context dependency of inequity aversion: when causing pain to the coplayer (i.e., guilt context), participants cared more about the advantageous inequity and became more tolerant of the disadvantageous inequity, compared with other conditions. Consistently, neuroimaging results suggested the two types of inequity were associated with differential neurocognitive substrates. While the context-dependent processing of advantageous inequity was associated with social- and mentalizing-related processes, involving left anterior insula, right dorsolateral prefrontal cortex, and dorsomedial prefrontal cortex, the context-dependent processing of disadvantageous inequity was primarily associated with emotion- and conflict-related processes, involving left posterior insula, right amygdala, and dorsal anterior cingulate cortex. These results extend our understanding of decision-making processes related to inequity aversion.

advantageous inequity | disadvantageous inequity | guilt context | insula | fMRI

Inequity aversion, or the preference for fairness, is an other-regarding preference observed widely in human society (1, 2). Individuals can be averse to inequity both when they receive more (i.e., advantageous inequity) and when they receive less (i.e., disadvantageous inequity) than others (2). The distinction between these two types of inequity aversion has been demonstrated in different disciplines. Behavioral studies showed that, although individuals dislike both types of inequity, their responses to advantageous inequity are usually not as strong as to disadvantageous inequity (2–4). Both evolutionary and developmental evidence demonstrated variations in the onset of the two types of inequity aversion. Disadvantageous-inequity aversion emerges at early stages of evolution and of human development, whereas advantageous-inequity aversion has only been seen in chimpanzees (5) and humans above 8 y old, who have relatively mature social and cognitive control abilities (6). These findings provide a theoretical motivation for investigating potentially differential psychological and neural mechanisms underpinning the two types of inequity aversion. Increased knowledge of the psychological and neural bases underlying individuals' attitudes toward inequity can provide valuable clues for understanding various social and economic phenomena, such as the asymmetrical responses to inequity when individuals are in advantageous vs. disadvantageous status in financial crises (3, 7). However, despite extensive research on disadvantageous inequity, little is known about advantageous inequity and whether these two types of inequity involve differential psychological and neural mechanisms.

Previous evidence has linked disadvantageous-inequity aversion with negative emotions (e.g., envy) elicited by receiving less than others (2, 6). Specifically, a number of studies have investigated the psychological and neural mechanisms of disadvantageous-inequity aversion using ultimatum game (UG), in which participants decide whether to accept a fair or unfair (i.e., disadvantageous) division of money suggested by a proposer (8). Participants'

Significance

Despite extensive research on disadvantageous inequity, little is known about advantageous inequity and whether these two types of inequity involve differential neurocognitive mechanisms. We address these questions from the perspective of context dependency and suggest that these two types of inequity are associated with differential neurocognitive substrates, subserved by different brain regions and in particular by the spatial gradient in insular activity. Our findings shed light on how social contexts (i.e., interpersonal guilt) are integrated into social decision making and suggest that the resistance to unequal situations when individuals are in disadvantageous status may primarily stem from their emotional responses, whereas the resistance to unequal situations when individuals are in advantageous status may involve advanced cognitive functions such as mentalizing.

Author contributions: X.G., H.Y., I.S., P.R.B., L.Z., M.H., and X.Z. designed research; X.G. and H.Y. performed research; I.S., L.Z., and M.H. contributed new reagents/analytic tools; X.G. and H.Y. analyzed data; and X.G., H.Y., P.R.B., L.Z., M.H., and X.Z. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. E.F. is a guest editor invited by the Editorial Board.

Data deposition: The unthresholded statistical images reported in this paper have been deposited in NeuroVault, https://neurovault.org/collections/3859/ (collection 3859).

[1]X.G. and H.Y. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: xz104@pku.edu.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1802523115/-/DCSupplemental.

Published online July 30, 2018.

self-reported negative feelings increased as the divisions became increasingly unfair (9); they had stronger skin conductance responses, indicating higher emotional arousal, when they rejected, as opposed to accepted, disadvantageous divisions (9, 10). In line with these findings, neuroimaging studies have demonstrated that the brain structures associated with processing negative and aversive emotions, such as the amygdala and anterior insula (aINS), are involved in disadvantageous-inequity processing (8, 11–13).

Compared with disadvantageous inequity, responses to advantageous inequity may involve more complex processes. Although receiving more than others may benefit oneself economically in the short term, the resulting fairness norm violation could frustrate partners and damage one's reputation, which is a threat to cooperation and benefits in the long run (5). Thus, the decision to accept getting more than others may require not only value representation of the relative economic gain but also advanced social cognitions, such as anticipating others' reactions to being treated unfairly (i.e., mentalizing), recognizing this norm violation of themselves (i.e., norm violation detection), and adjusting this violation to bring in long-term cooperation and benefits (i.e., cognitive control) (5, 6). In line with this notion, previous studies suggested a codevelopment of advantageous-inequity aversion with the ability of cognitive control (6) and mentalizing (14) in human development. In contrast to the existence of disadvantageous-inequity aversion across species, advantageous-inequity aversion has thus far only been observed in human and nonhuman species with extensive cooperation outside of kinship relationships (5). Therefore, it is conceivable that the processing of advantageous inequity may require brain structures involved in mentalizing [e.g., dorsomedial prefrontal cortex (DMPFC)] (15), norm violation detection (e.g., aINS) (16–20), and cognitive control [e.g., dorsolateral prefrontal cortex (DLPFC)] (21), in addition to regions representing values of relative gain (e.g., ventral striatum) (22).

Based on the observed behavioral differences and the proposed differences in psychological mechanisms, we propose two hypotheses regarding the neural mechanisms of inequity processing (Fig. 1). Hypothesis 1 (Fig. 1*A*) postulates a shared system for processing both advantageous and disadvantageous inequity (e.g., the value representation system for processing general inequity). This hypothesis is analogous to the neural common currency hypothesis in neuroeconomics, which posits that reward value from various modalities is encoded by the same neural computational process (23, 24). The shared system for processing general inequity could be modulated by separate systems involved in advantageous-inequity–specific processing (e.g., mentalizing, norm compliance, and cognitive control) and disadvantageous-inequity–specific processing (e.g., emotional processing), resulting in different attitudes toward inequity when one is in advantageous status vs. disadvantageous status. Hypothesis 2 (Fig. 1*B*) assumes that advantageous inequity and disadvantageous inequity rely on distinct psychological and neural mechanisms. Comparing the neural activations associated with advantageous and disadvantageous inequity can help tease apart these two hypotheses: if nonoverlapping brain networks are observed, it indicates differential neurocognitive substrates underlying these two types of inequity (hypothesis 2); if overlapping brain networks are identified, this hints at, but not necessarily demonstrates, shared neurocognitive processes underlying these two form of inequity, as different neural activation patterns (or representations) could be encoded in overlapping univariate activation networks (25). In this case, further multivariate analyses, such as multivariate pattern analysis (MVPA) (26) and representational similarity analysis (RSA) (27), are needed to formally examine whether the overlapping brain networks exhibit shared or distinct neural representations.

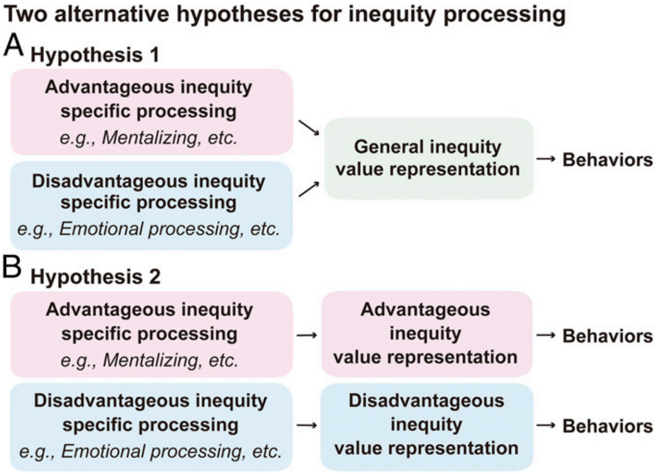
Two alternative hypotheses for inequity processing
A
Hypothesis 1
Advantageous inequity specific processing
e.g., Mentalizing, etc.
Disadvantageous inequity specific processing
e.g., Emotional processing, etc.
General inequity value representation
→ Behaviors
B
Hypothesis 2
Advantageous inequity specific processing
e.g., Mentalizing, etc.
Advantageous inequity value representation
→ Behaviors
Disadvantageous inequity specific processing
e.g., Emotional processing, etc.
Disadvantageous inequity value representation
→ Behaviors

Fig. 1. Two potential hypotheses regarding the psychological and neural mechanisms underlying inequity processing. (*A*) Hypothesis 1 postulates a shared system for processing both advantageous and disadvantageous inequity (e.g., the value representation system for processing general inequity). This hypothesis is analogous to the neural common-currency hypothesis in neuroeconomics, which posits that reward value from various modalities is encoded by the same neural computational process (23, 24). The shared system for processing general inequity could be modulated by separate systems involved in advantageous-inequity–specific processing (e.g., mentalizing, norm compliance, and cognitive control) and disadvantageous-inequity–specific processing (e.g., emotional processing), resulting in different attitudes toward inequity when one is in advantageous status vs. disadvantageous status. (*B*) Hypothesis 2 assumes that advantageous inequity and disadvantageous inequity rely on distinct psychological and neural mechanisms.

