The interplay between feedback-related negativity and individual differences in altruistic punishment: An EEG study

Hendrik Mothes[1,2] · Sören Enge[1] · Alexander Strobel[1]

Abstract To date, the interplay betwexen neurophysiological and individual difference factors in altruistic punishment has been little understood. To examine this issue, 45 individuals participated in a Dictator Game with punishment option while the feedback-related negativity (FRN) was derived from the electroencephalogram (EEG). Unlike previous EEG studies on the Dictator Game, we introduced a third party condition to study the effect of fairness norm violations in addition to employing a first person perspective. For the first time, we also examined the role of individual differences, specifically fairness concerns, positive/negative affectivity, and altruism/ empathy as well as recipients' financial situation during altruistic punishment. The main results show that FRN amplitudes were more pronounced for unfair than for fair assignments in both the first person and third party perspectives. These findings suggest that FRN amplitudes are sensitive to fairness norm violations and play a crucial role in the recipients' evaluation of dictator assignments. With respect to individual difference factors, recipients' current financial situation affected the FRN fairness effect in the first person perspective, indicating that when being directly affected by the assignments, more affluent participants experienced stronger violations of expectations in altruistic punishment decisions. Regarding individual differences in trait empathy, in the third party condition FRN amplitudes were more pronounced for those who scored lower in empathy. This may suggest empathy as another motive in third party punishment. Independent of the perspective taken, higher positive affect was associated with more punishment behavior, suggesting that positive emotions may play an important role in restoring violated fairness norms.

Keywords Altruistic punishment · Dictator Game · Third party · Fairness · Empathy · Feedback-related negativity

Introduction

Unlike other animal species, humans largely depend on cooperation among genetically unrelated individuals (Fehr & Fischbacher, 2004b). Existing explanations for cooperative and altruistic behavior usually refer to a genetic relationship or advantages emerging from altruistic behavior in repeated interactions such as fitness benefits through reputation building (Bowles & Gintis, 2004; Fehr & Fischbacher, 2003). However, an intriguing question is: Why do individuals show altruistic behavior in interactions with genetically unrelated people that they do not know and are unlikely to meet again?

Various studies have demonstrated that the pattern of human altruism is highly specific to our species as well as observable across diverse human societies (Fehr & Fischbacher, 2003; Henrich et al., 2006). In an attempt to find an explanation for human altruism, Bowles and Gintis (2004) proposed a model of cooperation and punishment called strong reciprocity. They suggest that unrelated groups sustain a high level of

---

✉ Hendrik Mothes
hendrik.mothes@sport.uni-freiburg.de

✉ Sören Enge
soeren.enge@tu-dresden.de

[1] Department of Psychology, Technische Universitaet Dresden, 01062 Dresden, Germany

[2] Institute of Sport and Sport Science, University of Freiburg, Freiburg im Breisgau, Germany

cooperation via two behavioral characteristics of human altruism, so-called altruistic rewarding and altruistic punishment. Altruistic rewarding is defined as the propensity to reward others for cooperative behavior. Similarly, altruistic punishment is defined as the tendency to punish norm violations of others without personal benefit for the punisher, but with potential benefit for other individuals. Although it is costly, humans engage in such behavior in order to sustain cooperation (Fehr & Fischbacher, 2003).

In recent years, evidence from various studies showed that cooperation is sustained by altruistic punishment both in one-shot interactions and in larger groups (Boyd, Gintis, Bowles, & Richerson, 2003). Despite several endeavors to identify underlying motives, brain processes, and genetic underpinnings (Fehr & Gächter, 2002; de Quervain et al., 2004; Strobel et al., 2011), our understanding of the driving forces for altruistic punishment is still limited, and thus further research on altruistic punishment is needed.

Typically, altruistic punishment is investigated in experimental games used in psychology and behavioral economics for evaluating aspects of social norms and preferences. For instance, both the Ultimatum Game and the Dictator Game can be used to investigate altruistic punishment. In a classical Dictator Game, one player (called the dictator) receives an amount of money that he or she can divide among him- or herself and another player (called the recipient) and the recipient cannot reject the proposed assignment; in an Ultimatum Game the assignment can be rejected (Güth, Schmittberger, & Schwarze, 1982). In order to measure altruistic punishment with a Dictator Game, a particular variant can be played, where the recipient has the opportunity to punish unfair allocations at the expense of his payoff. Additionally, the allocations in such a Dictator Game can either be examined when individuals are directly affected by the assignment – from a first person perspective – or when they observe someone receiving an assignment from a third party perspective (Fehr & Fischbacher, 2004a; Strobel et al., 2011).

Traditional game theory that is based on the concept of *Homo Oeconomicus* and that addresses the behavior of different actors in social conflict situations neither predicts substantial offers in the Ultimatum Game or Dictator Game scenarios nor first person or third party punishment (Holler & Illing, 2006). Empirically, however, it has been shown that dictators in the Dictator Game offer recipients on average 25 % of the total sum (Camerer, 1997). Moreover, it has been observed that dictators were punished for allocations which were less than half of the original sum both by players from a first person and third party perspective (Fehr & Fischbacher, 2004a; Strobel et al., 2011).

From a psychological perspective, social norms and emotions as well as other individual difference factors have been suggested to explain altruistic punishment behavior. On the one hand, altruistic punishment has been discussed as an act of social norm enforcement requiring self-control, most prominently the fairness norm (Fehr & Fischbacher, 2004b; Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006; Strobel et al., 2011). On the other hand, others proposed that altruistic punishment decision-making appears to be rather an impulsive act driven by emotions (Crockett, Clark, Lieberman, Tabibnia, & Robbins, 2010; Pillutla & Murnighan, 1996). Specifically, emotions of both negative and positive valence are suggested to play a role in impulsive altruistic punishment processes. For instance, negative emotions such as anger or disgust that emerge when fairness norms are violated were proposed as a reason for punishment (Mikula, Scherer, & Athenstaedt, 1998; van't Wout, Kahn, Sanfey, & Aleman, 2006; Xiao & Houser, 2005). In the same way, positive emotions in the form of anticipated satisfaction through revenge-like motives or a legitimate intervention were indicated to be decisive for altruistic punishment (de Quervain et al., 2004; Strobel et al., 2011). Furthermore, several studies emphasize the role of individual differences in personality in altruistic punishment behavior. For instance, Strobel and colleagues (2011) reported that higher scores of altruism measured by the Revised NEO Personality Inventory (Costa & McCrae, 1992) were accompanied by elevated third party punishment in a Dictator Game. Related to the concept of altruism, Singer and Steinbeis (2009) suggest empathy, particularly empathic concern/compassion, to be predictive for altruistic punishment from a third party perspective.

Altruistic punishment in neuroscience

Neuroscience studies recently started to investigate brain processes involved in playing social exchange games. Sanfey, Rilling, Aronson, Nystrom, and Cohen (2003) conducted one of the first functional magnetic resonance imaging (fMRI) studies examining neural correlates of economic decision making. Using an Ultimatum Game paradigm, they showed that activations in the dorsolateral prefrontal cortex (DLPFC), the anterior cingulate cortex (ACC), and the insula were associated with obtaining unfair offers. The authors interpreted their findings as a result of the elicited conflict between emotional ("rejecting") and cognitive ("accepting") motives to the unfair offers. This is underscored by a vast body of literature (for a review see Miller & Cohen, 2001) indicating the involvement of ACC and DLPFC in a common network of cognitive control, in which ACC serves as a monitor of response conflict that recruits DLPFC for the reduction of such conflicts. Similarly relevant for altruistic punishment, ACC activity is also considered to reflect aversive consequences in social contexts (Eisenberger, Lieberman, & Williams, 2003), and is critically involved in the processing of negative outcomes that do not meet expectations (reward prediction errors) (Amiez, Joseph, & Procyk, 2005; Matsumoto, Matsumoto, Abe, & Tanaka, 2007). Further, ACC activation has been found to predict the degree of decision-making effort (Paus, Koski, Caramanos, & Westbury, 1998).

