PsyCh Journal

INSTITUTE OF PSYCHOLOGY
Chinese Academy of Sciences

PsyCh Journal **3** (2014): 214–221
DOI: 10.1002/pchj.59

The processing course of conflicts in third-party punishment: An event-related potential study

Lulu Qu,[1,2]* Wei Dou,[1] Cheng You,[1] and Chen Qu,[1]*

[1]School of Psychology and Center for Studies of Psychological Application, South China Normal University, Guangzhou, China, [2]No. 3 Middle School of Binhai Weifang, Weifang, China

Abstract: In social decision-making games, uninvolved third parties usually severely punish norm violators, even though the punishment is costly for them. For this irrational behavior, the conflict caused by punishment satisfaction and monetary loss is obvious. In the present study, 18 participants observed a Dictator Game and were asked about their willingness to incur some cost to change the offers by reducing the dictator's money. A response-locked event-related potential (ERP) component, the error negativity or error-related negativity (Ne/ERN), which is evoked by error or conflict, was analyzed to investigate whether a trade-off between irrational punishment and rational private benefit occurred in the brain responses of third parties. We examined the effect of the choice type ("to change the offer" or "not to change the offer") and levels of unfairness (90:10 and 70:30) on Ne/ERN amplitudes. The results indicated that there was an ERN effect for unfair offers as Ne/ERN amplitudes were more negative for not to change the offer choices than for to change the offer choices, which suggested that participants encountered more conflict when they did not change unfair offers. Furthermore, it was implied that altruistic punishment, rather than rational utilitarianism, might be the prepotent tendency for humans that is involved in the early stage of decision-making.

Keywords: conflict; decision-making; Ne/ERN; punishment; third party
Correspondence: Miss Chen Qu, School of Psychology, South China Normal University, Guangzhou, China, 510631. Email: chenqu@scnu.edu.cn
Received 31 July 2013. Accepted 22 April 2014.
*These authors contributed equally to the work.

Experimental studies have indicated that people punish norm violators without any overt benefit, a strongly reciprocal behavior that is commonly termed altruistic punishment (Strobel et al., 2011). In the ultimatum game, for instance, responders frequently reject unfair offers from proposers (Fehr & Fischbacher, 2004; Hewig et al., 2011). The notion of altruistic punishment is often invoked to explain norm enforcement (Fehr & Fischbacher, 2004); that is, the presence of such sanctions may serve as a mechanism for protecting social norms. However, sanctions against a violator by a harmed party could be caused by nonnormative motives, such as a retaliatory impulse (Ohtsubo, Masuda, Watanabe, & Masuchi, 2010). Therefore, an uninvolved party is closer to social norms.

In one of the third-party punishment games, from the research of Fehr and Fischbacher (2004), uninvolved third-party participants observed other players behaving in an unfair manner in a Dictator Game (DG) and then decided to incur some cost to punish the unfair player. In the experiment, approximately 60% of the participants punished dictators at each unfair transfer level. So-called third-party punishment is caused by a social normative motive. Third-party punishment for violation of a fairness norm has been confirmed across a wide range of cultures with the same game (Henrich et al., 2006). These studies have demonstrated that strong reciprocity and social norms reliably induce sanctioning behavior of "unaffected" third parties (Fehr & Fischbacher, 2004).

Recently, neuroscientific experiments have provided more data for interpreting this economically irrational behavior. Research by Wu and Luo (2011) using event-related potential (ERP) to study the evaluative processes in the brain, when participants perceived the outcome of "punish-other-lose" versus "not punish-other-win," and "punish-self-lose" versus "not punish-self-not-lose" in the repeated trust game, found that not punish-other-win and not punish-self-not-lose elicited larger feedback-related negativities (FRNs) than punish-other-win and punish-other-lose. This suggested that the FRN amplitudes of not punish, which had a greater degree of negative emotions, were larger than those of punish, which had a smaller degree of negative emotions. Another study reported in the same research used a trust game in which participants were asked to evaluate their emotions when making a decision, and the results also showed that the degree of negative emotions caused by nonpunishment was greater than that caused by punishment (Wu & Luo, 2011). These results supported the view that emotion has a greater influence than economic rationality on participants who inflict altruistic punishment. Other research on the activation of brain reward regions during altruistic punishment in two-person one-shot exchange games suggested that satisfaction through the punishment of norm violations might be one underlying motivation (Delgado, Locke, Stenger, & Fiez, 2003). To address this issue in more detail and exclude the influence of revenge when acting as the second party, a neuroimaging study (Strobel et al., 2011) on the neural basis of altruistic punishment introduced a third-party punishment condition during a one-shot DG. Participants were either in the role of player B (the responder) who faced the decisions of player A (the dictator) or in the role of player C (a third party who had no direct beneficial relationship with the others) who observed interactions between the dictators and the responders. The task for participants was to decide whether or not to punish player A by assigning some punishment points. It was reported that reward regions, such as the nucleus accumbens, showed punishment-related activation even when participants were in the role of the third party, which suggested a common cognitive-affective-motivational network as the driving force for altruistic punishment (Strobel et al., 2011).