A few studies have investigated whether the processing of advantageous and disadvantageous inequity involve shared or distinct neural mechanisms (28–30). Given the inherent correlation between self-payoff and advantageous inequity and the correlation between other-payoff and disadvantageous inequity in experimental designs, the results of these studies are mixed. For example, one neuroimaging study (29) focused on participants' sharing decisions in advantageous and disadvantageous frames, in which participants were asked to choose between an equal split of money (1 coin for self and 1 coin for other) and an advantageous split (e.g., 2 coins for self and 0 coin for other) or a disadvantageous split (e.g., 1 coin for self and 2 coins for other). The brain activity for choosing the advantageous options in contrast to the brain activity for choosing the equal option was regarded as the neural correlates of advantageous inequity; similar analysis was conducted for disadvantageous inequity. However, it is hard to discern whether the brain activations revealed in these comparisons are driven by the "inequity," by the actual payoff, or by both.

While it is difficult to quantitatively dissociate the amount of self/other payoff and the degree of inequity, one way to dissociate the psychological and neural processes of payoff and inequity is to manipulate the social context in which the resource allocation occurs. If a certain social context modulates the salience of inequity but not the salience of absolute payoff, then manipulating such a context would allow us to observe changes in the processing of inequity but not the payoff (31). This logic is in line with the "separately modifiable" principle, which has been used to dissociate neurocognitive constructs that resemble each other psychologically and neurally (25, 32). Indeed, previous findings have provided the basis for us to apply this separately modifiable principle to the understanding of neurocognitive mechanisms underlying advantageous and disadvantageous inequity. On the one hand, the brain structures involved in nonsocial reward encoding (e.g., orbitofrontal cortex) are suggested to represent the relative/subjective value, which strongly depends on the context, rather than the absolute/objective value of an object (33). The existence of such contextual coding in nonsocial value and decision circuits provides a clue for the existence of similar value representation principles in social decision making

(34). On the other hand, similar to nonsocial valuation, individuals are capable of integrating social context-related information into decision-making processes to adjust their responses to inequity (4, 35–37). These flexible adjustments take place regularly in various contexts in everyday life (38), which enable us to maintain cooperative relationships, maximize personal welfare, and adapt to dynamic social situations (39, 40). Moreover, individuals' attitudes (or subjective values) toward advantageous and disadvantageous inequity may vary differently according to contexts. For example, when distributing resources as a dictator, individuals tend to avoid advantageous inequity when interacting with cooperative others (e.g., friends or neighbors) but are more tolerant of advantageous inequity when interacting with competitive others (e.g., competitors or salesmen). In contrast, the context change has no effect on disadvantageous-inequity aversion (4).

Investigating the neural correlates of advantageous and disadvantageous inequity from the perspective of context dependency requires a social context that modulates the salience of these two types of inequity simultaneously. One such context is the interpersonal guilt, an emotional state associated with the awareness of causing harm to a victim or violating perceived norms (41). Guilt is closely related to inequity aversion because it both signals and constitutes the obligation of wrongdoers to balance the inequity created by their wrongdoing or transgression (41–43). On the one hand, obtaining more or suffering less than the others is an important source of guilt (41, 44). In dictator game (DG), the extent to which a dictator is averse to advantageous inequity is regarded in some decision theories as reflecting anticipatory guilt (2, 45, 46). On the other hand, the experience of guilt motivates conciliatory gestures toward victims, which aim at reducing inequity and restoring the relationship back to an even footing (41, 47). Previous studies have shown that individuals show increased generosity to the victims when feeling guilty (47–49). This increased generosity may result from increased advantageous and/or decreased disadvantageous-inequity aversion when individuals are making decisions in a state of guilt (50).

We first tested this proposal in two behavioral studies. In each round, the participant played a dot estimation task with an anonymous coplayer (i.e., a confederate) who would receive pain stimulation with 50% probability when either the coplayer or the participant him/herself responded incorrectly (Fig. 2*A*). At the end of each round, the participant acted as the dictator and completed a continuous version (50) (experiment 1; Fig. 2*B*) or a binary choice version of DG (experiment 2; Fig. 2*C*), which determined the payoffs for him/herself and the coplayer (*Materials and Methods*). The DG gave the participant a chance to compensate for the coplayer in this trial. The participant was told that the coplayer did not know the existence of DG and could not see the DG choices during the experiment. Incorrect trials in the dot estimation task formed a 2 (Agent who performed dot estimation task: Self vs. Other) by 2 (Outcome for the coplayer: Pain vs. Nopain) within-participant design. The Self_Pain condition was the critical condition to induce guilt. The other three conditions controlled for confounding factors, such as empathy for the coplayer and regret for estimating incorrectly. Thus, the interaction between Agent and Outcome was the guilt effect that we focused on. We used the Fehr–Schmidt inequity aversion model (2, 50) to estimate the weights on advantageous (parameter α) and disadvantageous (parameter β) inequity aversion during DG for each of the four conditions. This model posits that when making resource allocation individuals trade self-interest against the two forms of inequity aversion. While the continuous version of DG in experiment 1 enabled us to conduct model fitting at the group level, the binary choice version of DG in experiment 2 enabled us to conduct model fitting at the individual level. We further investigated whether there are differential

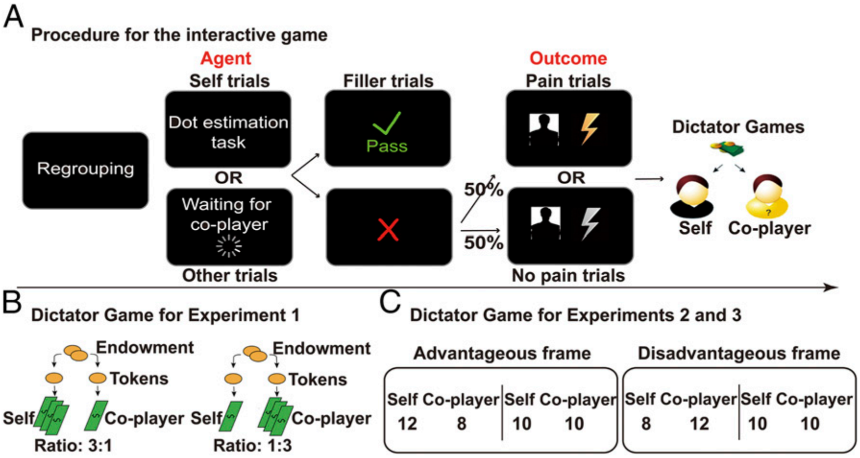

Fig. 2. Task display. (*A*) Each trial began by informing the participants that they were (randomly and anonymously) paired with one of three coplayers (i.e., confederates). In one-half of the trials, the participant performed a dot estimation task (Self trials), and in the other half of the trials he/she waited for his/her coplayer to make the estimation (Other trials). If the answer was correct, no one would receive pain stimulation, and the current trial terminated. If either of them responded incorrectly, the coplayer in the current trial had a 50% probability of receiving the pain stimulation (Pain trials and Nopain trials). At the end of each incorrect trial, the participant would act as a dictator in the dictator game (DG) to determine the payoffs for him/herself and the coplayer. This DG gave the participant a chance to compensate for the coplayer in this trial. This formed a 2 (Agent who performed dot estimation task: Self vs. Other) by 2 (Outcome for the coplayer: Pain vs. Nopain) within-participant design. The two independent variables were highlighted in red. The Self_Pain condition was the critical condition to induce guilt. The other three conditions controlled for the confounding factors, such as empathy for the coplayer and regret for wrong estimation. Thus, the interaction between Agent and Outcome [i.e., (Self_Pain − Other_Pain) > (Self_Nopain − Other_Nopain)] was the guilt effect that we focused on. Two versions of the DG were used. (*B*) In experiment 1, for each choice, the participant received an endowment and could unilaterally choose to give any integer amount of tokens (from 0 to the amount of endowment) to the coplayer. The relative cost and benefit of giving were manipulated by independently varying how much each token was worth to the participant and the coplayer (i.e., the exchange ratio) (50). There was no time limitation for each choice. (*C*) A binary choice version of DG was developed for experiments 2 and 3 to dissociate advantageous and disadvantageous frames. Each binary choice consisted of two options representing the payoffs that the participant and the coplayer would earn. One option was always "10 points for me, and 10 points for the coplayer," and the other option was an unequal option with different values, varied systematically across trials. The participants needed to make each choice within 4 s. Two types of unequal options were implemented corresponding to advantageous and disadvantageous inequity. In both versions of DG, one trial was selected randomly and actualized after the experiment, determining the final payoffs for the participant and the coplayer.

neural mechanisms underlying context-dependent advantageous- and disadvantageous-inequity aversion in experiment 3, in which participants performed the same task as experiment 2 in the fMRI scanner.

Results

Dissociable Contextual Effects on Advantageous- and Disadvantageous-Inequity Aversion at the Behavioral Level. For all of the three experiments, 2 (Agent: Self vs. Other) × 2 (Outcome: Pain vs. Nopain) repeated-measures ANOVAs on the self-reported guilt ratings in the postscan questionnaire yielded significant interactions between Agent and Outcome (Fig. 3*A* and *SI Appendix*, Table S3). Participants felt guiltier when they themselves inflicted the pain upon the coplayers (Self_Pain) than when the coplayers themselves inflicted the pain (Other_Pain); this Agent effect was decreased in Nopain conditions (i.e., Self_Nopain vs. Other_Nopain), demonstrating the robustness and validity of our paradigm to induce guilt.

