doi:10.1093/scan/nst051 SCAN (2014) 9, 731–736

A neural signature of the creation of social evaluation

Roman Osinsky,[1] Patrick Mussel,[1] Linda Öhrlein,[1] and Johannes Hewig[1,2]

[1]Department of Psychology I, Julius-Maximilians-University Würzburg, 97070 Würzburg, Germany and [2]Department of Psychology, Friedrich-Schiller-University Jena, 07743 Jena, Germany

Previous research has shown that receiving an unfair monetary offer in economic bargaining elicits also-called feedback negativity (FN). This scalp-recorded brain potential probably reflects a bad-*vs*-good evaluation in the medial frontal cortex and has been linked to fundamental processes of reinforcement learning. In the present study, we investigated whether the evaluative mechanism indexed by the FN is also involved in learning who is an unfair *vs* fair bargaining partner. An electroencephalogram was recorded while participants completed a computerized version of the Ultimatum Game, repeatedly receiving fair or unfair monetary offers from alleged other participants. Some of these proposers were either always fair or always unfair in their offers. In each trial, participants first saw a portrait picture of the respective proposer before the monetary offer was presented. Therefore, the faces could be used as predictive cues for the fairness of the pending offers. We found that not only unfair offers themselves induced a FN, but also (over the task) faces of unfair proposers. Thus, when interaction partners repeatedly behave in an unfair way, their faces acquire a negative valence, which manifests in a basal neural mechanism of bad-*vs*-good evaluation.

Keywords: social evaluation; feedback negativity; ultimatum game; evaluative conditioning

INTRODUCTION

Fairness plays a fundamental role in many domains of social life, including, for example, family, work, trading and politics. Accordingly, the consequence of being treated unfairly may vary in degree, ranging from a petty quarrel between siblings, to resentments between trading partners, to severe diplomatic tensions or even hostile acts between nations or ethnic groups. In many situations, it can be helpful to learn from direct experience who is an unfair *vs* fair interaction partner, enabling, for instance, adjustments in future social behavior (e.g. avoiding interactions with someone who has repeatedly behaved unfairly in the past). In the present study, we are interested in electrocortical correlates of this process of fairness-based reputation building.

In the past, researchers have used experimental economic tasks like the Ultimatum Game (UG; Güth *et al.*, 1982) for studying the influence of fairness in the distribution of limited resources. In the UG, one player (the proposer) divides a certain amount of money between him/herself and another player (the responder) and can do so in a fair (equal stakes) or unfair way (unequal stakes). The responder then decides whether to accept this offer or to reject it. In case of acceptance, the money is divided as proposed. However, if the responder rejects the offer, none of the players obtains any money. In accordance with the normative idea of economic rationality (von Neumann and Morgenstern, 1953), one would expect that responders accept all kind of offers to maximize their personal gain. In contrast, however, responders typically tend to reject extremely unfair offers (Güth *et al.*, 1982), a behavior probably mediated by a negative emotional reaction to inequity (Pillutla and Murnighan, 1996; Fehr and Gächter, 2002; Sanfey *et al.*, 2003; van't Wout *et al.*, 2006; Hewig *et al.*, 2011).

Moreover, it has been demonstrated that unfair compared with fair (or less unfair) offers in the UG evoke a frontocentral negative-going deflection in the event-related potential (ERP) of the electroencephalogram (Polezzi *et al.*, 2008; Boksem and De Cremer, 2010; Hewig *et al.*, 2011; van der Veen and Sahibdin, 2011; Wu *et al.*, 2011). This ERP component, which is commonly referred to as *feedback negativity* (FN), generally occurs ~300 ms following the onset of a stimulus signaling an unfavorable relatively to a favorable event (Miltner *et al.*, 1997). Functionally, it has been argued that the FN reflects a basal and context-dependent bad-*vs*-good evaluation (Gehring and Willoughby, 2002; Holroyd *et al.*, 2004a; Yeung and Sanfey, 2004; Hajcak *et al.*, 2006; Osinsky *et al.*, 2012a). As we will outline further below, this evaluative mechanism is probably linked to fundamental neural processes of reinforcement learning (Holroyd and Coles, 2002; Nieuwenhuis *et al.*, 2004). Moreover, source-locating approaches have identified the medial frontal cortex as the main generator of the FN (e.g. Miltner *et al.*, 1997; Gehring and Willoughby, 2002; Holroyd *et al.*, 2004b; Hewig *et al.*, 2007). There is multiple evidence that this brain region is crucially involved in social evaluation and decision making (for some overviews see, Rushworth *et al.*, 2007; Rilling and Sanfey, 2011), with some studies pointing to an important role in the generation of reputation during social interactions (King-Casas *et al.*, 2005; Krueger *et al.*, 2007; Behrens *et al.*, 2008; Pejic *et al.*, 2013). It may therefore be speculated that the basal evaluative mechanism indexed by the FN is also involved in learning who is a fair and who is an unfair interaction partner in the UG.