Above all, for altruistic punishment, the opinion that such punishment is irrational behavior is from the point of view of economic rationality, which holds that humans tend to maximize their benefit. However, it is possibly reasonable behavior when seen from the perspective of emotional motives or social reasons. In other words, emotion is more likely to be dominant at the stage of decision-making of altruistic punishment, particularly for unaffected third parties. Therefore, we used a third-party paradigm in the present study to examine this supposition.

Although previous studies have suggested that most participants acting as the third party would punish violators in the game (Fehr & Fischbacher, 2004; Strobel et al., 2011), third parties do not make the same decision all the time. It seems that people do not have a firm stand about the choice. Obviously, participants could not pursue the most money and protect social norms at the same time, so it is difficult to make a decision between the two choices, leading to a hesitation and a trade-off before the final determination. More specifically, when making the decision to punish or not, the trade-off in the psychological process between punishing at economic cost to oneself, versus the economically self-beneficial act of standing by but allowing unfairness to others go unsanctioned, induces a response conflict in the mind. As neither punishing nor standing by is a perfect decision, this response conflict occurs with either response. Obviously, the level of response conflict depends on the distance the determination deviates from the inner tendency (i.e., the prepotent tendency). If emotion plays a more important role in the decision-making process, punishment should be the prepotent tendency and the choice not to punish would induce greater conflict than the choice to punish. Conversely, given a more dominant tendency for benefit-oriented nonpunishment, the choice to punish would induce greater conflict. The purpose of the current study was to investigate the temporal course of conflicts evoked by third-party punishment during the executive stage using the ERP technique.

Error-related negativity (Ne/ERN) is a negative response-locked ERP component that peaks approximately 0–100 ms after the response and is most strongly pronounced at frontocentral midline scalp sites (Amodio et al., 2004; Amodio, Devine, & Harmon-Jones, 2008; Hirsh & Inzlicht, 2010). ERN occurs because of an erroneous button press in speeded response tasks, or in selection conflicts. Error detection theory, an early theory accounting for Ne/ERN, assumes that Ne/ERN is a neural correlate of mismatch detected by comparing representations of the intended and the actually performed actions (Bernstein, Scheffers, & Coles, 1995; Coles, Scheffers, & Holroyd, 2001). More recently, conflict monitoring theory holds that Ne/ERN reflects conflict during response selection and that errors are simply a specific example of response conflict that occurs between an erroneous and an error-correcting response (Holroyd et al., 2004; Holroyd, Yeung, Coles, & Cohen, 2005; Yeung, Botvinick, & Cohen, 2004; Yeung & Sanfey, 2004). This means that when we choose one action, the other choice is also activated, and the conflict between the two active, incompatible response tendencies generates Ne/ERN.

However, most previous studies about Ne/ERN have employed speeded response tasks in which Ne/ERN effects were intrinsically related to erroneous responses (Coles et al., 2001; Falkenstein, Hoormann, Christ, & Hohnsbein, 2000; Scheffers & Coles, 2000). These tasks, in which conflicts and response errors coexist at the same time, fuelled the debate between error detection theory and conflict monitoring theory. Recently, Yu and Zhou (2009) used a gambling task in which participants' responses were related more to a conflict than to response correctness. Still, an ERN effect was revealed.

In addition, the anterior cingulate cortex (ACC), which is reported to be the generator of Ne/ERN, is found to be involved in conflict monitoring (Botvinick, Cohen, & Carter, 2004; Bush, Luu, & Posner, 2000). The ACC monitors action outcomes and guides decision-making (Botvinick, 2007). We assume that conflicts in third-party punishment are detected by the ACC and reflected by Ne/ERN.

The present study employed a third-party punishment game (Fehr & Fischbacher, 2004) in which the participants played the role of third parties who observed an offer of the DG and decided whether to incur some personal cost to reduce the dictator's money. To avoid the impact of the word "punish" in the instructions to participants, "change" was used instead. The levels of offer fairness were manipulated as 90:10, 70:30 and 50:50, with 90:10 offers as highly unfair, 70:30 as moderately unfair, and 50:50 offers as fair (trials of 50:50 were used to increase the credibility of the experiment and were excluded from the main analysis, as the distribution of 50:50 was totally fair and there was no conflict between the choices). As a response conflict between punishing the norm violators or keeping the most money occurred in the case of unfair offers, we expected, according to conflict monitoring theory, that an ERN effect should be revealed for both choices in such cases, and if emotion and morality took precedence in the early stage of the decision-making process, that Ne/ERN amplitudes would be more negative for not punishing than for punishing.