Although research has begun to unveil the significance of the ACC in altruistic punishment, to date little is known about the role of the ACC during the decision-making process underlying altruistic punishment. Given that altruistic punishment-related decision processes are relatively fast, the use of electroencephalography (EEG) providing high temporal resolution would be useful to gain further insights into the underlying processes over the time course of an altruistic punishment decision.

In particular, a class of negative event-related potentials (ERPs) with a fronto-central scalp distribution that originate from activity in the ACC and that have been associated with the processing of conflict and reward prediction errors (Dehaene, Posner, & Tucker, 1994; Gehring & Willoughby, 2002) appear to be well suited for the investigation of altruistic punishment decisions. In the present context, this is particularly true for the so-called feedback-related negativity (FRN; also referred to as medial frontal negativity, or MFN) that is elicited around 250–300 ms after individuals obtained feedback about outcomes that are worse than expected (Holroyd & Krigolson, 2007) or received feedback about negative performance (Holroyd & Coles, 2002; Miltner, Braun, & Coles, 1997). Furthermore, more recent findings indicate that an FRN is elicited even when feedback is surprising or unexpected regardless of the affective valence of feedback (Hauser et al., 2014; Oliveira, McDonald, & Goodman, 2007). Moreover, recent studies show that by observing feedback on another person's decisions or actions an observer FRN is evoked (e.g., Fukushima & Hiraki, 2009; Yu & Zhou, 2006; see also Koban & Pourtois, 2014) demonstrating that the FRN can be measured reliably in a third person condition.

The Feedback-Related Negativity (FRN) in economic decision-making

Specifically, one functional interpretation of FRN is related to the motivational or affective evaluation of negative outcomes (Boksem, Tops, Kostermans, & De Cremer, 2008; Gehring & Willoughby, 2002). More recently, this line of research has been extended to decision making in the social domain where it has been shown that recipients in an Ultimatum Game exhibited a larger FRN to unfair as compared to fair offers made by one and the same proposer (Polezzi et al., 2008) or by different proposers, particularly in those individuals with high fairness concerns (Boksem & De Cremer, 2010), and that the FRN amplitude is related to the decision to reject unfair offers by the proposer (Hewig et al., 2011). This may indicate a relationship between FRN and the adherence of a fairness norm. Consequently, Boksem and De Cremer (2010) speculate that the primary concern of individuals in bargaining situations is fairness rather than their own outcome.

However, as the participants of the mentioned Ultimatum Game studies were directly affected by unfair assignments of the proposer, the influence of other motives such as revenge was not eliminated. While satisfaction through revenge is potentially relevant only for first person punishment (Strobel et al., 2011), previous research suggests that fairness norms play an important role both in first person and third party punishment games (Civai, Corradi-Dell'Acqua, Gamer, & Rumiati, 2010; Fehr & Fischbacher, 2004a). Thus, using the FRN as an indicator for deviations from a fairness norm, we were interested in whether a fairness effect of FRN is observable in a Dictator Game setting both from a first person perspective and from a third party perspective.

Furthermore, the FRN has been related to reinforcement learning. With regard to the reinforcement learning theory of FRN, it is suggested that more negative FRN amplitudes following outcomes that are worse than expected (i.e., a prediction error) – or even following face cues that signal negative outcomes – are accompanied by reduced reward expectations which in turn can lead to behavioral adjustments (Holroyd & Coles, 2002; Osinsky, Mussel, Ohrlein, & Hewig, 2014). Indeed, several results indicate an association between changes in FRN amplitude and subsequent behavioral changes (Gehring, Goss, Coles, Meyer, & Donchin, 1993; Yeung & Sanfey, 2004). Therefore, we expected that if altruistic punishment behavior is dependent on the violation of a fairness norm, larger FRN amplitudes following unfair assignments will predict altruistic punishment behavior in the context of the Dictator Game relatively independent from the perspective taken.

Furthermore, to better understand altruistic punishment behavior, it is important to advance knowledge of the interplay between relevant individual difference factors and FRN amplitude during altruistic punishment scenarios. Although there is initial research suggesting associations between FRN and individual differences (see Koban & Pourtois, 2014), it remains largely unknown how individual differences play a role in the modulation of the FRN during altruistic punishment situations. Since altruistic punishment behavior has been associated with individual fairness norms (Fehr & Fischbacher, 2004b), altruism (Strobel et al., 2011), and empathy (Singer & Steinbeis, 2009), we aimed at better understanding whether and how these individual difference factors are associated with FRN amplitude. Given that larger FRN amplitudes were observed to unfair offers in an Ultimatum Game when participants valued moral norms highly (Boksem & De Cremer, 2010), we expected to also find more pronounced FRN amplitudes in subjects with a higher fairness norm. Furthermore, Fukushima and Hiraki (2009) reported that empathy scores were positively correlated with the observer FRN amplitude in a gambling task paradigm. Because empathy is assumed to play a role in altruistic punishment from a third party perspective, we expected a positive relation between empathy and FRN amplitude, particularly in third party punishment scenarios.

Research questions

Overall, our main research questions concerned the investigation of the interrelationship between altruistic punishment behavior, individual difference variables, and the ACC-modulated FRN component. For this purpose, individuals participated in a Dictator Game with first person and third party punishment option while the EEG was recorded. With respect to the FRN following the initial assignments by the dictators, we expected (a) larger FRN amplitudes to unfair than to fair assignments independently of the perspective (first- person vs. third-party) taken, (b) larger FRN amplitudes following unfair assignments would predict altruistic punishment behavior both in first person and third party punishment, (c) more pronounced FRN amplitudes for individuals with higher fairness norms, and (d) for third party scenarios higher altruism and/or empathy to be related with increased FRN amplitudes. Furthermore, it was of interest to replicate previous behavioral findings concerning altruistic punishment. Specifically, we examined (a) whether participants would punish the more and evaluate fairness the less, the more money dictators retained for themselves, (b) whether altruistic punishment as well as fairness perceptions of different assignments would be relatively independent from the perspective taken by the recipient (first-person vs. third-party), and (c) whether a higher individual fairness norm would be linked to larger altruistic punishment. To elucidate the role of individual differences in altruistic punishment behavior, we tested whether altruistic punishment behavior can be explained by individual differences such as altruism and empathy. Particularly with respect to third party punishment, we expected a positive association between both altruism and/or the altruism-related trait of empathy with altruistic punishment. Lastly, to further investigate the role of emotions in altruistic punishment behavior, we examined whether more positive or negative affect would result in higher altruistic punishment independently of the perspective taken by the participant.

Methods

Participants

Forty-five right-handed participants between 18 and 34 years of age (12 males, M age $\pm$ SD: 22.4 $\pm$ 3.8) were recruited from the student population of the Technische Universitaet Dresden. They were remunerated for participating according to the decisions made in the Dictator Game (on average: 12.42 $\pm$ 1.18 €; range: 10.00–15.00 €). All participants had normal or corrected-to-normal vision, confirmed to be free of any relevant health problems, and did not abuse drugs. Additionally, we assessed each subject's sleep duration and nicotine, caffeine, and alcohol consumption in the last 24 h

via self-report. Note that an *a priori* power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that given four repeated measures (fair vs. unfair trials, first person vs. third party perspective), 45 participants would be required to achieve a power level of .80, assuming a medium effect size (Cohen's $d = .50$), with an alpha at .05.