Some evidence already suggests that during repetitive interpersonal bargaining an affective value is ascribed to the opponent, based on her/his fairness in the preceding interaction history (Singer *et al.*, 2006; Hofman *et al.*, 2012). This process is likely to be mediated by evaluative conditioning, i.e. the associative pairing between an affective stimulus (e.g. an unfair monetary offer) and a formerly neutral stimulus (e.g. the face of the proposer) (De Houwer *et al.*, 2001). We assume that this associative valence transfer should also manifest in form of the FN. This assumption is primarily based on the prominent Reinforcement Learning theory of FN generation (RL-FN theory; Holroyd and Coles 2002; Nieuwenhuis *et al.*, 2004), claiming that the FN reflects the dopaminergic signaling of reward prediction errors forwarded to the medial frontal cortex, in particular the anterior cingulate cortex. This RL-FN theory, in turn, was inspired by findings from intracortical recordings in monkeys, showing that midbrain dopamine neurons respond to an unexpected occurrence (phasic burst in activity) or omission (phasic dip) of a reward stimulus. Importantly, during learning, this phasic signal propagates back in time to an earlier stimulus, which is predictive for the upcoming reward (i.e. the conditioned stimulus; Schultz *et al.*, 1997). If the FN is an electrocortical index of

Received 14 November 2012; Accepted 28 March 2013
Advance Access publication 1 April 2013
This work was supported by the Volkswagen Foundation [Schumpeter-Fellowship to J.H.].
Correspondence should be addressed to Roman Osinsky, Department of Psychology I, Julius- Maximilians-University Würzburg, Marcusstr. 9-11, 97070 Würzburg, Germany. E-mail: roman.osinsky@uni-wuerzburg.de

these neural processes in reinforcement learning, as claimed in the RL-FN theory, it should not only be evoked by an outcome stimulus but also by a cue stimulus, which validly predicts the favorableness of the upcoming outcome. Indeed, this has been shown for simple visual cue stimuli in several recent studies (Dunning and Hajcak, 2007; Baker and Holroyd, 2009; Holroyd *et al.*, 2011; Liao *et al.*, 2011; Walsh and Anderson, 2011). Moreover, numerous molecular–genetic and pharmacological studies support the idea that the FN and the error-related negativity as its response-locked equivalent are directly associated with the magnitude of dopaminergic transmission (e.g. Santesso *et al.*, 2009; Mueller *et al.*, 2011; Smillie *et al.*, 2011; Osinsky *et al.*, 2012b).

Based on the aforementioned work on the FN and its link to reinforcement learning, we hypothesize that after repeatedly receiving unfair offers from one and the same proposer over multiple encounters in the UG, the face of this proposer starts to evoke a FN, reflecting a back propagation of the phasic prediction-error signal and, consequently, a learned bad-*vs*-good evaluation of this person. Moreover, after learning the contingency between the unfairness of the offer and the preceding face, the offer itself should not be unexpected anymore. Therefore, over the course of the UG, FN amplitude should become less sensitive to the offer and more sensitive to the face.

METHODS

Participants

Thirty-four student subjects [26 female; mean age = 23.2 years, standard deviation (s.d.) = 2.7] participated for a monetary compensation of €5. They were told that they could gain more money during the UG, depending on their task behavior. As, unknown to them, they played against the computer (see below), they finally received a fixed additional payout of €10 to keep any frustration about the deception as low as possible (€9.72 would be the maximum payout in the UG). Data of five participants were excluded from analyses due to low number of artifact-free trials (<20 per condition). All participants gave written informed consent and the study was approved by the local ethics committee of the University of Jena.

Experimental procedure

After arrival, participants received a general instruction about the UG, telling them that they would play with other participants. Unknown to them, however, they played against the computer. To enhance plausibility of the cover story, a picture of each participant was taken and she/he then played the UG in the role of the proposer, making 24 offers on a query sheet. In each offer, she/he could divide ¢10 into two shares: one for her/him and one for the other player. There were three predefined proposal options: 9/1 (¢9 for the proposer, ¢1 for the responder), 7/3 (¢7 for the proposer, ¢3 for the responder) and 5/5 (¢5 for the proposer, ¢5 for the responder). Participants were told that these offers would later be presented to other participants who could then decide whether to accept or reject each offer. They were also told that, for each proposal, they would later receive the respective amount of money if the offer is accepted by the other player.

