doi:10.1093/scan/nsr088

A reward prediction error for charitable donations reveals outcome orientation of donators

Katarina Kuss,[1,2] Armin Falk,[2] Peter Trautner,[3] Christian E. Elger,[1,2,3] Bernd Weber,[1,2,3] and Klaus Fliessbach[1,2,3]

[1]Department of Epileptology, University Hospital Bonn, Sigmund-Freud-Str. 25, [2]Center for Economics and Neuroscience, University of Bonn, Nachtigallenweg 86 and [3]Life & Brain Center, Department of NeuroCognition, University of Bonn, Sigmund-Freud-Str. 25, 53127 Bonn, Germany

The motives underlying prosocial behavior, like charitable donations, can be related either to actions or to outcomes. To address the neural basis of outcome orientation in charitable giving, we asked 33 subjects to make choices affecting their own payoffs and payoffs to a charity organization, while being scanned by functional magnetic resonance imaging (fMRI). We experimentally induced a reward prediction error (RPE) by subsequently discarding some of the chosen outcomes. Co-localized to a nucleus accumbens BOLD signal corresponding to the RPE for the subject's own payoff, we observed an equivalent RPE signal for the charity's payoff in those subjects who were willing to donate. This unique demonstration of a neuronal RPE signal for outcomes exclusively affecting unrelated others indicates common brain processes during outcome evaluation for selfish, individual and nonselfish, social rewards and strongly suggests the effectiveness of outcome-oriented motives in charitable giving.

Keywords: charitable donations; reward processing; social preferences; functional magnetic resonance imaging (fMRI); nucleus accumbens (NAc)

INTRODUCTION

There is compelling evidence that humans do not exclusively follow rational, self-interested motives in economic decision making (Camerer and Fehr, 2006). Perhaps the most striking exception to materialistic, self-interested behavior is giving one's own goods to unrelated others, as is in charitable donations.

Theoretical underpinnings of donation behavior

Economic theory suggests that different motives underlie donation decisions (Harbaugh, 1998). One possible motive is that a person has in fact a preference for the public good provided by donations. In this case, the money belonging to a charity organization carries a utility (or reward value) that is independent of the person's own contribution. Such a preference for a public good can be regarded as 'altruistic' and if this is the sole motivation for a donation, this behavior has been termed 'pure altruism' (Andreoni, 1989). Notably, pure altruism does not imply a noninterest in own belongings. According to Andreoni, pure altruism means that 'preferences depend only on private consumption and the total supply of the public good . . .' (Andreoni, 1990). Because the preference for the public good is directed toward the outcome of a donation, we refer to this motive as 'outcome orientation'.

Apart from this, it has been argued that other motives, such as guilt avoidance, reputation gain or a feeling of 'warm glow', can be associated with *the act of giving itself* ('action orientation'; Andreoni, 1989, 1990). Action- and outcome-oriented motives do not contradict each other. Rather, they are supposed to complement each other, empirical support for the effectiveness of both motives in charitable donations comes from behavioral studies (Andreoni, 1989, 1990; Konow, 2010). Andreoni refers to this as 'impure altruism'.

Neuroimaging findings in donation behavior

Recently, functional neuroimaging has been used to investigate brain processes underlying donation behavior. These studies have shown that donation decisions are associated with activations in the dopaminergic reward system (Moll *et al.*, 2006; Harbaugh *et al.*, 2007), providing support for action-associated positive feelings in the sense of a warm glow. At the same time, indirect support for outcome orientation has been found by showing increased reward-related brain activity during nonvoluntary transfers of money to a charity that can be used to predict subjects' donation behavior (Harbaugh *et al.*, 2007).

Reward prediction error induction as a test of outcome orientation

Our study aims at extending these findings by directly probing outcome-related reward activity in the context of

Received 22 July 2011; Accepted 14 November 2011

Advance Access publication 23 December 2011

K.K. and K.F. are funded by the German Research Council (Grant FL 715/1-1). B.W. is funded by the German Research Council with a Heisenberg Grant (Grant WE 4427/3-1). The authors thank Florian Mormann and Johannes Niediek for helpful comments on the manuscript.

Correspondence should be addressed to Klaus Fliessbach, MD, Department of Epileptology, University of Bonn Medical Center, Sigmund Freud-Str. 25, D-53127 Bonn, Germany.
E-mail: klaus.fliessbach@ukb.uni-bonn.de

charitable donations. For this purpose, we designed an experimental situation in which subjects make decisions about the allocation of money to themselves and to a charity. After their decisions have been made, some of these decisions are discarded while others are confirmed, which allows the definition of two different reward prediction errors (RPEs): one with respect to the own, personal payoff and one with respect to the charity's payoff. The term RPE originates from reinforcement learning, where RPEs are assumed to drive adaptive learning (Knutson *et al.*, 2000). In a broader sense, the term has been applied to all situations in which a mismatch between expected and actual outcome occurs, even in the absence of learning, such as in guessing (Yacubian *et al.*, 2006) or lottery tasks (Breiter *et al.*, 2001). We apply the term RPE in this broader sense, i.e. RPEs arise whenever rewards are not fully predictable. After subjects make their choice in our experiment, there is uncertainty whether this choice will be subsequently confirmed or discarded. Therefore, an RPE arises at the time when subjects are informed about the confirmation or discard of their choice.