Procedure

At the beginning of the experimental session in the laboratory, subjects were informed about the study, gave written informed consent for participation, and received several questionnaires assessing personality states and traits. In order to measure the current states of positive and negative affect, we employed the German state version of the Positive and Negative Affect Schedule (PANAS; Krohne, Egloff, Kohlmann, & Tausch, 1996) with four added items representing positive (cheerful and good-humored) and negative emotions (depressed and dejected) as proposed by Schimmack (2003). Moreover, we used the altruism facet scale of the German version of the Revised NEO Personality Inventory (NEO-PI-R; Ostendorf & Angleitner, 2003) to measure altruism. The German version of the Interpersonal Reactivity Index (SPF; Paulus, 2009) was employed to assess empathy, which contained the following four subscales: fantasy, perspective taking, empathic concern, and personal distress. As more general measures for empathy, we calculated an empathy score based on the sum of the subscales fantasy, empathic concern, and perspective taking, and an empathy overall score (empathy score – personal distress, see also Paulus, 2012). Note that one participant was excluded from all analyses concerning the altruism facet scale due to a missing questionnaire. Furthermore, we surveyed the participants' subjective financial situation ("How do you evaluate your current financial situation?") on a 4-point Likert scale (1 = very favorable, 4 = very unfavorable) and their objective financial situation ("How much money do you have available each month?") on a 4-point scale (1 = below 500 €; 2 = between 500 and 650 €; 3 = between 650 and 800 €; 4 = more than 800 €) based on official statistics on the financial situation of German students (Middendorf, Apolinarski, Poskowsky, Kandulla, & Netz, 2013). For further analysis subjective financial situation was recoded.

In the experimental session, all participants were seated in front of a computer screen in an acoustically shielded EEG cabin. Participants were instructed that the first EEG recording was to be done during a 4-min resting period with closed (2 min) and open (2 min) eyes. Subsequently, participants obtained written instruction for the Dictator Game. They were accurately informed that the dictators in the Dictator Game were real students who had participated earlier and that both (participants and dictators) would be compensated according to their financial decisions made (see supplementary material). Subsequently, they had the opportunity to ask questions and to

complete 12 practice trials before the experimental Dictator Game started. A session consisted of a Dictator Game-run in first person perspective and a Dictator Game-run in third party perspective as well as three short cognitive tasks that, however, were part of another study and will be reported elsewhere. The order of Dictator Game-runs was counterbalanced across subjects, so that half of the subjects started with first person perspective and the other half with third party perspective.

Dictator Game

In both first person and third party perspective of the Dictator Game, participants encountered 100 assignments, in which the dictators (named player A) had decided how to allocate 20 € between themselves and a recipient. These assignments had been obtained earlier from students of the Technische Universitaet Dresden. These students were asked how they would divide an amount of 20 € between themselves and another participant of the study or a third, uninvolved person. Out of N = 131 participating students, we selected for each perspective a set of 100 students. The basis for this selection was to have just as many fair assignments than unfair assignments (50:50) in both the first person and third party perspective while ensuring that the chosen assignments largely mirror the distribution of assignments in the original sample. Assignments to recipients of more than 6 € out of 20 € were considered as fair, the remaining ones as unfair according to the classification used in previous studies (Fehr & Fischbacher, 2004a; Strobel et al., 2011).

Whereas in the first person perspective, the participant (acting as player B) was the recipient, in the third party perspective a fictitious, uninvolved player (named player C) was the recipient. Hence, in the first person perspective, subjects (players B) directly received the offered assignments of dictators (players A), whereas in the third party perspective, participants only observed the allocation of player A to player C. Each dictator assignment constituted one trial. During 50 out of 100 trials in each perspective, subjects had the opportunity to assess the degree of fairness of the assignment using a scale ranging from −2 (very unfair) to 2 (very fair). Note that we used subjective fairness evaluations of each individual to calculate a hypothetical assignment that would be assessed as neither fair nor unfair in each perspective. More specifically, for each participant (and each perspective) we first calculated a non-linear function that best describes the responses to the discrete assignments by the Dictator. Then, we computed zeros of these functions in order to obtain such hypothetical assignments for each individual. In our opinion, these measures can serve as a behavioral operationalization for an individual fairness norm. In the other 50 trials, subjects were able to punish player A for unfair assignments to themselves (first person perspective) or player C (third party perspective) by allocating zero to four punishment points (the occurrence

of fairness and punishment trials was pseudo-randomized within both the blocks of first person and third party perspectives). Note that player C (third party perspective) was fictitious and did not have an opportunity to punish unfair assignments; only player B observing this assignment had the chance to punish unfair decisions. Subjects were informed that in each punishment trial, they would receive four punishment points and that each punishment point assigned would reduce the outcome of the corresponding player A by 2.50 € (resulting in a 10 € reduction for player A when four punishment points were assigned). Furthermore, participants were briefed that their compensation would be determined according to the averaged amount of punishment points *not assigned* (specifically, they would receive twice the value of averaged punishment points not assigned, each punishment point worth 1 €) plus the average assigned amount from player A (for an example, see supplementary material).

Each trial proceeded as described in Fig. 1. In general, there were 50 trials within each of the four conditions: first person perspective/fairness evaluation, first person perspective/punishment option, third party perspective/fairness evaluation, and third party perspective/punishment option. Each condition comprised 25 fair and 25 unfair trials. Please note that within each perspective, fair and unfair trials were randomly allocated to the conditions *fairness evaluation* or *punishment option*. Average assignments of fair and unfair trials did not differ between first person and third party perspectives. Table 1 illustrates the number of dictator: recipient assignments (in €) in both the first person and the third party perspectives. Note that in both perspectives distributions of assignments were not equal as they reflect the natural assignments of the dictators. Moreover, in each perspective a few dictators assigned themselves less than the recipient (e.g., in first person perspective the 7:13 assignment). In order to use realistic scenarios in this study, we decided to maintain all of the assignments as allocated by the dictators.

EEG recording and pre-processing

EEG data were continuously recorded (band-pass 0.1–30 Hz, sampling rate 500 Hz) with Ag/AgCl electrodes on 29 standardized scalp sites (Fp1, Fp2, AF3, AF4, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, O2), affixed to an electrode cap from EasyCap (EASYCAP GmbH, Herrsching, Germany) using BrainVision Recorder software (Version 1.3 Brainproducts, GmbH, Munich, Germany). Additionally, two linked electrodes at left and right mastoids served as reference and AFz as ground. In order to measure ocular artifacts, vertical electro-oculogram (VEOG) was recorded with an electrode placed under the right eye and horizontal EOG from two electrodes placed at the outer canthi of both eyes. Impedances were kept below 5 kΩ.
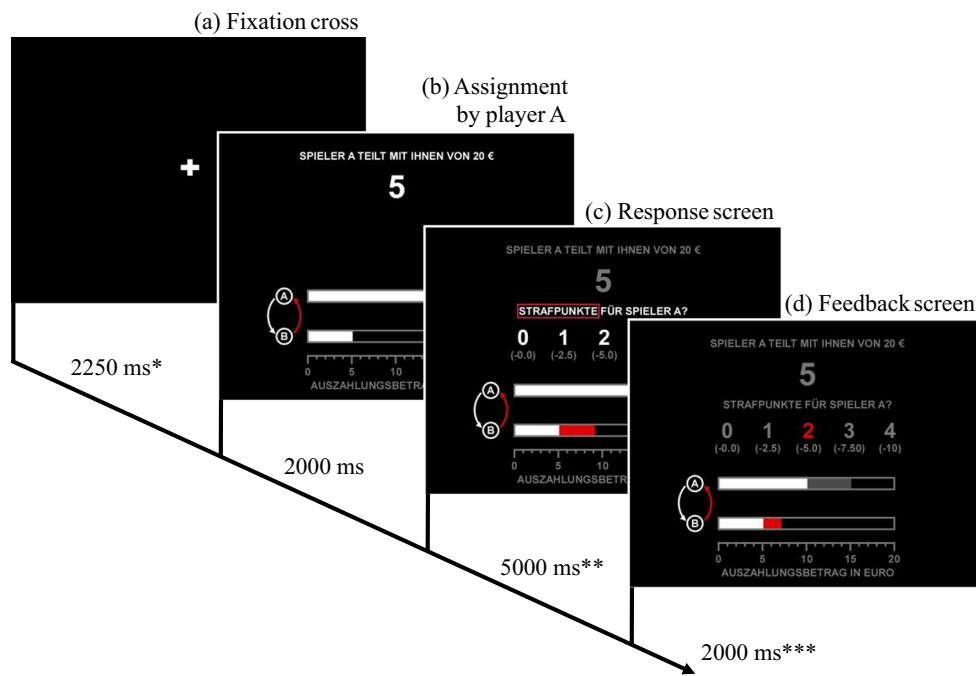
Fig. 1 Example of the time course for a trial (15:5 assignment) with punishment option in the Dictator Game from first person perspective. (**a**) Each trial started with the presentation of a fixation cross for 2,250 ms*. (**b**) Sequentially, subjects saw the assignment (here 5 €) of one player A for 2,000 ms. (**c**) Following, subjects had the opportunity for 5,000 ms** either to assign punishment points (shown in this example) or to evaluate the fairness of the assignment (not shown in this example) by pressing the appropriately labeled button on the keyboard (e.g., red "1" for assigning one punishment point). For better differentiability and analogous to the button labeling, punishment trials displayed red frames on the screen (see screen **c**), while trials asking for fairness evaluations displayed green frames on the screen. Finally, subjects were presented for 2,000 ms*** with a feedback screen (**d**) revealing the outcomes of player A, player B, and player C (only in the third party perspective, not shown here), then the next trial containing the decision of another player A started. * 2,250 ms (range 2,000–2,500 ms). ** 5,000 ms is the maximal decision time for player B. *** 2,000 ms + (5,000 ms – decision time used) (for colored illustration see electronic version of this article)