Afterward they played a computerized version of the UG in the role of the responder while electroencephalography (EEG) was recorded. In this critical phase, they were first told that they would receive monetary offers, which have been made by six other participants before. In truth, however, they received offers from six pseudo-proposers (three female, three male), portrait photographs of whom were taken from a standardized stimulus set (Langner *et al.*, 2010). The UG consisted of 216 trials in which all stimuli were centrally presented on a computer screen. Each trial started with a fixation cross presented for a variable duration of 500–1000 ms followed by a photograph of a single proposer for a duration of 1500 ms (see Figure 1). Afterward a fixation cross was presented again for 500–1000 ms before participants received an offer about splitting ¢10. This offer was presented in form of a pie chart, representing the proposed shares (light-gray portion = share of the proposer, dark-gray portion = share of the responder). In addition, the proposal was also presented in written form above the pie chart. The offer could either be fair (5/5), slightly unfair (7/3) or highly unfair (9/1). By pressing the left or right response button, participants accepted or rejected the offer. After button press, a fixation cross was presented again for 500 ms. Finally, participants were informed about the amount of money booked to their game account (duration = 1250 ms) before the next trial started. The whole task was divided into three blocks (72 trials each), which were separated by short breaks of 15 s. In each block, participants received 12 offers from each proposer. Two of the proposers always made fair offers (fair proposers: each proposer made 12 offers of a 5/5 split) and two always made highly unfair offers (unfair proposers: each proposer made 12 offers of a 9/1 split). Additionally, in 24 filler trials per block, two proposers made all kinds of offers with an equal frequency (each proposer made four offers of a 5/5 split, four offers of a 7/3 split and four offers of a 9/1 split). Thus, the relative offer probability was 0.44 for the 5/5 split, 0.44 for the 9/1 split and 0.11 for the 7/3 split. For each proposer type, one male and one female face were presented, with assignment of individual pictures to proposer categories being balanced across participants. Data from the filler trials were not analyzed. Thus, all analyses are based only on data obtained from trials with fair and unfair proposers. For each participant, a pseudo-random trial order was generated with the restriction that each proposer occurs twice in smaller subblocks of 12 trials, guaranteeing an equal distribution of single proposers across the task. Moreover, there was no direct proposer repetition.

After completing the task, all proposers were presented again and subjects were asked to rate each of them with regard to their fairness and kindness on two respective 7-point scales, ranging from very unfair/unkind (1) to very fair/kind (7). Finally, participants were informed about the deception and paid out.

All stimuli were presented on a 17" screen with a black background. Stimulus presentation and response recordings were controlled by Presentation 12.2 software (Neurobehavioral Systems Inc., Albany, CA, USA). During the task, participants were seated in a comfortable chair with a distance of ~70 cm between the head and the screen. Each of the face pictures was 18.8 cm high and 14.8 cm wide, resulting in a visual angle of $\sim 15.3^\circ \times 12.1^\circ$. The pie charts had a diameter of 9 cm (7.4° visual angle). Subjects performed the task in an acoustically and electrically shielded chamber.

EEG recordings and analyses

While subjects performed the UG, EEG (analog bandpass: 0.1–80 Hz, sampling rate: 250 Hz) was recorded from 31 scalp sites according to the 10–20 system (Fp1, Fp2, F9, F7, F3, Fz, F4, F8, F10, FC5, FC1, FC2, FC6, T7, C3, C4, T8, TP9, CP1, CP2, TP10, P7, P3, Pz, P4, P8, PO9, O1, O2, PO10 and Iz), using Ag/AgCl electrodes and a BrainAmpDC amplifier (Brain Products GmbH, Gilching, Germany). During recording, impedances were kept below 10 kΩ and electrodes were referenced to the vertex (Cz). Data were processed offline, using Brain Vision Analyzer 2.0 software (Brain Products GmbH, Gilching, Germany). First, data were re-referenced to the averaged mastoid electrodes, and electrode Cz was reinstated. Data were then filtered, using a 20 Hz low-pass filter. For detection of blinks and eye movements, a horizontal (bipolar channel pair of F9 and F10) and vertical electrooculogram (Fp1 and left IO) was used. Ocular artifacts were corrected with the method suggested by Gratton *et al.* (1983). Subsequently, the

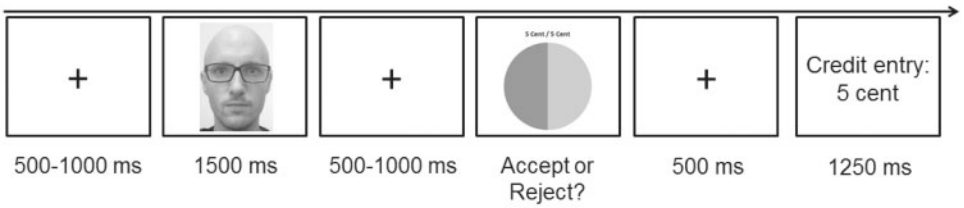

Fig. 1 Schematic depiction of a single trial in the Ultimatum Game (UG). In the original task, color portrait photos were used (note that the example photo was not used in the study).