To test whether our context manipulation modulated individuals' advantageous- and disadvantageous-inequity aversion, we used the Fehr–Schmidt inequity aversion model (2, 50) to capture individuals' weights on advantageous (α) and disadvantageous (β) inequity aversion for each of the four conditions, at both the group level (experiments 1–3) and the individual level (experiments 2 and 3) (*Materials and Methods*). The Fehr–Schmidt inequity aversion model explained participants' choices significantly better than six other plausible models (*SI Appendix*, *SI Methods* and Table S11). These results indicate that participants' behavioral changes in the guilt context were derived, to a large extent, from increased advantageous-inequity aversion and decreased disadvantageous-inequity aversion, but not from increased subjective values of other-payoff, changes in participants' attitudes toward inequity aversion in general, or changes in their perceived fairness norms.

Both group-level model fitting in the three experiments (Fig. 3 *B*–*D*) and individual-level model fitting (*SI Appendix*, Fig. S1) in experiments 2 and 3 demonstrated that, compared with the Other_Pain condition, participants' α increased and β decreased in the Self_Pain condition, while these effects were absent or decreased in the Nopain conditions (i.e., Self_Nopain vs. Other_Nopain) (*SI Appendix*, Tables S3 and S4). Patterns of model-free results in all of the three experiments were consistent with the model-based results (*SI Appendix*, *SI Results* and Fig. S2).

To test whether the contextual effects on advantageous and disadvantageous inequity are dissociable at the behavioral level, we examined the correlation between the contextual effects (i.e., the interactions between Agent and Outcome) on α and β in both experiments 2 and 3. To this end, we estimated α and β for each participant in each condition, and estimated the contextual effect [i.e., (Self_Pain − Other_Pain) − (Self_Nopain − Other_Nopain)] on α and β for each participant. Results showed that the contextual effects on the two types of inequity aversion were uncorrelated (experiment 2: $r = -0.023$, $P = 0.911$; experiment 3: $r = -0.152$, $P = 0.479$), indicating that the individuals with higher contextual effects on α did not necessarily exhibit higher or lower contextual effects on β (Fig. 3 *E* and *F*).

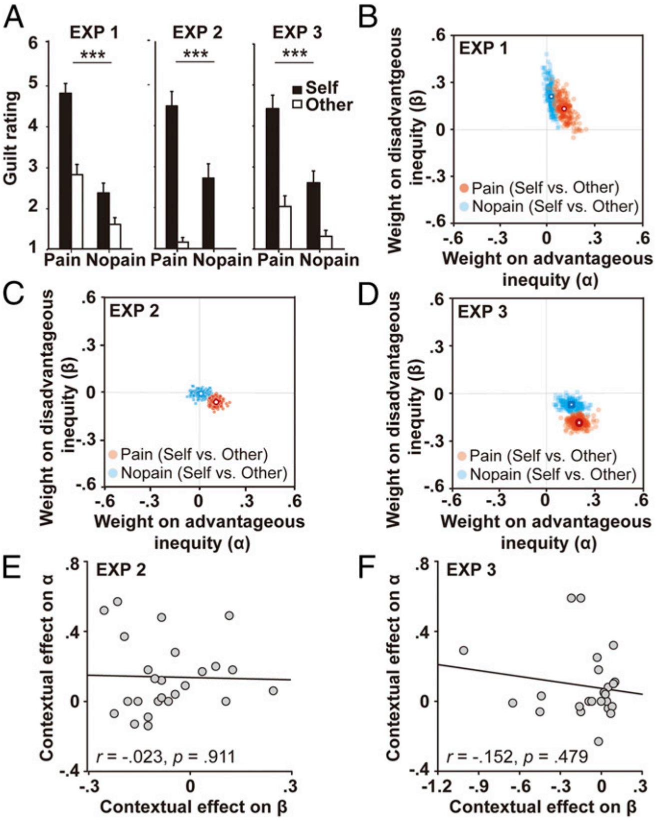

Fig. 3. Behavioral results. (*A*) Significant 2 (Agent: Self vs. Other) × 2 (Outcome: Pain vs. Nopain) interaction effects were observed for postscan guilt ratings in all of the three experiments. Participants felt guiltier when they themselves inflicted the pain upon the coplayers (Self_Pain) than when the coplayers themselves inflicted the pain (Other_Pain); this Agent effect was decreased in Nopain conditions (Self_Nopain vs. Other_Nopain), demonstrating the robustness and validity of our paradigm to induce guilt. (*B*–*D*) Group-level model-based results for participants' choices during DG in experiments 1, 2, and 3, respectively. The *x* axis and *y* axis represent the weight on advantageous inequity (α) and the weight on disadvantageous inequity (β), respectively. The red dots represent the difference between bootstrap pseudosample estimates (*Materials and Methods*) for the Self_Pain condition and the Other_Pain condition, while the blue dots represent the difference between bootstrap pseudosample estimates for the Self_Nopain condition and the Other_Nopain condition. Thus, the location of red dots relative to blue dots captures the interaction effect between Agent and Outcome [i.e., the guilt effect: (Self_Pain − Other_Pain) > (Self_Nopain − Other_Nopain)] on α and β. In all three experiments, red dots move down–right relative to the blue dots (i.e., increased α and decreased β), indicating that when participants felt guilty, their advantageous-inequity aversion increased and their disadvantageous-inequity aversion decreased, compared with other conditions. (*E* and *F*) Individual-level model-based results for participants' choices during DG in experiments 2 and 3, respectively. The *x* axis represents the contextual effect [i.e., (Self_Pain − Other_Pain) − (Self_Nopain − Other_Nopain)] on individual weight on disadvantageous inequity (β). The *y* axis represents the contextual effect on individual weight on advantageous inequity (α). Results showed that the contextual effects on α and β were uncorrelated with each other in both experiments, suggesting that individuals with a higher contextual effect on α did not necessarily have the same trend for β. Error bars represent SEM. ***$P < 0.001$.

Neural Responses to Inequity in Advantageous and Disadvantageous Frames. To identify brain regions involved in advantageous- and disadvantageous-inequity processing, we classified binary choices shown to the participants into the advantageous frame and the disadvantageous frame according to the relative status of self-payoff compared with other-payoff implemented in the unequal option of each binary choice. Then in each frame, all of the binary choices were further median split into the high-inequity condition (HI) and the low-inequity condition (LI) according to the amount of self/other-payoff differences implemented in the unequal option of each binary choice (*SI Appendix*, Table S2). This procedure of choice classification was independent of participants' actual choices, resulting in a balanced design for fMRI analysis. In close correspondence to the behavioral analysis, we established hypotheses for fMRI analysis showing the potential response patterns for brain regions that were involved in advantageous- and disadvantageous-inequity aversion processing (Fig. 4). Specifically, given the increased advantageous-inequity aversion observed in the Self_Pain condition, we hypothesized that brain regions involved in context-dependent processing of advantageous inequity would show greater sensitivity to advantageous inequity in this condition than in the other three conditions, which would in turn result in boosted activity differences between HI and LI conditions in these regions (Fig. 4*A*). Thus, the activity of these brain regions would show significant Agent × Outcome × Inequity level three-way interactions [(Self_Pain_HI > Self_Pain_LI) − (Self_Nopain_HI > Self_Nopain_LI) − (Other_Pain_HI > Other_Pain_LI) + (Other_Nopain_HI > Other_Nopain_LI)] in the advantageous frame. Similarly, brain regions in context-dependent processing of disadvantageous inequity would show less sensitivity to disadvantageous inequity in the Self_Pain condition than in the other three conditions, which would in turn result in decreased activity difference between HI and LI conditions in these regions (Fig. 4*B*). Thus, the activity of these regions would also show significant Agent × Outcome × Inequity level three-way interactions, but the direction of this effect would be opposite to that in the advantageous frame [(Self_Pain_LI > Self_Pain_HI) − (Self_Nopain_ LI > Self_Nopain_ HI) − (Other_Pain_ LI > Other_Pain_ HI) + (Other_Nopain_ LI > Other_Nopain_ HI)]. Based on these hypotheses, we focused on the neural responses during the decision phase at which participants decided self- and other-payoffs and established general linear model 1 (GLM1) to reveal brain regions that were separately involved in context-dependent processing of advantageous- and disadvantageous-inequity aversion (*SI Appendix*, *SI Methods*).

For the advantageous frame, significant activations were found in the left aINS [−30, 21, −20], rDLPFC [39, 20, 37], and DMPFC [−12, 47, 40] (Fig. 5*A*, *I* and *VI*, and *SI Appendix*, Table S7). Compared with other conditions, when the participant inflicted pain upon the coplayer, the activity difference between HI and LI choices on left aINS, rDLPFC, and DMPFC increased in the advantageous frame (Fig. 5*A*, *II*, *IV*, and *VII*, and *SI Appendix*, Fig. S3 *A*–*C*), indicating the increased sensitivity of these regions to advantageous inequity. This pattern on left aINS, rDLPFC, and DMPFC was absent for the disadvantageous frame (Fig. 5*A*, *III*, *V*, and *VIII*).

