

# Fairness influences early signatures of reward-related neural processing

Bart Massi<sup>1</sup> · Christian C. Luhmann<sup>2</sup>

Published online: 12 May 2015  
© Psychonomic Society, Inc. 2015

**Abstract** Many humans exhibit a strong preference for fairness during decision-making. Although there is evidence that social factors influence reward-related and affective neural processing, it is unclear if this effect is mediated by compulsory outcome evaluation processes or results from slower deliberate cognition. Here we show that the feedback-related negativity (FRN) and late positive potential (LPP), two signatures of early hedonic processing, are modulated by the fairness of rewards during a passive rating task. We find that unfair payouts elicit larger FRNs than fair payouts, whereas fair payouts elicit larger LPPs than unfair payouts. This is true both in the time-domain, where the FRN and LPP are related, and in the time-frequency domain, where the two signals are largely independent. Ultimately, this work demonstrates that fairness affects the early stages of reward and affective processing, suggesting a common biological mechanism for social and personal reward evaluation.

**Keywords** Inequality-aversion · Feedback-related-negativity · Late-positive-potential

Human social behavior is defined by a strong regard for the welfare of others. Decades of work in behavioral economics have shown that many individuals exhibit a preference for fairness, even when maintaining equality among decision-

makers is risky or costly (Fehr & Fishbacher, 2003). Although mutual cooperation tends to emerge in tasks where it is possible to build a reputation (e.g., Kurzban & Houser, 2005), not all other-regarding behavior is consistent with reputation building or even selfish payoff maximization. For example, human decision-makers tend to prefer equitable wealth distributions when anonymously distributing funds between themselves and another individual (Eckel & Grossman, 1996; Engelmann & Strobel, 2004). Anonymity prevented reciprocity during these games, so selections of fair allocations are likely mediated by a genuine preference for fair outcomes. The finding that human decision-makers prefer fair outcomes inspired econometric models that describe decision-makers' preferences as a function of both expected payout and fairness (Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000), thereby suggesting that fairness-seeking mediates apparently altruistic-behavior.

A plausible hypothesis for the neural analog of such models is that reward-related neural activity elicited by vicarious rewards mediates fairness-seeking behavior. It is well-established that voluntary donation elicits activity in regions of the brain that represent rewards (Harbaugh, Mayr, & Burghart, 2007; Moll et al., 2006), suggesting that processes responsible for evaluating others' gains use the same neural circuitry that is responsible for processing personal rewards. In fact, single-unit recordings from non-human primates show that neurons in the anterior cingulate cortex carry information about both the subject's own rewards and the rewards of other monkeys (Chang, Gariépy, & Platt, 2013). Collectively, these findings demonstrate that the brain represents others' rewards vicariously, suggesting a biological basis for the effect of fairness on behavior.

Studies using functional magnetic resonance imaging (fMRI) have established that fairness affects neural responses in regions involved in reward-related processing. For example, the neural activity in reward- and emotion-related regions

---

✉ Bart Massi  
bart.massi@yale.edu

<sup>1</sup> Interdepartmental Neuroscience Program, Yale University, New Haven, CT 06520-8074, USA

<sup>2</sup> Department of Psychology, Stony Brook University, Stony Brook, NY, USA

differed following unfair offers during an ultimatum game - a game in which a recipient (i.e., the participant) rejects or accepts putative splits of a pool of money made by a proposer - specifically when the offers were made by human proposers rather than computers (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). Moreover, neural activity in regions involved in emotion-related processing predicted decisions to accept or reject offers, suggesting that emotional processing may influence decision-making strategies in social contexts. Later work found that activity in the ventral striatum and ventromedial prefrontal cortex, two regions with known roles in the processing of subjective reward (Kable & Glimcher, 2007), were differentially affected by payouts if they increased or decreased the wealth inequality between two subjects (Tricomi, Rangel, Camerer, & O'Doherty, 2010). As neither subject could influence the expected value of the outcomes, fairness-related differences in neural responses could not be explained by strategic reasoning. Together, these studies show that fairness and equity affect reward processing in the brain. Unfortunately, due to the poor temporal resolution of fMRI, it is unclear whether the previously observed effects of fairness on reward and affective neural processing are mediated by compulsory bottom-up processes or slower cognitive deliberation.

Previous investigations have attempted to address the timing of the effects of fairness on reward processing by using the superior temporal resolution of electroencephalography (EEG). These studies have treated the feedback-related-negativity (FRN) component of the event-related potential (ERP) as a proxy for early reward processing, as the FRN occurs approximately 200–400 milliseconds after stimulus onset and tends to be more negative for bad outcomes than for good outcomes (Holroyd & Coles, 2002; Hajcak, Moser, Holroyd, & Simons, 2006). The FRN can also be modulated by social context. For example, greater FRNs are typically observed for unfair offers than for fair offers during the ultimatum game (Boksem & De Cremer, 2010; Mussel, Hewig, Allen, Coles, & Miltner, 2014; Polezzi et al., 2008). Unfortunately, reward and fairness were confounded in all but one of these studies as the least fair offers favored the proposer, and were thus the least valuable offers. The results of the remaining study suggest that the FRN may be larger for unfair offers than for fair offers even when the fair offer is less valuable, although they failed to find a significant effect (Mussel et al., 2014). It is therefore unclear from the results of previous studies alone whether physiological signatures of early reward processing are affected by fairness. Moreover, as the ultimatum game requires a response from the participant, it is unclear whether the observed effects are truly due to outcome evaluation or if they are related to upcoming decisions.