EEG was segmented into event-locked epochs of 1000 ms (−200 to 800 ms). Larger artifacts were automatically detected by an algorithm implemented in Brain Vision Analyzer 2.0 software, and epochs were discarded if applicable. For this purpose, the following exclusion criteria were applied: maximal voltage difference within the epoch >150 μV and maximal voltage step of 20 μV/ms. At least 20 artifact-free trials were available for averaging per participant and condition. The 200 ms interval before the stimulus was used for baseline correction. The FN was quantified as mean amplitude in the time window 270–370 ms following stimulus onset at electrodes F3, Fz, F4, C3, Cz, C4, P3, Pz and P4. The chosen time window was based on visual inspection of the ERP waves and is similar to previous studies on the FN in the UG (Boksem and De Cremer, 2010; Hewig *et al.*, 2011; Wu *et al.*, 2011). At this point, it should be noted that FN latency may strongly vary between studies what might be caused, for example, by differences in stimulus complexity or task-specific parameters. Such factors can also influence the general shape of wave forms (compare, for instance, Wu *et al.*, 2011, Hewig *et al.*, 2011 and the present study with regard to ERP waveforms and offer stimuli). Moreover, the FN should be considered as a relative negativity defined by the difference between conditions rather than an absolute negativity in a certain single condition (see, for example, Dunning and Hajcak, 2007).

In addition to electrophysiological data, we also analyzed acceptance rates (ARs, percentage of accepted offers in relation to the total number of offers per offer type) as well as explicit ratings of proposers' fairness and kindness after the task. Statistical analyses were conducted using SPSS software (IBM, Armonk, NY, USA). For analyses including more than two levels repeated-measures analyses of variance (ANOVA) were conducted. In case of violation of sphericity assumption, epsilon (E; Greenhouse-Geisser correction) and corrected P values are reported. For pairwise comparisons, paired t-tests were conducted. All tests were two-tailed, and P values $\leq$0.05 were considered significant.

RESULTS

Acceptance rates and explicit ratings

ARs were entered into a 2 × 3 repeated-measures ANOVA with the within-subject factors 'offer type' (5/5 split or 9/1 split) and 'task block' (1, 2 and 3). Fair offers (mean AR = 99.23%, s.d. = 2.18) were significantly more often accepted than unfair offers (mean AR = 23.37%, s.d. = 32.26), $F(1,28) = 159.58$, $P < 0.001$, $\eta_p^2 = 0.85$. Neither the main effect of 'task block' nor the interaction of 'task block' × 'offer type' were significant, all $F(1,56) < 1.06$, all $P > 0.32$, indicating that ARs were stable across the UG.

The post-task fairness and kindness ratings were analyzed with pairwise t-test. Fair proposers were rated as being more fair [mean = 6.40, s.d. = 0.75; $t(28) = 13.58$, $P < 0.001$] and kind [mean = 5.57, s.d. = 1.22; $t(28) = 5.93$, $P < 0.001$] than unfair proposers (fairness: mean = 2.24, s.d. = 1.22; kindness: mean = 3.24, s.d. = 1.47). Thus, participants have learned to discriminate between the two proposer types during the UG.

Offer-locked feedback negativity

Offer-locked mean amplitudes in the FN range were entered into a 3 × 3 × 3 × 2 repeated-measures ANOVA with the within-subject factors 'position' (frontal, central, parietal), 'electrode' (left, middle, right), 'task block' (1, 2, 3) and 'offer type' (5/5 split or 9/1 split). Unfair compared with fair offers elicited a clear FN as a relative negative deflection 270–370 ms after offer onset, $F(1,28) = 11.91$, $P = 0.002$, $\eta_p^2 = 0.30$ (see Figure 2A). As indicated by a significant three-way interaction of 'offer' × 'position' × 'electrode', $F(4,112) = 6.43$, $P < 0.001$, $E = 0.80$, $\eta_p^2 = 0.19$, this FN effect had a frontocentral maximum with a slight left shift. Across task blocks, the amplitude in the FN time range generally decreased, especially at left frontal and central electrodes as indicated by a significant interaction of 'task block' × 'position' × 'electrode', $F(8,224) = 5.43$, $P < 0.001$, $E = 0.67$, $\eta_p^2 = 0.16$. However, none of the interaction terms containing both the factors 'offer type' and 'task block' approached significance (all $P > 0.28$). Thus, the offer-locked FN did not substantially vary across the task.

Face-locked feedback negativity

Again we calculated a 3 × 3 × 3 × 2 repeated-measures ANOVA, containing the within-subject factors 'position' (frontal, central, parietal), 'electrode' (left, middle, right), 'task block' (1, 2, 3) and 'proposer type' (fair or unfair). Most important, we observed a significant 'task block' × 'proposer type' interaction, $F(2,56) = 5.02$, $P = 0.011$, $E = 0.96$, $\eta_p^2 = 0.15$. As can be seen in Figure 2B and as post hoc pairwise comparisons revealed, for unfair proposers, the amplitude in the FN range slightly increased from the first to the second task block [$t(28) = 2.20$, $P = 0.036$] before it finally strongly decreased in the last task block [block 3 *vs* block 2: $t(28) = 5.06$, $P < 0.001$; block 3 *vs* block 1: $t(28) = 2.97$, $P = 0.006$]. For faces of fair proposers, no significant amplitude changes in the FN range across task blocks could be observed (all $P > 0.92$). Moreover, the difference between fair and unfair proposer faces in FN amplitude was significant only in the third task block [$t(28) = 4.21$, $P < 0.001$; see Figure 2C] but not in the first [$t(28) = 0.94$, $P = 0.357$] or second task block [$t(28) = 0.57$, $P = 0.576$]. In sum, across the task, a FN developed in response to faces of unfair as compared with fair proposers.