For the disadvantageous frame, we observed only activation in the left pINS [−42, −4, 7] (Fig. 5*B*, *I*, and *SI Appendix*, Table S7), which showed reduced sensitivity to disadvantageous inequity in the Self_Pain condition, relative to other conditions (Fig. 5*B*, *III*, and *SI Appendix*, Fig. S3*D*). The effect of left pINS was absent for the advantageous frame (Fig. 5*B*, *II*). Moreover, no cluster survived statistical correction in either the conjunction analysis for the two frames or if we combined the data from the two frames to conduct contrast analysis. Thus, different brain regions were involved in processing advantageous and disadvantageous frames: while left aINS, rDLPFC, and DMPFC were involved in the context-dependent processing of advantageous inequity, left pINS was involved in the context-dependent processing of disadvantageous inequity.

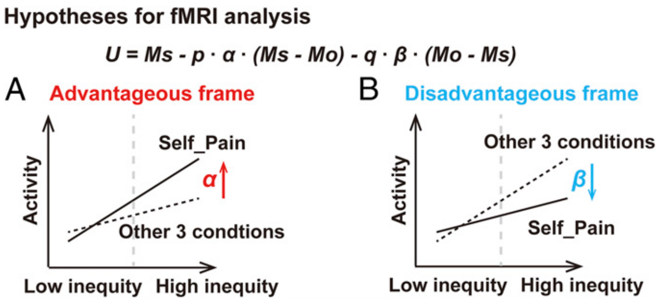

Fig. 4. Hypotheses for fMRI analysis. In close correspondence to the behavioral analysis, we established the hypotheses for fMRI analysis. (*A*) Given the increased advantageous-inequity aversion (*α*) observed in the Self_Pain condition, we hypothesized that brain regions involved in advantageous-inequity processing would show greater sensitivity to advantageous inequity in the Self_Pain condition than in other three conditions, which would in turn result in boosted activity difference between the high-inequity conditions (HI) and low-inequity conditions (LI) in these regions. Thus, the activity of these brain regions would show significant Agent by Outcome by Inequity level three-way interactions in the advantageous frame. (*B*) Similarly, given the decreased disadvantageous-inequity aversion (*β*) observed in the Self_Pain condition, brain regions in the processing of disadvantageous inequity would show less sensitivity to disadvantageous inequity in the Self_Pain condition than in the other three conditions, which would in turn result in decreased activity difference between high-inequity conditions (HI) and low-inequity conditions (LI) in these regions. Thus, the activity of these brain regions would also show significant Agent by Outcome by Inequity level three-way interactions, but the direction of these effects would be opposite to that in the advantageous frame.

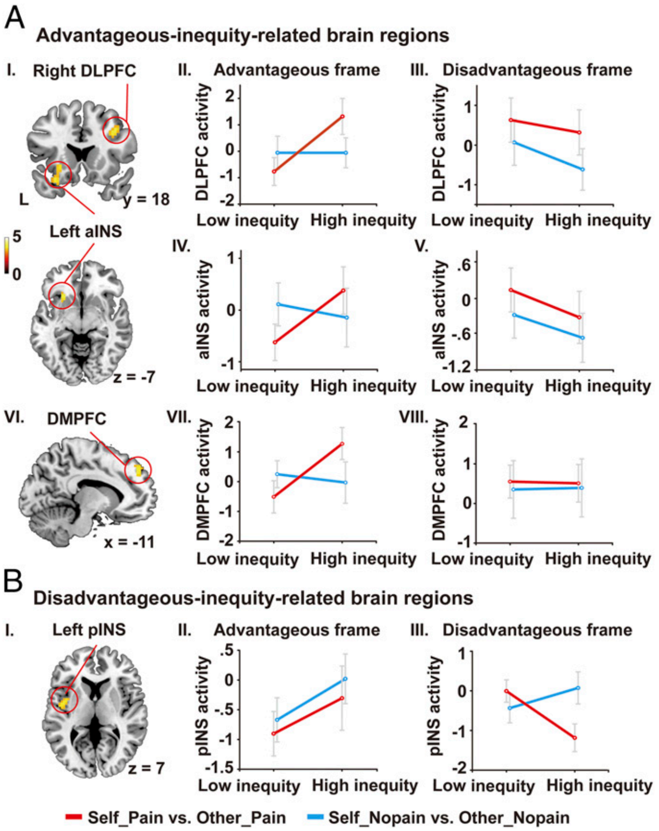

Fig. 5. Neural correlates of context-dependent advantageous- and disadvantageous-inequity processing. (*A*) Compared with other conditions, when the participant inflicted pain upon the coplayer (i.e., Self_Pain condition), the activity difference between high-inequity and low-inequity choices in left aINS, rDLPFC, and DMPFC (*I* and *VI*) increased in the advantageous frame, indicating increased sensitivity of these regions to advantageous inequity (*II*, *IV*, and *VII*). This pattern of effects was absent on left aINS, rDLPFC, and DMPFC for the disadvantageous frame (*III*, *V*, and *VIII*). (*B*) Compared with other conditions, when the participant inflicted pain upon the coplayer (i.e., Self_Pain condition), the activity difference between high-inequity and low-inequity choices in left pINS (*I*) decreased in the disadvantageous frame, indicating decreased sensitivity of these regions to disadvantageous inequity (*III*). The effect was absent for the advantageous frame (*II*). Error bars represent SEM.

Spatial Gradient Within Insula for Context-Dependent Advantageous-Inequity vs. Disadvantageous-Inequity Aversion Processing. To formally test whether the contextual effects on neural responses to advantageous and disadvantageous inequity were modulated by rostrocaudal position, we defined five regions of interest (ROIs) along the left anterior to posterior insula (Fig. 6*A*, *Left*) (51, 52). Specifically, in addition to the ROIs of the aINS and the pINS identified in the whole-brain analysis, we drew a straight line between these two ROIs and selected another three coordinates equally distributed along this line. The specific locations of each ROI on the coronal axis were slightly adjusted to ensure that

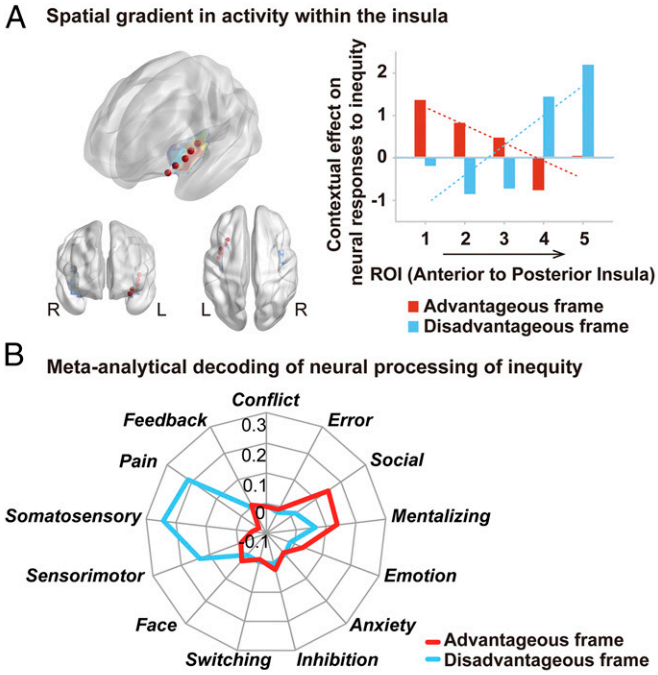

Fig. 6. (*A*) Spatial gradient for context-dependent inequity processing. Five regions of interest (ROIs) were defined along the axis from the left aINS to the left pINS (*Left*), which were identified in the whole-brain analysis for the advantageous and disadvantageous frames. For each ROI, we extracted the β estimates for each condition in the advantageous frame and the disadvantageous frame, respectively, using fMRI data without smoothing in data preprocessing (*Right*). In each frame and for each ROI, we computed the contextual effect on neural responses to inequity, taking it as the absolute value of the Agent by Outcome by Inequity level three-way interaction effect [(Self_Pain_HI > Self_Pain_LI) − (Self_Nopain_HI > Self_Nopain_LI) − (Other_Pain_HI > Other_Pain_LI) + (Other_Nopain_HI > Other_Nopain_LI)] (HI represents high-inequity condition, and LI represents low-inequity condition). We then put the values into a 2 (frame: Advantageous vs. Disadvantageous) × 5 (ROI locations) repeated-measures ANOVA. Results showed that the contextual effect on neural responses to inequity in the disadvantageous frame became stronger in more posterior ROIs, whereas the contextual effect on neural responses to inequity in the advantageous frame became stronger in more anterior ROIs. Dotted lines indicate linear fits of the spatial gradient for advantageous (red) and disadvantageous (blue) frames. (*B*) Metaanalytical decoding of neural processing of inequity. Results of metaanalytical decoding using Neurosynth Image Decoder (54) demonstrated that the processing of advantageous inequity is associated with terms related to "social" and "mentalizing," while the processing of disadvantageous inequity is associated primarily with "somatosensory," "pain," and "sensorimotor" terms.

these ROIs were located on the insula template of automated anatomical labeling (53) due to the curvy shape of the insular cortex (final MNI coordinates of the five ROIs: [−30, 20, −17], [−33, 14, −11], [−39, 8, −5], [−40, 2, 1], [−42, −4, 7]). For each ROI, we extracted the β estimates for each condition in the advantageous frame and the disadvantageous frame, respectively, using fMRI data without smoothing in data preprocessing. In each frame and for each ROI, we computed the contextual effect on neural responses to inequity, taking it as the absolute value of the Agent × Outcome × Inequity level three-way interaction effect [(Self_Pain_HI > Self_Pain_LI) − (Self_Nopain_HI > Self_Nopain_LI) − (Other_Pain_HI > Other_Pain_LI) + (Other_Nopain_HI > Other_Nopain_LI)]. We then put the values into a 2 (frame: Advantageous vs. Disadvantageous) × 5 (ROI locations) repeated-measures ANOVA. Results showed a clear spatial distinction in the context-dependent processing of advantageous- vs. disadvantageous-inequity aversion, with a significant interaction between frame and ROI location [$F_{(4,\ 100)} = 4.319$, $P = 0.003$] (Fig. 6*A*, *Right*). The contextual effect on neural responses to inequity in the disadvantageous frame became stronger in more posterior ROIs, with a linear increase from left aINS to left pINS [$F_{(1,\ 25)} = 5.084$, $P = 0.033$]; the contextual effect on neural responses to inequity in the advantageous frame became stronger in more anterior ROIs, but the linearity did not reach significance [$F_{(1,\ 25)} = 1.946$, $P = 0.175$]. The results remained the same using smoothed data (*SI Appendix*, Fig. S4).