Questions and hypotheses

At the neural level, RPEs for one's own material goods are represented in the dopaminergic mesolimbic system (Schultz, 1998), and human fMRI studies reliably detect a corresponding BOLD signal in the nucleus accumbens (NAc; Pagnoni *et al.*, 2002). Therefore, in our study, we expected to detect a signal corresponding to the RPE for personal payoffs in the NAc. The key question was whether we would observe an equivalent signal for payoffs concerning the charity organization. We expected such a signal in subjects that made donations at their own costs. If outcome orientation has motivated these subjects' donation decisions, they should attribute a reward value to the charity's payoff, and consequently our RPE manipulation should be associated with an RPE signal similar to that for personal rewards. In addition to the NAc, we tested for outcome-related activity in other brain areas relevant to reward processing in a prosocial context, such as the subgenual area (Moll *et al.*, 2006), the medial orbitofrontal cortex (mOFC; Hare *et al.*, 2010) and the ventral tegmental area (VTA) of the midbrain (Krueger *et al.*, 2007). Furthermore, we tested whether at the time of decision making we could detect activation in any of the areas of interest that would further support the idea of action orientation.

METHODS

Participants and procedure

Prior to the fMRI experiment, each of our 33 subjects (17 female, mean age = 25.6 years, range: 21–35 years) chose one charity from a list of six organizations (Supplementary Table S1) that would benefit from her decisions. In each of the 180 trials of the fMRI experiment, subjects chose one of two alternatives, each consisting of a payoff for themselves and for the charity (Figure 1). Subjects were informed that the payoff of one randomly chosen trial would be implemented after the experiment (*actual payoff*). There was no deception: the selection of the implemented trial was random and the *actual payoffs* were transferred to the subjects and the charities. Subjects were guaranteed anonymity of their decisions (see Methodological Details in Supplementary Data).

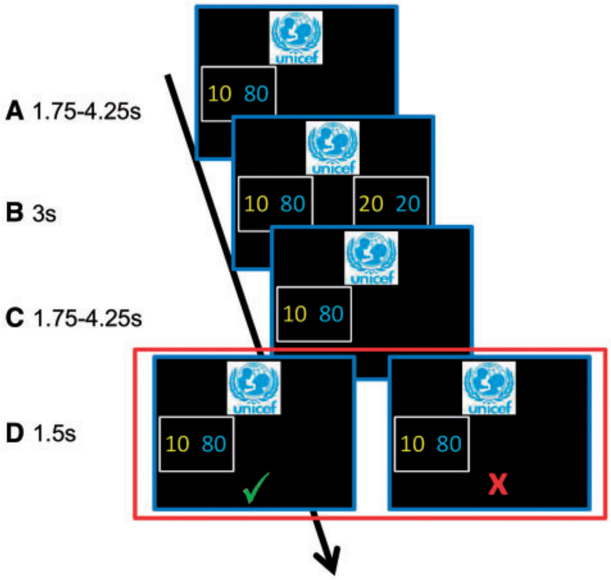

Fig. 1 Single trial settings. On the first screen, subjects saw the first alternative comprising of a payoff for the subject in yellow and for the charity in blue (A). After a jittered time interval, the second alternative appeared. Subjects had up to 3 s to select one alternative by button-press (B). Note that the trials are randomly drawn from all possible combinations of decision situations. This results in a randomization of the order of payoff alternatives (see Methodological Details in Supplementary Material). The selected alternative was presented as a response feedback (C). After a jittered time interval, a fourth screen appeared, informing subjects whether the trial was discarded (red cross) or confirmed (green check) to be among the trials from which the actual payoff trial would be chosen (D). In the second part of the experiment, confirmed outcomes from part 1 were presented (screen one in part 2; C in Figure 1), subjects were informed on a second screen whether the trial was discarded or confirmed (D).

Immediately after their decisions in each trial, subjects were informed whether this trial would be considered for the selection of the *actual payoff* (first RPE induction). At this point, 50% of the trials were randomly discarded. After the decision experiment, a second fMRI session followed, during which subjects saw each of the 90 chosen alternatives that had been confirmed during the first session. Fifty percent of these alternatives were again discarded (second RPE induction). At the time of the first RPE induction, reactions to the presented outcome might still be influenced by the previous decision. In contrast, the outcomes presented in the second session cannot be related to the decision they were based on. We therefore regard the point of the second RPE induction as the one that best represents pure outcome evaluation.