The current study aims to extend the aforementioned work by determining the effect of the fairness of potential outcomes on physiological correlates of reward processing using EEG. If early physiological signatures of reward processing are

affected by the outcome fairness, then it is likely that social outcome evaluation is carried out by bottom-up reward evaluation processes and is not entirely mediated by slow deliberation or strategic consideration. Both the FRN and the late-positive potential (LPP), an ERP component that occurs approximately 500 ms after the onset of stimuli with affective valence (Hajcak, Dunning, & Foti, 2009), provide a measurable substrate for early reward- and affect-related neural processing of outcomes. A passive outcome-evaluation task was used that required participants to rate the pleasantness of potential payouts that they and another individual could receive. The distribution of wealth associated with each payout was systematically manipulated so that offers were either equitable, inequitable favoring the participant, or inequitable favoring the other individual. Crucially, participants could not influence their expected payout at all, so reward-related neural responses in our task cannot be contaminated by signals related to decision making or performance monitoring.

## Methods

### Participants

Twenty Stony Brook University undergraduates (13 females, mean age of 21 years) with no history of neurologic damage were recruited from the university's participant pool. All participants received partial course credit and monetary compensation. Informed consent was obtained from all participants before the start of the experiment. Two participants were discarded due to excessive movement during EEG data acquisition, leaving 18 participants for analysis.

### Data acquisition and stimulus presentation

Sixty-four Ag/Ag-Cl electrodes were attached prior to the experiment with a Compumedics Neuroscan Quick-Cap arranged with a standard 10–20 layout. Recordings were referenced to the left and right mastoids. Electrodes placed on the outer canthi of the left and right eyes and above and below the right eye recorded the vertical and horizontal electrooculogram. Impedances were kept below 10 kOhms at all electrode sites. Electroencephalogram (EEG) data were recorded with SynAmp amplifiers (Neuroscan Inc.) at a sample rate of 500 Hz and digitized using SCAN 4.3 software (Neuroscan Inc.). All stimuli were presented on LCD monitors and responses were collected from a small button box mounted on the arm of the chair in which participants were seated. Stimuli were generated and presented using custom software built using the PsychoPy toolbox in Python (Peirce, 2007).

## Task and stimuli

Participants rated monetary payouts that they and a second, anonymous individual (henceforth “the other”) could potentially receive. Participants were told that the other was a second participant who would come in during a future experimental session, and belief in the existence of the other was confirmed verbally after the experiment. The sequence of task events is given in Fig. 1a. At the beginning of each trial, a payout to the participant and the other appeared. Two thermometers with labeled tick marks indicated the value of the payout to the participant (left) and the other (right). Participants viewed the payout for 3 s followed by a 500-ms fixation interval. Participants then rated the pleasantness of the payout using a scale ranging from  $-3$  (very displeasing) to  $3$  (very pleasing). The next trial began after a 500-ms inter-trial interval.

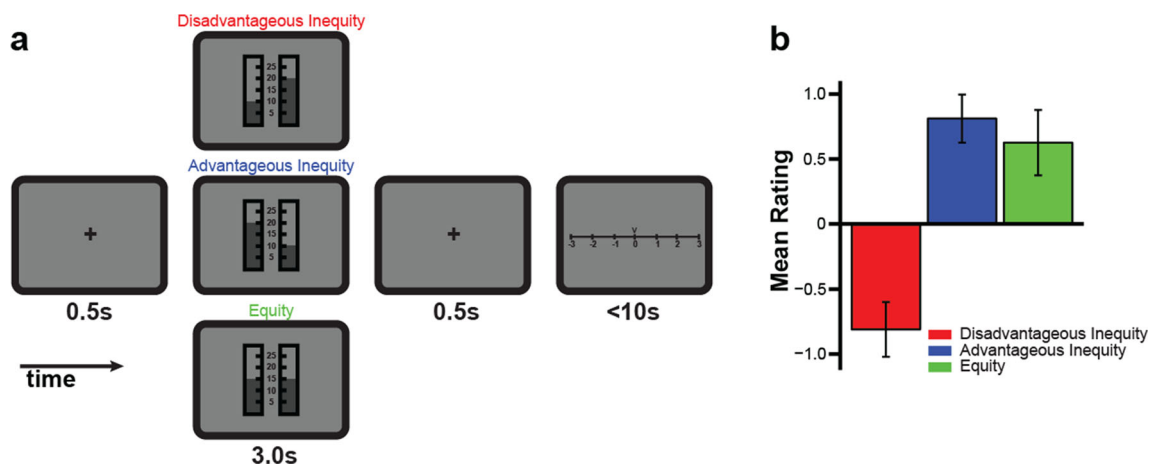
Both the payouts to both the participants were drawn from a low interval of  $[0,9]$  or a high interval of  $[17,26]$ . The other's payout was drawn from the same interval. Trials were split into three conditions corresponding to the wealth distribution between the participant and the other: disadvantageously inequitable payouts in which the other was given more than the participant, advantageously inequitable payouts in which the other was given less than the participant, and equitable payouts in which the participant and the other were given the same amount of money. To ensure that payout differences were readily apparent in the visual stimuli, we required the participant's payout to be at least \$4 larger than the other's during advantageously inequitable trials and at least \$4 smaller in disadvantageously inequitable trials. We presented each of the 20 unique equitable trial types, 42

