frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 08 June 2012
doi: 10.3389/fnhum.2012.00165

Neural responses to advantageous and disadvantageous inequity

Klaus Fliessbach[1,2,3]*, Courtney B. Phillipps[1], Peter Trautner[2], Marieke Schnabel[4], Christian E. Elger[1,2,3], Armin Falk[3,4] and Bernd Weber[1,2,3]

[1] Department of Epileptology, University Hospital Bonn, Bonn, Germany
[2] Life and Brain Center, Department of NeuroImaging/NeuroCognition, University of Bonn, Bonn, Germany
[3] Center for Economics and Neuroscience, University of Bonn, Bonn, Germany
[4] Department of Economics, University of Bonn, Bonn, Germany

Edited by:
Chris Frith, University College London, UK

Reviewed by:
Rebecca Elliott, University of Manchester, UK
Lutz Jäncke, University of Zurich, Switzerland

***Correspondence:**
Klaus Fliessbach, Department of Epileptology, University of Bonn Medical Center, Sigmund Freud-Street 25, D-53105 Bonn, Germany.
e-mail: klaus.fliessbach@ukb.uni-bonn.de

In this paper we study neural responses to inequitable distributions of rewards despite equal performance. We specifically focus on differences between advantageous inequity (AI) and disadvantageous inequity (DI). AI and DI were realized in a hyperscanning functional magnetic resonance imaging (fMRI) experiment with pairs of subjects simultaneously performing a task in adjacent scanners and observing both subjects' rewards. Results showed (1) hypoactivation of the ventral striatum (VS) under DI but not under AI; (2) inequity induced activation of the right dorsolateral prefrontal cortex (DLPFC) that was stronger under DI than under AI; (3) correlations between subjective evaluations of AI evaluation and bilateral ventrolateral prefrontal and left insular activity. Our study provides neurophysiological evidence for different cognitive processes that occur when exposed to DI and AI, respectively. One possible interpretation is that any form of inequity represents a norm violation, but that important differences between AI and DI emerge from an asymmetric involvement of status concerns.

Keywords: equity norm, social preferences, functional magnetic resonance imaging (fMRI), ventral striatum

INTRODUCTION

It is a widely accepted principle of distributive justice that goods should be distributed to individuals according to their contribution, i.e., people should receive equal pay for equal work (equity principle) (Homans, 1961). There are numerous recent examples for the relevance and pursuit of this form of equity, such as resistance to pay cuts, efforts to abolish gender discrimination in salary, or the public debate about the appropriateness of extremely high wages for managers. Evidence for the behavioral importance of the equity principle comes from a large body of behavioral economics experiments (Fehr and Schmidt, 1999) and has been demonstrated even during early childhood in humans (Fehr et al., 2008). Behavioral effects of inequity manipulations have also been demonstrated in non-human species such as capuchin monkeys (Brosnan and De Waal, 2003) and dogs (Range et al., 2009).

From an individual perspective, the equity principle can be violated in two forms, to one's advantage or to one's disadvantage, respectively. Previous evidence suggests that reactions to inequity typically differ greatly, depending on whether it is advantageous or disadvantageous. In a questionnaire study by Loewenstein et al. (1989) it is shown, e.g., that most subjects strongly oppose disadvantageous inequity (DI) while reactions to advantageous inequity (AI) are relatively modest (see also Falk and Fischbacher, 2006). Moreover, it has been shown that evaluating AI requires more cognitive resources than does evaluating DI (van den Bos et al., 2006). Finally, while several studies on non-human primates have demonstrated rejections of DI, reports of animals rejecting AI (i.e., abandoning their own advantage) are scarce [for a review, see Brosnan (2009)]. These findings reveal a fundamental asymmetry between positive and negative violations of the equity principle, which cannot be explained solely in terms of inequity aversion: conceptually, inequity aversion implies an increase of dissatisfaction with increasing inequity, no matter whether this is to one's advantage or disadvantage. In light of the evidence it is, therefore, likely that other motives are involved in the evaluation of distributional inequity, in particular status concerns (Heffetz and Frank, 2008) and material self-interest. Consider, for example, the Ultimatum Game (UG) (Guth et al., 1982): In the UG, the first player (the proposer) suggests a division of a given amount of money to the second player (the responder). The responder then decides whether to accept or reject the proposal. In case of a rejection none of the players receives any money. Now consider an unequal proposal (of a pie of, say 10 monetary units, MU) so that the responder receives less than the proposer (say 2:8 MU). Compared to an equitable distribution (5:5 MU), such an offer simultaneously violates the equity principle (because it is an unequal distribution), status concerns of the responder (because getting less puts him in an inferior position) and material self-interest of the responder (because he receives less money compared to the equitable distribution). Like in the UG, in many experiments as well as outside the laboratory, DI simultaneously violates equity norms, status concerns and notions of self-interest. In contrast,

in case of AI the equity norm is in conflict with status-related interests and material self-interest: while AI violates the equity principle, it implies a higher status than in an equitable state. These theoretical considerations and previous empirical findings suggest that (1) DI elicits greater dissatisfaction than AI; (2) DI violates both the equity norm and status concerns/self-interest, whereas AI violates only the equity norm; and (3) given that in AI, equity-oriented norms and status concerns/self-interest are in conflict, AI places higher cognitive demands on evaluative processing than DI.

Recently, neuroscientific studies have begun to address neural processes underlying social and economic phenomena like e.g., reactions to norm violations (Hsu et al., 2008), status concerns (Zink et al., 2008), and reactions to unfair behavior (De Quervain et al., 2004). These studies have convergingly identified brain regions that are important for these aspects of social behavior. One consistent finding is that activations of the dopaminergic mesolimbic ("reward") system, especially the nucleus accumbens (NAcc) do not exclusively reflect material self-interest, but also social aspects. For example, NAcc activity is responsive to social rewards (Izuma et al., 2008, 2010), to status differences (Zink et al., 2008; Ly et al., 2011) and to outcomes of others (Fliessbach et al., 2007; Tricomi et al., 2010). More generally, it has been shown that NAcc activity is context dependent in many ways, i.e., it is influenced not only by the social context, but also by personal characteristics like own financial background (Tobler et al., 2007), by the previous reward history (Elliott et al., 2000; Akitsuki et al., 2003), or the by set of alternative outcomes (Breiter et al., 2001; Nieuwenhuis et al., 2005). Additionally, specific prefrontal brain regions have been shown to mediate responses in economic transactions. Specifically, the dorsolateral prefrontal cortex (DLPFC) seems to play a critical role in overriding material self-interest in favor of punishing unfair behavior in the UG (Knoch et al., 2006; Baumgartner et al., 2011). Additionally, ventrolateral prefrontal cortex (VLPFC) and the anterior insula have been implicated in emotion regulation as an important component of reactions to unfairness (Sanfey et al., 2003; Tabibnia et al., 2008).