Subject's and charity's payoffs varied independently among 5, 10, 20, 40 and 80€, and payoff alternatives were randomly chosen from all possible unique combinations of different alternatives (Supplementary Figure S1). This led to four qualitatively distinguishable decision situations: 'pure self-interest' (PSI), 'noncostly donation' (NCD), 'efficiency' (E) and 'costly donation' (CD) situations. In the costly donation situations, subjects could choose to forgo monetary advantages to allocate more money to the charity (e.g. favoring 10€ for themselves and 80€ for the charity over 20€ for both; see Figure 1). In this situation, subjects had to trade material self-interest with altruistic preferences. The other three situations do not entail a conflict between different motives because one alternative is unequivocally advantageous with respect to self-interest motives (PSI), efficiency (NCD) or both motives (E). These different decision situations parallel to those implemented in a study by Moll *et al.* (2006). In contrast to Moll *et al.*, the subjects in our experiment chose between two payoff alternatives instead of having a choice to accept or reject a single payoff distribution. This allowed us to test whether at the time of presentation of the first alternative, there was a brain signal predicting the upcoming decision in the sense of a value computation. However, we did not observe such a signal (not reported). The most important extension in comparison to the study by Moll *et al.* (2006) consists of the RPE induction manipulations, which allow us to observe outcome-related effects without confounding decision-related effects.

fMRI analysis

General linear model

We included three events for the first part of the experiment: onset of the appearance of the first alternative (event 1), different onset-regressors, depending on the decision situations (event 2), onset of RPE-induction, including two parametric modulators representing the RPE of the subject's payoff, and the RPE of the charity's payoff (event 3). For the second part of the experiment, two events were included: the appearance of the chosen alternative (event 4) and the onset of the second RPE induction (event 5), including the two parametric RPE regressors (Supplementary Table S2). The parametric regressors for the RPEs were collapsed over all decision types; the rationale for this was to test whether there is an RPE signal for an outcome which is independent of the decision it is based on. This is based on the assumption that the money belonging to the charity organization carries a reward value for subjects with prosocial preferences independent of their own contribution (see 'Introduction' section).

Importantly, the parametric regressor for the charity's payoff RPE was entered after the subject's payoff RPE regressor and regressors were orthogonalized in ascending order. This means that in case of shared variance between these regressors, all commonly explained BOLD variance was attributed to the subject's payoff RPE regressor, yielding an independent and conservative estimate for the effect of the charity's payoff RPE (for Preprocessing and further information, see Details of Analysis in Supplementary Data).

Reward prediction error model

The RPE is defined as the difference between the reward magnitude (RM) of an outcome and the expected value (EV). In our experiment, the RM is not simply reflected by the respective absolute payoff of a choice (x), but must be computed with respect to a reference point (RP): $\mathrm{RM} = x\text{–RP}$. The RP depends on the subjects' previous experiences from the experiment through facing the range of possible payoffs. We assumed that subjects attribute a negative RM to payoffs that are lower than their RP. We defined the RP as the median of all previous experienced payoffs, since this resulted in the strongest effects for the subject's payoff RPE. (We also tested alternative RPs and resulting RPEs. For a discussion, see Details of Analysis in Supplementary Data and Supplementary Table S6). Thus RPE is defined as RM–EV when a choice was confirmed, and 0–EV when a choice was discarded, respectively, with $\mathrm{EV} = 0.5 \times \mathrm{RM}$, given a 50% chance that a choice is confirmed or discarded.

Region of interest definition

We focused the analysis on the NAc and derived anatomical masks of this region from the Harvard–Oxford cortical and subcortical structural atlases (http://www.cma.mgh.harvard.edu), applying a probability of 0.5. In the same way, we generated anatomical masks for the subgenual area and the mOFC. Note, that these two ROIs are not fully covered by the EPI images (Supplementary Figure S2). To cover the VTA, we used an anatomical mask of the entire midbrain posteriorly cut off at MNI coordinate $y = -22$. For the analysis of the RPE induction in the NAc, we extracted parameter estimates for the subjects' and the charity's payoff RPE from the NAc masks, averaged across all voxels. For the other regions, small volume corrections for multiple comparisons [family wise error (FWE)] were applied because we did not expect an average effect across these relatively large and functionally heterogeneous areas. In addition, we performed a whole brain conjunction analysis to test for overlapping effects of the subject's payoff RPE and the charity's payoff RPE regressor in the donator group [minimum statistic against conjunction null at $P < 0.001$, uncorrected for each individual contrast, see Nichols *et al.* (2005)].

RESULTS

Behavioral results

In situations PSI, NCD and E, all subjects consistently chose the advantageous alternative (>98% of all trials over all subjects; Table 1). We classified the few instances of deviant choices as implausible decisions and assumed that they were based on accidental errors. High inter-individual variance was only observed for the CD situation, with 17.4 % (s.d. = 25.17%) of all trials over all subjects being costly

Table 1 The four decision situations and their underlying payoff-structures (A1/B1 A2/B2) including percentages of subjects choosing A1/B1 in each situation

Decision situation	Payoff structure	Percentage of subjects choosing A1/B1 (mean ± SD)
Pure self-interest (PSI)	A1>A2, B1 = B2 e.g. 10/20 5/20	98.64 (±2.49)
Noncostly donation (NCD)	B1>B2, A1 = A2 e.g. 5/40 5/10	97.9 (±3.22)
Efficiency (E)	A1>A2, B1>B2 e.g. 20/40 10/20	99.48 (±0.93)
Costly donation (CD)	A1<A2, B1>B2 e.g. 5/80 20/40	17.49 (±25.17)

Notes: A: subject's payoff, B: charity's payoff, A1/B1: first alternative, A2/B2: second alternative. Note that the trials are randomly drawn from all possible combinations of decision situations. This results in a randomization of the order of payoff alternatives (see Methodological Details in Supplementary Data).