Differential Psychological Components Associated with Advantageous- and Disadvantageous-Inequity Processing. To investigate whether the context-dependent processing of advantageous and disadvantageous inequity were associated with differential psychological components, we metaanalytically decoded these two processes using the Neurosynth Image Decoder (neurosynth.org; ref. 54); this allowed us to quantitatively evaluate the representational similarity (27) between any Nifti-format brain image and selected metaanalytical images generated by the Neurosynth database. Using this online platform, we compared the unthresholded contrast maps in advantageous and disadvantageous frames against the reverse inference metaanalytical maps for 13 terms related to the processing of inequity (11, 12) and the function of insula (55) generated from this database. Results demonstrated that the processing of advantageous inequity was associated with terms related to "social" and "mentalizing," while the processing of disadvantageous inequity was associated primarily with "somatosensory," "pain," and "sensorimotor" terms (Fig. 6*B*).

Functional Connectivity of the pINS in the Disadvantageous Frame. To test the probability that the processing of context-dependent advantageous- and disadvantageous-inequity aversion relies not only on neural activities but also functional connectivities between brain regions, we performed a psychophysiological interaction analysis (PPI) (56) focusing on left aINS, rDLPFC, and DMPFC identified for the advantageous frame and left pINS identified for the disadvantageous frame. When left pINS was used as the seed, results revealed significant Agent by Outcome by Inequity level three-way interactions on the functional connectivity between left pINS and dACC (peak coordinate extracted from Neurosynth metaanalysis for "conflict" term: [−6, 20, 34]; max T value = 3.59; cluster size = 79 voxels; Fig. 7*A*, I, and *SI Appendix*, Table S8), and between left pINS and right amygdala (peak coordinate extracted from Neurosynth metaanalysis for "emotion" term: [27, 2, −20]; max T value = 3.09; cluster size = 18 voxels; Fig. 7*A*, *II*, and *SI Appendix*, Table S8) [small volume correction (SVC), $P_{\mathrm{FWE}} < 0.05$, following an initial threshold of $P < 0.005$, uncorrected]. Specifically, compared with Self_Nopain condition, the contrast between the high-inequity (HI) and the low-inequity (LI) conditions in left pINS–dACC connectivity and left pINS–right amygdala connectivity decreased in the Self_Pain condition [dACC: $F_{(1,\ 25)} = 5.068$, $P = 0.033$, right amygdala: $F_{(1,\ 25)} = 5.767$, $P = 0.024$], while no difference was observed between the Other_Pain and Other_Nopain conditions [dACC: $F_{(1,\ 25)} < 0.001$, $P = 0.983$, right amygdala: $F_{(1,\ 25)} = 0.275$, $P = 0.604$]. These results were consistent with the decreased disadvantageous-inequity aversion in Self_Pain condition suggested by behavioral results. PPI analyses with seeds identified in the advantageous frame failed to survive the whole-brain cluster-level threshold and SVC.

Neural Correlates of Individual Differences in Contextual Effects on Advantageous- and Disadvantageous-Inequity Aversion. We further investigated the neural correlates of individual differences in the contextual effects on advantageous- and disadvantageous-inequity aversion. Here, the strength of neural adjustment was defined as the value of the Agent by Outcome by Inequity level three-way interaction [(Self_Pain_HI > Self_Pain_LI) − (Self_Nopain_HI > Self_Nopain_LI) − (Other_Pain_HI > Other_Pain_LI) + (Other_Nopain_HI > Other_Nopain_LI)]. Given the opposite behavioral and neural interaction effects observed in the advantageous and disadvantageous frames, here, for the advantageous frame, the larger this value of

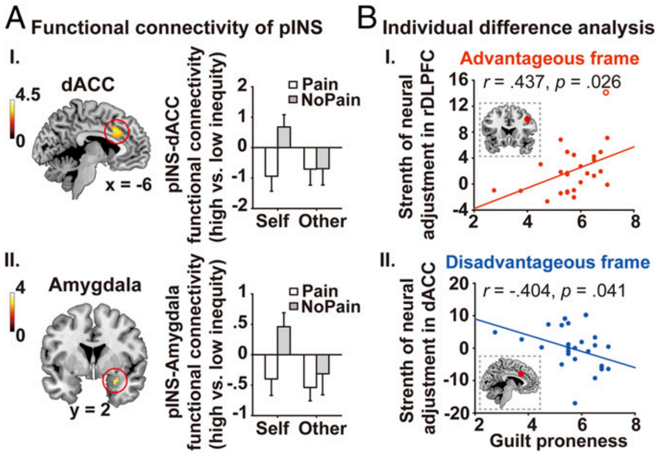

Fig. 7. (*A*) Functional connectivity of pINS in the disadvantageous frame. When left pINS was used as the seed, results revealed significant Agent by Outcome by Inequity level three-way interactions on the functional connectivity between left pINS and dACC (*A*, *I*) and between left pINS and right amygdala (*A*, *II*). For illustrative purposes, each bar represents the connectivity difference between high-inequity and low-inequity conditions in the disadvantageous frame, which indicates the neural responses to disadvantageous inequity. Compared with the Self_Nopain condition, the neural responses to disadvantageous inequity in left pINS–dACC connectivity and in left pINS–right amygdala connectivity decreased in the Self_Pain condition, while no difference was observed between the Other_Pain and Other_Nopain conditions. Error bars represent SEM. (*B*) Neural correlates of individual differences in contextual effects on advantageous- and disadvantageous-inequity aversion. Here, the strength of neural adjustment across contexts was defined as the value of the Agent by Outcome by Inequity level three-way interaction [(Self_Pain_HI > Self_Pain_LI) – (Self_Nopain_HI > Self_Nopain_LI) – (Other_Pain_HI > Other_Pain_LI) + (Other_Nopain_HI > Other_Nopain_LI)] (HI represents high-inequity condition, and LI represents low-inequity condition). (*B*, *I*) Results demonstrated that individual differences in sensitivity to context (i.e., guilt proneness measured by the GASP) predicted the strength of neural adjustment in rDLPFC identified in the advantageous frame, and the regression remained significant after one extreme value, which could be considered as an outlier, was removed. This effect was not observed for the disadvantageous frame. (*B*, *II*) Moreover, individuals' guilt proneness predicted the strength of neural adjustment in dACC identified in the disadvantageous frame. This effect was not observed for the advantageous frame.

interaction, the stronger the neural adjustment; for the disadvantageous frame, the smaller this value of interaction, the stronger the neural adjustment. The individual difference in sensitivity to guilt context [i.e., guilt proneness assessed by the Guilt and Shame Proneness scale (GASP) (*Materials and Methods*)] was related to the strength of neural adjustment in brain regions involved in advantageous-inequity aversion processing (i.e., left aINS, rDLPFC, and DMPFC) and brain regions involved in disadvantageous-inequity aversion processing (i.e., left pINS, dACC, and right amygdala) (*SI Appendix*, *SI Methods*). Results demonstrated that the individual's sensitivity to context predicted the strength of neural adjustment in rDLPFC in the advantageous frame ($r = 0.437$, $P = 0.026$), and the regression remained significant after one extreme value, which could be considered an outlier, was excluded ($r = 0.402$, $P = 0.047$) (Fig. 7*B*, *I*). This effect was not observed for the disadvantageous frame ($r = 0.018$, $P = 0.929$). Moreover, the individual's guilt proneness could predict the strength of neural adjustment in dACC in the disadvantageous frame ($r = -0.404$, $P = 0.041$) (Fig. 7*B*, *II*). This effect was not observed for the advantageous frame ($r = 0.129$, $P = 0.529$).

We established two more GLMs to rule out the possibility that the observed activations of brain regions were driven by the value representation of self-payoff (GLM2) and other-payoff (GLM3) (*SI Appendix*). For the fMRI analyses on the contextual effects on self-payoff and other-payoff representation, no cluster survived the whole-brain cluster level threshold. Moreover, no three-way interaction was found for neural responses to self-payoff or other-payoff in left aINS, rDLPFC, DMPFC, and left pINS (*SI Appendix*, Fig. S5), indicating that the observed effects of these regions in inequity processing were not driven by the value representations of self-payoff or other-payoff per se.