To directly test whether the face-locked FN in the third task block resembles the offer-locked FN observed across the whole task, we performed an omnibus repeated-measures ANOVA with the factors 'stimulus type' (offer or face), 'position' (frontal, central, parietal), 'electrode' (left, middle, right) and 'fairness' (unfair offer/face or fair offer/face). Importantly, neither the two-way interaction of 'stimulus type' × 'fairness' [$F(2, 28) = 0.01$, $P = 0.96$] nor any higher interaction with 'position' or 'electrode' (all $P > 0.11$) reached significance. Thus, the FN effect evoked by the faces in the third task block did not significantly differ from the one observed for the offers across all blocks.

We further examined whether the face-locked FN effect was related to the ratings after the task. For this purpose, we analyzed correlations (Kendalls τ b) between FN difference scores (difference score 1: block 3

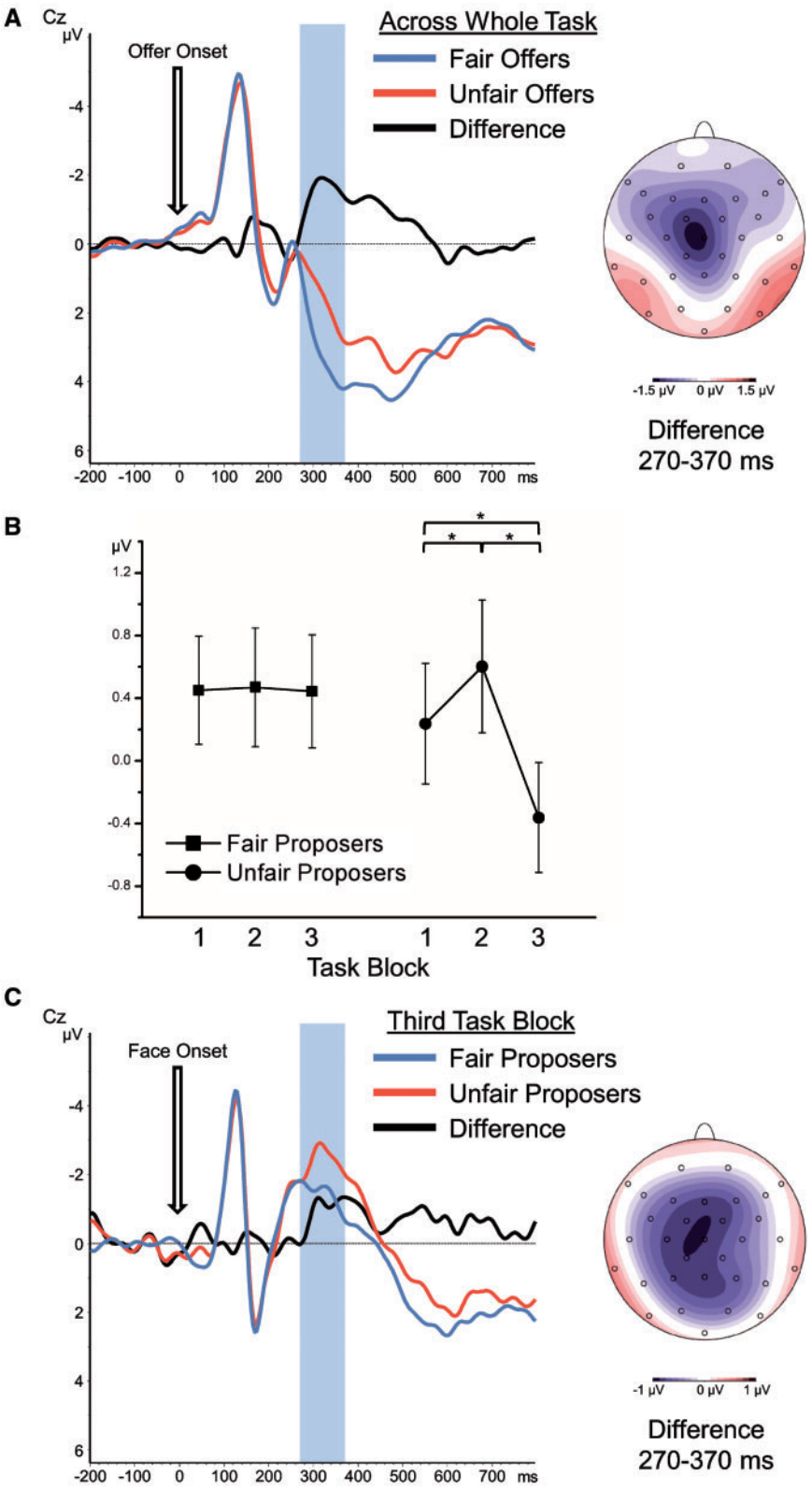

Fig. 2 Event-related potentials (ERP). (**A**) Offer-locked ERP waveforms at electrode Cz and scalp distribution of the unfair minus fair difference in the FN time window (270–370 ms after offer onset). (**B**) Mean amplitudes (±S.E.M.) for the time window 270–370 ms after face onset across electrode positions F3, Fz, F4, C3, Cz, C4, P3, Pz, P4. Asterisks indicate significant pairwise differences ($P < 0.05$). (**C**) Face-locked ERP waveforms at electrode Cz in task block 3 for fair and unfair proposers. The topographical map shows the scalp distribution of the unfair minus fair difference in the FN time window (270–370 ms after face onset).

minus block 1 for unfair proposers; difference score 2: unfair proposers minus fair proposers for task block 3) and explicit fairness and kindness ratings of unfair proposers. We observed only small and statistically insignificant correlations ($\tau = -0.01$ to -0.21; all $P > 0.13$), indicating that the FN evoked by the faces of unfair proposers at the end of the UG was unrelated to the explicit ratings of those proposers following the task.

DISCUSSION

In the present study, we aimed to investigate whether the evaluative mechanism reflected by the FN is involved in the creation of social evaluation during repeated interpersonal bargaining. Not surprisingly, we observed higher ARs for fair compared with unfair offers (cf. Güth *et al.*, 1982). Moreover, the explicit ratings of proposer fairness and kindness after the UG were strongly modulated by proposer type, indicating that our participants learned to discriminate between fair and unfair proposers. Most important, we found that over multiple encounters, faces of unfair proposers start to evoke a similar FN response as an unfair offer itself. Thus, when a person repeatedly behaves in an unfair way, her/his face becomes a valid indicator for the inequity of the pending offer. This learned contingency does then manifest in a bad-*vs*-good evaluation of the face as reflected by the FN (cf. (Gehring and Willoughby, 2002; Holroyd *et al.*, 2004a; Yeung and Sanfey, 2004; Hajcak *et al.*, 2006). Given that the FN originates from the medial frontal cortex (Miltner *et al.*, 1997; Gehring and Willoughby, 2002; Holroyd *et al.*, 2004b; Hewig *et al.*, 2007), our results support other prior findings, indicating that this area is crucially involved in the generation of social reputation in terms of associative learning (King-Casas *et al.