Methods

Participants

For the study, 18 healthy participants (nine males, nine females) between 19 and 28 years of age ($M = 22.61$ $SD = 1.58$) were recruited from South China Normal University, China. All participants had normal or corrected-to-normal vision and had no history of neurobiological or psychiatric disorders. The study was conducted in accordance with the guidelines of the authors' academic institution and informed consent was obtained from all participants.

Task and procedure

Participants sat in an electrically shielded room in front of a 17-in. CRT display approximately 1 m away. Simultaneous electroencephalogram (EEG) recordings were conducted during their participation. Participants were told that they would play games repeatedly in a series of one-shot trials in the role of the third party of the DG. In the DG, money is always assigned by the dictator to both themselves and the responder, and the responder cannot reject the offer. Actually, the dictator and the responder were virtual roles. For each trial, the participants had 50 points per person and had to decide whether to change the offer (from a total of 100 points) assigned by the dictator. They could spend 15 points to cut the dictator's total by 45 points, which is regarded as a punishment for the dictator, or they could just keep their 50 points by making no change. They were also informed that distribution assignments were chosen from experimental consequences with real players who participated in the DG and that payoffs for the players in the DG would be based on their decisions as to whether to change the offer or not. At the same time, participants were informed that one point in the experiment was worth 0.01 yuan, and that they would be paid according to 80 randomly selected trials from their decisions. Therefore, they would be paid between 28 and 40 yuan (approximately $4.43–6.33 U.S.).

Before the start of the experiment, participants were given the task instructions. To increase the credibility of the experiment, they were asked whether they wanted to play the DG for further research after the current experiment. If they agreed, they were asked to provide their personal photos for a later study. To become familiar with the task, the participants completed six trials as practice. The entire task consisted of 270 trials and each trial contained an offer which was an allocative decision created by the dictator and had three conditions: 70:30, 90:10, or 50:50. Each trial started with the presentation of the sentence, "The new trial is coming" (800 ms), and then a red fixation cross (700 ms). This was followed by a 1400-ms presentation of a color bar, which contained two kinds of colors that indicated the con-