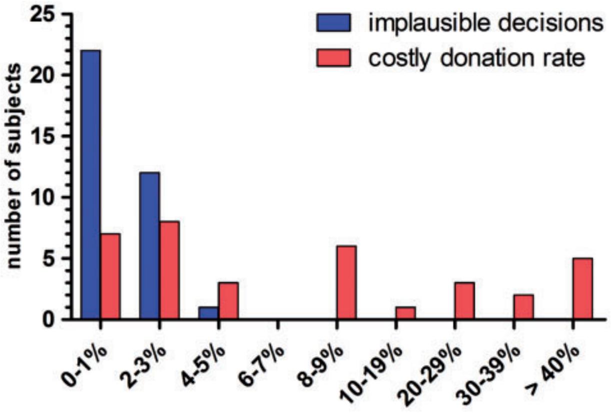

Fig. 2 Donation behavior. Distribution of subjects with respect to donation rates (CD$^+$) and rates of implausible decisions; CD$^+$: choosing donation-alternative in CD situations.

donation decisions. To confidently identify subjects who intentionally donated money in the costly donation situation, we applied the following statistical criterion: Subjects with a donation rate (CD$^+$) significantly (Fisher's exact test, $P < 0.05$) higher than the same subjects' rate of implausible decisions were classified as 'donators' ($n = 16$) and the others as 'nondonators' ($n = 17$) (Figure 2). Support for this separation comes from the fact that donators engaged more frequently in real-life prosocial activities than nondonators, according to a self-report questionnaire (Additional Results in Supplementary Data). The rates of costly donations within the donator group ranged from 8% (5 out of 60) to 97% with a mean of 33.7%. There were significant reaction time (RT) differences between the decision situations [main-effect of decision situation: $F(3, 93) = 44.848$, $P > 0.001$]. In the subgroup of donators RT in costly donation situations (CD$^+$) were longer than in noncostly donations (NCD$^+$), pure self-interest situations (PSI$^+$) and efficiency situations (E$^+$). There was also a significant decision situation × subgroup interaction $F(3, 93) = 16.686$, $P < 0.001$. *Post hoc t*-tests reveal faster reactions of the nondonators in pure self-interest situations (PSI$^+$), and for self-interest decisions in costly donation situations (CD$^-$, Supplementary Table S4). Donators and nondonators did not differ in demographic variables, such as age, monthly income or gender (Supplementary Table S3).

fMRI results

fMRI data analysis focused on the two RPE induction events. Further, we tested whether different decision types revealed different activation levels in the NAc at the time of decision making. We did not find any association: irrespective of the decision type, there was a positive BOLD signal in the NAc during decision making compared to unmodeled baseline activity (Supplementary Figures S3 and S4). Further, there were no significant differences between donation decision trials and any other conditions in any of the ROIs or between donators and nondonators (reported in Supplementary Data). These results do not support the existence of differential decision-related reward system activity for donation decisions.

For the outcome phase, BOLD activity in the NAc was highly correlated with the subject's payoff RPE at the first and second RPE induction for the entire group of subjects. This sets the stage for the main question of whether we can detect a co-localized, equivalent signal for the charity's payoff RPE. Figure 3 shows our main result (averaged over both RPE induction time-points and all voxels in the NAc masks): in the NAc, there was indeed a highly significant positive modulation of BOLD activity by the charity's payoff RPE in donators [$t(14) = 4.644$, $P = 0.00018$, one-tailed] but not in the nondonators [$t(16) = 1.195$, $P = 0.125$, one tailed]. The group difference between donators and nondonators was significant [$t(30) = 2.164$, $P = 0.02$, one tailed]. The relation between RPE signal and donation behavior was further corroborated by a significant correlation between the costly donation rate as a continuous, behavioral variable and the charity's payoff RPE signal (Spearman's $\rho = 0.309$, $P = 0.043$, one tailed).

The comparison between first and the second RPE induction reveals several differences: the effect for the charity RPE regressor in donators at the first RPE induction was only significant in the right but not in the left NAc and is on average significantly weaker than at the second RPE induction (Supplementary Table S5). In addition, at the first RPE induction, the RPE signal for *personal* payoff for the nondonators was significantly lower than that of donators (Supplementary Figure S5). To further elucidate these differences, we ran an additional analysis for the first RPE induction in which all trials in the costly donation situations were excluded. The rationale for this is that we expected the evaluation of costly donation situations in donators to be