Discussion

Combining an interpersonal interactive game (49) that modulates individuals' advantageous- and disadvantageous-inequity aversion simultaneously and a variant of DG (50) that enables us to characterize individuals' changes in inequity aversion, we provide evidence showing that the context-dependent processing of advantageous inequity and disadvantageous inequity are associated with activities of different brain structures, demonstrating the existence of differential neurocognitive substrates underlying these two types of inequity aversion. While advantageous-inequity aversion is associated with the social- and mentalizing-related processes, which involve left aINS, rDLPFC, and DMPFC, disadvantageous-inequity aversion is associated primarily with somatosensory, emotional, and conflict processing, which involves left pINS, amygdala, and dACC. These results are consistent with previous evidence from behavioral economics (2, 4) and from the developmental (6) and evolutionary (5) perspectives.

Distinct Roles of Insular Subregions in Inequity-Aversion Processing.

The insula is involved in a circuit responsible for the detection of salience (for a review, see ref. 57). Previous studies employing a range of approaches, such as cytoarchitectonic mapping, tractography, and functional connectivity analysis, reveal an anteroposterior organization of the insula: the posterior region is involved in primary interoceptive representation, whereas the anterior region is involved in motivational, social, and cognitive processing (55, 58). Extending this functional segregation, our results demonstrate that the context-dependent processing of inequity exhibits a spatial gradient within the insula, such that anterior parts are predominantly involved in advantageous-inequity aversion and posterior parts are predominantly involved in disadvantageous-inequity aversion.

The involvement of pINS in disadvantageous-inequity aversion processing in the current study might seem puzzling, considering the primary role of pINS in interoceptive representations (58) and the role of aINS in UG responders' responses to unfair (i.e., disadvantageous) offers (for reviews and metaanalyses, see refs. 11–13). However, by mapping peak coordinates of insula regions identified in studies focused on UG responders (*SI Appendix*, Fig. S6), we found that unfair offers were always associated with increased activity in the dorsal anterior insula, but the context-related processing of unfair offers also involved the midposterior parts of insula, consistent with the pINS identified here. The involvement of pINS in inequity processing is further supported by Hsu et al. (59), who demonstrated that pINS represents aversion to inequity in a third-party resource distribution task. These results are also in line with the involvement of pINS in other high-level computations related to intertemporal choice (60, 61) and language perception (62).

Here, we did not observe aINS activity for the disadvantageous frame as suggested by previous UG studies. In those studies, detecting the fairness norm violation by others is assumed to be the main role of aINS in UG (16–20). In the current study, however, the participants made monetary allocations voluntarily as a dictator, rather than as a passive receiver in the game. Voluntarily giving others more than oneself was not treated as a form of norm violation, and hence aINS was not involved. Instead, we observed pINS activity in this situation. We believe that this pINS activity reflects the involvement of emotional processing in disadvantageous monetary allocation. It has been shown that there are both structural (63) and functional (64) connections between pINS and amygdala, a brain region that plays an important role in emotional processing (65). Patients with the post-traumatic stress disorder symptom of hyperarousal have hyperconnectivity between pINS and amygdala, indicating the involvement of pINS–amygdala connectivity in emotional responses (66). Interoceptive awareness, which is mainly represented in pINS (58, 67), modulates the emotional responses to unfair proposals in UG (68). Here, we found that not only pINS activity but also its functional connectivity with amygdala showed significant interactions between context and disadvantageous-inequity level.

In contrast to the absence of aINS activity in the disadvantageous frame, we observed context-dependent responses in aINS for the advantageous frame. In DG, the advantageous-inequity aversion is usually interpreted as resulting from the anticipated "guilt" feeling, that is, the negative feeling induced by norm violation (e.g., earning more than others) (2, 45, 46). Chang et al. (69) suggested the role of aINS in minimizing anticipated guilt and motivating adherence to the perceived social norm in trust game. Consistently, previous work on social conformity has revealed the involvement of aINS (70, 71), indicating that one function of aINS is to track deviations from the perceived social norm and bias actions to maintain adherence to this norm. This proposal can also be applied to a series of studies demonstrating the involvement of aINS in deciding to reject unfair offers in UG (11–13). Our results extend the role of aINS and demonstrate its involvement in adjusting advantageous-inequity aversion (or the anticipated guilt) according to social contexts. Taking into account the functions of aINS suggested by the aforementioned studies, we suggest that the increased aINS responses to advantageous inequity when individuals inflict pain upon others reflect their increased sensitivity to anticipated norm violation for choosing advantageous options; this anticipation might prevent them from actually choosing these options.

Neural Correlates of Context-Dependent Advantageous-Inequity Aversion. In addition to aINS, two regions that play critical roles in social decision making, DMPFC and rDLPFC, were identified for the advantageous frame. DMPFC is primarily related to the understanding of other's mental states (i.e., mentalizing) (for a review, see ref. 15). Given that the ability of mentalizing (i.e., understanding the other's feeling of being hurt in the interpersonal transgression context) is a foundation of guilt experience (41, 72), it seems natural to extend the mentalizing process to the processing of the anticipated feeling of "guilt," that is, the advantageous-inequity aversion (2, 45, 46). The recruitment of DMPFC in the context-dependent processing of advantageous inequity here may help individuals to accurately anticipate the coplayer's feelings of disappointment in getting less across different contexts and adjust their behaviors according to contexts.

Disadvantageous-inequity aversion emerges in early childhood, whereas advantageous-inequity aversion emerges in late childhood, as the latter may require the development of behavioral-control–related brain regions to support norm compliance (for a review, see ref. 6). Consistently, a behavioral study demonstrated that rejecting advantageous inequity requires more cognitive resources than rejecting disadvantageous inequity (73). Here, we provide neural evidence that DLPFC, a region implicated in cognitive control (21) and social norm compliance (74–76), contributes to the adjustment of advantageous-inequity aversion to social contexts. Moreover, individuals with greater neural adjustments in DLPFC activity were associated with higher sensitivity to guilt context in daily life. These findings are congruent with the suggestion that robust cognitive control allows for responding to the dynamically changing environments with increased flexibility (77). Taken together, our findings demonstrate the "social" nature underpinning the context-dependent processing of advantageous inequity, which recruits the processes of norm violation detection, cognitive control, and mentalizing, reflected by the neural adjustments in aINS, DLPFC, and DMPFC.

Neural Correlates of Context-Dependent Disadvantageous-Inequity Aversion. In addition to the pINS–amygdala functional connectivity discussed above, context-dependent disadvantageous-inequity aversion processing was also associated with the functional connectivity between pINS and dACC, a region implicated in conflict monitoring (78). Previous studies have shown increased activity of dACC in response to unfair offers in UG, which may reflect the conflict between the unfairness-evoked aversive responses and the self-interest to gain monetary reward (8, 11–13). We suggest that, during the experience of guilt, the decreased sensitivity of pINS–dACC functional connectivity in response to disadvantageous inequity might indicate reduced conflict between aversive responses and self-interest. The role of dACC in the context-dependent disadvantageous-inequity aversion was further confirmed by the correlation between individuals' sensitivity to contexts and the strength of neural adjustment in dACC. In sum, the context-dependent processing of disadvantageous-inequity aversion recruits the interoceptive system, emotional system, and conflict monitoring system, reflected by the involvements of pINS, amygdala, and dACC.

Implications and Future Directions. First, in the current study, we have leveraged the benefit of combining an interactive game in social psychology and computational models in neuroeconomics. While interactive games enable us to observe participants' behaviors and neural responses in real-life–like contexts, the application of sophisticated economic models enable us to quantify psychological constructs mathematically and examine these psychological constructs at the neural level (79, 80). Although we manipulated only a single context (i.e., interpersonal guilt) in the current study, our interdisciplinary paradigm may promote future studies to investigate related issues in various contexts.

Second, our findings of nonoverlapping brain regions for advantageous- and disadvantageous-inequity aversion provide valuable evidence for distinguishing between the two potential hypotheses regarding the neural mechanisms of inequity processing and support hypothesis 2 (Fig. 1*B*). It is worth noting, however, that the observed difference in regions involved in processing advantageous vs. disadvantageous inequity is not sufficient for us to draw the conclusion that advantageous inequity and disadvantageous inequity are separate, nonoverlapping psychological constructs. In the current study, we investigated the neural correlates of two types of inequity aversion under the framework of univariate analysis. However, it is possible that these two types of inequity aversion are also encoded by spatially distributed activation patterns in the brain (26), which may not be detected in the univariate analysis. Future studies using multivariate analyses [e.g., MVPA (26) or RSA (27)] may provide more evidence for these questions. Moreover, it is also possible that there are overlapping brain regions, which process both types of inequity, but fail to survive thresholding in our study. Therefore, future independent replications, as well as further studies using other context manipulations or other paradigms, are needed to address this issue.

Third, we did not observe the involvement of the reward system [e.g., ventral striatum (VS) or ventromedial prefrontal cortex (VMPFC)] in either the advantageous frame or the disadvantageous frame, which seems inconsistent with the activation of VS and VMPFC for both frames suggested by previous studies (30, 81). Nevertheless, in the later paradigms, participants either were presented with the redistributions that reduced the initial inequity in the original distributions (81) or had the right to change the unequal distributions after seeing these distributions (30). Therefore, we suggest these activations of reward-related areas actually reflected the experienced or expected pleasure of the changes in the unequal distributions, not the reactions to inequity itself. Despite this, the absence of reward-related regions in the advantageous frame seems puzzling, given that a number of studies on social comparison showed the activation of VS and VMPFC when participants performed better or attained better outcomes than others in competitive environments (22). One possible explanation is that, in those studies, individuals were satisfied with the advantageous outcomes due to the context of competition. In contrast, in the current study, the setting that the coplayer may pay for the participant's mistakes created a cooperative relationship, and the participant put negative values on advantageous inequity (positive parameter α) in all of the four conditions, which may explain the absence of reward-related regions. Thus, an empirical question for future research is which psychological and neural mechanisms support the transition from advantageous status enjoyment to advantageous status aversion.