*, 2005; Krueger *et al.*, 2007; Behrens *et al.*, 2008). Generally, our findings are consistent with previous studies showing that faces as complex visual stimuli can acquire an affective value (e.g. Baeyens *et al.*, 1992; Todrank *et al.*, 1995; Schneider *et al.*, 1999; Petrovic *et al.*, 2008; Pejic *et al.*, 2013), a process that can be subsumed under the broader concept of evaluative conditioning (De Houwer *et al.*, 2001, Walther *et al.*, 2005). More specifically, the observed cross-task development of a FN in response to faces of unfair proposers is consistent with findings from several prior studies, which have reported a similar pattern for simpler visual cue stimuli (Dunning and Hajcak, 2007; Baker and Holroyd, 2009; Holroyd *et al.*, 2011; Liao *et al.*, 2011; Walsh and Anderson, 2011). Our study therefore contributes to a growing literature supporting one main tenet of the RL-FN theory (Holroyd and Coles, 2002), according to which reward prediction error signals (as indexed by the FN) propagate back to outcome-predictive cues during learning. To the best of our knowledge, the present study is the first to demonstrate that this mechanism is also involved in the experience-based creation of social reputation.

However, it should also be noted that our data are not in line with another central assumption of the RL-FN theory: when the favorableness of an outcome becomes predictable, the outcome-locked FN should decrease. Accordingly, the magnitude of the offer-locked FN should have declined over task blocks in our study. Contrary to this, we observed a rather stable offer-locked FN over the course of the UG. Our results are therefore in contrast to previous studies, which have indeed reported a reduction in outcome-locked FN when the outcome was validly predicted by preceding cue stimuli (Potts *et al.*, 2006; Baker and Holroyd, 2009; Holroyd *et al.*, 2011; Liao *et al.*, 2011; Martin and Potts, 2011). At this point, some important differences between these other studies and the present one should be considered. First, in some of the aforementioned studies, participants were explicitly informed about the meaning of the cue stimuli so that cue-based outcome expectancies were established before the respective task began (Baker and Holroyd, 2009; Liao *et al.*, 2011). In the other studies, simple images of gold bars (reward) and lemons (non-reward) were used both as outcome stimuli *and* as cue stimuli in a passive S1–S2 paradigm (Potts *et al.*, 2006; Holroyd *et al.*, 2011; Martin and Potts, 2011). Therefore, the cue-outcome contingencies in these studies were probably learned fast. In sum, in all mentioned studies, simple visual cues were used, the predictive meaning of which was either known to the participants at the beginning of the task or could be easily learned. In contrast, the six faces of strangers used in our study are much more complex and, therefore, learning the face-offer contingency may have progressed slower. Thus, although a back propagation of the reward-prediction error signal to the faces may have taken place before the end of our UG, the learning process must not necessarily have been completed at this time, and therefore, unfair offers still evoked a substantial FN. Future studies may further investigate the potential influence of stimulus complexity on FN dynamics in cue-outcome contingency learning.