In the present study, we applied functional brain imaging in order to investigate whether the assumptions outlined above are supported by neurophysiological data, i.e., whether and how the observed asymmetry between AI and DI is reflected by differential activation of brain regions that are essential for the processing of rewards and norm violations, specifically the NAcc, the DLPFC, VLMPFC, and the anterior insula. In a previous study on 32 male subjects, we have demonstrated that payment inequity principally affects brain activity in the ventral striatum (VS) (Fliessbach et al., 2007). To address the questions underlying the present study, i.e., to investigate differences between different types of inequity, we applied the same experimental procedure and obtained additional data from a large sample of female subjects (resulting in a total of 64 subjects). Additionally, we surveyed pleasantness ratings for the different payment conditions from our subjects using an 11-point Likert scale reaching from −5 to 5. This allowed us to correlate brain activity with individual evaluations of different distributions.

Based on the three outlined assumptions, we hypothesized that:

(i) Activity in the VS is reduced more in DI than in AI, reflecting a higher level of dissatisfaction resulting from DI than from AI.
(ii) Regions that process norm violations, as well as status concerns and self-interest violations (DLPFC, anterior insula) are differentially affected by DI and AI. While both types of inequity should activate these areas because they involve a norm violation, the additional violation of status concerns in DI should lead to an enhanced increase in activation of these areas.
(iii) There should be a dissociation of areas associated with the evaluation of DI and AI (i.e., displaying correlations between subjects' ratings and BOLD signal strength). The higher cognitive demands placed by the simultaneous weighing of equity and status concerns in AI should require higher order cortical processing. In contrast, the evaluation of DI should predominantly rely on subcortical structures involved in reward and emotion processing.

METHODS

SUBJECTS

Eighteen pairs of male subjects and 18 pairs of female subjects participated in the experiment. All subjects were native German-speakers without any history of neurological or psychiatric disease (one subject was subsequently excluded because of a previously unknown history of schizophrenic psychosis). The study was approved by the Ethics Committee of the University of Bonn and all subjects gave written informed consent. Eight subjects were excluded from the analysis for various technical reasons (e.g., excessive head movement, scanner dysfunction), so that the final analysis included data from 32 female individuals (mean age 25.8, SD 3.9) and 32 male individuals (mean age 29.2, SD 4.9) (30 of which were scanned on the 1.5 Tesla (T) scanner and 34 on the 3T scanner). All analyses included covariates for the between subject parameters "scanner type" and "gender" as potential nuisance factors, which did not show any significant interaction with the reported constrasts.

EXPERIMENTAL PROCEDURE

Two subjects were simultaneously placed in two MR scanners situated at opposite sides of the same control room at the research center. The two subjects saw each other when being led to the scanners, but they did not have the opportunity to talk to each other or to become acquainted before the experiment began. The task was presented via video goggles (Nordic NeuroLab, Bergen, Norway) using Presentation© software (NeuroBehavioural Systems, Inc.). During scanning, both subjects performed 300 trials of the following task (**Figure 1**): they saw a screen with a varying number (4–55) of blue dots for 1.5 s. The time of the appearance of this screen defined the task onset. Immediately, thereafter, a number was presented that differed by 20 percent from the number of dots previously shown. Subjects had to decide whether the number of dots

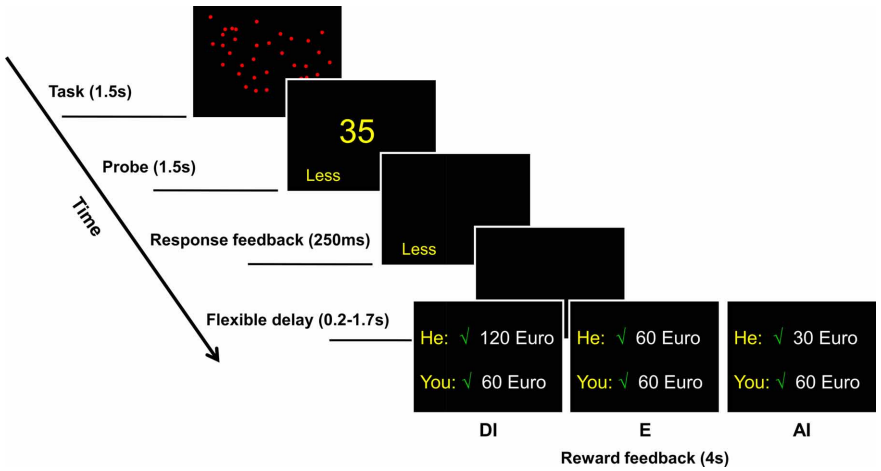


FIGURE 1 | Single-trial settings. Subjects saw a number of blue dots for 1500 ms (screen 1). Immediately afterwards, a number was presented and subjects had to decide by pressing a button whether the number of dots on the first screen was less than or greater than this number within a time limit of 1500 ms (screen 2). After a response feedback (250 ms, screen 3) and a short delay (blank screen 4), a feedback screen informed subjects about their own and the other subject's performance (correct or incorrect) together with the respective monetary rewards (screen 5). Here, three alternative outcomes representing the main conditions for this study are depicted.

shown first was greater or less than the number presented second. They indicated their answers with the help of response grips (NordicNeuroLab, Bergen) within a time limit of 1.5 s. Later responses were counted as incorrect. A response terminated the screen and the selected option was highlighted for 250 ms as response feedback. The timing parameters were derived from pretests that showed that on average about 80% of trials were solved correctly at this level of difficulty, resulting in a sufficient number of events for each experimental condition. The presentation stopped when both subjects had responded, introducing a variable delay of at most 1500 ms for the subject who responded faster. Once both responses were available, and following an approximately 200 ms delay for the exchange of response information between the two presentation computers, a feedback screen was displayed for 4 s. This screen revealed to both players whether they were correct (indicated by a green check-mark) or not (indicated by a red X), as well as the amount of money they earned for that trial. The next trial started following a jittered time interval of 4.5–7 s.