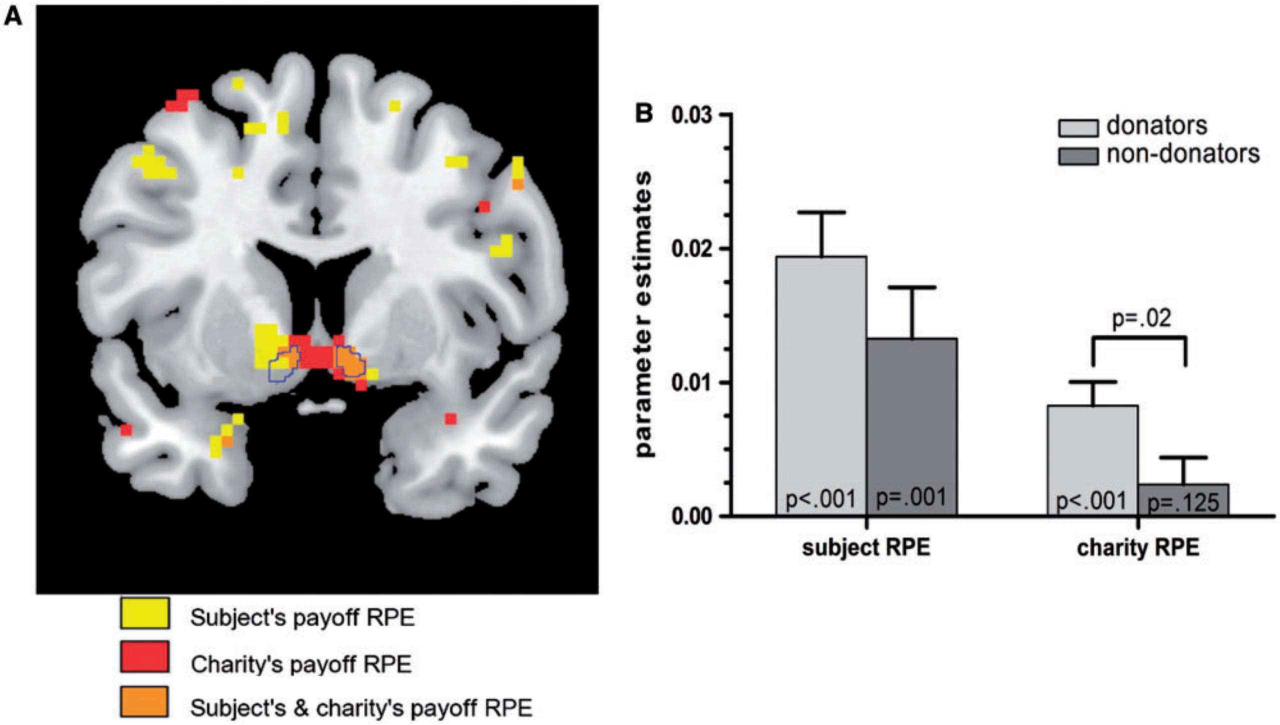

Fig. 3 FMRI results. Nucleus accumbens signals reward prediction errors for one's own payoff in all subjects and for charity's payoff in the donator group. Results are averaged across the two time points of RPE induction. (A) Coronal brain section ($y = 8$) showing voxels with a significant modulation of the BOLD signal by the subject's payoff RPE (yellow, thresholded at $t > 5.99$ corresponding to $P < 0.0001$, uncorrected) and by the charity's payoff RPE (red, $t > 2.98$, $P < 0.005$, uncorrected) for the donators ($n = 15$) (one subject did not complete part 2 of the experiment, see Methodological Details in Supplementary Material). The NAc region of interest (ROI) is framed in blue. The effect of charity's payoff RPE is significant after small-volume correction for multiple comparisons within this ROI ($P_{\text{FWE-corrected}}$<0.05). (B) Bar plot showing mean parameter estimates (± *SEM*) for the bilateral NAc. *Ps* (one tailed) for one-sample *T*-tests against zero for each regressor and for a two-sample *T*-test are shown. For separate results at each stage of RPE induction, see Supplementary Figures S5 and S6 and Table S5.

more deliberate (reflected by longer RTs) than in the other situations which might have influenced the following outcome phase. After the exclusion of these trials, we observe a highly significant charity's payoff RPE in the donator group, which does not differ from that of the second RPE induction (Supplementary Figure S7). For separate results of the first and second RPE induction, see Supplementary Figures S5 and S6, Supplementary Table S5.

Remarkably, the activation peaks and the majority of activated voxels for the charity's payoff RPE in donators lie slightly medial outside our predefined NAc region of interest toward the septal region (activation peak at MNI coordinates: $X = -3$, $Y = 8$, $Z = -8$, $t = 6.45$, $P_{\text{uncorrected, whole brain}}$ <0.001). The activation cluster extends anteriorly into the subgenual area [MNI coordinates for peak voxel in the subgenual area ROI: $X = 3$, $Y = 11$, $Z = -8$, $t = 4.95$, $P_{\text{FWE(small-volume corrected)}}$ <0.05]. Conversely, activation peaks for the personal payoff RPE lie more laterally, within the predefined NAc masks [MNI coordinates: $X = 9$, $Y = 11$, $Z = -8$, $t = 8.79$, $P_{\text{FWE(small-volume corrected)}} < 0.05$]. For details on the spatial relations between activation clusters and ROI, see Supplementary Figures S8 and S9.

Within the other ROI, there was a significant positive modulation of BOLD activity by the subject's payoff RPE in the mOFC and ventral midbrain (because of methodological limitations of fMRI in the localization of brainstem activity we refrain from using the term VTA), but no corresponding signal for the charity's RPE (Supplementary Figures S10 and S11).