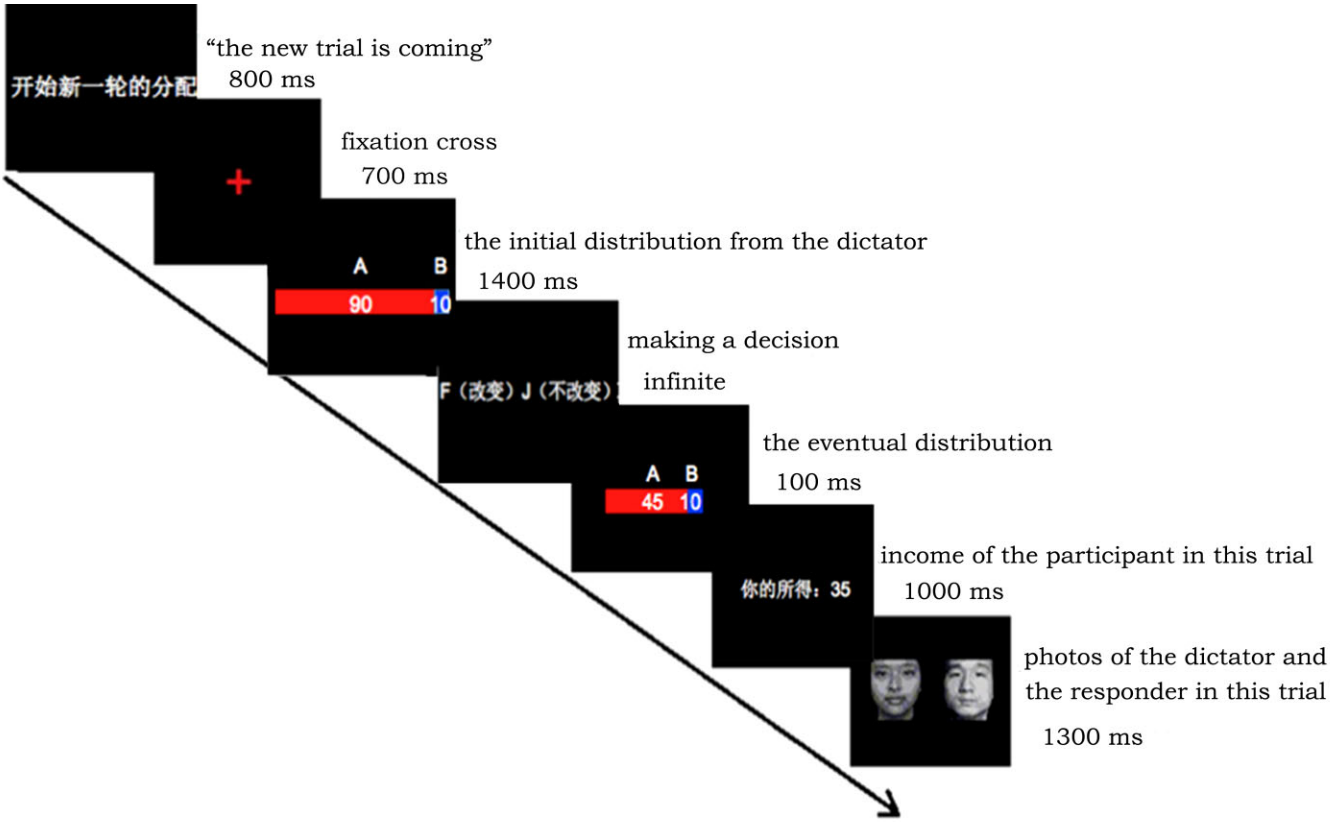

Figure 1. The time line of a single trial. Participants were shown the offer from the dictator and made a decision with a button press to select whether or not to change it. Then the participants were informed of the eventual distribution between the two players in the Dictator Game and their own income on this trial.

dition of the offer (Hewig et al., 2011). For half of the participants, the number in the blue part indicated the points taken by the dictator himself and the red part was for the responder. The participant then pressed a key on a keyboard to indicate a decision for either changing the offer (the F key), or for standing by (the J key). There was no limit on the response time. The offer of the other two players would remain for 1 s after they pressed the key, and the income of the participants on this trial would then remain for another 1 s. At this point, the participant was shown personal photos of both the dictator and the responder for this trial for 1300 ms. In fact, the pictures of dictators and responders were taken from the Chinese Affective Picture System and the offers were predetermined. Figure 1 shows a single trial of the game.

Following the experiment, participants completed a subjective rating of the level of unfairness (from 1–7) they had experienced for the 50:50 offers, the 70:30 offers, and the 90:10 offers while playing the role of the third party. Participants also self-reported their reasons for changing the offers.

Electrophysiological recording and analysis

EEGs were recorded from 32 scalp sites using tin electrodes mounted in an elastic cap (Brain Products, Munich, Germany) with the reference on the left mastoid. The vertical electrooculograms (EOGs) were monitored with electrodes located in four places: above and below the right eye and 1.5 cm lateral to the left and right external canthi. All electrode recordings were referenced to an electrode placed on the left mastoid and electrode impedances were kept below 5 kΩ for all recordings.

Offline analysis was performed using Brain Vision Analyzer software (Brain Products). The electrophysiological signals were amplified with a band pass of 0.01–100 Hz and continuously digitized at a rate of 500 Hz. Ocular artifacts were identified and corrected with an eye-movement correction algorithm. Trials in which EEG voltages exceeded a threshold of ±80 μV during the recording epoch were excluded from further analysis. EEG data were digitally filtered below 30 Hz (24 dB/octave) for all recordings. The response-locked ERPs were averaged for epochs of

Table 1
Mean and Standard Deviation of the Numbers of Choice Type on Different Unfairness Levels

	Unfairness level			
	70:30		90:10	
Choice type	*M*	*SD*	*M*	*SD*
To change	86.7	53.4	191.7	58.1
Not to change	181.8	52.7	75.3	59.4

800 ms starting 200 ms prior to the recorded response as a baseline.

For all trials, the mean Ne/ERN amplitudes were calculated for Fz, FCz, and Cz, as Ne/ERN is typically measured at midline frontal or central sites (Olvet & Hajcak, 2008). We quantified the ERN/Ne as the most negative peak occurring within the 100 ms following the response. For statistical analyses, we used the average amplitude of these peaks in a time window starting 20 ms before the peak and terminating 20 ms after the peak (Boksem, Tops, Kostermans, & De Cremer, 2008). A 2 (unfairness level: 70:30 vs. 90:10) × 2 (choice type: "to change the offer" vs. "not to change the offer") repeated-measures ANOVA was performed on Ne/ERN amplitudes and behavior results.