Payoff conditions were as follows: when both subjects were incorrect, both received nothing. When only one subject was correct, this subject received either an amount of approximately €30 (low-level) or approximately €60 (high-level) while the other subject received nothing. When both subjects were correct, one of six possible payoff conditions was randomly selected, generated by a 2 × 3 factorial design that varied the absolute amount of money (factor 1) and the amount relative to the other subject (factor 2) (see **Table 1**). In order to reduce boredom that could result from repeatedly seeing the same monetary figures, we varied the reward amount in each condition within a 10 percent interval from the mean (i.e., for the €30 trial, the amounts varied from €27 to €33). At the end of the experiment, one trial was randomly selected and paid out according to the respective outcomes in that trial. On average, subjects received an additional €45 resulting from the experiment, together with a show-up fee of €15, i.e., payoffs from the experiment were relatively large compared to the show-up fee. The purpose of this was to ensure a relatively high salience of the reward events during the experiment.

Table 1 | Payoff conditions.

Task performance (A/B)	Relative reward level (A:B)	Absolute reward level	Payoffs (€) (A–B)	Condition
–/–			0–0	C1
+/–		High	60–0	C2 ("win alone")
		Low	30–0	C3
–/+		High	0–60	C4 ("no win")
		Low	0–30	C5
+/+	1: 2	High	60–120	C6 (DI)
		Low	30–60	C7
	1: 1	High	60–60	C8 (E)
		Low	30–30	C9
	2: 1	High	120–60	C10
		Low	60–30	C11 (AI)

"–" = incorrect task performance, "+" = correct task performance. Conditions of interest are highlighted.

SCANNING PROCEDURE

Scanning was conducted using a 1.5T Avanto Scanner and a 3T Trio Scanner (Siemens, Erlangen, Germany) using standard eight channel head coils. Slices were in axial orientation and covered all of the brain including the midbrain but not the entire cerebellum. Scan parameters for the 1.5T scanner were as follows: Slice thickness: 3 mm; interslice gap 0.3 mm; matrix size: 64 × 64; field of view: 192 × 192 mm; echo time (TE): 50 ms; repetition time

(TR): 2.91 s. Scan parameters for the 3T scanner were as follows: Slice thickness: 2 mm; interslice gap 1 mm; matrix size: 128 × 128; field of view: 230 × 230 mm; TE: 33 ms; TR: 2.5 s.

fMRI DATA ANALYSIS

Functional magnetic resonance imaging (fMRI) data analysis was performed using Statistical Parametric Mapping 8 (SPM8, www.fil.ion.ucl.ac.uk/spm/). For preprocessing, the functional images were realigned to the first image of each time series and again realigned to the mean image after the first realignment. Images were then normalized to the canonical EPI template used in SPM8 and smoothed with an 8 mm Gaussian kernel. After normalization, images were re-sampled to a voxel size of 3 × 3 × 3 mm for both scanners, allowing for a combined analysis of data from both scanners.

For modeling the BOLD response, 11 types of events were defined according to the payoff conditions C1–C11. The onset times (defined by the appearance of the feedback screen informing the subjects about the outcome) was convolved with the canonical hemodynamic response function (HRF) used in SPM8 and its temporal derivative. Additionally, a regressor for the onset times of the task was included in the model as well as movement parameters derived from the motion correction procedure. Parameter images for the contrasts for each single condition were generated for each subject and were then subjected to a second-level random effects analysis. We (1) investigated the main effects of inequity [F-contrast for conditions C6, C8, C11, and differential T-contrasts C6 > C8 (DI > E) and C11 > C8 (AI > E)] on brain activity; and (2) tested for correlations of the self-reported inequity aversion measures and the respective BOLD contrasts. According to our hypotheses these were conducted for the VS and for the whole brain.

REGION OF INTEREST DEFINITIONS

Based on a priori considerations, we were specifically interested in the VS. A region of interest in the VS was defined functionally by contrasting the conditions in which one subject received a reward and the other did not (C2, C3) with the conditions in which a subject did not receive any reward at all (C1, C4, C5) on a relative conservative statistical threshold (Voxelwise FWE-whole brain corrected $P < 0.05$). This resulted in a bilateral ventral striatal ROI with peak voxels at $X = 18$, $Y = 11$, $Z = -8$ (number of voxels: 120) and $X = -9$, $Y = 8$, $Z = -5$ ($n = 144$), respectively. We assumed that this ROI definition would ensure that we would consider all striatal areas that show (under our study and pre-processing conditions) a clear sensitivity to rewards. Alternatively, we applied an anatomical mask for the NAcc from the Harvard-Oxford cortical and subcortical structural atlas (http://www.cma.mgh.harvard.edu) applying a probability of 0.5. For these two regions of interest (shown in **Figure 2**), parameter estimates were extracted and averaged over all voxels in the entire ROI, allowing for statistics based on conventional statistical thresholds without the need to correct for multiple comparisons (except the number of conducted tests).

THRESHOLDING

For the whole-brain analyses of the main effects of inequity and the correlational analysis we used a cluster corrected $P_{FWE} < 0.005$ (in order to correct for the number of contrasts), after an inclusion threshold of $P < 0.001$ unc). Here, the diagrams depicting mean parameter estimates (**Figure 7**) or scatter plots (**Figure 8**) were derived from the peak voxels (plus a surrounding 5 mm sphere) of the so identified clusters. Note that this serves demonstration purposes only and that no statistical inferences rely on these analyses. Note further, that the *post-hoc* calculation of correlation coefficients for the so identified regions bears the danger of overestimation, because the regions revealing highest effects are selectively analyzed.

