Effort increases sensitivity to reward and loss magnitude in the human brain

Julen Hernandez Lallement,[1,2] Katarina Kuss,[1,3] Peter Trautner,[1,4] Bernd Weber,[1,3,4] Armin Falk,[1] and Klaus Fliessbach[1,3,5,6]

[1]Center for Economics and Neuroscience, University of Bonn, Nachtigallenweg, 53113 Bonn, Germany, [2]Department of Comparative Psychology, Institute of Experimental Psychology, Heinrich-Heine University Düsseldorf, 40225 Düsseldorf, Germany, [3]Department of Epileptology, University Hospital Bonn, Sigmund-Freud-Str. 25, 53105 Bonn, Germany, [4]Life & Brain Center, Department of NeuroCognition, University of Bonn, Sigmund-Freud-Str. 25, 53105 Bonn, Germany, [5]Department of Psychiatry, University Hospital Bonn, Sigmund-Freud-Str. 25, 53105 Bonn, Germany, and [6]German Center for Neurodegenerative Diseases (DZNE), Sigmund-Freud-Str. 25, 53105 Bonn, Germany

It is ecologically adaptive that the amount of effort invested to achieve a reward increases the relevance of the resulting outcome. Here, we investigated the effect of effort on activity in reward and loss processing brain areas by using functional magnetic resonance imaging. In total, 28 subjects were endowed with monetary rewards of randomly varying magnitude after performing arithmetic calculations that were either difficult (high effort), easy (low effort) or already solved (no effort). Subsequently, a forced donation took place, where a varying part of the endowment was transferred to a charity organization, causing a loss for the subject. Results show that reward magnitude positively modulates activity in reward-processing brain areas (subgenual anterior cingulate cortex and nucleus accumbens) only in the high effort condition. Furthermore, anterior insular activity was positively modulated by loss magnitude only after high effort. The results strongly suggest an increasing relevance of outcomes with increasing previous effort.

Keywords: effort; reward processing; functional magnetic resonance imaging; locus of control

INTRODUCTION

Reward processing is essential for adaptive behavior. Rewards initiate approach behavior, induce reward-related learning, guide decision and lead to positive hedonic feelings (Schultz, 2006). In natural environments, rewards rarely occur without previous effort, such as effort invested in foraging activities. Ecological theories hold that effort is incorporated into the processing of the resulting outcome to guide future decisions (Stephens and Anderson, 2001; Kolling et al., 2012). In humans (Prévost et al., 2010) and in animals (Rudebeck et al., 2006; Walton et al., 2007), recent findings indicate that effortful options are devaluated, suggesting that effort is incorporated as a cost when a decision is faced (effort discounting during valuation of choice alternatives). Only a few studies, however, have addressed the effect of prior effort on the processing of a subsequent reward. Animal studies have demonstrated that expenditure of effort increases the hedonic value of the resulting reward in a variety of species (Lewis, 1964; Clement et al., 2000; Johnson and Gallagher, 2011). In humans, prior actions or skill demand can increase the valuation of a subsequent reward (Zink et al., 2004; Alessandri et al., 2008; Vostroknutov et al., 2012). Further insights on the effect of effort on outcome valuation come from behavioral economic studies: The willingness to spend money is decreased for earned gains in comparison with non-earned ('windfall') gains (Muehlbacher and Kirchler, 2009). In contrast to these studies showing an increase in value for rewards following effort, Botvinick et al. (2009) report a neural activity pattern consistent with the concept of effort discounting; i.e. a devaluation of rewards following effort. In this article, we investigate how prior effort influences the valuation of different reward magnitudes. We do so under the assumption that effort does not constantly increase or decrease the valuation of a subsequent reward, but that it has differential effects on different levels of rewards: relatively low rewards might become devaluated, whereas the value of relatively high rewards might increase. In other words, the sensitivity to reward magnitude might increase with increasing effort. Such a mechanism appears plausible for organisms to assess whether an effort 'paid off'. As we assumed that this applies to outcomes in general we also investigated the effect of effort on subsequent losses of varying magnitude.

Because the neurophysiological correlates of outcome processing are well characterized, we investigated the effect of effort on reward processing using functional magnetic resonance imaging (fMRI). While the subjects were being scanned, we provided them with monetary rewards and varied the amount of effort they had to invest to solve the prior task. Subsequently, we induced a forced donation event, in which the endowment was randomly divided between the subject and a charity organization, causing a personal loss for the subject. Our study is able to provide new insights into the effect of effort on the neural processing of outcomes, because both reward and loss magnitudes were varied over a large range.