Results

Behavior results

Among the 270 trials, the number of trials (mean ± standard error) for changing the 70:30 offer was 86.7 ± 53.4, and for changing the 90:10 offer it was 191.7 ± 58.1 (indicated in Table 1). There were no main effects of unfairness level, $F(1, 17) = 1.00$, $p > .1$, or choice type, $F(1, 17) = 0.33$, $p > .1$, on the rate of choices. However, we found a significant interaction, $F(1, 17) = 31.63$, $p < .01$, between unfairness level and choice type, suggesting that offers of 90:10 were significantly changed more often ($p < .01$), and offers of 70:30 were accepted more often ($p < .01$). Additionally, there were no main effects of unfairness level, $F(1, 17) = 0.01$, $p > .1$, or choice type, $F(1, 17) = 0.465$, $p > .1$, on reaction times (RTs). We also did not find a significant interaction between the two factors on RTs, $F(1, 17) = 2.41$, $p > .1$.

ERPs

The ANOVA revealed a main effect of choice type that was significant at Fz, FCz, and Cz (*p*s < .05), and the average ERN amplitudes were significantly more negative for the not

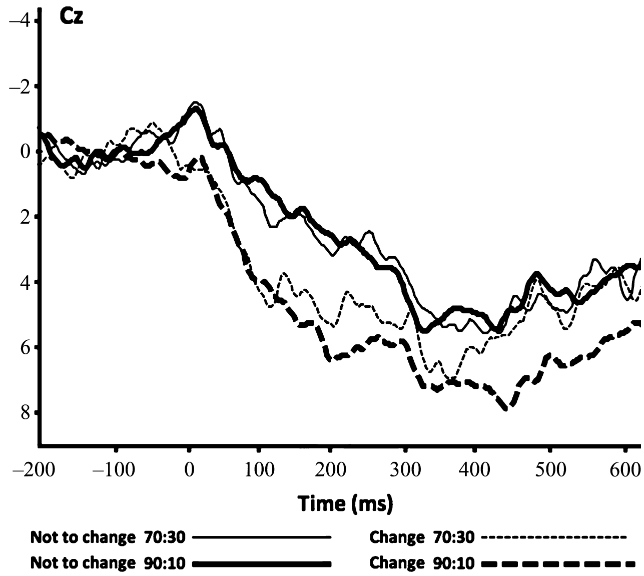

Figure 2. Response-locked grand-average event-related potential (ERP) waveforms at electrode site Cz. Error negativity or error-related negativity (Ne/ERN) waveforms for "to change" choices and "not to change" choices on different levels of unfairness were shown.

to change the offer choices (−2.39 μV, −1.68 μV, −1.03 μV) than for the to change the offer choices (−0.77 μV, 0.08 μV, 0.53 μV), while unfairness level had no significant main effect (as indicated in Figure 2, which shows the grand average response-locked ERPs at Cz, and in Figure 3, which shows the topographic maps). However, there was no interaction of choice type and unfairness level ($p > .1$). Although there was no further analysis of the fair conditions, ERP data for not to change the 50:50 offers were recorded and the independent *t*-test with other conditions showed that the ERN amplitude of not to change the 50:50 offers was significantly more positive than the others (indicated in Table 2). Data for to change the 50:50 offers were not analyzed because there were only three participants who made this decision more than six times, which was the least trials for ERN analysis (Olvet & Hajcak, 2009).

Self-report ratings

Participants provided subjective ratings of the unfair valence from 1 (*very fair*) to 7 (*very unfair*) for the monetary offers of 70:30 and 90:10. In total, 90:10 offers ($M = 6.76$, $SD = 0.44$) were considered more significantly unfair than 70:30 offers ($M = 4.12$, $SD = 0.86$, $p < .01$). Participants also indicated their reasons for changing. For 90:10 offers, most participants thought the offers were too unfair to accept, even though they would pay for the decision. For the distri-