unique advantageously inequitable trials, and 42 unique disadvantageously inequitable trials a total of three times across six blocks, resulting in a total of 312 trials per subject. The mean payout to the participant was \$13 in the equitable condition, and \$10.17 and \$15.83 in the disadvantageous and advantageously inequitable conditions, respectively. Trials were presented in a random order, and participants were informed that the payout from one randomly selected trial would be actualized at the end of the experiment.

## Physiological data analysis

Continuous EEG data were down-sampled to 250 Hz, band-pass filtered between .1 and 20 Hz with a FIR filter, and partitioned into trials aligned at stimulus onset. We aligned all trials to a common baseline for each channel by subtracting the average of the 100-ms prestimulus interval from each trial's waveform. Trials with non-stereotypical artifacts were excluded from analysis by a semi-automated artifact detection routine and remaining ocular artifacts were corrected using the algorithm developed by Miller, Gratton, and Yee (1988). Finally, mean waveforms were computed for each condition and each participant, and these waveforms were used in all further analyses.

Averaging in the time-domain was used to characterize the FRN and LPP at a cluster of nine electrodes centered at CZ (FC1, FCZ, FC2, C1, CZ, C2, CP1, CPZ, and CP2). For each participant and each condition, the FRN and LPP were scored as the mean voltage in the 300- to 400-ms and 500- to 800-ms post-stimulus intervals, respectively. Participants' scores were computed separately for each wealth distribution condition.



**Fig. 1** Task and behavior. **(a)** Temporal sequence of the outcome rating task. Subjects completed a social outcome evaluation task in which potential monetary payouts from the experimenter to the subject and an anonymous other were rated. After a .5-s ITI, subjects fixated a central cross for .5 s. A potential monetary payout was presented for 3 s. Payouts fell into one of three categories: those in which the subject received less than the other (disadvantageous inequity), those in which the subject

received more than the other (advantageous inequity), and those in which the subject and the other received the same amount (equity). Then, after another .5-s fixation period, subjects had 10 s to rate the pleasantness of the payout. Exactly one payout was actualized at the end of the experiment. **(b)** Mean ratings for each equity condition. Ratings were averaged within subject, and then within-subject mean ratings for each condition were averaged across subjects

Time-frequency analysis is a powerful tool for disambiguating temporally coincident ERP components based on their spectral profile. Thus, to disambiguate the components of the ERP, we used a wavelet-based approach to decompose event-related potentials into time-frequency scalograms. A Morse wavelet was chosen for the analysis due to its exact analyticity and approximate symmetry in the frequency domain (Lilly & Olhede, 2012), ultimately producing a more accurate estimate of spatiotemporal energy than other, similar wavelets (e.g., Morlet). Morse wavelets have been previously demonstrated to be suitable for analysis of EEG time series (Brittain, Halliday, Conway, & Nielsen, 2007). A complex Morse wavelet ( $\gamma = 3$ ,  $\beta = 5$ ) with 20 scales (linearly spaced from 1 to 20 Hz) was convolved with each participant's mean time-domain ERP waveform for each wealth distribution condition to extract a scalogram of time-varying frequency components comprising the ERP. Each three-dimensional scalogram was converted into multiple two-dimensional time series by averaging across scales corresponding to obtain time series of power in frequency bands of interest. The average power in a 100-ms pre-stimulus period was used to align all scales in all scalograms to a common baseline. Analogous to the time domain ERP scores, time-frequency domain scores were obtained by averaging across scales in a given frequency band and time points during a period of interest. Previous work has found that the FRN is characterized by a burst of power in the theta frequency range (Gehring & Willoughby, 2004), suggesting that theta power several hundred ms after stimulus onset is a valid time-frequency domain proxy for the FRN (Bernat, Nelson, Steele, Gehring, & Patrick, 2011). Thus, the average power in the 4–8 Hz (theta) range during the 250- to 350-ms post-stimulus interval was taken as an index of the FRN. Because the LPP is a sustained change in scalp potential that can last up to several seconds (Schupp et al., 2000), it is likely to result from neural processes that fluctuate infrequently. Thus, power in the 1–3 Hz (delta) range during the 500- to 800-ms post-stimulus intervals was treated as an index of the LPP.