BrainVision Analyzer 2.0 software was used for stimulus-locked FRN analyses. The EEG data were segmented for 2,000 ms in the epochs "fair assignments" and "unfair assignments" starting 200 ms before presentation of the assignment. Ocular movements' artifacts were corrected using the Gratton and Coles algorithm implemented in Brain Vision Analyzer 2.0 and artifactual channels exceeding ± 100 μV were selected and excluded from averaging (on average < 10 %). The epochs were aligned to the 200 ms baseline before onset of the stimuli. Lastly, EEG data were averaged separately for each participant and each experimental condition. For the FRN analysis, mean amplitudes were used. Visual inspection of grand averages elicited by fair and unfair assignments suggested that the FRN obtained its maximum around 300 ms after presentation of an assignment over multiple frontal electrodes concentrated around Fz (see Fig. 3B). This fits with previous observations that showed maximum FRN amplitudes at frontal electrodes (e.g., Gehring & Willoughby, 2002), notably in studies examining altruistic punishment using an Ultimatum Game or Dictator Game paradigm (Boksem & De Cremer, 2010; Hewig et al., 2011). Thus, data from the Fz electrode in a 270–330 ms time window were used for further analyses.

Statistical analyses

All analyses were performed using SPSS 18. First, using Friedman tests, we examined the expected relationship between increasing monetary assignments and raising fairness evaluations as well as decreasing invested punishment points.

Table 1 Dictator-recipient assignments for the first person and third party perspective (in €)

Perspective	5:15	7:13	8:12	9:11	10:10	11:9	12:8	13:7	14:6	15:5	16:4	17:3	18:2	20:0	Σ
First person	-	1	1	-	22	2	17	7	3	26	-	3	2	16	100
Third party	1	-	-	1	32	3	10	3	6	13	4	4	3	20	100

Note. Numbers represent the quantity of dictator: recipient assignments in the respective perspective

- indicates that no assignment was obtained

To investigate whether differences in altruistic punishment and in subjective fairness evaluations between the first person and third party perspectives can be found, we secondly performed Wilcoxon tests using the average of punishment points assigned over all trials with punishment option for each perspective. Moreover, to examine the relationship between altruistic punishment and individual fairness norms as well as other individual differences, we calculated Spearman correlations between these variables. With regard to the electrophysiological hypotheses, we conducted a 2 (fairness: unfair, fair trials) × 2 (perspective: first person, third party) repeated-measures analysis of variance (ANOVA) for FRN amplitudes. Out of the 100 assignments per perspective (conditions *fairness evaluation* and *punishment option* combined), on average, 98.6 of unfair and 98.4 of fair artifact-free trials at Fz electrode entered FRN analysis for the first person perspective (for the third party perspective 98.8 and 98.6, respectively). FRN amplitudes were further used in nonparametric correlation analyses (Spearmans' *rho*) with subsequent punishment behavior as well as individual differences variables such as NEO altruism, empathy, and state positive/negative affectivity.

Furthermore, we controlled for confounding effects such as sex, participants' financial situation, strategy use, sleep duration, and nicotine, caffeine, and alcohol consumption. Nonparametric Spearman correlation analyses revealed a significant correlation between subjective financial situation and punishment behavior averaged across the two perspectives, $r = .33$, $p = .027$. Additionally, FRN amplitudes across fair and unfair assignments during the first person perspective were negatively associated with both the subjective financial situation, $r = -.35$, $p = .017$, and similarly, the objective financial situation, $r = -.36$, $p = .015$, indicating that a better financial background leads to larger FRN amplitudes. As this points to the relevance of an individual's financial background for altruistic punishment behavior, we controlled for financial situation in the FRN repeated-measures ANOVA and the correlation analyses, when indicated. None of the other variables were significantly associated with independent and/or dependent variables (all $ps > .05$). Throughout the study, two-tailed p-values are reported.

Results

Behavior

As expected, participants invested the less punishment points, the more money was assigned to them (first person perspective), $\chi^2 = 227.32$, $df = 7$, $p < .001$, or to the fictitious uninvolved player (third party perspective), $\chi^2 = 345.59$, $df = 10$, $p < .001$ (see Fig. 2A). In general, these results are in
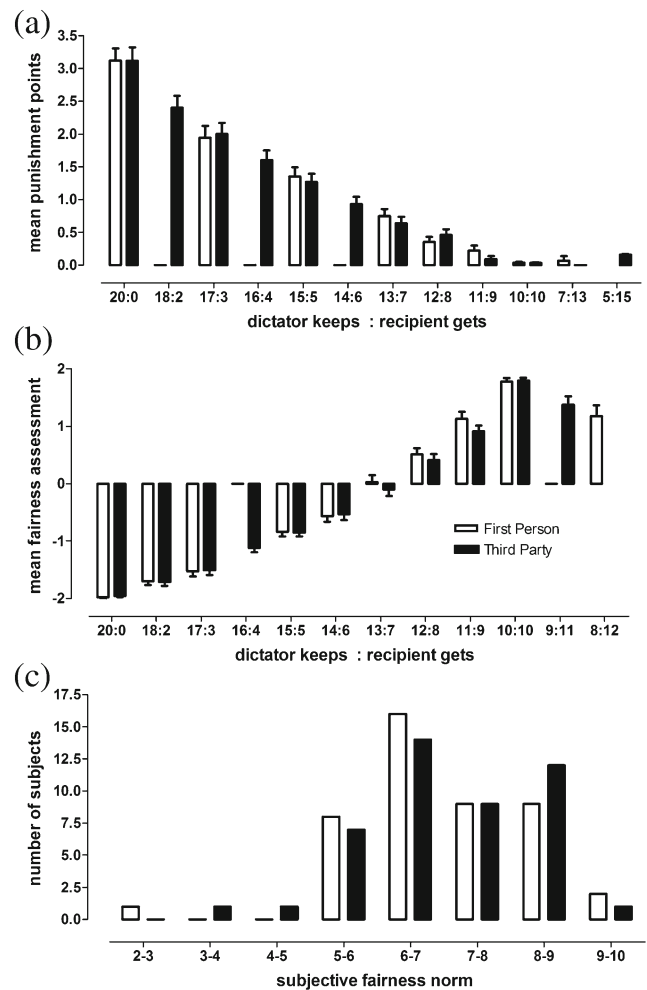

Fig. 2 Summary of behavioral results. (**a**) Mean punishment points (± S.E.M.) assigned to player A by player B and (**b**) mean fairness assessment (± S.E.M.) for different assignments by player B, separately for the first person and third party perspectives. (**c**) Illustration of behaviorally measured subjective fairness norms among the sample, separately for the first person and third party perspectives, for example 5–6 represents that a player B still considers an assignment of 5 € as unfair, but an assignment of 6 € as fair

accordance with recent results (Strobel et al., 2011). Moreover, consistent with the findings of Groß and Schuricht (2009), the more money the dictators retained, the more participants rated assignments as unfair in the first person perspective, $\chi^2 = 354.62$, $df = 9$, $p < .001$, as well as in the third party perspective, $\chi^2 = 400.84$, $df = 10$, $p < .001$ (see Fig. 2B).