PLEASANTNESS RATING

Three to six months following the fMRI scanning session, the same subjects were asked to rate the pleasantness of each experimental payoff condition on an 11-point Likert scale ("On a scale from from −5 (this bothers me very much) to +5 (this makes me very happy) how would you evaluate these events?"). These pleasantness ratings allowed us to define two measures of inequity aversion analogously to the BOLD contrasts, i.e., (E-DI) as a measure for the aversion to DI and (E-AI) as a measure for AI, respectively. Pleasantness ratings from two male subjects were not obtainable.

Our experiment offers an ideal setting to test for the neural consequences of violations of the equity principle for several reasons: first, it allows us to disentangle effects of equity norm violations from status concerns and self-interest violations. The realization of DI and AI within person allows us to assess the effect of equity norm violation with or without simultaneous status concern violations (DI vs. AI). It also allows ruling out the influence of material self-interest, because each of the conditions, DI, AI, and equity (E), were realized in the same subject, keeping the subject's own *absolute* income constant. Second, allocations

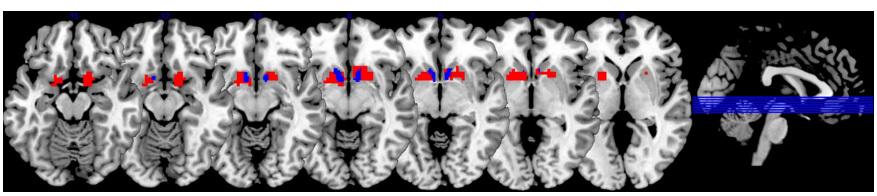


FIGURE 2 | ROI Masks. In red: Voxels within the striatum which show a reward-related signal (derived from the contrast C2, C3 > C1, C4, and C5) on a $P_{FWE} < 0.05$ (per voxel), whole brain corrected. In blue: anatomically defined NAcc mask (according to Harvard-Oxford brain atlas).

were not implemented by another person [such as in the UG (Guth et al., 1982)] but were randomly assigned by a computer. Therefore, reactions to inequity were not confounded by the perceived fairness or fairness intention of the other person. Third, subjects could not take action to reduce inequity (again unlike in the UG or other fairness experiments). This means that brain activity did not reflect experience or expectation of behavioral reactions to the observed inequity.

RESULTS

PLEASANTNESS RATINGS (FIGURE 3)

Our first result uses data from the post-experimental questionnaire where subjects had to rate the pleasantness of different allocations. On average, subjects strongly preferred E over DI (mean ratings ± SEM: 4.0 ± 1.2 for E vs. 1.3 ± 3.0 for DI, t_{63} = 7.1, $p < 10^{-8}$) demonstrating a strong and systematic aversion toward DI. Not a single subject preferred DI over E (i.e., no subject preferred a better outcome for the other person for a given own absolute income level, which could be interpreted as an altruistic preference). On average, there was also a preference for E over AI (mean ratings ± SEM: 4.0 ± 1.2 for E vs. 3.5 ± 1.7 for AI, $t_{63} = 2.3$, $p = 0.024$), but here, differences were much smaller and less consistent between subjects, i.e., 16 subjects preferred E over AI, nine preferred AI over E, and the rest of subjects was indifferent in this respect. These findings closely match previous results (Loewenstein et al., 1989; Falk and Fischbacher, 2006).

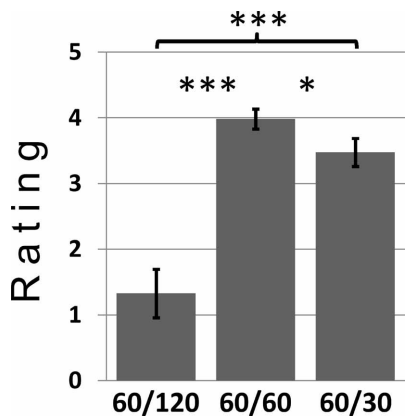


FIGURE 3 | Mean pleasantness ratings for the conditions DI (own income/other's income: 60/120), E (60/60), and AI (60/30). Error bars indicate the standard error of the mean (SEM). *$p < 0.05$, ***$p < 0.001$ (dependent samples *t*-tests).

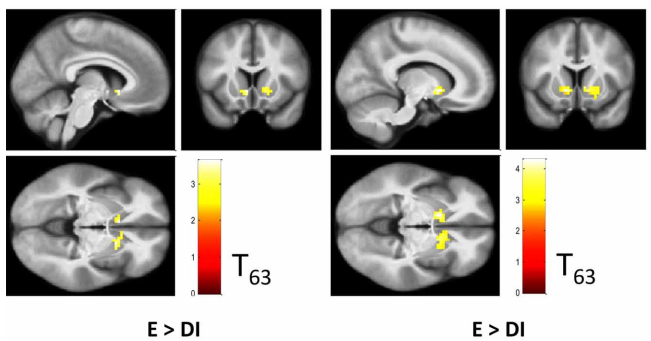
HYPOTHESIS 1: EFFECTS OF INEQUITY ON VS ACTIVITY (FIGURES 4, 5, 6)

There was a significant main effect of inequity in the VS ROI [within-subject ANOVA: ($F_{(2,\ 59)} = 8.26$, $P < 0.001$)]. Pairwise comparisons showed that the DI condition was associated with a significantly weaker activity compared to both conditions E ($t_{63} = 2.76$, $p = 0.007$) and AI ($t_{63} = 3.56$, $p < 0.001$). Interestingly, activity in the VS was *higher* for AI than for E, albeit insignificantly. Similar results are obtained when applying an anatomically defined ROI mask (**Figure 5**). We also tested for associations between BOLD signal changes in the VS and individual pleasantness ratings and observed a significant but relatively weak relationship (**Figure 6**).

HYPOTHESIS 2: EFFECTS OF INEQUITY IN OTHER BRAIN REGIONS (FIGURE 7)

Outside the VS, a significant effect of DI (contrast DI > E) of reward was observed in the right DLPFC (**Figure 6**). *Post-hoc* paired *t*-tests for parameter estimates derived from the peak voxel of this activation shows increased activation in this area also for AI, but this activity was significantly lower than for DI.

On the statistical threshold used for the whole brain analyses there was no significant effect observed for AI > E.