To compute scalp topographies in the time domain, we used a current source density (CSD) analysis, in which we took the difference between the CSD estimates for the equitable and inequitable conditions. This allowed a between-condition comparison of the relative source of current changes that ultimately produce the observed ERPs. For time-frequency domain signals, wavelet power was averaged across time intervals of interest separately for the theta and delta bands at each electrode for each subject in each condition. Then, the difference between the spatial distribution of signals in the equitable and disadvantageously inequitable conditions was plotted. For visualization, points between channels were interpolated.

All EEG data preprocessing was done with the EEGLAB toolbox (Delorme & Makeig, 2004), and subsequent ERP and

wavelet analyses were conducted using custom MATLAB (Mathworks) software and the JLAB signal processing package (Lilly, 2011). Time-domain scalp CSD maps were computed with the CSD toolbox (Kayser & Tenke, 2006). All statistical analyses were conducted using R (R Core Team), and p-values of all post-hoc pairwise tests were corrected using the Bonferroni-Holm method.

## Results

### Behavior

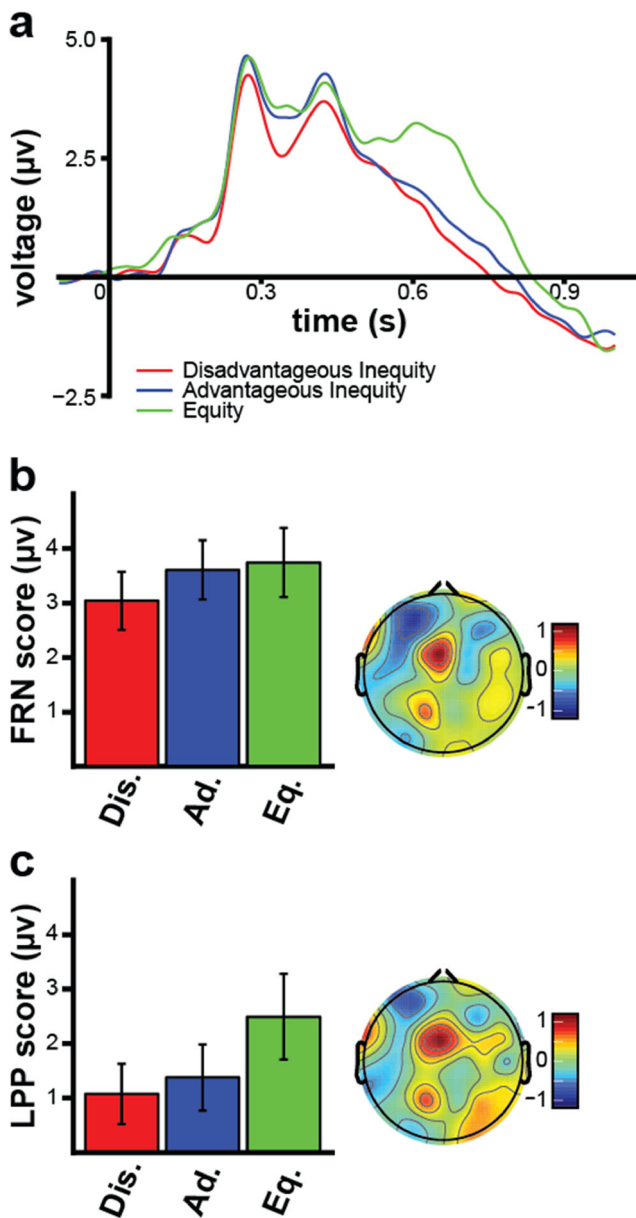
Figure 1b illustrates the mean ratings given to the payouts. There was a significant effect of wealth distribution on mean rating ( $F(2,34) = 22.4482$ ,  $p < 10^{-7}$ ). Post-hoc pairwise tests show that advantageously inequitable payouts were given higher ratings than disadvantageously inequitable payouts on average ( $t(17) = 6.41$ , corrected  $p < 10^{-6}$ ), and equitable payouts were rated more pleasing than disadvantageously inequitable payouts ( $t(17) = 5.70$ , corrected  $p < 0.001$ ), although advantageously inequitable payouts were indistinguishable from equitable payouts ( $t(17) = .644$ , corrected  $p > 0.5$ ). These results demonstrate that payouts were judged on the basis of the underlying distribution of wealth, which is ultimately independent of the monetary value of the payout to the participant in the present experiment.