Our results may have some important implications for future research on the role of fairness not only in economic bargaining but daily social interactions in general. Whether we have repeatedly treated someone else fairly or unfairly in the past appears to predefine how the neurocognitive system of this person evaluates us, or more precisely our face as a unique and visible physical feature of our identity. The latency of the FN in relation to the face stimulus suggests that some aspects of this evaluation are completed in a blink of an eye, i.e. ~320 ms following the onset of face presentation. This is also in line with other recent ERP evidence showing that implicit social attitudes toward others are activated automatically at an early stage of information processing (van der Lugt *et al.*, 2012). Such automatic evaluations are probably distinguishable from more elaborated and controlled processes of social judgment (cf. Satpute and Lieberman, 2006; Cunningham and Zelazo, 2007; Adolphs, 2009), as is also indicated by the absence of substantial correlations between the face-evoked FN and the explicit ratings in our study. It is up to future research to further specify the neural processes underlying more reflective mechanisms in social evaluation. For the moment, however, we can conclude that the basal bad-*vs*-good evaluation that is mirrored by the FN is involved in learning who is a fair *vs* unfair interaction partner. Therefore, the present study also clearly underlines the fundamental role of fairness in the creation of social evaluation and personal reputation.

Conflict of Interest

None declared.

REFERENCES

Adolphs, R. (2009). The social brain: neural basis of social knowledge. *Annual Review of Psychology*, *60*, 693–716.

Baker, T.E., Holroyd, C.B. (2009). Which way do I go? Neural activation in response to feedback and spatial processing in a virtual T-Maze. *Cerebral Cortex*, *19*, 1708–22.

Behrens, T.E.J., Hunt, L.T., Woolrich, M.W., Rushworth, M.F.S. (2008). Associative learning of social value. *Nature*, *456*(7219), 245–9.

Boksem, M.A.S., De Cremer, D. (2010). Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Social Neuroscience*, *5*(1), 118–28.

Cunningham, W.A., Zelazo, P.D. (2007). Attitudes and evaluations: a social cognitive neuroscience perspective. *Trends in Cognitive Sciences*, *11*(3), 97–104.

De Houwer, J., Thomas, S., Baeyens, F. (2001). Associative learning of likes and dislikes: a review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, *127*(6), 853–69.

Dunning, J.P., Hajcak, G. (2007). Error-related negativities elicited by monetary loss and cues that predict loss. *Neuroreport*, *18*(17), 1875–8.

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

Gehring, W.J., Willoughby, A.R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science*, *295*(5563), 2279–82.

Gratton, G., Coles, M.G.H., Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–84.

Güth, W., Schmittberger, R., Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, *3*(4), 367–88.

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), 148–54.

Hewig, J., Kretschmer, N., Trippe, R.H., et al. (2011). Why humans deviate from rational choice. *Psychophysiology*, *48*(4), 507–14.

Hewig, J., Trippe, R., Hecht, H., Coles, M.G.H., Holroyd, C.B., Miltner, W.H.R. (2007). Decision-making in Blackjack: an electrophysiological analysis. *Cerebral Cortex*, *17*(4), 865–77.

Hofman, D., Bos, P.A., Schutter, D.J.L.G., van Honk, J. (2012). Fairness modulates non-conscious facial mimicry in women. *Proceedings. Biological Sciences/The Royal Society*, *279*(1742), 3535–9.

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.

Holroyd, C.B., Krigolson, O.E., Lee, S. (2011). Reward positivity elicited by predictive cues. *NeuroReport*, *22*, 249–52.

Holroyd, C.B., Larsen, J.T., Cohen, J.D. (2004a). Context dependence of the event-related brain potential associated with reward and punishment. *Psychophysiology*, *41*(2), 245–53.

Holroyd, C.B., Nieuwenhuis, S., Yeung, N., et al. (2004b). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nature Neuroscience*, *7*(5), 497–8.

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

Krueger, F., McCabe, K., Moll, J., et al. (2007). Neural correlates of trust. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50), 20084–9.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, *24*(8), 1377–88.

Liao, Y., Gramann, K., Feng, W., Deák, G.O., Li, H. (2011). This ought to be good: brain activity accompanying positive and negative expectations and outcomes. *Psychophysiology*, *48*(10), 1412–19.

Martin, L.E., Potts, G.F. (2011). Medial frontal event-related potentials and reward prediction: do responses matter? *Brain and Cognition*, *77*(1), 128–34.