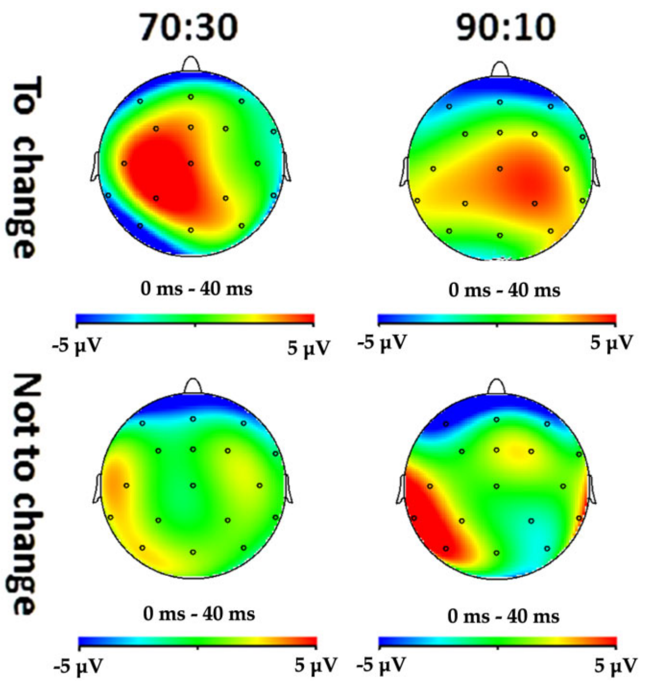

Figure 3. Topographic maps. The scalp distribution of the error negativity or error-related negativity (Ne/ERN) for "to change" choices and "not to change" choices on different levels of unfairness.

Table 2
Average Error-Related Negativity (ERN) Amplitude in Five Conditions at Three Sites and the Independent t*-Tests between the ERN Amplitude of "Not to Change 50:50" Offers and that of Each Other Condition*

	Sites		
	Fz	FCz	Cz
ERN amplitudes (μV)			
Not to change 50:50 offers	2.3 ± 2.1	1.8 ± 1.6	1.5 ± 1.1
To change 70:30 offers	−0.3 ± 3.5	0.4 ± 2.5	0.6 ± 2.3
t-value (with no to change 50:50 offers)	2.939	1.984	1.524
Significance	.003**	.027*	0.068
To change 90:10 offers	−1.2 ± 1.8	−0.3 ± 1.4	0.5 ± 1.1
t-value (with no to change 50:50 offers)	5.374	4.260	2.824
Significance	.000***	.000***	.004**
Not to change 70:30 offers	−2.1 ± 2.3	−1.7 ± 2.4	−2.4 ± 2.0
t-value (with no to change 50:50 offers)	6.117	5.198	7.140
Significance	.000***	.000***	.000***
Not to change 90:10 offers	−2.6 ± 3.4	−1.7 ± 3.5	−2.5 ± 3.6
t-value (with no to change 50:50 offers)	5.27	3.907	4.537
Significance	.000***	.000***	.000***

Note. From the *t*-test, the ERN amplitude of not to change 50:50 offers was significantly more positive than that of each other condition at almost every site. *$p < .05$. **$p < .01$. ***$p < .001$. One-tailed.

bution of 70:30, thinking that it was not as unfair as the 90:10 offers, they took into account their own monetary loss and decreased the frequency of their decision to change.

Discussion

The current study employed a third-party punishment task to investigate the processing course of conflicts evoked by altruistic punishment in the executive stage. As participants will face intense economic and emotional conflicts when making decisions, we assumed that conflicts induced by altruistic punishment choices might evoke Ne/ERN. The experimental results supported our assumption.

Behaviorally, for the 90:10 trials participants made more to change the offer choices, while for the 70:30 trials participants made more not to change the offer choices. It indicated that when confronted with increasingly unfair offers, participants made to change the offer choices more often, which was consistent with previous studies.

For the ERP results, we found an Ne/ERN-like component occurred for both choices after the response, which peaked at approximately 0–40 ms and was most strongly pronounced at frontocentral midline scalp sites. The not to change the offer choice elicited a larger negative component than the to change the offer choices (see Figure 2).

In the present study, the outcome for either to change the offer or not to change the offer was predictable prior to feedback. For the participants, without external judgment criteria, the decision to press the F or J key was subjectively intended, so it seems difficult to interpret the Ne/NRN-like component in terms of error detection theory. However, the conflict between the to change with monetary loss and the not to change, while not punishing the violation of a fairness norm, was evident during the decision-making process. In line with an extended version of conflict monitoring theory (van Veen, Cohen, Botvinick, Stenger, & Carter, 2001), the ACC could monitor conflicts between different internal desires or plans in response selection, except for conflicts occurring between different S-R mappings (Yu & Zhou, 2009).