The whole-brain conjunction analysis for both regressors confirmed overlapping areas bilaterally within the NAc-ROI [$P_{\text{FWE (small-volume corrected)}} < 0.05$] (Supplementary Figure S12 and Table S7). There were no overlapping activations surviving correction for multiple comparisons (whole brain or small volume correction for other ROIs) elsewhere. Areas surviving the inclusion threshold are reported in Additional Results in Supplementary Data.

DISCUSSION

Our study investigates decision- and outcome-related reward system activity in the context of charitable donations. For this purpose, our subjects took part in an fMRI experiment in which they made decisions affecting both their personal and a charity organization's payoff. This decision experiment served two purposes: it allows us to observe brain activation accompanying donation decisions (similar to Moll *et al.*, 2006), and it allows us to identify subjects which behaviorally express their willingness to donate. The main focus, however, was on the processing of outcomes. For this reason, we introduced a novel experimental manipulation by discarding part of the outcomes, thus introducing one RPE for the subject's personal payoff, and another RPE for the payoff to the charity organization.

In our subject group, ~50% of subjects made costly donation decisions. This is consistent with recent data from a worldwide survey that shows that 49% of the German population sometimes donates money (Charities Aid Foundation, 'The World Giving Index 2010'). Higher donation rates in other neuroimaging studies (Moll *et al.*, 2006; Harbaugh *et al.*, 2007; Hare *et al.*, 2010) might partly be explained by cultural differences. In the USA, for example, the percentage of donators in the average population is about 64%, according to the same source. Within the group of donators, there was large variability in the rate of costly donations, which is presumably due to the fact that within the costly donation situations there is large variation in the efficiency associated with each decision alternative. In many trials that involved a costly donation, subjects would have to give up more money than the charity would gain. Consequently, only one subject chose to donate in almost every case. Others only donated in situations where they had to give up little of their own to make a large donation, yielding only 5 out of 60 donation decisions. In contrast, in other neuroimaging studies (Moll *et al.*, 2006; Hare *et al.*, 2010), the payoff to the charity in costly donation decision was consistently enhanced in relation to the personal losses, which presumably promotes higher donation rates. Finally, in contrast to another neuroimaging study (Izuma *et al.*, 2010), our subjects made their decisions anonymously, so that social reputation gains were unlikely to contribute to higher donation rates.

Our main analysis addressed the question whether similar reward signals can be detected related to personal and the charity's payoff in reward processing areas of the brain; it thereby addressed the fundamental question of whether social and nonsocial cognition share common underlying brain processes (Adolphs, 2003). A number of previous studies have shown overlapping neural substrates at different stages of reward processing for nonsocial (e.g. own monetary) and social rewards (Izuma *et al.*, 2008; Zink *et al.*, 2008; Smith *et al.*, 2010; Lin *et al.*, 2011). In these studies, social rewards included rewards with a nonmonetary benefit for the subjects themselves (such as appraisal or reputation gains) and thus, a selfish reward. In contrast, our study examined processing of events that are exclusively relevant to someone else, i.e. a charity organization. Along with a limited number of previous studies, our study specifically addressed the processing of nonselfish, prosocial preferences. For such prosocial decisions, Hare *et al.* (2010) have demonstrated overlapping activations during value computations for personal and charity's money in the medial ventral prefrontal cortex at the decision stage. In our study, we also tested whether we could detect a value signal that would predict the subsequent choice but did not observe such a signal. It is likely that the higher variance in the values attributed to different charity organizations (in Hare *et al.* (2010), negatively evaluated organizations were included) contributed to the better detection of such signals. In addition, we tested for the effect of different decision types but did not find NAc activity or activity in the other ROIs specifically linked to donation decisions. Thus, our results do not offer support for a specific rewarding effect of the act of donating itself in the sense of a 'warm glow effect' as described previously (Moll *et al.*, 2006; Harbaugh *et al.*, 2007). This discrepancy might partly be due to the fact that we compared different types of active decisions, whereas in Harbaugh *et al.* (2007) a condition with active decision making (voluntary transfer) was compared with passive observations of transfers initiated by a computer. Our results do show activation above baseline level during all kinds of decisions. This finding is principally compatible with results of Moll *et al.* (2006), who describe common striatal activation during both pure reward decisions and NCD decisions. However, in our study activity during donation decisions was not higher than during pure reward or nondonation decisions, which leaves open the possibility that NAc activity reflects processes related to decisions in general. The comparison of different decision types in our study might be aggravated by response time differences between different decision types. In our study, donators took longer to decide in costly donation situations. In contrast, nondonators took longest in NCD situations. Such reaction time differences post a considerable problem for imaging analyses, because differences in BOLD signal might simply reflect RT differences. In our general linear model (GLM), we have included RT as duration of the respective events [which is in line with recommendations derived from empirical work by Grinband *et al.* (2008)] but nevertheless, the comparison of decisions with different RT might be problematic. Furthermore, the number of costly donation decisions varied substantially within the donator group, and in several subjects we observed less than 10 occurrences of these events, which limits the power of contrasts between costly donation decisions and other decisions. The lack of specific activity related to donation decisions might additionally be explained by the absence of social approval due to the anonymization of decisions since social approval enhances NAc activity during donation decisions (Izuma *et al.*, 2010). Finally, as a null finding, our results do not principally contradict the assumption of action orientation in charitable donations.