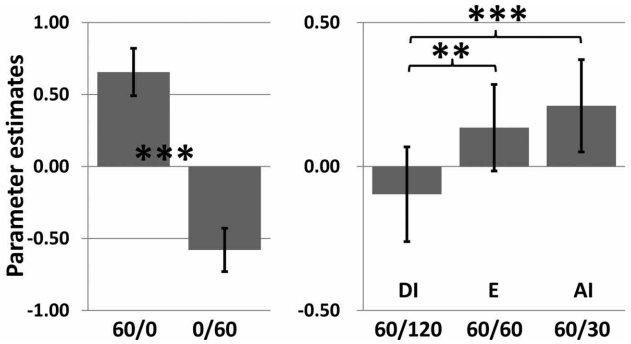


FIGURE 4 | Results for the ventral striatum. Left: Brain images showing significantly ($P < 0.005$ for demonstration purposes) higher activation for E than DI (peakvoxel MNI-coordinates: $X = -6$, $Y = 14$, $Z = -5$), and for AI than for DI ($X = -12$, $Y = 11$, $Z = -5$), within the functionally defined ROI. **Right:** The barplot shows mean parameter estimates for the different conditions averaged *across all voxels* of the functionally defined ROI. The left side demonstrates the strong responsiveness of the area to rewards *per se* (which was the selection criterion for the ROI, implying circularity of this result). The right side shows a significant main effect of relative payoff on activation in this area [within-subject ANOVA: ($F_{(2,\ 59)} = 8.26$, $P < 0.001$)] with stronger acvtivation in the equity ($t_{63} = 2.76$, $P = 0.007$) and advantageous inequity ($t = 3.56$, $P < 0.001$) condition than for the disadvantageous inequity condition. Note that these contrasts are independent of the ROI defining contrasts. Error bars indicate standard error of means and are not informative with regard to (within-subjects) statistical inference.

HYPOTHESIS 3: CORRELATIONS BETWEEN THE BOLD SIGNAL AND PLEASANTNESS RATINGS OUTSIDE THE VS (FIGURE 8)

Outside the VS, significant correlations between pleasantness ratings and BOLD contrasts were observed only for the AI-E comparison and that bilaterally in the VLPFC and in the left insula (**Figure 5**).

For the DI-E comparison no significant correlations were found at the statistical threshold defined for the whole brain analysis.

In all cases correlations were positive, i.e., greater activation was associated with higher pleasantness ratings. There were no significant negative correlations between BOLD activations and pleasantness ratings.

GENDER DIFFERENCES

The equally sized groups of male and female subjects in our study sample provide a good basis to analyze gender effects of inequity processing. However, neither the behavioral results (ratings of inequity conditions) nor the described neuroimaging findings showed any significant interaction between inequity conditions and gender.

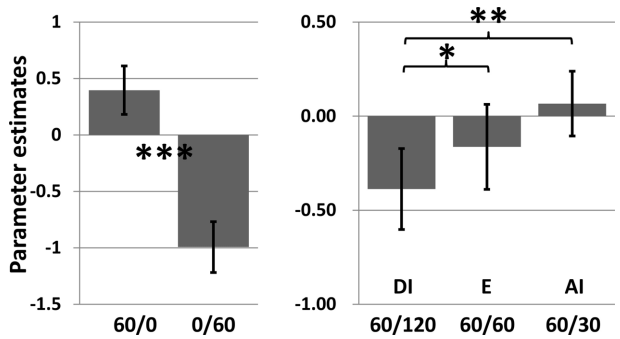


FIGURE 5 | Results for the anatomically defined ROI: Mean parameter estimates for the different conditions averaged *across all voxels* of the ROI (cf. Figure 3). The left side demonstrates the strong responsiveness of the area to rewards *per se*. The right side shows a significant main effect of relative payoff on activation in this area [within-subject ANOVA: ($F_{(2,\ 59)} = 6.556$, $P = 0.008$)] with stronger activation in the E ($t_{63} = 2.32$, $P = 0.023$) and advantageous inequity ($t = 2.92$, $P < 0.001$) condition than for the disadvantageous inequity condition. Error bars indicate standard error of means and are not informative with regard to (within-subjects) statistical inference.

DISCUSSION

The present study investigated the neural consequences of violations of the equity principle. Specifically, we tested for differences between responses to DI and AI. Our results show that:

(i) VS activity is reduced for conditions of DI but not for AI.
(ii) Disadvantageous (and to a lower extend also advantageous) inequity increases activation in the right DLPFC.
(iii) The evaluation of AI is related to ventrolateral prefrontal and insular regions.

ad (i): On the basis of converging evidence that VS activity increases with increasing expected value of events [for a summary, see Knutson et al. (2009)] this finding is consistent with the assumption that relative to E, DI causes dissatisfaction which is reflected by lower VS activity. Furthermore, there was no indication of a lower level of satisfaction with AI in the VS. VS activity was actually slightly higher in the AI condition than in the E condition, despite the significantly lower pleasantness ratings for AI than for E. Notably, subjects did not explicitly rate the pleasantness of the different outcomes while in the scanner. Although the correlations of brain activity with the ratings (obtained later) suggest that implicit evaluation processes took place during scanning, it seems likely that subjects did not extensively reflect and evaluate the outcomes at that time, in part given the limited time available to do so. Therefore, the discrepancy between the VS activity during scanning and pleasantness ratings, which were acquired outside the scanner without any time limitations, might reflect the fact that longer periods of reflection lead to more negative assessments of AI, a finding that is in line with results from van den Bos et al., 2006. In addition one may speculate that the questionnaire ratings reflect an element of social desirability

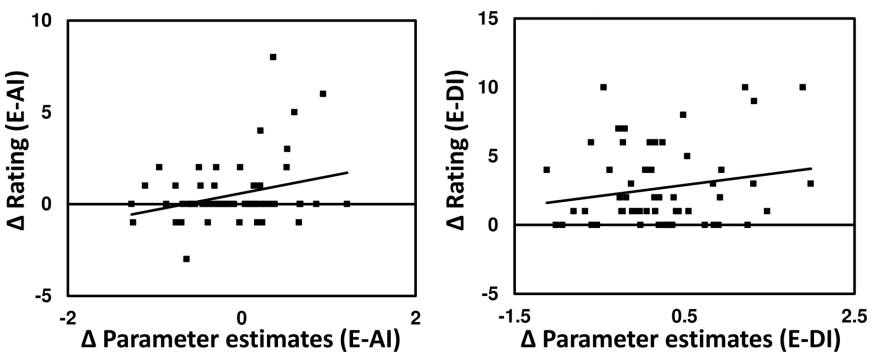


FIGURE 6 | Association of BOLD responses to ratings in the VS (averaged across all voxels of the functionally defined ROI) and subjective pleasentness ratings demonstrate a significant positive association between pleasantness and signal for AI evaluation ($r = 0.29$, $P = 0.01$, one-sided) and a non-significant trend for DI evaluation ($R = 0.18$, $P = 0.08$, one-sided).