With respect to subjective fairness evaluations, we found that across all assignments they did not differ according to the perspective, $Z = -.75$, $p = .45$. With respect to altruistic punishment, however, participants punished the dictators more when being directly affected by the dictators' decisions (first person perspective) than when observing them interacting with other persons (third party perspective), $Z = -2.77$, $p = .006$. Nonetheless, this difference mainly

results from the natural distribution of assignments made by the dictators used (see Table 1); the descriptive distribution (see Fig. 2A) shows that there is no general difference in punishment behavior between first person and third party perspectives. The statistical difference between perspectives as outlined above may be mainly driven by the 11:9, $p = .074$, and the 12:8 assignments, $p = .049$.

To test the hypothesis that a higher individual fairness norm leads to increased altruistic punishment for both first person and third party perspectives, we calculated Spearman correlations between the means of punishment points assigned and subjective fairness norms in each perspective. As expected, altruistic punishment behavior was significantly correlated with subjective fairness norms in both conditions, but was somewhat more pronounced in the first person perspective, $r = .39$, $p = .009$, than in the third party perspective, $r = .36$, $p = .017$. The important role of fairness norms for altruistic punishment behavior is underscored by the fact that even after controlling for subjective or objective financial situation, the pattern of medium-sized correlations remained stable (in the first person perspective: all $r \geq .36$, all $p \leq .016$; in the third party perspective: all $r \geq .28$, all $p \leq .062$). Figure 2C shows the distribution of subjective fairness norms.

In order to analyze effects of individual differences in the states of positive and negative affectivity, NEO altruism, and empathy on altruistic punishment behavior of player B, we computed Spearman correlations between these variables and mean punishment points in the first person and third party perspectives, respectively. While state positive affectivity showed medium-sized positive correlations with altruistic punishment behavior in the first person perspective, $r = .33$, $p = .029$, and in the third party perspective, $r = .41$, $p = .005$, state negative affectivity did not reveal any significant effect (all $p > .78$). Contrary to our expectations, NEO altruism and empathy were not significantly related to altruistic punishment behavior, neither in the first person nor in the third party perspective (all $p > .08$).

FRN and individual differences

In order to examine whether there were differences in FRN amplitude between unfair and fair trials both in the first person and the third party perspectives, we conducted a 2 (fairness: unfair trials, fair trials) × 2 (perspective: first person, third party) repeated-measures analysis of variance (ANOVA) for the FRN amplitudes. As expected, the analysis yielded a fairness main effect, $F(1, 44) = 4.12$, $p = .048$, $\eta^2_{partial} = .086$, with more pronounced FRN amplitudes for unfair than for fair assignments (see Fig. 3A for Fz electrode, graphic representations of additional midline locations can be found in the supplementary material).

Neither a perspectives main effect, $F(1, 44) = 0.06$, $p = .801$, $\eta^2_{partial} = .001$, nor an interaction effect fairness × perspective, $F(1, 44) = 0.71$, $p = .404$, $\eta^2_{partial} = .016$, was found. The latter results indicate that the fairness effect described above appears to be independent of the perspective taken.

When including financial situation as covariate to the ANOVA model, a main effect was revealed for objective financial situation, $F(1, 43) = 4.68$, $p = .036$, $\eta^2_{partial} = .098$, and also but less pronounced for subjective financial situation, $F(1, 43) = 3.51$, $p = .068$, $\eta^2_{partial} = .076$. However, in both models the influence of fairness was no longer significant ($p > .05$). Additionally, an interaction effect with perspective was found for objective financial situation, $F(1, 43) = 4.46$, $p = .041$, $\eta^2_{partial} = .094$. That is, more pronounced FRN amplitudes were associated with a better financial situation predominantly in first person perspective, indicating that this variable appears to play a critical role in the FRN modulation to fair and unfair assignments when someone is directly affected.

In order to address the question of a potential relation between FRN amplitudes and subsequent altruistic punishment behavior, nonparametric correlation analyses (Spearmans' rho) were performed separately for unfair and fair trials between FRN amplitudes and the respective punishment behavior both for first person and third party perspectives. Only for the first person perspective, consistent with our hypothesis, did the analysis yield a medium-sized negative correlation between FRN amplitude and altruistic punishment in unfair trials, $r = -.27$, $p = .076$; this, however, narrowly missed significance. There was no correlation revealed for fair trials, $r = -.16$, $p = .285$. This may suggest that for unfair assignments more pronounced (negative) FRN amplitudes were related to increased subsequent punishment behavior. However, when controlling for objective and/or subjective financial situation, the correlation of FRN and altruistic punishment in unfair trials decreased (all $p > .490$); similarly, there was no statistical significance for fair trials (all $p \geq .068$). In the third party perspective, no significant associations of FRN amplitudes with altruistic punishment behavior emerged, neither for fair nor for unfair trials (all $p > .322$).

To investigate the relation between FRN and individual differences, we calculated nonparametric correlations (Spearmans' rho) between FRN amplitudes and subjective fairness norms, empathy, NEO altruism, and state positive/negative affectivity. Contrary to our expectations, neither subjective fairness norms (all $p > .33$) nor NEO altruism (all $p > .19$) showed significant correlations with FRN amplitudes in any condition. With respect to empathy, in the third party perspective, a moderate though non-significant positive association was revealed between FRN amplitudes to unfair trials and the empathy score, $r = .27$, $p = .069$. This correlation gained
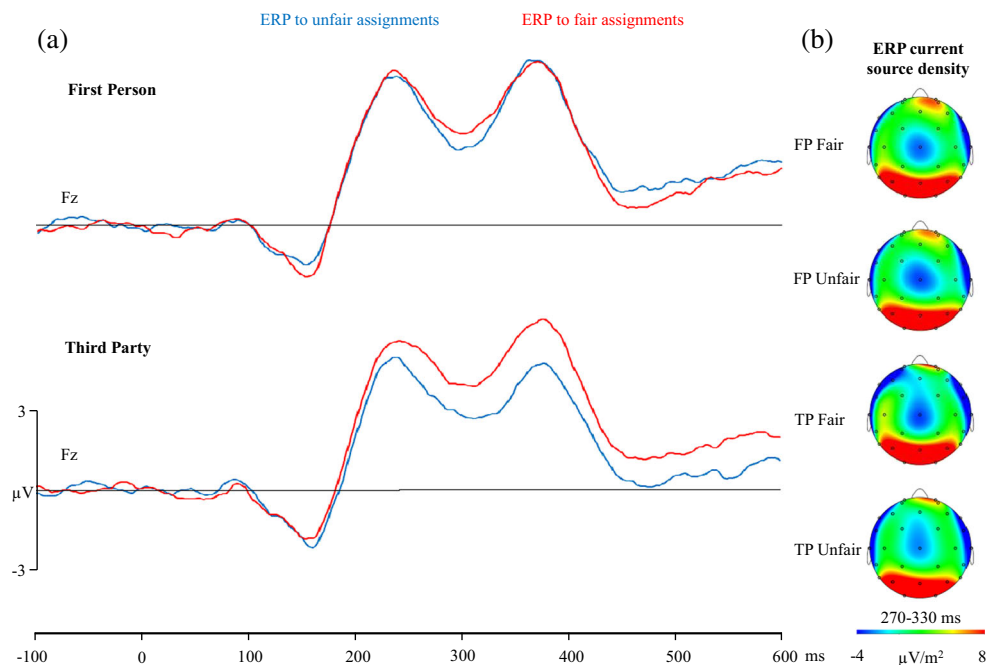
Fig. 3 Grand average event-related potentials (ERP) waveforms depicting the Feedback Related Negativity (FRN) fairness effect as well as the topography following the presentation of fair versus unfair assignments both in the first person and the third-party perspectives, respectively. (**A**) ERPs at electrode Fz for unfair (blue) and fair assignments (red) for the first person (upper panel) and the third party perspective (lower panel). A greater negativity at fronto-central sites is observable for unfair as compared to fair assignments. (**B**) The topographic current source density maps of the ERP peaks for fair and unfair trials for both perspectives, respectively, demonstrate the distribution of the FRN over the cortex (for colored illustration see electronic version of this article)