### Time-domain analysis

Figure 2 depicts the physiological effects of each type of payout. Grand-average ERPs across all participants at a CZ-centered cluster of electrodes are given in Fig. 2a. While the shape of the FRN waveform and the spatial distribution of its sources are consistent with previously reports (e.g., Holroyd, Krigolson, & Lee, 2011), the FRN occurred later during the present study than is typically observed. We attribute its late arrival to the complexity of our compound stimuli and the fact that subjects' behavior suggests that they computed the relative magnitude of the two payouts. Accordingly, a later (i.e., 300–400 ms) post-stimulus window was used to compute the FRN. There was a significant effect of wealth distribution on FRN magnitude ( $F(2,34) = 3.298$ ,  $p < 0.05$ ; Fig. 2b), though no post-hoc tests survived correction for multiple comparisons.

A significant effect of wealth distribution on LPP magnitude was also found ( $F(1.31,22.19) = 7.474$ ,  $p < 0.01$ ; Fig. 2c). Mauchly's test indicated that participants' LPP scores violated the assumption of sphericity ( $X^2 = 0.468$ ,  $p < 0.005$ ), so the degrees of freedom of the reported ANOVA were adjusted with a Greenhouse-Geisser correction. LPPs elicited by equitable offers were more positive than those elicited by disadvantageously inequitable offers ( $t(17) = 2.480$ , corrected





**Fig. 2** Time-domain electrophysiological results. **(a)** Event-related potentials at a CZ-centered cluster relative to the presentation of the payout. Both the feedback-related negativity (FRN) and late positive potential (LPP) were scored as the mean potential evoked by trials of each condition for each subject in the 300- to 400-ms and 500- to 800-ms range, respectively. **(b)** Mean FRN scores and **(c)** mean LPP scores across subjects with associated current source density (CSD) maps

$p < 0.05$ ) and advantageously inequitable offers ( $t(17) = 3.131$ , corrected  $p < 0.05$ ).

To look at the overall relationship of LPP and FRN scores, we z-scored each across participants separately for each condition, then combined all scores in all conditions into two large vectors for each ERP component. Overall, LPP and FRN scores were highly correlated ( $r = 0.54$ ,  $p < 0.001$ ,  $n = 54$ ). Because we scored the FRN and LPP so close to one-another, it is highly plausible that this relationship is an artifact of

averaging in the time-domain rather than a functional coupling of the neural processes underlying the LPP and FRN. Thus, time-frequency analysis was used to isolate the two components in time-frequency space.

### Time-frequency domain analysis

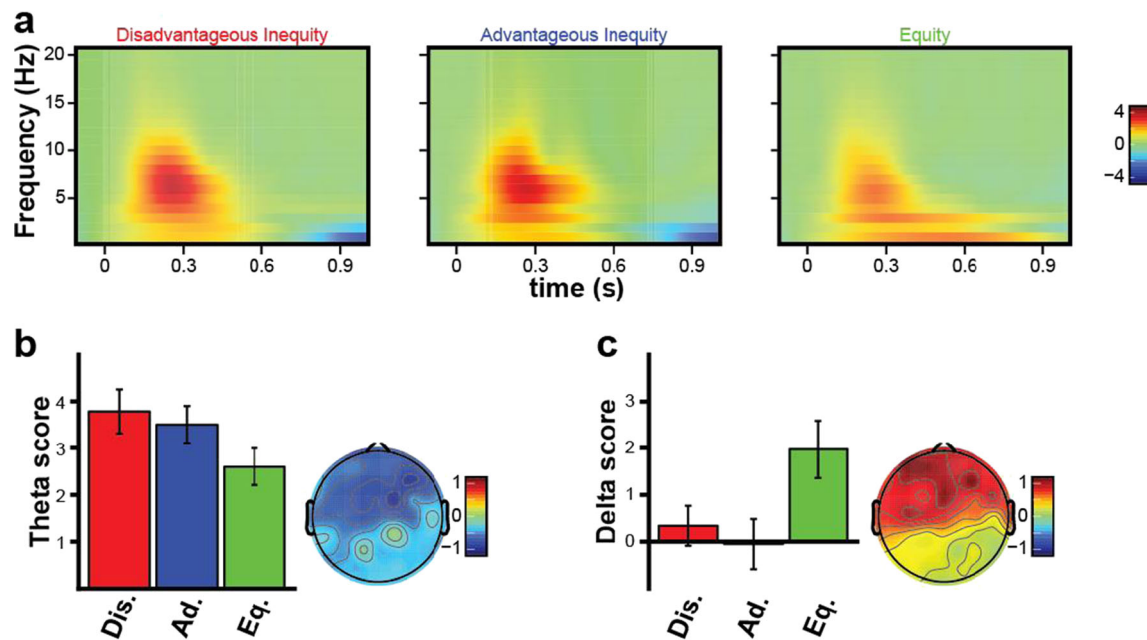
Figure 3 illustrates the results from a time-frequency analysis of participants' responses to each type of payout. First, scalograms were obtained for each participant (grand averages illustrated in Fig. 3a). Power in the theta band (4–8 Hz) between 250 and 350 ms was used to index the FRN, and power in the delta band (1–3 Hz) between 500 and 800 ms was used to index the LPP. There was a significant effect of wealth distribution on evoked power in the theta band ( $F(2,34) = 6.836$ ,  $p < 0.005$ ; Fig. 3b). Equitable payouts elicited significantly less theta power than disadvantageously inequitable payouts ( $t(17) = 3.363$ , corrected  $p < 0.005$ ) and advantageously inequitable payouts ( $t(17) = 2.626$ , corrected  $p < 0.05$ ), but theta responses evoked by advantageously and disadvantageously inequitable payouts were not different ( $t(17) = 0.918$ , corrected  $p > 0.2$ ). Similar to the FRN, the theta signal was greatest at central electrodes. There was also a significant effect of wealth distribution on delta power ( $F(2,34) = 5.844$ ,  $p < 0.01$ ; Fig. 3c), although no post-hoc tests achieved significance. Delta was more frontal than the LPP, but both signals were strong at central electrodes.