Reward processing is a major function of the mesolimbic dopaminergic system, which includes the projection sites of dopaminergic midbrain neurons to the nucleus accumbens (NAcc) (Breiter et al., 2001) and to the medial prefrontal cortex including the subgenual anterior cingulate cortex (sgACC) and the anterior medial orbitofrontal cortex (mOFC) (O'Doherty et al., 2003). Neuroimaging studies suggest a functional specialization in the components of this network, with the NAcc rather encoding reward prediction errors (RPEs, i.e. the difference between actual and expected outcome) and the medial prefrontal cortex rather encoding absolute reward value (Rangel and Hare, 2010). This functional specialization is of minor concern in this study because our paradigm neither includes reinforcement learning (which would demand RPE representations) nor decision making (demanding value representations). As both RPE and reward value increase with increasing reward magnitude we expected a positive correlation between

reward magnitude and the blood oxygen level-dependent (BOLD) signal in both areas. Our research question was whether BOLD responses to the same rewards depend on the history of their acquisition (e.g. effortful or not). The neural correlates of loss processing are less well defined, but some key areas have consistently been implicated in loss processing: the anterior insula (AI), the anterior cingulate cortex and the amygdala (see Liu *et al.*, 2011 for a meta-analysis).

We hypothesized that neural activity in reward (loss) processing areas should depend more strongly on reward (loss) magnitude in the high effort condition. Importantly, this does not necessarily imply that a reward as such is valued more after supplying effort than after providing no effort, but that large rewards are valued more in comparison to low rewards especially after high effort (i.e. effortful gains have a steeper utility function). Accordingly, we expected that increased effort leads to: (i) stronger effects of reward magnitude on activity in the NAcc and basal medial prefrontal areas at reward receipt and (ii) stronger effects of the loss magnitude on activity during the forced donation event in loss-related structures, such as AI, dorsal anterior cingulate cortex (dACC) and amygdala.

MATERIALS AND METHODS
Participants and procedure
Thirty subjects (14 female, mean age = 25.4 years, s.d. = 4.03 years) participated in the fMRI experiment, two participants were excluded from analysis due to excessive head movement. All subjects were native German speakers, right-handed and had no history of psychiatric or neurological disorders. During the experiment, subjects were presented with a total of 120 arithmetical tasks that were either difficult (40, high effort condition), easy (40, low effort condition) or already solved (40, no effort condition). After 8 s time for the calculation, subjects were presented with four possible solutions and had to quickly (3 s) choose the correct answer by button press (Figure 1). The limited time for the answer aimed at forcing subjects to actually perform the calculation when the task was displayed.

In case of an incorrect response, a negative visual feedback (red 'X') was displayed, and the next trial started. After a correct response, the subjects were first presented with a positive visual feedback (green check mark). Displaying a correctness informing feedback at this point is crucial to avoid uncertainty about reward reception for the following endowment event. After this visual feedback, subjects were endowed with money ('endowment' event). Subsequently, the endowment was split between the subject and a charity organization ('forced donation' event).

Subjects had no control over the monetary donations; instead they were presented with a forced donation. All endowments and splits were randomly determined by the experiment software. The endowments ranged from 5€ to 35€, in steps of 1€. Subjects were informed about the maximum endowment before the experiment. The amount taken away during the forced donation event ranged from 0€ to the entire endowment amount. Crucially, randomization of endowments and splits was across the three effort conditions, so that the average reward and loss magnitude was randomly distributed between conditions. At the end of the experiment, one trial was randomly selected and implemented: charity and participant (additionally to a 15€ show up fee) received the payoffs of this trial, but only if this trial was solved correctly. Subjects received detailed written and verbal instructions and gave informed written consent. The study was approved by the Ethics committee of the University of Bonn.

Additionally, 30 different subjects performed a behavioral study (see Supplementary Methods). Contrary to the fMRI paradigm, these subjects were offered the possibility to donate money obtained through the same conditions presented as in the fMRI paradigm. This manipulation allowed us to test whether our experimental manipulation would actually influence monetary decisions.