Finally, the ability to flexibly integrate contextual information and adjust decisions and behaviors accordingly is a crucial skill underlying successful social interactions (39, 40, 77). Understanding how individuals make behavioral and neural adjustments to the social context provides valuable insights regarding certain social dysfunctions, such as autism (82) or psychopathy (83), which are associated with reduced sensitivity to social signals. Our results suggest that the strength of neural adjustments in rDLPFC in the advantageous frame and dACC in the disadvantageous frame are correlated with individuals' sensitivity to the guilt context in daily life. Future research is needed to test whether these individual differences in neural adjustment can be applied to other social contexts that influence inequity aversion or social dysfunction.

In summary, our findings shed light on how social and economic contexts are taken into account in social decision making and suggest that the resistance to unequal situations when individuals are in disadvantageous or low status may stem from their emotional responses, whereas the resistance to unequal situations when individuals are in advantageous or high status may require the involvement of advanced cognitive functions, such as mentalizing and cognitive control.

Materials and Methods

Participants. In total, 37 (25 females), 28 (17 females), and 34 (21 females) healthy graduate and undergraduate students were recruited for experiments 1, 2, and 3, respectively. All of the experiments were carried out in accordance with the Declaration of Helsinki and were approved by the Ethics Committee of the School of Psychological and Cognitive Sciences, Peking University. Informed written consent was obtained from each participant before each experiment.

DGs. Two versions of DG were used. In experiment 1, for each choice, the participant received an endowment and could unilaterally choose to give any integer amount of tokens (from 0 to the amount of endowment) to the coplayer (50). The relative cost and benefit of giving were manipulated by independently varying how much each token was worth to the participant and the coplayer (*SI Appendix*, Table S1). This DG enabled us to conduct model fitting at the group level. To further conduct model fitting at the individual level and to dissociate advantageous and disadvantageous frames, we developed a binary choice version of DG in experiments 2 and 3. Each binary choice consisted of two options representing the payoffs that the participant and the coplayer would earn. One option was always "10 points for me, and 10 points for the coplayer," and the other option was an unequal option with different values in each trial. Two types of unequal options were implemented corresponding to the two types of inequity (*SI Appendix*, Table S2). In both versions of the DG, one trial was selected randomly and actualized after the experiment, determining the final payoffs for the participant and the coplayer.

Computational Modeling. In the Fehr–Schmidt inequity aversion model (2, 50), we defined the subjective value function as follows:

$$U = Ms - p \cdot \alpha \cdot (Ms - Mo) - q \cdot \beta \cdot (Mo - Ms),$$

where Ms and Mo refer to self- and other-payoff, respectively, and p and q are indicator functions: $p = 1$ if $Ms \geq Mo$ (advantageous inequity), and 0 otherwise; and $q = 1$ if $Ms < Mo$ (disadvantageous inequity), and 0 otherwise. Thus, α and β quantify subjective aversion to inequity under advantageous and disadvantageous frames, respectively.

Additional Measures. Each participant was asked to complete the GASP (84) after the fMRI experiment. This scale measures individual differences in the proneness to experiencing guilt and shame across a range of personal transgressions in daily life. Individuals with higher scores in the guilt–negative-behavior evaluation (NBE) subscale of GASP feel guiltier after harming others and are more empathic and altruistic than those with lower guilt–NBE scores. In the current study, participants' guilt proneness, reflected by scores on guilt–NBE in GASP, was used as an index for individual's sensitivity to interpersonal guilt in daily life; these guilt proneness scores were used to investigate the neural correlates of individual differences in the contextual effects on advantageous- and disadvantageous-inequity aversion.

fMRI Data Acquisition and Analysis. Images were acquired using a GE Healthcare 3.0-T Medical Systems Discovery MR 750 with a standard head coil. We used standard preprocessing in SPM8 (Wellcome Trust Centre for Neuroimaging) and estimated three GLMs for each participant that focused on the neural responses during DG. For whole-brain analyses, all results were corrected for multiple comparisons using the threshold of voxel-level $P < 0.001$ (uncorrected) combined with cluster-level threshold $P < 0.05$ [familywise error (FWE) corrected]. This threshold provides an acceptable family error control (85, 86). SVC was conducted using a cluster-level threshold $P < 0.05$ (FWE corrected), following an initial threshold of $P < 0.005$ (uncorrected). The small volumes of dACC and amygdala were defined as spheres with 10-mm radius, centered on the peak MNI coordinates extracted from the metaanalyses on the "emotion" and "conflict" terms in the Neurosynth database. A detailed description of methods including participants, procedures, computational modeling, and fMRI data analyses are given in *SI Appendix*.

Note. The behavioral part of this study was presented in a poster at the annual meeting of the Social and Affective Neuroscience Society 2016 (New York, April 28 to May 2, 2016). The whole study was presented as a talk at the Society for Neuroeconomics Annual Conference 2017 (Toronto, October 6–8, 2017).

ACKNOWLEDGMENTS. We thank Dr. Christian C. Ruff, Dr. Peter Sokol-Hessner, Dr. Luke J. Chang, and two anonymous reviewers for their comments on this article. This work was supported by National Natural Science Foundation of China (31630034) and by National Basic Research Program of China (973 Program: 2015CB856400). H.Y. was supported by a Newton International Fellowship from the British Academy (NF160700).

1. Decety J, Yoder KJ (2017) The emerging social neuroscience of justice motivation. *Trends Cogn Sci* 21:6–14.
2. Fehr E, Schmidt KM (1999) A theory of fairness, competition, and cooperation. *Q J Econ* 114:817–868.
3. Bechtel MM, Liesch R, Scheve KF (2018) Inequality and redistribution behavior in a give-or-take game. *Proc Natl Acad Sci USA* 115:3611–3616.
4. Loewenstein GF, Bazerman MH, Thompson L (1989) Social utility and decision-making in interpersonal contexts. *J Pers Soc Psychol* 57:426–441.
5. Brosnan SF, de Waal FB (2014) Evolution of responses to (un)fairness. *Science* 346: 1251776.
6. McAuliffe K, Blake PR, Steinbeis N, Warneken F (2017) The developmental foundations of human fairness. *Nat Hum Behav* 1:0042.
7. Marthinsen M (2016) Inequality, redistribution and growth—interrelations and directions. Master's thesis (University of Oslo, Oslo).
8. Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD (2003) The neural basis of economic decision-making in the ultimatum game. *Science* 300:1755–1758.
9. Civai C, Corradi-Dell'Acqua C, Gamer M, Rumiati RI (2010) Are irrational reactions to unfairness truly emotionally-driven? Dissociated behavioural and emotional responses in the ultimatum game task. *Cognition* 114:89–95.
10. van 't Wout M, Kahn RS, Sanfey AG, Aleman A (2006) Affective state and decision-making in the ultimatum game. *Exp Brain Res* 169:564–568.
11. Aoki R, Yomogida Y, Matsumoto K (2015) The neural bases for valuing social equality. *Neurosci Res* 90:33–40.
12. Feng C, Luo YJ, 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.
13. Gabay AS, Radua J, Kempton MJ, Mehta MA (2014) The ultimatum game and the brain: A meta-analysis of neuroimaging studies. *Neurosci Biobehav Rev* 47:549–558.
14. Frith U, Frith CD (2003) Development and neurophysiology of mentalizing. *Philos Trans R Soc Lond B Biol Sci* 358:459–473.
15. Isoda M, Noritake A (2013) What makes the dorsomedial frontal cortex active during reading the mental states of others? *Front Neurosci* 7:232.
16. Chang LJ, Sanfey AG (2013) Great expectations: Neural computations underlying the use of social norms in decision-making. *Soc Cogn Affect Neurosci* 8:277–284.
17. Civai C, Crescentini C, Rustichini A, Rumiati RI (2012) Equality versus self-interest in the brain: Differential roles of anterior insula and medial prefrontal cortex. *Neuroimage* 62:102–112.
18. Corradi-Dell'Acqua C, Civai C, Rumiati RI, Fink GR (2013) Disentangling self- and fairness-related neural mechanisms involved in the ultimatum game: An fMRI study. *Soc Cogn Affect Neurosci* 8:424–431.
19. Gu X, et al. (2015) Necessary, yet dissociable contributions of the insular and ventromedial prefrontal cortices to norm adaptation: Computational and lesion evidence in humans. *J Neurosci* 35:467–473.