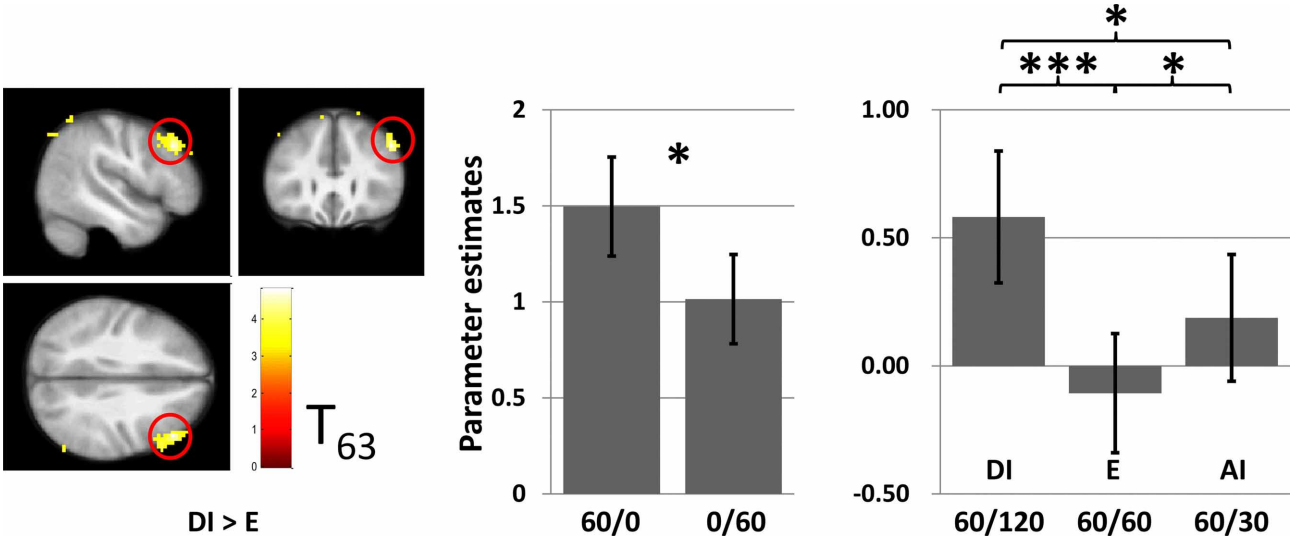


FIGURE 7 | Whole-brain analysis of the contrast DI > E. Left: Significant activation cluster (cluster corrected $P_{FWE} < 0.005$, inclusion threshold $P < 0.001$ unc.) in the rDLPFC (Peak voxel MNI coordinates $X = 48$, $Y = 29$, $Z = 37$). The majority of the voxels lie in Brodman Area 9. **Right:** The bar plot depicts effects of the different conditions at the peak voxel (plus surrounding 5 mm). Note that this diagram only serves demonstration reasons. Error bars indicate standard error of means and are not informative with regard to (within-subjects) statistical inference.

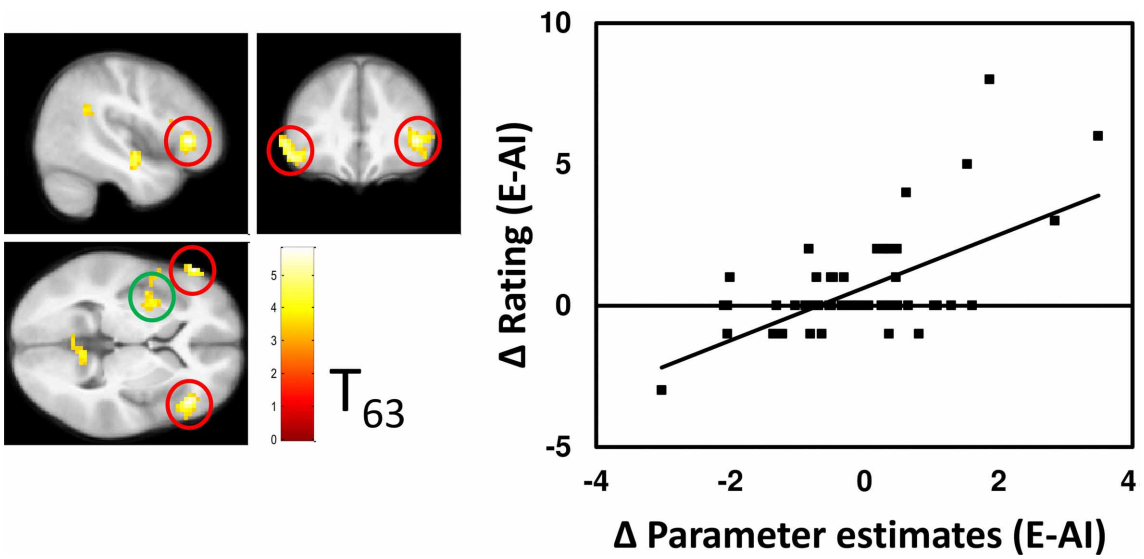


FIGURE 8 | Whole-brain correlational analysis of the contrast E-AI with the respective difference in the pleasantness rating. Significant clusters (cluster corrected $P_{FWE} < 0.005$, inclusion threshold $P < 0.001$ unc.) showing this relation lie bilaterally in the VLPFC (Peak voxel MNI coordinates: $X = 45$, $Y = 32$, $Z = 1$, and $X = -57$, $Y = 35$, $Z = 1$) and in the left insula ($X = -33$, $Y = 2$, $Z = -2$). Right: The scatterplot depicts the relation between contrast and ratings for the peak voxels averaged across both sides of the VLMPFC clusters, and serves demonstration reasons only.