Locus of control questionnaire
After the fMRI experiment, subjects completed a German version of the locus of control questionnaire (Krampen, 1981). The questionnaire consists of three subscales which asses beliefs related to self-efficiency
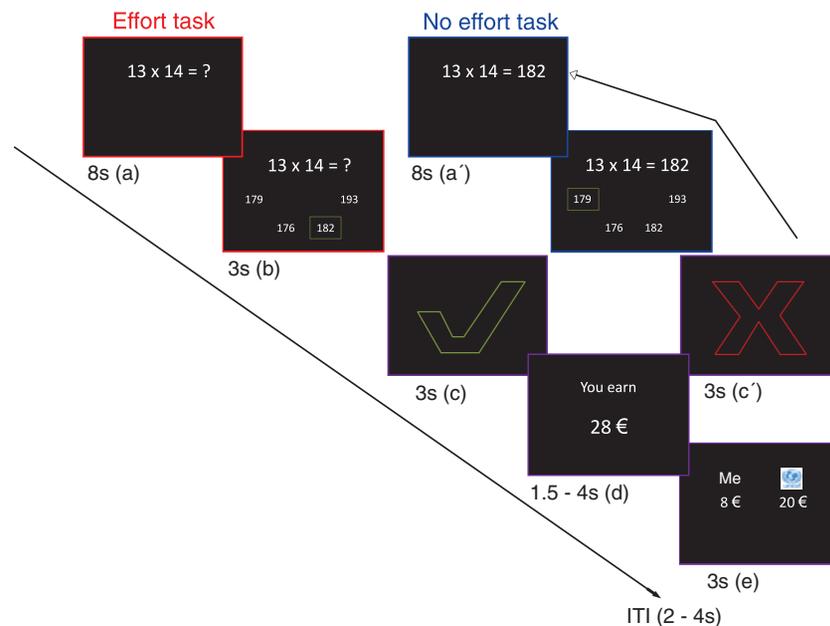

Fig. 1 Timeline of the experiment. Subjects were confronted with a calculation task, which they either solved on their own (effort task) or which was already presented with the correct solution (no effort task) (a). Subjects chose one solution out of four options. The chosen option was highlighted with a rectangle (b). A feedback signaled the correctness of the answer (c). Only in case of correct solution, subjects received an endowment ('endowment' event) (d); otherwise (c') a new trial started. A random split of the endowment (between subject and charity) was presented ('forced donation' event) (e). A jittered inter-trial interval (2–4 s) followed. Note that the sequence was identical for effort and no effort, except for the already displayed solution in the no-effort condition.

and causal attribution (for details on the subscales, see Methodological Details in Supplementary Material). We implemented this questionnaire because our experimental design aims at inducing an internal attribution of a reward in the high effort condition. Therefore, the effects of effort on neural reward- and loss-processing should depend on the predominant attribution style of the subjects.

fMRI scanning and analysis

Scanning was performed on a 1.5-T Avanto Scanner (Siemens, Erlangen, Germany) using an 8-channel head coil. Functional data were acquired using Echo Planar Imaging (EPI)-sequences (for more information, see Methodological Details in Supplementary Material).

Brain imaging analyses focused on changes in activation during endowment presentation (Figure 1d) and during forced donation (Figure 1e). For both events, we included parametric modulators in the first level general linear model (GLM), that represent the total reward magnitude (endowment event) and the relative loss magnitude (forced donation) of a given trial, respectively. The relative loss was defined as the ratio of the charity's payoff to the reward magnitude of the endowment (i.e. the fraction of money taken away from the subject). See Methodological Details in Supplementary Material for a detailed description of the GLM.

Single-subject contrasts were computed for each regressor. These were subjected to a random effects second-level analysis to test main effects of the parameters 'reward magnitude' and 'loss magnitude' for each single condition (one sample t-tests), to test differences between conditions (within subjects ANOVAs) and to test for correlations with questionnaire data (see Methodological Details in Supplementary Material).

For all group level analyses, we applied an inclusion threshold of $P = 0.005$ uncorrected, and performed Family wise error (FWE) correction for multiple comparisons restricted to the regions of interest (ROI) (endowment event) or for the whole brain (forced donation event). To test for overlapping effects of the contrasts high effort > no effort and high effort > low effort, we performed an inclusive masking procedure as implemented in SPM8. We tested the effect of high effort > no effort on $P = 0.005$ (uncorrected) and masked this with the effect of high effort > low effort [mask image $P = 0.05$ (uncorrected)]. The same masking procedure was conducted, using a mask derived from the contrast low effort > no effort. This masking procedure resembles a conjunction analysis with a logical 'AND'. In contrast to a standard conjunction analysis as implemented in SPM8, it uses different thresholds for different contrasts. Note that the resulting clusters of activation were tested on a FWE-corrected P-value of 0.05, either small volume corrected for the ROI or on the whole-brain level.