in significance when controlling for subjective financial situation, $r = .31$, $p = .047$, or objective financial situation, $r = .38$, $p = .015$. Considering the empathy overall score instead, similarly a significant positive association with FRN amplitudes to unfair trials in third party perspective was observed, $r = .36$, $p = .017$. That is, low scores in empathy were related to more pronounced (negative) FRN amplitudes in unfair trials. Hence, these results suggest an association between FRN in the third party condition and individual differences in trait empathy. Exploratory correlation analyses between FRN amplitudes and state positive or negative affectivity did not reveal any significant effect (all $p \geq .098$).

Discussion

Research on the behavioral and neurophysiological underpinnings of altruistic punishment has been of growing interest in recent years. Previous studies suggested that modulation in ACC and DLPFC activity following fair and unfair offers of proposers as well as individual differences in fairness concerns and traits like altruism appear to play an important role in altruistic punishment decisions (Boksem & De Cremer, 2010; Strobel et al., 2011). In order to further the understanding of possible explanatory and modulatory factors of altruistic punishment behavior, our study aimed at examining the interplay between altruistic punishment, individual difference

variables, and the ACC-related FRN component that is assumed to reflect feedback or outcome evaluations, i.e. whether outcomes match expectations (e.g. reward prediction errors). Recent EEG studies have begun to examine the modulation of FRN in altruistic punishment scenarios in the first person perspective (Boksem & De Cremer, 2010; Hewig et al., 2011). Similarly, a few studies investigated behavioral effects in third party punishment games (e.g., Fehr & Fischbacher, 2004a), but only one study so far has examined associated ERP correlates (Qu, Dou, You, & Qu, 2014). Finally, while previous research examined the effects of individual differences in altruistic punishment behavior (e.g., Strobel et al., 2011), our study provides the first investigation on the modulatory role of individual differences in ERP correlates of altruistic punishment.

In line with previous research and our expectations (e.g., Groß & Schuricht, 2009; Strobel et al., 2011), the behavioral data demonstrated that punishment behavior varied by the assigned amount of money by the dictators and the perceived unfairness of that assignment. Additionally, overall fairness perceptions of different assignments were independent from the perspective taken, suggesting that the fairness norm applied did not rely on the degree of personal involvement. Similar to the results of Strobel et al. (2011), punishment behavior was rather independent of personal involvement. Overall, our behavioral data support the notion of altruistic punishment as a fairness norm-enforcing behavior.

Regarding individual differences, in neither perspective could we detect a relationship between punishment points assigned and altruism (unlike Strobel and colleagues, 2011) nor a relationship between altruistic punishment and empathy scores. However, more punishment points assigned in both the first person and third party perspectives were predicted by higher positive affectivity at the beginning of the experiment as assessed using the state version of the PANAS (Krohne et al., 1996). This finding is consistent with the literature suggesting that positive affect can promote helpfulness (Isen & Lewin, 1972) as well as concerns for fairness (Labroo, 2004). One may therefore expect that in the context of altruistic punishment, positive emotions play an important role in initiating actions to restore violated fairness norms. Alternatively, it could be that positive affect fosters deviations from economical rational choices, as suggested by the research of Hirsh, Guindon, Morisano, and Peterson (2010), who showed a positive association between positive affectivity and impulsive responses, specifically with impulsive, reward-driven behavior in extraverts. In sum, these findings substantiate the importance of impulsive, emotional processes besides cognitive processes of social norm enforcement in altruistic punishment.

FRN

As outlined above, we examined the relation between altruistic punishment behavior, individual differences, and the FRN component that has been associated with ACC-dependent responses towards unexpected outcomes (Hauser et al., 2014; Holroyd, Pakzad-Vaezi, & Krigolson, 2008). Consistent with previous research indicating that the FRN is evoked by negative outcomes in social games (Boksem & De Cremer, 2010; Hewig et al., 2011; Polezzi et al., 2008) and in agreement with our hypothesis, we observed that FRN amplitude was more pronounced for unfair than for fair assignments in both the first person and third party conditions. Further, because larger FRN amplitudes have been observed for unexpected or surprising events, one might assume that fair assignments were more likely to be expected by the participants than unfair assignments. This interpretation corroborates the notion that individuals indeed adhere to a fairness norm in altruistic punishment situations as for example proposed by Boksem and De Cremer (2010), considering that the existence of a fairness norm implies a higher frequency of occurrence for fair than for unfair events in daily life. Taken together, these results support the hypothesis that violations of fairness norms may play an important role in the initial evaluation of assignments in the Dictator Game.

Regarding individual fairness norms, Boksem and De Cremer (2010) reported that individuals who value moral norms such as fairness highly (according to self-report) showed enhanced FRN amplitudes to unfair offers in an Ultimatum Game. This, however, is contrary to the results in the present study as subjective fairness norms empirically assessed via behavioral measures did not affect FRN amplitudes following dictator assignments. One possible reason for these conflicting results may relate to the different measures used (questionnaires vs. behavior). For example, Bicchieri and Chavez (2010) showed that fairness perception in a social game may critically vary by the situational context and expectations of the participants. This suggests that a fairness norm measured by behavioral data in the present Dictator Game can be different from fairness scores of a questionnaire-based dispositional measure of moral identity as used by Boksem & De Cremer (2010).

Our observation of fairness-dependent FRN deflections in the third party perspective appears to be in accordance with findings by Hewig et al. (2008) who reported larger FRN amplitudes also when others are observed experiencing disadvantageous outcomes (however, see also for diverging results: Alexopoulos, Pfabigan, Lamm, Bauer, & Fischmeister, 2012; Alexopoulos, Pfabigan, Göschl, Bauer, & Fischmeister, 2013). One could speculate that such FRN modulations occur due to the observed violation of fairness norms. Based on previous results suggesting an positive association between empathy and (observer) FRN amplitude (Fukushima & Hiraki, 2009), we expected FRN fairness effects in the third party perspective to be potentially modulated by empathy for the uninvolved player C. However, the data point to a different direction. The lower someone's empathy scores were, the more pronounced FRN amplitudes for unfair trials in the third party perspective occured. In comparison with individuals scoring high in empathy, it seems that for low empathizers unfair assignments were more surprising or unexpected in the rather uninvolved third party perspective, suggesting that fairness considerations are more important for them in such situations. For high empathizers, however, unfair dictator assignments might not have been as unexpected or surprising as for low empathizers, potentially because they are better able to put themselves in the position of the dictator and to reflect how they would behave if being in his position. While such an interpretation remains speculative until replication, the findings underscore the importance of fairness considerations and suggest empathy as a further motive of altruistic punishment from a third party perspective.