At the stage of outcome processing, previous studies have demonstrated increased activation in the ventral striatum for *socially* preferred outcomes, e.g. in the context of ultimatum bargaining (de Quervain *et al.*, 2004; Tabibnia *et al.*, 2008), cooperation tasks (Phan *et al.*, 2010) or inequity treatments (Fliessbach *et al.*, 2007; Tricomi *et al.*, 2010). These activations are observed even if events are neutral or negative with respect to personal monetary belongings. In the context of charitable giving, Harbaugh *et al.* (2007) have demonstrated ventral striatal activation during passive observations of money transfers to a charity. Beyond the mere existence of such outcome-related activation, our results show that NAc activity in response to payoffs to charity is parametrically modulated by an RPE term in the same way that BOLD signals respond to one's individual rewards. Remarkably, our paradigm allowed us to observe this modulation in the same subjects and at the same time as the corresponding signal for their own monetary outcomes. Our results further demonstrate that the amount of modulation of NAc activity by the charity RPE differed inter-individually, depending on a subject's donation behavior, which in turn was linked to everyday prosocial activities. This finding is in line with results showing that NAc activity generally reflects subjective rather than objective value of rewards (Tobler *et al.*, 2007) and with more specific findings linking social value orientations and reward-related brain activity (Haruno and Frith, 2010).

For our main analysis, we averaged activity across all voxels from an anatomically defined region of interest (NAc). This was done under the assumption that the NAc is functionally homogenous. Recent data suggest a functional specialization between the NAc and the adjacent septal region, with the latter being more strongly related to social aspects of rewarding events (Moll *et al.*, 2006; Hsu *et al.*, 2008) and social attachment (Krueger *et al.*, 2007). Although the relatively low spatial resolution of fMRI must be considered, it is interesting that the peak voxels and the majority of voxels showing activations for the charity's payoff RPE were located slightly medial to the predefined NAc ROI within the septal region, whereas the peak for the personal rewards is located more laterally. The activation cluster for the charity RPE extends anterior into the subgenual area (BA 25). This finding nicely complements a previous finding by Moll *et al.* (2006), who found that decision-related reward activity in charitable giving was specifically associated with activity in the septal/subgenual region. Conversely, we found a personal payoff RPE signal in the ventral midbrain and more anterior parts of the mOFC, and here no equivalent signal for the charity's payoff was observed. Together with previous findings, our results thus suggest commonalities in the processing of not only personal and social rewards (overlapping RPE signal in the NAc), but also specific reward signals in the context of prosocial behavior, with involvement of the septal area and the subgenual part of the cingulate cortex. The specific contributions of these brain areas to social cognition are a promising target for future research.

In decision experiments, it is notoriously difficult to disentangle the effects of action and outcomes because they are typically observed simultaneously. We propose a simple but innovative manipulation to selectively test outcome-related effects: by discarding part of the subject's decisions, we introduced an RPE for given outcomes. We propose such a procedure as a generally useful method for testing outcome values in decision tasks. This approach makes explicit use of reverse inference: the observed brain signal is used to infer underlying psychological constructs such as preferences. Although many studies suggest that reward-associated signals in the NAc can serve as a surrogate marker for subjects' preferences (Knutson *et al.*, 2008), reverse inferential conclusions always need to be drawn with caution because, obviously, brain signals observed by fMRI are never unambiguous in their meaning [for a comprehensive discussion, see Poldrack (2006)]. In our case, the observed RPE signal for the charity's payoff thus suggests the effectiveness of outcome-related motives but it cannot be regarded as direct proof of such motives.

Our study design included two different events of RPE induction. One took place immediately after each decision (Session 1), the other took place after all decisions had been made (Session 2). Only the outcomes from the preceding decision experiment were shown, and they were either discarded or confirmed. We expected to detect RPEs at both time points. The second RPE induction was implemented to test whether RPE signals during Session 1 might be influenced by the previous decision. The results differ between these two RPE events. Unexpectedly, during the first RPE induction, subjects in the donator group had a higher RPE signal for their own payoff than nondonators, and only a marginally significant RPE effect for the charity payoff. The results for the second RPE event appear much clearer, with a significant effect for the charity payoff in donators and no such effect in nondonators, with a significant group difference. On the other hand, there is a similar, highly significant effect for the personal payoff RPE in both groups. We can only speculate about the reasons for these differences between the two time points. Generally, we assume that during the first part of the experiment, subjects might spend less attention to the outcomes than during the second part, where these outcomes are all they are presented with. This does not fully explain why the different RPE signals seem to be differentially affected in the two groups, i.e. why the charity RPE is lower in the donator group (compared to the second RPE induction) and the personal payoff RPE is lower in the nondonator group. Interestingly, the charity's payoff RPE observed during the decision part was higher after exclusion of costly donation condition trials. As mentioned before, the time-point of the second RPE induction appears to be the clearest test of outcome-related activity. For this event (and for the average of the two RPE inductions), there is a highly significant effect for the charity RPE in the donator group, which constitutes our main finding.