Regions of interest definition

We defined areas that are known to be involved in reward processing (NAcc, sgACC and mOFC) as ROI for the analysis of the endowment event. As the processing of loss events is less clearly restricted to singular regions, we refrained from defining a priori ROI and analyzed this data on the whole-brain level. After we identified a whole-brain correctable effect for the high effort condition in the frontal operculum/insula, we post hoc included an anatomically defined mask of the insula, to focus on differential effects of the conditions in this region.

RESULTS
Task effects on performance and brain activation

Accuracy was 81.36% (s.d. = 12.12%) in the high effort condition, 94.17% (s.d. = 6.2%) in the low effort condition and 98.49% (s.d. = 2.07%) in the no effort condition, yielding a significant condition effect on accuracy ($F = 55.098$, $P < 0.001$). The lowest number of

correct solutions in the high effort condition was 22, yielding a sufficient number of events entering the fMRI analysis. Comparing task related activity for the effort conditions (low + high) with the no effort condition yielded activity in areas known to be related to arithmetic processing, such as the intraparietal sulci, inferior prefrontal cortices and precentral cortices (Dehaene et al., 2004). In a widely overlapping network activity was higher for the high effort than for the low effort condition (see Tables S1–S3 and Figure S3).

Activity in the sgACC and NAcc scales with reward magnitude after high effort

At the time of the endowment presentation, we observed a significant positive modulation of the BOLD signal by the amount of money earned for the high effort condition in the sgACC [MNI coordinates of peak voxel: $X = 6$, $Y = 23$, $Z = -23$, $t = 4$, pFWE (small-volume corrected) < 0.05]. There was no such association in the other conditions. Based on the hypothesis that reward sensitivity is highest after high effort, we tested for differences between conditions (high effort > no effort, high effort > low effort and low effort > no effort). We observed a significant stronger modulation of BOLD signal by the reward amount after high effort compared with no effort in the sgACC [$X = 3$, $Y = 20$, $Z = -23$, $t = 4.53$, pFWE (small-volume corrected) < 0.05]. There was no such effect for the contrast high effort > low effort or low effort > no effort, when we applied the same threshold of $P = 0.005$, uncorrected. To test whether differences between the other conditions exist on a more lenient threshold and to test for overlapping activity for the contrasts, we performed an inclusive masking procedure (as described in the 'Materials and Methods' section). This procedure demonstrates overlapping effects for the contrast (high effort > no effort and high effort > low effort) in the sgACC surviving small volume correction [$X = 3$, $Y = 20$, $Z = -23$, $t = 4.53$, pFWE (small-volume corrected) < 0.05], see Figure 2. For the contrast low effort > no effort, no overlapping effect emerged. Figure 3a shows the mean parameter estimates for the parametric modulator 'reward magnitude' in the three experimental conditions averaged across a 5-mm sphere in the sgACC ($X = 3$, $Y = 20$, $Z = -23$). This figure serves demonstration reasons only, no statistical inferences are based on these averaged parameters.