Crucially, we observed no relationship between theta and delta scores using the same approach that we used to determine the relationship between FRN and LPP scores in the time-domain ( $r = -.0021$ ,  $p > 0.9$ ,  $n = 54$ ), suggesting that the neural processes generating power in the theta and delta frequency bands (i.e., the processes generating the FRN and LPP) are largely independent.

### Discussion

The present study characterizes the effects of fairness, and equity specifically, on physiological signatures of hedonic processing. The FRN and the LPP, two ERP components known for their role in reward and affective evaluation, were modulated by the fairness of monetary outcomes. This was apparent both in the time domain, where the magnitude of both components was correlated, and in the time-frequency domain where the components were largely uncorrelated. Overall, our findings suggest that two distinct neural mechanisms of hedonic processing are modulated by the overall fairness of monetary outcomes.

The results of our experiments are readily interpretable in the framework of behavioral economics. Both the theory of inequality aversion put forth by Fehr and Schmidt (1999) and



**Fig. 3** Time-frequency domain results. **(a)** Morse wavelet scalograms for evoked responses at a CZ-centered cluster in each payout condition. Theta scores were computed as the mean power in the 4- to 8-Hz range in the 250- to 350-ms post-stimulus interval, and delta was scored as the average

in the 1- to 3-Hz range in the 500- to 800-ms post-stimulus intervals. **(b)** Mean theta scores and **(c)** mean delta scores across subjects with associated scalp distributions of wavelet power

behavior during wealth allocation experiments (Engelmann & Strobel, 2004) suggest that inequitable outcomes are less subjectively valuable than equitable outcomes, particularly when the decision-maker has the lowest payoff. Although not strictly required by the theory, Fehr and Schmidt also suggest that advantageously inequitable outcomes are generally more favorable than disadvantageously inequitable outcomes, as negative emotions brought on by the inequality are accompanied by perceived losses in the case of disadvantageous inequality. Our findings are grossly consistent with the predictions of this theory, as the neural signatures of reward-related neural processing considered by the present study favor equitable outcomes over inequitable outcomes, and tend to favor advantageous inequality over disadvantageous inequality. Additionally, our findings suggest a potential neural substrate for inequality aversion, as the influence of equity on early signatures of reward processing suggest that fairness exerts its effects on reward processing during bottom-up outcome evaluation processes. While this does not exclude the possibility that top-down processing can influence fairness-seeking behavior, it does suggest that the brain is capable of rapidly integrating social context into outcome evaluations.

The finding that the FRN is modulated by fairness is consistent with findings from experiments conducted during fMRI, given the hypothesized origin of the FRN. Early studies hypothesized that the FRN originates from a cortical region that receives dopaminergic projections from the basal ganglia, such as the anterior cingulate cortex (Holroyd & Coles, 2002). More recent work suggests that the FRN may originate

directly from activity in the basal ganglia (Foti, Weinberg, Dien, & Hajcak, 2011). Thus, the present findings suggest that social factors influence the early stages of reward processing in the basal ganglia. This interpretation is consistent with previous work showing that reward-related activity in the ventral striatum is affected by social inequality (Tricomi et al., 2010). The current study adds to this previous work by demonstrating that social factors exert their effects on early stages of mesolimbic hedonic processing, suggesting that social information is already present in neural processes reflecting bottom-up outcome evaluation. It is possible that humans process social factors by default, and may have difficulty inhibiting evaluations of others' welfare even when it does not affect one's own payoff. This interpretation is consistent with the altruistic behavior observed in anonymous games (e.g., Eckel & Grossman, 1996), as the recommendations of reflexive social evaluation processes could contradict selfish strategies that may require deliberation (Rand, Greene, & Nowak, 2012).

We interpret fairness-modulated activity in the theta band of the evoked spectral response as a separate index of the FRN given its similar latency and duration. Previous reports have implicated both early theta power within the evoked spectral response and the FRN in performance monitoring and reinforcement learning (Cavanagh, Frank, Klein, & Allen, 2010; Christie & Tata, 2009; Holroyd & Coles, 2002; Holroyd, Krigolson, & Lee, 2011), therein identifying a functional analogy between theta and the FRN that strengthens our conclusion that the two signals reflect the same neural processes.