The current results are consistent with conflict monitoring theory, which holds that ERN reflects conflict during response selection. In the present study, Ne/ERN was revealed in both choices, and the not to change the offer choices elicited a more negative Ne/ERN than did the to change the offer choices, which was in accordance with our expectation. First of all, in our study, when making to change the offer choices, participants confronted a conflict due to their monetary loss, while when making not to change the offer choices they confronted a conflict due to the fairness norm. As a result, both choices elicited ERN effects. Second and more importantly, as we supposed in the introduction, if the emotion and morality that drove the participants to punish the violator of social norms dominated in the early stage of the decision-making process, a choice not to punish in violation of these emotions would induce greater conflict and evoke more negative Ne/ERN amplitudes, as we found. Conversely, if economic benefit-oriented thinking was more dominant, the choice of punishment with benefit loss would induce greater conflict and more negative Ne/ERN amplitudes. The ERP results of a larger ERN being elicited by the choice not to punish supported the view that the satisfaction of emotion rather than economic benefit was more important at the stage of decision making. This was in accordance with neuroimaging studies in which participants viewed the third-party punishment, connected with the emotion, as a reward instead of as punishment, connected with benefit (Delgado et al., 2003; Strobel et al., 2011). Behavioral studies have found that participants were more willing to punish their partners who made unfair proposals, even at a personal cost. More than half the third parties in the trust game punished the trustees (who returned less than 50% of the total resource to the trustors; Fehr & Fischbacher, 2004; Ohtsubo et al., 2010). Another study on altruistic third-party punishment found that third-parties would like to punish the dictator strongly when they became the richest even for cases of equitable allocation (Leibbrandt & López-Pérez, 2011). This showed a dark side of altruistic third-party punishment, in that third-parties perhaps reduced the money of the dictators on account of the comparison. Some studies also found that altruistic punishment reflected the absence rather than the presence of self-control (Crockett, Clark, Lieberman, Tabibnia, & Robbins, 2010; Koenigs & Tranel, 2007; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003; Tabibnia, Satpute, & Lieberman, 2008), which indicated altruistic punishment was an uncontrolled behavior. Such evidence suggests that emotional factors are uncontrolled and automatic when people make a decision, providing great power to punish and making third-party punishment decisions spontaneous and unconscious. In contrast, not-punishment behaviors are cogitative and conscious. In other words, punishment was the automatic and prepotent tendency when making the decision. In our experiment, when making the choice not to change the offer, participants inevitably confronted high response conflict between the intended not-punishment responses and their automatic punishment response tendencies, reflected by more negative Ne/ERN amplitudes. However, participants faced low response conflict when making punishment key presses because consciously intended responses were consistent with prepotent response tendencies, reflected by less negative Ne/ERN amplitudes.

In addition, Ne/ERN amplitudes in the current study were not modulated by the level of unfairness. According to conflict monitoring theory, Ne/ERN amplitudes for changing 70:30 offers should have been larger than those for changing 90:10 offers, because the rate of changing 70:30 offers was smaller than that of not changing them. Conversely, participants preferred to change 90:10 offers. Indeed, the ERP result did not reflect the behavior completely and the present data could not provide an explanation for this finding. A likely interpretation, we argue, was that in addition to prompting response conflicts, processing unfairness information cost both cognitive and emotional resources, making it too complicated to handle for Ne/ERN. At the moment of making the decision, pressing the F or J key that represented changing the offer or not changing the offer was the direct response selection for participants. So the conflicts of response selection were detected by the ACC and reflected by Ne/ERN. However, the level of unfairness was a further factor behind the selection, becoming too difficult to be detected by the ACC in the early phase.

In conclusion, the current study paid close attention to third-party experiences during the decision-making process, particularly focusing on selection conflicts. The ERP result showed that Ne/ERN amplitudes were more negative for the not to change the offer choice, which might support the view that punishing a violation was the prepotent tendency for third parties, and people were more driven by emotion in the early stage of response judgment.

References

Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2008). Individual differences in the regulation of intergroup bias: The role of conflict monitoring and neural signals for control. *Journal of Personality and Social Psychology*, *94*(1), 60–74. doi:10.1037/0022-3514.94.1.60

Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartley, S. L., & Covert, A. E. (2004). Neural signals for the detection of unintentional race bias. *Psychological Science*, *15*(2), 88–93. doi:10.1111/j.0963-7214.2004.01502003.x

Bernstein, P. S., Scheffers, M. K., & Coles, M. G. H. (1995). "Where did I go wrong?" A psychophysiological analysis of error detection. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(6), 1312–1322. doi:10.1037/0096-1523.21.6.1312

Boksem, M. A. S., Tops, M., Kostermans, E., & De Cremer, D. (2008). Sensitivity to punishment and reward omission: Evidence from error-related ERP components. *Biological Psychology*, *79*(2), 185–192. doi:10.1016/j.biopsycho.2008.04.010

Botvinick, M. M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive, Affective, and Behavioral Neuroscience*, *7*(4), 356–366. doi:10.3758/CABN.7.4.356

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*(12), 539–546. doi:10.1016/j.tics.2004.10.003

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, *4*(6), 215–222. doi:10.1016/S1364-6613(00)01483-2

Coles, M. G. H., Scheffers, M. K., & Holroyd, C. B. (2001). Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. *Biological Psychology*, *56*(3), 173–189. doi:10.1016/S0301-0511(01)00076-X