We observed the same pattern in the NAcc: positive modulation of BOLD signal with the amount of money earned only in the high effort condition [MNI coordinates of peak voxel: $X = 9$, $Y = 11$, $Z = -5$, $t = 3.5$, pFWE (small-volume corrected) < 0.05], see also Figure 3b. Single voxels in the right NAcc showed significant differences surviving small volume correction for the following contrasts: high effort > no effort [$X = 6$, $Y = 11$, $Z = -5$, $t = 2.8$, pFWE (small-volume corrected) < 0.05] and high effort > low effort [$X = 6$, $Y = 8$, $Z = -5$, $t = 3.07$, pFWE (small-volume corrected) < 0.05]. Applying our inclusive masking procedure, we identified overlapping effects of both contrasts surviving small volume correction in the right NAcc [$X = 6$, $Y = 11$, $Z = -5$, $t = 2.8$, pFWE (small-volume corrected) < 0.05]. Again, no overlapping effects of the contrast low effort > no effort was observed. There was no significant effort related effect in the mOFC although on a more lenient threshold ($P < 0.005$, uncorrected) the same pattern emerged. In sum, these findings demonstrate a stronger association of BOLD signal with reward magnitude in reward-processing areas (sgACC and NAcc) specifically in the high effort condition. These results do not imply that the BOLD signal is in absolute terms higher in the high effort condition. To illustrate that point, we ran additional analysis splitting up the endowment event in four onset regressors according to the reward amount (very low, low, medium and high) for each experimental condition (see Additional Results and Figure S6 in Supplementary Material), showing a linear
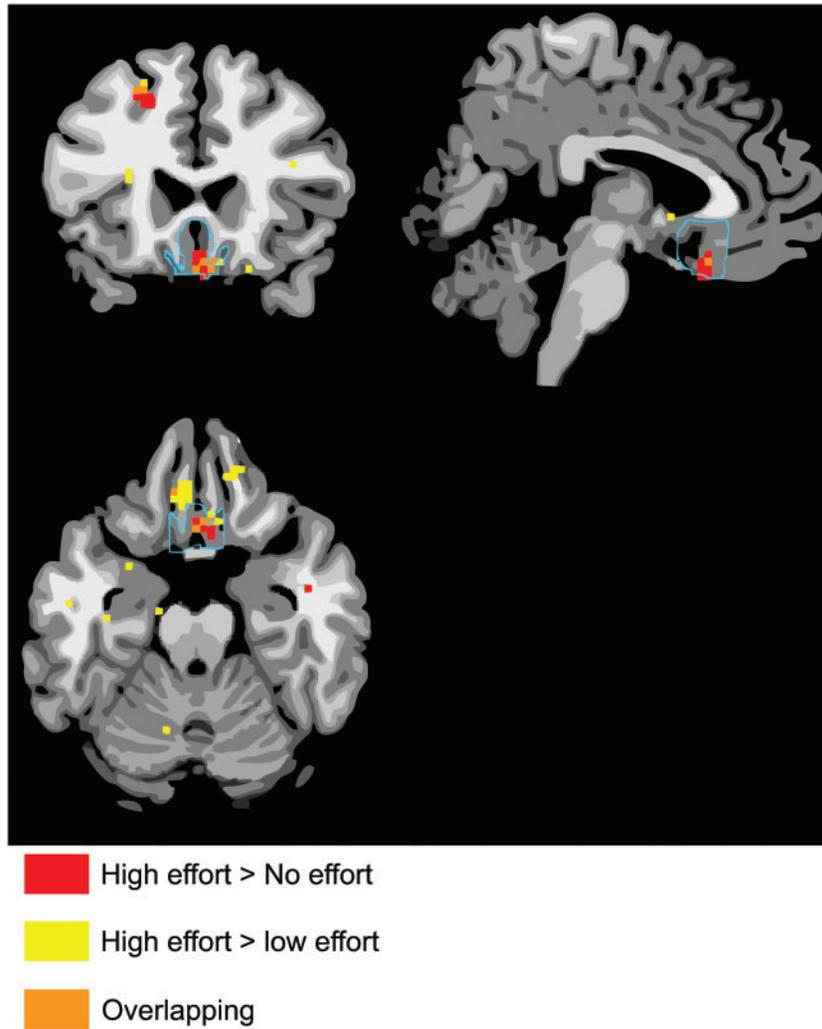
Fig. 2 Subgenual ACC. Overlapping effects (orange) of the contrasts high effort > no effort (red) and high effort > low effort (yellow). The sgACC is framed in blue. There was a stronger positive modulation of the BOLD signal by the endowment after high effort (high effort > no effort in red, thresholded at $t > 2.67$, corresponding to $P < 0.005$, uncorrected; high effort > low effort in yellow, thresholded at $t > 2.39$, corresponding to $P < 0.01$, uncorrected). MNI coordinates: $X = 3$, $Y = 23$, $Z = -21$.
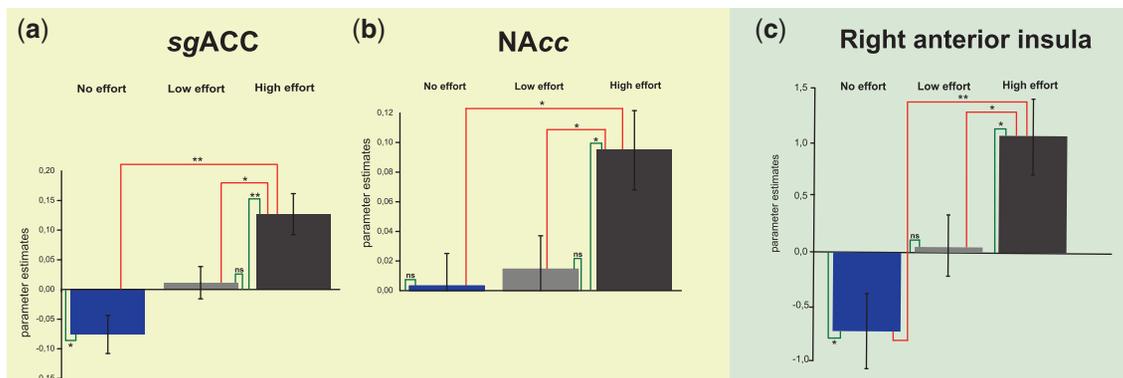