Moreover, previous research suggests that the initial assessment of a situation is crucial for subsequent decisions (Gehring & Willoughby, 2002). Thus, we were interested in whether punishment behavior can be predicted by the initial assessment of the assignments reflected in the FRN. Considering the importance both of fairness evaluations (Boksem & De Cremer, 2010) and of emotional involvement (Crockett et al., 2010) when being exposed to unfair assignments, we particularly expected FRN amplitudes in unfair trials to predict punishment behavior both in first person and in third party perspectives. Although the conventional

significance level was missed, medium-sized correlations indicated that in unfair trials (when participants supposedly experience the most unexpected events), larger FRN amplitudes were associated with punishment behavior only in first person perspective (e.g., Hewig et al., 2011). One may speculate that such an effect would have passed the conventional significance level in case of a larger sample size. In the third party perspective, FRN amplitudes did not predict punishment behavior. These perspective-dependent results may be explained in two ways. First, they support the notion that besides fairness concerns, altruistic punishment could also be driven by other sources of variance such as impulsive emotional responses induced by unfairness (Crockett et al., 2010). That is, the level of involvement (to be directly affected by fairness or not) may play a critical role for the experience of emotions. As suggested by previous studies (e.g., Civai et al., 2010), emotions such as anger and frustration may play a more dominant role in first person punishment and thus are more likely to trigger altruistic punishment behavior when fairness norm violations are personally perceived (e.g., Civai et al., 2010). Secondly, these results can be explained with regard to fairness concerns. While FRN fairness effects in this study were similarly observable not only for the first person condition but also for the third party condition, other studies suggest that FRN modulations may be stronger in cases of higher personal relevance (see Koban & Pourtois, 2014). Hence, it is conceivable that particularly FRN fairness modulations in scenarios of high personal relevance (such as in the first person perspective) would be predictive of altruistic punishment behavior. Although more research is needed, both arguments provide a preliminary explanation for why the level of personal relevance or involvement might influence the relationship between FRN and punishment behavior.

Another important finding of our study was that the FRN effect of fairness was affected by financial situation in the first person perspective. The relevance of financial situation is underscored by the fact that it reduced the effect of fairness on FRN amplitudes when being included in the ANOVA model. Specifically, the observed relationship between FRN and financial situation may indicate that participants in a better financial situation experienced stronger violations of expectations in altruistic punishment decisions, particularly when being directly affected from the assignment. In general, this fits to a large body of research suggesting that neural mechanisms are indeed shaped by socioeconomic status (Hackman, Farah, & Meaney, 2010). The role of financial situation is further underlined by the behavioral finding that participants with a more privileged financial background showed more altruistic punishment. One could argue that such individuals may be less concerned about the costs of punishment, but are more focused on fairness during the game. Thus, they are possibly more likely to act according to their fairness norms due to a different framing of costs and benefits of cooperative behavior

and punishment (Fehr & Fischbacher, 2004b; see also Benenson, Pascoe, & Radmore, 2007; for conflicting results on socioeconomic status see Cameron, 1999; Güth, Schmidt, & Sutter, 2007). Given that financial situation might be an important factor in altruistic punishment, future research investigating the interplay of financial situation and fairness norms on altruistic punishment would be a fruitful endeavor.

Limitations

The present study has several potential limitations. First, some effects of our study would not withstand a strong correction for multiple testing. However, because the main hypotheses of our study were carefully derived from a large set of studies on FRN and social decision making, our research is not exploratory in nature. This is supported by the fact that our results are in many aspects consistent with this previous research, thereby suggesting a low probability of Type I error. Nonetheless, as we refrained from such a correction, the weaker results have to be interpreted with caution and additional research replicating these findings is warranted.

Moreover, because we sought to have both punishment behavior and fairness evaluations at a randomized trial-wise level, general punishment behavior might have been restricted. Nevertheless, we used such a design because of the advantages of measuring fairness judgments of dictator assignments directly in punishment relevant situations over offline fairness judgments. Further, we were not able to disentangle completely the influence of emotions such as anger or disgust from fairness concerns in altruistic punishment. Future studies may use physiological measures such as skin conductance responses (van't Wout et al., 2006). Another drawback of our study may have emerged from the uneven distribution of males and females. However, no sex main or interaction effect was detected in our study.

Conclusion

At the behavioral level our results suggest that altruistic punishment primarily depends on violations of individual fairness norms in both first person and third party settings. At the level of individual differences, we showed that altruistic punishment is influenced by positive affectivity. At the neurophysiological level, the present study adds to previous research indicating an important role of ACC-dependent processes of feedback or outcome evaluations in altruistic punishment. Specifically, our findings indicate that the ACC-related FRN component is sensitive to fairness norm violations during the evaluation of initial assignments and is potentially predictive for subsequent punishment behavior, thereby improving the understanding of neurophysiological processes underlying altruistic punishment. Importantly, we also observed that behavioral and neurophysiological effects are mediated by the

participants' fairness concerns, financial situation, and empathy scores. However, future studies are necessary to shed more light on the complex cognitive-affective-motivational interplay during altruistic punishment behavior.

References

Alexopoulos, J., Pfabigan, D. M., Göschl, F., Bauer, H., & Fischmeister, F. P. S. (2013). Agency matters! Social preferences in the three-person ultimatum game. *Frontiers in Human Neuroscience, 7,* 312. doi:10.3389/fnhum.2013.00312

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

Amiez, C., Joseph, J.-P., & Procyk, E. (2005). Anterior cingulate error-related activity is modulated by predicted reward. *The European Journal of Neuroscience, 21*(12), 3447–3452. doi:10.1111/j.1460-9568.2005.04170.x

Benenson, J. F., Pascoe, J., & Radmore, N. (2007). Children's altruistic behavior in the dictator game. *Evolution and Human Behavior, 28*(3), 168–175. doi:10.1016/j.evolhumbehav.2006.10.003

Bicchieri, C., & Chavez, A. (2010). Behaving as expected: Public information and fairness norms. *Journal of Behavioral Decision Making, 23*(2), 161–178. doi:10.1002/bdm.648

Boksem, M. A. S., & De Cremer, D. (2010). Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Social Neuroscience, 5*(1), 118–128. doi:10.1080/17470910903202666

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

Bowles, S., & Gintis, H. (2004). The evolution of strong reciprocity: Cooperation in heterogeneous populations. *Theoretical Population Biology, 65*(1), 17–28. doi:10.1016/j.tpb.2003.07.001

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences, 100*(6), 3531–3535. doi:10.1073/pnas.0630443100

Camerer, C. (1997). Progress in behavioral game theory. *The Journal of Economic Perspectives, 11*(4), 167–188. Retrieved from http://www.jstor.org/stable/2138470

Cameron, L. A. (1999). Raising the stakes in the ultimatum game: Experimental evidence from Indonesia. *Economic Inquiry, 37*(1), 47–59. doi:10.1111/j.1465-7295.1999.tb01415.x

Civai, C., Corradi-Dell'Acqua, C., Gamer, M., & Rumiati, R. I. (2010). Are irrational reactions to unfairness truly emotionally-driven? Dissociated behavioural and emotional responses in the Ultimatum Game task. *Cognition, 114*(1), 89–95. doi:10.1016/j.cognition.2009.09.001

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Odessa: Psychological Assessment Resources.