Though our task does not involve learning an action-outcome contingency, the neural loci of reinforcement learning signals (e.g., the basal ganglia or the anterior cingulate cortex) are typically modulated by reward even when the payoff contingencies are fixed (Kable & Glimcher, 2007; Kolling, Behrens, Mars, & Rushworth, 2012). Moreover, the FRN has also been observed in outcome evaluation tasks that do not require learning (Hajcak et al., 2006). The discrepancy between the scalp distributions for the time and time-frequency domain indices of the FRN probably result from convolution of the time-domain FRN with unrelated evoked responses, such as the LPP.

Our results suggest that fairness also affects the LPP. Given that the LPP is modulated by emotional intensity rather than emotional valence (Schupp et al., 2000), the observed findings are consistent with the interpretation that equitable outcomes are more affectively salient than inequitable outcomes. This interpretation is also consistent with a recent study that observed relationships between LPP magnitude and BOLD activity in a number of regions, including the ventral aspects of prefrontal cortex (Liu, Huang, McGinnis-Deweese, Keil, & Ding, 2012). Parts of the ventral prefrontal cortex project to the limbic system (Carmichael & Price, 1995) and are involved in both subjective value representation (Kable & Glimcher, 2007) and processing of outcome inequality (Tricomi et al., 2010). Thus, it is possible that the LPP observed in the current study reflects sustained affective processing of equitable outcomes by cortical regions involved in outcome evaluation. Despite the central distribution of the LPP in the time-domain, spectrotemporal delta power was distributed frontally. Previous work has shown that the topography of the LPP tends to shift frontally following affective picture presentation despite an initial central/parietal distribution (e.g., Hajcak & Olvet, 2008), which may be due to a stationary but increasingly strong frontal delta component. Nonetheless, the latency and duration of the observed delta power provide converging evidence that it is simply a spectrotemporal measure of the LPP, thereby strengthening our conclusion that the LPP observed in the present study reflects sustained prefrontal reward processing.

Although the FRN appears to suggest that equitable outcomes are more favorable than inequitable outcomes, participants' explicit ratings suggested otherwise. Superficially, this mismatch appears inconsistent with the idea that the FRN reflects subjective evaluation and casts doubt on an interpretation of our findings that is centered on reward circuitry. Interestingly, a similar discrepancy between neural activity and subjective ratings was found by Tricomi et al. (2010). Those authors speculate that overt, behavioral evaluations (much like our participants' trial-by-trial ratings) reflect both the basic reward-related and affective processing that are captured by physiological measures, but also the complex cognitive

processes that might lead to strategic or self-serving responses (e.g., a desire to appear fair). Given the superior temporal resolution permitted with EEG, our results suggest that the FRN and LPP reflect early stages of the evaluations, with explicit behavior potentially reflecting additional, later stages of evaluation. Future work exploring factors that predict discrepancies between the ERP and ratings during social reward evaluation could offer insights on the nature of these later stages of evaluative processing.

Given that the expected value of equitable offers was greater than the expected value of disadvantageously inequitable offers, one could argue that our observed effects are explained entirely by personal evaluation. However, expected value alone is not sufficient to explain the effects of equity on time-frequency signals. Specifically, equitable offers elicited significantly less theta power than advantageously inequitable offers even though advantageously inequitable offers had a higher expected value to the participant. If the gains of the other were ignored, then advantageously inequitable payouts should have elicited more power than equitable payouts. It is therefore not likely that the participants only considered personal value when evaluating each payout, and that the equity of the payout had a strong influence on reward-related processing.

The basal ganglia and the ventral prefrontal cortex have been implicated in the processing of social inequality (Tricomi et al., 2010), and are hypothesized to underlie the FRN (Foti et al., 2011) and LPP (Liu et al., 2012), respectively. Thus, the present study offers a temporal characterization of social outcome evaluation that is distributed throughout the brain. We find that ERP components as early as several hundred ms following stimulus presentation carry information about equity, suggesting that the brain rapidly integrates social information into existing reward circuitry during outcome evaluation. Thus, it is possible that self- and other-oriented rewards use a common neural currency, consistent with econometric models that collapse self and other oriented gains onto the same subjective dimension (e.g., Fehr & Schmidt, 1999). Future experiments could test the common-currency hypothesis using source localization or simultaneous fMRI to determine whether ERPs elicited by equitable and inequitable gains are anatomically congruent.

## References

- Bernat, E. M., Nelson, L. D., Steele, V. R., Gehring, W. J., & Patrick, C. J. (2011). Externalizing psychopathology and gain-loss feedback in a simulated gambling task: Dissociable components of brain response revealed by time-frequency analysis. *Journal of Abnormal Psychology, 120*(2), 352.
- Boksem, M. A., & De Cremer, D. (2010). Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Social Neuroscience, 5*(1), 118–128.