In conclusion, our results provide a first demonstration of an RPE signal in the NAc for a monetary outcome that is exclusively relevant to unrelated others. This provides additional evidence for the assumption that common brain mechanisms underlie the processing of nonselfish, social and nonsocial rewards. The pattern of activation furthermore suggests an involvement of the septal region and the subgenual area in the processing of such rewards. Our results suggest that money belonging to a charity organization carries a reward value for subjects who are willing to donate and thereby provide neurophysiological support for the assumption of outcome-oriented motives in charitable giving.

SUPPLEMENTARY DATA

Supplementary data are available at *SCAN* online

Conflict of Interest

None declared.

REFERENCES

Adolphs, R. (2003). Investigating the cognitive neuroscience of social behavior. *Neuropsychologia*, *41*, 119–26.

Andreoni, J. (1989). Giving with impure altruism—applications to charity and Ricardian equivalence. *Journal of Political Economy*, *97*, 1447–58.

Andreoni, J. (1990). impure altruism and donations to public-goods—a theory of warm-glow giving. *Economic Journal*, *100*, 464–77.

Breiter, H.C., Aharon, I., Kahneman, D., Dale, A., Shizgal, P. (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron*, *30*, 619–39.

Camerer, C.F., Fehr, E. (2006). When does "economic man" dominate social behavior? *Science*, *311*, 47–52.

de Quervain, D.J., Fischbacher, U., Treyer, V., et al. (2004). The neural basis of altruistic punishment. *Science*, *305*, 1254–8.

Fliessbach, K., Weber, B., Trautner, P., et al. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. *Science*, *318*, 1305–8.

Grinband, J., Wager, T.D., Lindquist, M., Ferrera, V.P., Hirsch, J. (2008). Detection of time-varying signals in event-related fMRI designs. *Neuroimage*, *43*, 509–20.

Harbaugh, W.T. (1998). What do donations buy? A model of philanthropy based on prestige and warm glow. *Journal of Public Economics*, *67*, 269–84.

Harbaugh, W.T., Mayr, U., Burghart, D.R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science*, *316*, 1622–5.

Hare, T.A., Camerer, C.F., Knoepfle, D.T., Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *Journal of Neuroscience*, *30*, 583–90.

Haruno, M., Frith, C.D. (2010). Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nature Neuroscience*, *13*, 160–1.

Hsu, M., Anen, C., Quartz, S.R. (2008). The right and the good: distributive justice and neural encoding of equity and efficiency. *Science*, *320*, 1092–5.

Izuma, K., Saito, D.N., Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron*, *58*, 284–94.

Izuma, K., Saito, D.N., Sadato, N. (2010). Processing of the incentive for social approval in the ventral striatum during charitable donation. *Journal of Cognitive Neuroscience*, *22*, 621–31.

Knutson, B., Delgado, M.R., Phillips, P.E.W. (2008). Representation of subjective value in the striatum. In: Glimcher, P.W., Camerer, C.F., Fehr, E., Poldrack, R.A., editors. *Neuroeconomics*. London, San Diego, Burlington: Academic Press.

Knutson, B., Westdorp, A., Kaiser, E., Hommer, D. (2000). FMRI visualization of brain activity during a monetary incentive delay task. *Neuroimage*, *12*, 20–7.

Konow, J. (2010). Mixed feelings: theories of and evidence on giving. *Journal of Public Economics*, *94*, 279–97.

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

Lin, A., Adolphs, R., Rangel, A. (2011). Social and monetary reward learning engage overlapping neural substrates. *Social Cognitive and Affective Neuroscience*, doi:10.1093/scan/nsr006.

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 Acadamy of Sciences United States of America*, *103*, 15623–8.

Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J.B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage*, *25*, 653–60.

Pagnoni, G., Zink, C.F., Montague, P.R., Berns, G.S. (2002). Activity in human ventral striatum locked to errors of reward prediction. *Nature Neuroscience*, *5*, 97–8.

Phan, K.L., Sripada, C.S., Angstadt, M., McCabe, K. (2010). Reputation for reciprocity engages the brain reward center. *Proceedings of the National Acadamy of Sciences United States of America*, *107*, 13099–104.

Poldrack, R.A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*, 59–63.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*, 1–27.

Smith, D.V., Hayden, B.Y., Truong, T.K., Song, A.W., Platt, M.L., Huettel, S.A. (2010). Distinct value signals in anterior and posterior ventromedial prefrontal cortex. *Journal of Neuroscience*, *30*, 2490–5.

Tabibnia, G., Satpute, A.B., Lieberman, M.D. (2008). The sunny side of fairness: preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science*, *19*, 339–47.

Tobler, P.N., Fletcher, P.C., Bullmore, E.T., Schultz, W. (2007). Learning-related human brain activations reflecting individual finances. *Neuron*, *54*, 167–75.

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

Yacubian, J., Glascher, J., Schroeder, K., Sommer, T., Braus, D.F., Buchel, C. (2006). Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. *Journal of Neuroscience*, *26*, 9530–7.

Zink, C.F., Tong, Y., Chen, Q., Bassett, D.S., Stein, J.L., Meyer-Lindenberg, A. (2008). Know your place: neural processing of social hierarchy in humans. *Neuron*, *58*, 273–83.