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

De Quervain, D. J. F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science, 305*(5688), 1254–1258. doi:10.1126/science.1100735

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science, 5*(5), 303–305. doi:10.1111/j.1467-9280.1994.tb00630.x

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An FMRI study of social exclusion. *Science, 302*(5643), 290–292. doi:10.1126/science.1089134

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature, 425*(6960), 785–791. doi:10.1038/nature02043

Fehr, E., & Fischbacher, U. (2004a). Social norms and human cooperation. *Trends in Cognitive Sciences, 8*(4), 185–190. doi:10.1016/j.tics.2004.02.007

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

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*(6868), 137–140. doi:10.1038/415137a

Fukushima, H., & Hiraki, K. (2009). Whose loss is it? Human electrophysiological correlates of non-self reward processing. *Social Neuroscience, 4*(3), 261–275. doi:10.1080/17470910802625009

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4*(6), 385–390. doi:10.1111/j.1467-9280.1993.tb00586.x

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science, 295*(5563), 2279–2282. doi:10.1126/science.1066893

Groß, J., & Schuricht, F. (2009). *Neuronale und behaviorale Erklärungsansätze für das Phänomen der altruistischen Bestrafung [Neuronal and behavioral explanations for the phenomenon of altruistic punishment]. Unpublished Thesis.* Johann Wolfgang Goethe Universitaet, Frankfurt, Germany.

Güth, W., Schmidt, C., & Sutter, M. (2007). Bargaining outside the lab - A newspaper experiment of a three-person ultimatum game. *Economic Journal, 117*(518), 449–469. doi:10.1111/j.1468-0297.2007.02025.x

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization, 3*(4), 367–388. doi:10.1016/0167-2681(82)90011-7

Hackman, D. A., Farah, M. J., & Meaney, M. J. (2010). Socioeconomic status and the brain: Mechanistic insights from human and animal research. *Nature Reviews Neuroscience, 11*(9), 651–659. doi:10.1038/nrn2897

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

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

Hewig, J., Trippe, R. H., Hecht, H., Coles, M. G. H., Holroyd, C. B., & Miltner, W. H. R. (2008). An electrophysiological analysis of

coaching in Blackjack. *Cortex, 44*(9), 1197–1205. doi:10.1016/j.cortex.2007.07.006

Hirsh, J. B., Guindon, A., Morisano, D., & Peterson, J. B. (2010). Positive mood effects on delay discounting. *Emotion, 10*(5), 717–721. doi:10.1037/a0019466

Holler, M. J., & Illing, G. (2006). *Einführung in die Spieltheorie [Introduction to game theory]*. Berlin: Springer Verlag.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109*(4), 679–709. doi:10.1037//0033-295X.109.4.679

Holroyd, C. B., & Krigolson, O. E. (2007). Reward prediction error signals associated with a modified time estimation task. *Psychophysiology, 44*(6), 913–917. doi:10.1111/j.1469-8986.2007.00561.x

Holroyd, C. B., Pakzad-Vaezi, K. L., & Krigolson, O. E. (2008). The feedback correct-related positivity: Sensitivity of the event-related brain potential to unexpected positive feedback. *Psychophysiology, 45*(5), 688–697. doi:10.1111/j.1469-8986.2008.00668.x

Isen, A. M., & Lewin, P. F. (1972). Effect of feeling good on helping: Cookies and kindness. *Journal of Personality and Social Psychology, 21*(3), 384–388. doi:10.1037/h0032317

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science, 314*(5800), 829–832. doi:10.1126/science.1129156

Koban, L., & Pourtois, G. (2014). Brain systems underlying the affective and social monitoring of actions: An integrative review. *Neuroscience and Biobehavioral Reviews, 46*(Pt 1), 71–84. doi:10.1016/j.neubiorev.2014.02.014

Krohne, H. W., Egloff, B., Kohlmann, C.-W., & Tausch, A. (1996). Untersuchungen mit einer deutschen form der positive and negative affect schedule (PANAS) [Investigations with a German version of the positive and negative affect schedule (PANAS)]. *Diagnostica, 42*, 139–156.

Labroo, A. A. (2004). *Positive affect increases concern for fairness: The case of restrictions to promotional offers. Unpublished Dissertation.* Cornell University, Ithaca.

Matsumoto, M., Matsumoto, K., Abe, H., & Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. *Nature Neuroscience, 10*(5), 647–656. doi:10.1038/nn1890

Middendorf, E., Apolinarski, B., Poskowsky, J., Kandulla, M., & Netz, N. (2013). *Die wirtschaftliche und soziale Lage der Studierenden in der Bundesrepublik Deutschland 2012. 20. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch das HIS-Institut für Hochschulforschung.* Bonn, Berlin: BMBF. Retrieved from http://www.sozialerhebung.de/erhebung_20/soz_20_haupt

Mikula, G., Scherer, K. R., & Athenstaedt, U. (1998). The role of injustice in the elicitation of differential emotional reactions. *Personality and Social Psychology Bulletin, 24*(7), 769–783. doi:10.1177/0146167298247009

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*, 167–202. doi:10.1146/annurev.neuro.24.1.167

Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "Generic" neural system for error detection. *Journal of Cognitive Neuroscience, 9*(6), 788–798. doi:10.1162/jocn.1997.9.6.788

Oliveira, F. T. P., McDonald, J. J., & Goodman, D. (2007). Performance monitoring in the anterior cingulate is not all error related:

Expectancy deviation and the representation of action-outcome associations. *Journal of Cognitive Neuroscience, 19*(12), 1994–2004. doi:10.1162/jocn.2007.19.12.1994

Osinsky, R., Mussel, P., Ohrlein, L., & Hewig, J. (2014). A neural signature of the creation of social evaluation. *Social Cognitive and Affective Neuroscience, 9*(6), 731–736. doi:10.1093/scan/nst051

Ostendorf, F., & Angleitner, A. (2003). *NEO-Persönlichkeitsinventar, revidierte Form, NEO-PI-R nach Costa und McCrae [Revised NEO Personality Inventory, NEO-PI-R of Costa and McCrae].* Göttingen: Hogrefe.

Paulus, C. (2009). Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie. Retrieved from http://psydok.sulb.uni-saarland.de/volltexte/2009/2363/

Paulus, C. (2012). Ist die Bildung eines Empathiescores in der Deutschen Fassung des IRI sinnvoll? Retrieved from http://scidok.sulb.uni-saarland.de/volltexte/2012/4889/

Paus, T., Koski, L., Caramanos, Z., & Westbury, C. (1998). Regional differences in the effects of task difficulty and motor output on blood flow response in the human anterior cingulate cortex: A review of 107 PET activation studies. *Neuroreport, 9*(9), R37–R47. doi:10.1097/00001756-199806220-00001

Pillutla, M., & Murnighan, J. (1996). Unfairness, anger, and spite: emotional rejections of ultimatum offers. *Organizational Behavior and Human, 68*(3), 208–224. doi:10.1006/obhd.1996.0100

Polezzi, D., Daum, I., Rubaltelli, E., Lotto, L., Civai, C., Sartori, G., & Rumiati, R. (2008). Mentalizing in economic decision-making. *Behavioural Brain Research, 190*(2), 218–223. doi:10.1016/j.bbr.2008.03.003

Qu, L., Dou, W., You, C., & Qu, C. (2014). The processing course of conflicts in third-party punishment: An event-related potential study. *Psychology Journal, 3*(3), 214–221. doi:10.1002/pchj.59

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

Schimmack, U. (2003). Affect measurement in experience sampling research. *Journal of Happiness Studies, 4*(1), 79–106. doi:10.1023/A:1023661322862

Singer, T., & Steinbeis, N. (2009). Differential roles of fairness- and compassion-based motivations for cooperation, defection, and punishment. *Annals of the New York Academy of Sciences, 1167*, 41–50. doi:10.1111/j.1749-6632.2009.04733.x

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

van't Wout, M., Kahn, R. S., Sanfey, A. G., & Aleman, A. (2006). Affective state and decision-making in the Ultimatum Game. *Experimental Brain Research, 169*(4), 564–568. doi:10.1007/s00221-006-0346-5

Xiao, E., & Houser, D. (2005). Emotion expression in human punishment behavior. *Proceedings of the National Academy of Sciences, 102*(20), 7398–7401. doi:10.1073/pnas.0502399102

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. (2006). Brain responses to outcomes of one's own and other's performance in a gambling task. *Neuroreport, 17*(16), 1747–1751. doi:10.1097/01.wnr.0000239960.98813.50