- Bolton, G. E., & Ockenfels, A. (2000). ERC: a theory of equity, reciprocity, and competition. *The American Economic Review*, *90*(1), 166–193.
- Brittain, J. S., Halliday, D. M., Conway, B. A., & Nielsen, J. B. (2007). Single-trial multiwavelet coherence in application to neurophysiological time series. *Biomedical Engineering, IEEE Transactions on*, *54*(5), 854–862.
- Carmichael, S., & Price, J. (1995). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *Journal of Comparative Neurology*, *363*(4), 615–641.
- Cavanagh, J. F., Frank, M. J., Klein, T. J., & Allen, J. J. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *NeuroImage*, *49*(4), 3198–3209.
- Chang, S. W., Gariépy, J.-F., & Platt, M. L. (2013). Neuronal reference frames for social decisions in primate frontal cortex. *Nature Neuroscience*, *16*(2), 243–250.
- Christie, G. J., & Tata, M. S. (2009). Right frontal cortex generates reward-related theta-band oscillatory activity. *NeuroImage*, *48*(2), 415–422.
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.
- Eckel, C. C., & Grossman, P. J. (1996). Altruism in anonymous dictator games. *Games and Economic Behavior*, *16*(2), 181–191.
- Engelmann, D., & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *The American Economic Review*, *94*(4), 857–869.
- Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785–791.
- Fehr, E., & Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868. doi:10.1162/003355399556151
- Foti, D., Weinberg, A., Dien, J., & Hajcak, G. (2011). Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: Temporospatial principal components analysis and source localization of the feedback negativity. *Human Brain Mapping*, *32*(12), 2207–2216.
- Gehring, W. J., & Willoughby, A. R. (2004). Are all medial frontal negativities created equal? Toward a richer empirical basis for theories of action monitoring. *Errors, conflicts, and the brain. Current opinions on performance monitoring* (pp. 14–20). Leipzig, Germany: Max Planck Institute for Human Cognitive and Brain Sciences.
- Hajcak, G., Dunning, J. P., & Foti, D. (2009). Motivated and controlled attention to emotion: Time-course of the late positive potential. *Clinical Neurophysiology*, *120*(3), 505–510. doi:10.1016/j.clinph.2008.11.028
- Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2006). The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biological Psychology*, *71*(2), 148–154.
- Hajcak, G., & Olvet, D. M. (2008). The persistence of attention to emotion: brain potentials during and after picture presentation. *Emotion*, *8*(2), 250.
- Harbaugh, W. T., Mayr, U., & Burghart, D. R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science*, *316*(5831), 1622–1625.
- Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679.
- Holroyd, C. B., Krigolson, O. E., & Lee, S. (2011). Reward positivity elicited by predictive cues. *Neuroreport*, *22*(5), 249–252.
- Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, *10*(12), 1625–1633.
- Kayser, J., & Tenke, C. E. (2006). Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks. *Clinical Neurophysiology*, *117*(2), 348–368.
- Kolling, N., Behrens, T. E., Mars, R. B., & Rushworth, M. F. (2012). Neural mechanisms of foraging. *Science*, *336*(6077), 95–98.
- Kurzban, R., & Houser, D. (2005). Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(5), 1803–1807. doi:10.1073/pnas.0408759102
- Lilly, J. (2011). JLAB: Matlab freeware for data analysis, Version 0.94
- Lilly, J. M., & Olhede, S. C. (2012). Generalized Morse Wavelets as a Superfamily of Analytic Wavelets. *Signal Processing, IEEE Transactions on*, *60*(11), 6036–6041. doi:10.1109/tsp.2012.2210890
- Liu, Y., Huang, H., McGinnis-Deweese, M., Keil, A., & Ding, M. (2012). Neural substrate of the late positive potential in emotional processing. *The Journal of Neuroscience*, *32*(42), 14563–14572.
- Miller, G. A., Gratton, G., & Yee, C. M. (1988). Generalized implementation of an eye movement correction procedure. *Psychophysiology*, *25*(2), 241–243.
- Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., & Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences*, *103*(42), 15623–15628.
- Mussel, P., Hewig, J., Allen, J. J., Coles, M. G., & Miltner, W. (2014). Smiling faces, sometimes they don't tell the truth: Facial expression in the ultimatum game impacts decision making and event-related potentials. *Psychophysiology*, *51*(4), 358–363.
- Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13.
- 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.
- Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*(7416), 427–430.
- 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.
- Schupp, H. T., Cuthbert, B. N., Bradley, M. M., Cacioppo, J. T., Ito, T., & Lang, P. J. (2000). Affective picture processing: the late positive potential is modulated by motivational relevance. *Psychophysiology*, *37*(2), 257–261.
- Tricomi, E., Rangel, A., Camerer, C. F., & O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature*, *463*(7284), 1089–1091.