Miltner, W.H., Braun, C.H., Coles, M.G. (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), 788–98.

Mueller, E.M., Makeig, S., Stemmler, G., Hennig, J., Wacker, J. (2011). Dopamine effects on human error processing depend on catechol-O-methyltransferase VAL158MET genotype. *Journal of Neuroscience*, *31*, 15818–25.

Nieuwenhuis, S., Holroyd, C.B., Mol, N., Coles, M.G. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neuroscience and Biobehavioral Reviews*, *28*, 441–8.

Osinsky, R., Hewig, J., Alexander, N., Hennig, J. (2012a). COMT Val 158Met genotype and the common basis of error and conflict monitoring. *Brain Research*, *1452*, 108–18.

Osinsky, R., Mussel, P., Hewig, J. (2012b). Feedback-related potentials are sensitive to sequential order of decision outcomes in a gambling task. *Psychophysiology*, *49*, 1579–89.

Pejic, T., Hermann, A., Vaitl, D., Stark, R. (2013). Social anxiety modulates amygdala activation during social conditioning. *Social Cognitive and Affective Neuroscience*, *8*(3), 267–76.

Petrovic, P., Kalisch, R., Pessiglione, M., Singer, T., Dolan, R.J. (2008). Learning affective values for faces is expressed in amygdala and fusiform gyrus. *Social Cognitive and Affective Neuroscience*, *3*(2), 109–18.

Pillutla, M.M., Murnighan, J.K. (1996). Unfairness, anger, and spite: emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes*, *68*(3), 208–24.

Polezzi, D., Daum, I., Rubaltelli, E., et al. (2008). Mentalizing in economic decision-making. *Behavioural Brain Research*, *190*(2), 218–23.

Potts, G.F., Martin, L.E., Burton, P., et al. (2006). When things are better or worse than expected: the medial frontal cortex and the allocation of processing resources. *Journal of Cognitive Neuroscience*, *18*(7), 1112–9.

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

Rushworth, M.F.S., Buckley, M.J., Behrens, T.E.J., Walton, M.E., Bannerman, D.M. (2007). Functional organization of the medial frontal cortex. *Current Opinion in Neurobiology*, *17*, 220–7.

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

Santesso, D.L., Evins, A.E., Frank, M.J., Schetter, E.C., Bogdan, R., Pizzagalli, D.A. (2009). Single dose of a dopamine agonist impairs reinforcement learning in humans: evidence from event-related potentials and computational modeling of striatal-cortical function. *Human Brain Mapping*, *30*, 1963–76.

Satpute, A.B., Lieberman, M.D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research*, *1079*(1), 86–97.

Schneider, F., Weiss, U., Kessler, C., et al. (1999). Subcortical correlates of differential classical conditioning of aversive emotional reactions in social phobia. *Biological Psychiatry*, *45*, 863–71.

Schultz, W., Dayan, P., Montague, P.R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–9.

Singer, T., Seymour, B., O'Doherty, J.P., Stephan, K.E., Dolan, R.J., Frith, C.D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, *439*(7075), 466–9.

Smillie, L.D., Cooper, A.J., Pickering, A.D. (2011). Individual differences in reward–prediction–error: extraversion and feedback-related negativity. *Social Cognitive and Affective Neuroscience*, *6*, 646–52.

Todrank, J., Byrnes, D., Wrzesniewski, A., Rozin, P. (1995). Odors can change preferences for people in photographs: a cross-modal evaluative conditioning study with olfactory USs and visual CSs. *Learning and Motivation*, *26*, 116–40.

van der Lugt, A.H., Banfield, J.F., Osinsky, R., Münte, T.F. (2012). Brain potentials show rapid activation of implicit attitudes towards young and old people. *Brain Research*, *1429*, 98–105.

van der Veen, F.M., Sahibdin, P.P. (2011). Dissociation between medial frontal negativity and cardiac responses in the ultimatum game: effects of offer size and fairness. *Cognitive, Affective and Behavioral Neuroscience*, *11*(4), 516–25.

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), 564–8.

von Neumann, J., Morgenstern, O. (1953). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Walsh, M.W., Anderson, J.R. (2011). Learning from delayed feedback: neural responses in temporal credit assignment. *Cognitive Affective and Behavioral Neuroscience*, *11*(2), 131–43.

Walther, E., Nagengast, B., Trasselli, C. (2005). Evaluative conditioning in social psychology: facts and speculations. *Cognition and Emotion*, *19*(2), 175–96.

Wu, Y., Zhou, Y., van Dijk, E., Leliveld, M.C., Zhou, X. (2011). Social comparison affects brain responses to fairness in asset division: an ERP study with the ultimatum game. *Frontiers in Human Neuroscience*, *5*, 131.

Yeung, N., Sanfey, A.G. (2004). Independent coding of reward magnitude and valence in the human brain. *Journal of Neuroscience*, *24*(28), 6258–64.