Crockett, M. J., Clark, L., Lieberman, M. D., Tabibnia, G., & Robbins, T. W. (2010). Impulsive choice and altruistic punishment are correlated and increase in tandem with serotonin depletion. *Emotion*, *10*(6), 855–862. doi:10.1037/a0019861

Delgado, M. R., Locke, H. M., Stenger, V. A., & Fiez, J. A. (2003). Dorsal striatum responses to reward and punishment: Effects of valence and magnitude manipulations. *Cognitive, Affective, and Behavioral Neuroscience*, *3*(1), 27–38. doi:10.3758/CABN.3.1.27

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: A tutorial. *Biological Psychology*, *51*(2–3), 87–107. doi:10.1016/S0301-0511(99)00031-9

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*(2), 63–87. doi:10.1016/S1090-5138(04)00005-4

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., . . . Ziker, J. (2006). Costly punishment across human societies. *Science*, *312*(5781), 1767–1770. doi:10.1126/science.1127333

Hewig, J., Kretschmer, N., Trippe, R. H., Hecht, H., Coles, M. G. H., Holroyd, C. B., & Miltner, W. H. R. (2011). Why humans deviate from rational choice. *Psychophysiology*, *48*(4), 507–514. doi:10.1111/j.1469-8986.2010.01081.x

Hirsh, J. B., & Inzlicht, M. (2010). Error-related negativity predicts academic performance. *Psychophysiology*, *47*(1), 192–196. doi:10.1111/j.1469-8986.2009.00877.x

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., Nystrom, L., Mars, R. B., Coles, M. G. H., & Cohen, J. D. (2004). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nature Neuroscience*, *7*(5), 497–498. doi:10.1038/nn1238

Holroyd, C. B., Yeung, N., Coles, M. G. H., & Cohen, J. D. (2005). A mechanism for error detection in speeded response time tasks. *Journal of Experimental Psychology: General*, *134*(2), 163–191. doi:10.1037/0096-3445.134.2.163

Koenigs, M., & Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: Evidence from the Ultimatum Game. *The Journal of Neuroscience*, *27*(4), 951–956. doi:10.1523/JNEUROSCI.4606-06.2007

Leibbrandt, A., & López-Pérez, R. (2011). The dark side of altruistic third-party punishment. *Journal of Conflict Resolution*, *55*(5), 761–784. doi:10.1177/0022002711408010

Ohtsubo, Y., Masuda, F., Watanabe, E., & Masuchi, A. (2010). Dishonesty invites costly third-party punishment. *Evolution and Human Behavior*, *31*(4), 259–264. doi:10.1016/j.evolhumbehav.2009.12.007

Olvet, D. M., & Hajcak, G. (2008). The error-related negativity (ERN) and psychopathology: Toward an endophenotype. *Clinical Psychology Review*, *28*(8), 1343–1354. doi:10.1016/j.cpr.2008.07.003

Olvet, D. M., & Hajcak, G. (2009). Reliability of error-related brain activity. *Brain Research*, *1284*, 89–99. doi:10.1016/j.brainres.2009.05.079

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), 1755–1758. doi:10.1126/science.1082976

Scheffers, M. K., & Coles, M. G. H. (2000). Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 141–151. doi:10.1037/0096-1523.26.1.141

Strobel, A., Zimmermann, J., Schmitz, A., Reuter, M., Lis, S., Windmann, S., & Kirsch, P. (2011). Beyond revenge: Neural and genetic bases of altruistic punishment. *NeuroImage*, *54*(1), 671–680. doi:10.1016/j.neuroimage.2010.07.051

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), 339–347. doi:10.1111/j.1467-9280.2008.02091.x

van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger, V. A., & Carter, C. S. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing. *NeuroImage*, *14*(6), 1302–1308. doi:10.1006/nimg.2001.0923

Wu, Y. 吴燕, & Luo, Y.-J. 罗跃嘉. (2011). 利他惩罚中的结果评价——ERP 研究 [The outcome evaluation in the altruistic punishment: An ERP Study]. *Acta Psychologica Sinica* [心理学报], *43*(6), 661–673.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, *111*(4), 931–959. doi:10.1037/0033-295X.111.4.931

Yeung, N., & Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *The Journal of Neuroscience*, *24*(28), 6258–6264. doi:10.1523/JNEUROSCI.4537-03.2004

Yu, R., & Zhou, X. (2009). To bet or not to bet? The error negativity or error-related negativity associated with risk-taking choices. *Journal of Cognitive Neuroscience*, *21*(4), 684–696. doi:10.1162/jocn.2009.21034.