or normative pressure: when interviewed subjects may feel they "should" dislike AI when in fact they don't. In this sense our finding provides an interesting case where valuations from BOLD signals lead to different and perhaps more reliable conclusions than valuations derived from interviews. A similar discrepancy between ventral striatal responses and behavioral data concerning distributional inequity was recently reported by Tricomi et al. (2010), although in this case subjects had *reduced* ventral striatal activations in a self-AI condition despite more *favorable* ratings, i.e., results were seemingly opposite to ours. The discrepancy between the two findings can probably explained by differences in the experimental design. In our study the two subjects were principally in the same situation when they were faced with the unequal distributions. In contrast, Tricomi et al. applied

a strong inequity manipulation *prior* to the evaluated events. Thus, in Tricomi et al.'s study the principal status of the subjects was defined before scanning and the monetary transfers during scanning did not compromise the superiority of the high-pay subject. We assume that under these circumstances the superior subjects are more likely to pay attention to equity concerns explaining more negative responses also to AI. This means that the asymmetry that we assume to underlie DI and AI processing can be reversed by the induction of a stronger prior asymmetry between the subjects, comparable with a shift in the reference point. Future studies should address the highly interesting relation between status and inequity aversion by manipulating status in inequity experiments independently from monetary distributions.

Our study provides one example of relative reward processing in the human VS, i.e., it demonstrates that responses to a given reward size critically depend on contextual, in this case social, factors. It is important to note that relative reward processing in the VS occurs in many ways, e.g., the response to a given reward depends on the set of possible alternatives (Breiter et al., 2001; Nieuwenhuis et al., 2005) or on reward history (Elliott et al., 2000; Akitsuki et al., 2003). Therefore, it will be an interesting challenge for future studies to investigate even more specific effects of social comparison on reward processing, by e.g., adressing social comparison with regard to performance measures instead of monetary rewards. As another example of relative reward processing, it has been demonstrated that processing of monetary rewards depends on the financial status of the subjects (Tobler et al., 2007). We did not explicitly control for this factor, but our subjects group is supposedly relative homogenous (and not representative) in this regard (stemming from a typical student population) so that we assume that this factor does not introduce significant noise. However, relating inequity processing to (socioeconomic) status provides another promising goal for future research.

ad (ii): Generally, the DLPFC is assumed to play a role in goal maintenance and cognitive control (Mansouri et al., 2009). In the present study, activation in the DLPFC of was attributable to the experience of inequity. Previous studies on ultimatum bargaining suggest that these activations can be interpreted in terms of registering both norm and status concern/self-interest violations (Sanfey et al., 2003; Knoch et al., 2006; Singer et al., 2006). A well-known imaging study revealed activation of the MPFC, right DLPFC, and anterior insula when responders were confronted with unfair offers in the UG (Sanfey et al., 2003). A recent study combined fMRI with TMS and suggested that the right DLPFC is specifically and causally involved (together with the ventromedial prefrontal cortex) in the rejection of unfair offers in the UG (Baumgartner et al., 2011). Therefore, to clarify the role of the DLPFC in such complex social behaviors appears highly interesting.

In the UG, being confronted with an unfair offer implies several important aspects. First, as outlined in the introduction, it violates equity norms as well as self-interest and status concerns. Second, because the unfair allocation has been intentionally proposed by another person, it is likely to induce negative feelings toward the proposer. Third, because the responder must decide whether to accept or reject the offer, it involves active decision making, e.g., in the form of negative reciprocity. Different to the UG, our experimental design rules out the second and third aspect; it controls for the self-interest aspect by keeping own income constant; and it allows to test the effect of equity norm violation with or without violations of status-related interests (DI vs. AI). Our results showed strong responses to DI in the DLPFC. Additionally, DLPFC was also activated by AI but significantly less than by DI. It did not show activations when the rewards were equally distributed. This finding is consistent with the assumption that both forms of inequity represent some kind of norm violation, which is registered in the DLPFC. Further, this result is consistent with the conjunction that the additional violation of status motives leads to a further increase of activity in case of DI.

ad (iii): Our results demonstrate the importance of ventromedial prefrontal areas along with the insular cortex *in the evaluation* of AI. Generally, these areas have implicated in emotion processing (insula) (Nitschke et al., 2006) and with cognitive regulation of emotions (VLPFC) (Wager et al., 2008). Both regions have been shown to be involved in the processing of unfair offers in the UG (Tabibnia et al., 2008).

In both areas, greater activity correlated with subjects' satisfaction with the corresponding outcome. In other words, people who preferred E over AI (according to the post-experimental survey) showed greater activity in the VLPFC and insula during E trials than during AI trials, and those subjects who stated a preference for AI over E showed greater activity here during AI trials than during E trials. The data, therefore, do not support the assumption that a conflict between fairness-based cognitive processes and a situation of AI leads to a modification of an immediate positive evaluation of such an event. If that was the case, one would expect to find a negative correlation between the level of brain activity for AI trials and evaluation of AI trials. An alternative assumption could be that VLPFC and insula activation during E trials reflects a positive cognitive appraisal of these trials in subjects who show a strong preference for E. This interpretation would be in line with previous reports that the VLPFC plays an important role in evaluating norm compliance (Spitzer et al., 2007). It would be interesting to complement our findings with methods that allow causal inferences such as transcranial magnetic stimulation. In light of our findings we would expect that disturbance of right VLPFC function should alter the evaluation of one's own social advantages more than the evaluation of one's disadvantages.

In conclusion, our study provides neurophysiological evidence for the existence of different cognitive processes involved in the confrontation with DI and AI. Our data are consistent with the idea that any form of inequity represents a norm violation, but that differences between DI and AI emerge from the additional involvement of status-related motives.

ACKNOWLEDGMENTS

Klaus Fliessbach is funded by the German Research Council (Grant FL 715/1-1). Bernd Weber is funded by the German Research Council with a Heisenberg Grant (Grant WE 4427/3-1). We thank Florian Mormann for helpful comments on the manuscript.