20. Xiang T, Lohrenz T, Montague PR (2013) Computational substrates of norms and their violations during social exchange. *J Neurosci* 33:1099–108a.
21. Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202.
22. Luo Y, Eickhoff SB, Hétu S, Feng C (2018) Social comparison in the brain: A coordinate-based meta-analysis of functional brain imaging studies on the downward and upward comparisons. *Hum Brain Mapp* 39:440–458.
23. Levy DJ, Glimcher PW (2012) The root of all value: A neural common currency for choice. *Curr Opin Neurobiol* 22:1027–1038.
24. Sugrue LP, Corrado GS, Newsome WT (2005) Choosing the greater of two goods: Neural currencies for valuation and decision making. *Nat Rev Neurosci* 6:363–375.
25. Woo C-W, et al. (2014) Separate neural representations for physical pain and social rejection. *Nat Commun* 5:5380.
26. Haynes JD, Rees G (2006) Decoding mental states from brain activity in humans. *Nat Rev Neurosci* 7:523–534.
27. Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis—connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:4.
28. Fliessbach K, et al. (2012) Neural responses to advantageous and disadvantageous inequity. *Front Hum Neurosci* 6:165.
29. Güroğlu B, Will GJ, Crone EA (2014) Neural correlates of advantageous and disadvantageous inequity in sharing decisions. *PLoS One* 9:e107996.
30. Yu R, Calder AJ, Mobbs D (2014) Overlapping and distinct representations of advantageous and disadvantageous inequality. *Hum Brain Mapp* 35:3290–3301.
31. Konovalov A, Hu J, Ruff CC (2018) Neurocomputational approaches to social behavior. *Curr Opin Psychol* 24:41–47.
32. Sternberg S (2001) Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychol (Amst)* 106:147–246.
33. Seymour B, McClure SM (2008) Anchors, scales and the relative coding of value in the brain. *Curr Opin Neurobiol* 18:173–178.
34. Ruff CC, Fehr E (2014) The neurobiology of rewards and values in social decision making. *Nat Rev Neurosci* 15:549–562.
35. Güroğlu B, van den Bos W, Rombouts SA, Crone EA (2010) Unfair? It depends: Neural correlates of fairness in social context. *Soc Cogn Affect Neurosci* 5:414–423.
36. Tversky A, Simonson I (1993) Context-dependent preferences. *Manage Sci* 39: 1179–1189.
37. Wright ND, Symmonds M, Fleming SM, Dolan RJ (2011) Neural segregation of objective and contextual aspects of fairness. *J Neurosci* 31:5244–5252.
38. Wang Y, Yang LQ, Li S, Zhou Y (2015) Game theory paradigm: A new tool for investigating social dysfunction in major depressive disorders. *Front Psychiatry* 6:128.
39. Louie K, De Martino B (2013) The neurobiology of context-dependent valuation and choice. *Neuroeconomics: Decision Making and the Brain*, eds Glimcher PW, Fehr E (Elsevier, London), pp 455–476.
40. Rilling JK, Sanfey AG (2011) The neuroscience of social decision-making. *Annu Rev Psychol* 62:23–48.
41. Baumeister RF, Stillwell AM, Heatherton TF (1994) Guilt: An interpersonal approach. *Psychol Bull* 115:243–267.
42. Owens D (2008) Rationalism about obligation (D. Davidson). *Eur J Philos* 16:403–431.
43. Tangney JP, Stuewig J, Mashek DJ (2007) Moral emotions and moral behavior. *Annu Rev Psychol* 58:345–372.
44. Kubany ES, Watson SB (2003) Guilt: Elaboration of a multidimensional model. *Psychol Rec* 53:51–90.
45. Rey-Biel P (2008) Inequity aversion and team incentives. *Scand J Econ* 110:297–320.
46. Krajbich I, Adolphs R, Tranel D, Denburg NL, Camerer CF (2009) Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *J Neurosci* 29:2188–2192.
47. Izard CE (1977) *Human Emotions* (Plenum, New York).
48. Reed LI (2010) The effect of guilt on altruism in the one-shot anonymous prisoner's dilemma game, PhD dissertation (University of Pittsburgh, Pittsburgh).
49. Yu H, Hu J, Hu L, Zhou X (2014) The voice of conscience: Neural bases of interpersonal guilt and compensation. *Soc Cogn Affect Neurosci* 9:1150–1158.
50. Sáez I, Zhu L, Set E, Kayser A, Hsu M (2015) Dopamine modulates egalitarian behavior in humans. *Curr Biol* 25:912–919.
51. McNamee D, Rangel A, O'Doherty JP (2013) Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nat Neurosci* 16:479–485.
52. Yu H, Li J, Zhou X (2015) Neural substrates of intention—consequence integration and its impact on reactive punishment in interpersonal transgression. *J Neurosci* 35: 4917–4925.
53. Tzourio-Mazoyer N, et al. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15:273–289.
54. Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD (2011) Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods* 8:665–670.
55. Chang LJ, Yarkoni T, Khaw MW, Sanfey AG (2013) Decoding the role of the insula in human cognition: Functional parcellation and large-scale reverse inference. *Cereb Cortex* 23:739–749.
56. Friston KJ, et al. (1997) Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* 6:218–229.
57. Menon V, Uddin LQ (2010) Saliency, switching, attention and control: A network model of insula function. *Brain Struct Funct* 214:655–667.
58. Craig AD (2009) How do you feel—now? The anterior insula and human awareness. *Nat Rev Neurosci* 10:59–70.
59. Hsu M, Anen C, Quartz SR (2008) The right and the good: Distributive justice and neural encoding of equity and efficiency. *Science* 320:1092–1095.
60. Tanaka SC, et al. (2004) Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat Neurosci* 7:887–893.
61. Wittmann M, Leland DS, Paulus MP (2007) Time and decision making: Differential contribution of the posterior insular cortex and the striatum during a delay discounting task. *Exp Brain Res* 179:643–653.
62. Jones CL, Ward J, Critchley HD (2010) The neuropsychological impact of insular cortex lesions. *J Neurol Neurosurg Psychiatry* 81:611–618.
63. McDonald AJ (1998) Cortical pathways to the mammalian amygdala. *Prog Neurobiol* 55:257–332.
64. Roy AK, et al. (2009) Functional connectivity of the human amygdala using resting state fMRI. *Neuroimage* 45:614–626.
65. Pessoa L, Adolphs R (2010) Emotion processing and the amygdala: From a "low road" to "many roads" of evaluating biological significance. *Nat Rev Neurosci* 11:773–783.
66. Sripada RK, et al. (2012) Altered resting-state amygdala functional connectivity in men with posttraumatic stress disorder. *J Psychiatry Neurosci* 37:241–249.
67. Craig AD (2011) Significance of the insula for the evolution of human awareness of feelings from the body. *Ann N Y Acad Sci* 1225:72–82.
68. Dunn BD, Evans D, Makarova D, White J, 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.
69. Chang LJ, Smith A, Dufwenberg M, Sanfey AG (2011) Triangulating the neural, psychological, and economic bases of guilt aversion. *Neuron* 70:560–572.
70. Berns GS, Capra CM, Moore S, Noussair C (2010) Neural mechanisms of the influence of popularity on adolescent ratings of music. *Neuroimage* 49:2687–2696.
71. Klucharev V, Hytönen K, Rijpkema M, Smidts A, Fernández G (2009) Reinforcement learning signal predicts social conformity. *Neuron* 61:140–151.
72. Basil DZ, Ridgway NM, Basil MD (2008) Guilt and giving: A process model of empathy and efficacy. *Psychol Mark* 25:1–23.
73. Van den Bos K, Peters SL, Bobocel DR, Ybema JF (2006) On preferences and doing the right thing: Satisfaction with advantageous inequity when cognitive processing is limited. *J Exp Soc Psychol* 42:273–289.
74. Knoch D, Pascual-Leone A, Meyer K, Treyer V, Fehr E (2006) Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314:829–832.
75. Ruff CC, Ugazio G, Fehr E (2013) Changing social norm compliance with noninvasive brain stimulation. *Science* 342:482–484.
76. Zhu L, et al. (2014) Damage to dorsolateral prefrontal cortex affects tradeoffs between honesty and self-interest. *Nat Neurosci* 17:1319–1321.
77. Steinbeis N, Crone EA (2016) The link between cognitive control and decision-making across child and adolescent development. *Curr Opin Behav Sci* 10:28–32.
78. Heilbronner SR, Hayden BY (2016) Dorsal anterior cingulate cortex: A bottom-up view. *Annu Rev Neurosci* 39:149–170.
79. Crockett MJ (2016) How formal models can illuminate mechanisms of moral judgment and decision making. *Curr Dir Psychol Sci* 25:85–90.
80. O'Doherty JP, Hampton A, Kim H (2007) Model-based fMRI and its application to reward learning and decision making. *Ann N Y Acad Sci* 1104:35–53.
81. Tricomi E, Rangel A, Camerer CF, O'Doherty JP (2010) Neural evidence for inequality-averse social preferences. *Nature* 463:1089–1091.
82. Palmer CJ, Paton B, Kirkovski M, Enticott PG, Hohwy J (2015) Context sensitivity in action decreases along the autism spectrum: A predictive processing perspective. *Proc Biol Sci* 282:20141557.
83. Domes G, Hollerbach P, Vohs K, Mokros A, Habermeyer E (2013) Emotional empathy and psychopathy in offenders: An experimental study. *J Pers Disord* 27:67–84.
84. Cohen TR, Wolf ST, Panter AT, Insko CA (2011) Introducing the GASP scale: A new measure of guilt and shame proneness. *J Pers Soc Psychol* 100:947–966.
85. Eklund A, Nichols TE, Knutsson H (2016) Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proc Natl Acad Sci USA* 113: 7900–7905.
86. Flandin G, Friston KJ (2017) Analysis of family-wise error rates in statistical parametric mapping using random field theory. *Hum Brain Mapp*, 10.1002/hbm.23839.