REFERENCES

Akitsuki, Y., Sugiura, M., Watanabe, J., Yamashita, K., Sassa, Y., Awata, S., Matsuoka, H., Maeda, Y., Matsue, Y., Fukuda, H., and Kawashima, R. (2003). Context-dependent cortical activation in response to financial reward and penalty: an event-related fMRI study. *Neuroimage* 19, 1674–1685.

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., and Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nat. Neurosci.* 14, 1468–1474.

Breiter, H. C., Aharon, I., Kahneman, D., Dale, A., and Shizgal, P. (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron* 30, 619–639.

Brosnan, S. F. (2009). "Responses to inequity in non-human primates," in *Neuroeconomics. Decision Making and the Brain,* eds P. W. Glimcher, C. Camerer, E. Fehr, and R. A. Poldrack (London, San Diego, Burlington: Elsevier), 285–302.

Brosnan, S. F., and De Waal, F. B. (2003). Monkeys reject unequal pay. *Nature* 425, 297–299.

De Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., and Fehr, E. (2004). The neural basis of altruistic punishment. *Science* 305, 1254–1258.

Elliott, R., Friston, K. J., and Dolan, R. J. (2000). Dissociable neural responses in human reward systems. *J. Neurosci.* 20, 6159–6165.

Falk, A., and Fischbacher, U. (2006). A theory of reciprocity. *Games Econ. Behav.* 54, 293–315.

Fehr, E., Bernhard, H., and Rockenbach, B. (2008). Egalitarianism in young children. *Nature* 454, 1079–1083.

Fehr, E., and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Q. J. Econ.* 114, 817–868.

Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C. E., and Falk, A. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. *Science* 318, 1305–1308.

Guth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental-analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* 3, 367–388.

Heffetz, O., and Frank, R. H. (2008). "Preferences for status: evidence and economic implications," in *Handbook of Social Economics,* eds J. Benhabib, A. Bisin, and M. Jackson (Amsterdam: Elsevier), 69–91.

Homans, G. C. (1961). *Social Behavior: Its Elementary Forms.* New York, NY: Harcourt, Brace.

Hsu, M., Anen, C., and Quartz, S. R. (2008). The right and the good: distributive justice and neural encoding of equity and efficiency. *Science* 320, 1092–1095.

Izuma, K., Saito, D. N., and Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron* 58, 284–294.

Izuma, K., Saito, D. N., and Sadato, N. (2010). Processing of the incentive for social approval in the ventral striatum during charitable donation. *J. Cogn. Neurosci.* 22, 621–631.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., and Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832.

Knutson, B., Delgado, M. R., and Phillips, P. E. M. (2009). "Representation of subjective value in the striatum," in *Neuroeconomics. Decision Making and Brain,* eds P. W. Glimcher, C. Camerer, E. Fehr, and R. A. Poldrack (London San Diego, Burlington: Academic Press), 389–406.

Loewenstein, G. F., Bazerman, M. H., and Thompson, L. (1989). Social utility and decision-making in interpersonal contexts. *J. Pers. Soc. Psychol.* 57, 426–441.

Ly, M., Haynes, M. R., Barter, J. W., Weinberger, D. R., and Zink, C. F. (2011). Subjective socioeconomic status predicts human ventral striatal responses to social status information. *Curr. Biol.* 21, 794–797.

Mansouri, F. A., Tanaka, K., and Buckley, M. J. (2009). Conflict-induced behavioural adjustment: a clue to the executive functions of the prefrontal cortex. *Nat. Rev. Neurosci.* 10, 141–152.

Nieuwenhuis, S., Heslenfeld, D. J., Von Geusau, N. J., Mars, R. B., Holroyd, C. B., and Yeung, N. (2005). Activity in human reward-sensitive brain areas is strongly context dependent. *Neuroimage* 25, 1302–1309.

Nitschke, J. B., Sarinopoulos, I., Mackiewicz, K. L., Schaefer, H. S., and Davidson, R. J. (2006). Functional neuroanatomy of aversion and its anticipation. *Neuroimage* 29, 106–116.

Range, F., Horn, L., Viranyi, Z., and Huber, L. (2009). The absence of reward induces inequity aversion in dogs. *Proc. Natl. Acad. Sci. U.S.A.* 106, 340–345.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science* 300, 1755–1758.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., and Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature* 439, 466–469.

Spitzer, M., Fischbacher, U., Herrnberger, B., Gron, G., and Fehr, E. (2007). The neural signature of social norm compliance. *Neuron* 56, 185–196.

Tabibnia, G., Satpute, A. B., and Lieberman, M. D. (2008). The sunny side of fairness: preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychol. Sci.* 19, 339–347.

Tobler, P. N., Fletcher, P. C., Bullmore, E. T., and Schultz, W. (2007). Learning-related human brain activations reflecting individual finances. *Neuron* 54, 167–175.

Tricomi, E., Rangel, A., Camerer, C. F., and O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature* 463, 1089–1091.

van den Bos, K., Peters, S. L., Bobocel, D. R., and Ybema, J. F. (2006). On preferences and doing the right thing: satisfaction with advantageous inequity when cognitive processing is limited. *J. Exp. Soc. Psychol.* 42, 273–289.

Wager, T. D., Davidson, M. L., Hughes, B. L., Lindquist, M. A., and Ochsner, K. N. (2008). Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron* 59, 1037–1050.

Zink, C. F., Tong, Y., Chen, Q., Bassett, D. S., Stein, J. L., and Meyer-Lindenberg, A. (2008). Know your place: neural processing of social hierarchy in humans. *Neuron* 58, 273–283.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 26 December 2011; accepted: 22 May 2012; published online: 08 June 2012.

Citation: Fliessbach K, Phillipps CB, Trautner P, Schnabel M, Elger CE, Falk A and Weber B (2012) Neural responses to advantageous and disadvantageous inequity. Front. Hum. Neurosci. ***6****:165. doi: 10.3389/fnhum.2012.00165*

Copyright © 2012 Fliessbach, Phillipps, Trautner, Schnabel, Elger, Falk and Weber. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.