Fig. 3 Reward-related activity in the subgenual ACC (a) and NAcc (b), as well as loss-related activity in the AI (c) for each condition. (a) Mean parameter estimates ($\pm$s.e.m.) for the parametric modulator of the endowment averaged across a 5-mm sphere in the subgenual ACC ($X = 3$, $Y = 20$, $Z = -23 + 5$ mm). (b) Mean parameter estimates ($\pm$s.e.m.) for the parametric modulator of the endowment averaged across a 5-mm sphere in the NAcc ($X = 6$, $Y = 11$, $Z = -5 + 5$ mm). (c) Mean parameter estimates ($\pm$s.e.m.) for the parametric modulator of the relative loss averaged across 5-mm sphere in the AI/frontal operculum ($X = 39$, $Y = 8$, $Z = -8 + 5$ mm). The respective voxels are the peakvoxels of the overlapping effect of the contrast high effort > no effort AND high effort > low effort (identified by masking procedure). These barplots serve demonstration reasons only, no statistical inferences are based on those barplots. $**P < 0.001$; $*P < 0.05$; P's are two-tailed.

relation of the BOLD signal with the reward/loss magnitude only in the high effort condition.

Activity in the insula scales with relative loss after high effort

At the presentation of the forced donation, we observed a significant positive modulation of the BOLD signal by the relative amount of money taken away from the subject (relative loss) in the AI/frontal operculum only for the high effort condition [Left peak voxel: $X = -51$, $Y = 11$, $Z = 4$, $t = 4.53$, pFWE (whole brain, cluster level) < 0.05 with 106 voxels; right peak voxel: $X = 42$, $Y = 2$, $Z = -11$, $t = 4.00$, pFWE (whole brain, cluster level) < 0.05 with 111 voxels]. These two activation clusters are the only clusters surviving whole brain correction for multiple comparisons on cluster-level [P (FWE) < 0.05]. In the other conditions, no such association was found. We tested an alternative GLM with the absolute loss as the parametric modulator of the split event for each condition (instead of the relative loss). There is no significant modulation of BOLD signal by the absolute loss amount in neither condition (high, low and no effort).

Figure 3c shows the mean parameter estimates for the parametric modulator (relative loss) in the three experimental conditions averaged across a 5-mm sphere in the AI ($X = 39$, $Y = 8$, $Z = -8$). We tested for differences between conditions (high effort > no effort; high effort > low effort). We observed a significantly stronger modulation of the BOLD signal by the relative loss after high effort than after low effort in the AI/frontal operculum [MNI coordinates: $X = -51$, $Y = 11$, $Z = 4$, $t = 5.07$, pFWE (whole brain/cluster level) < 0.05, 189 voxels in the cluster]. The contrast high effort > no effort yields similar activation clusters in the bilateral frontal operculum and insula, surviving small volume correction for multiple comparisons in the insula [MNI coordinates: $X = 39$, $Y = 8$, $Z = -8$, $t = 4.47$, pFWE (small-volume corrected) < 0.05]. Figure 4 shows overlapping activity for both contrasts (high effort > low effort; high effort > no effort). We applied the same inclusive masking procedure as for the payoff event to test for overlapping activity for the contrasts (high effort > no effort and high effort > low effort). This procedure demonstrates overlapping effects in an anatomically defined insula mask surviving small volume correction [$X = 39$, $Y = 8$, $Z = -8$, $t = 4.47$, pFWE (small-volume corrected) < 0.05]. No such effects emerged for the contrast low effort > no effort. In sum, these findings indicate that an increase in neural activity with the magnitude of monetary losses in the insula is
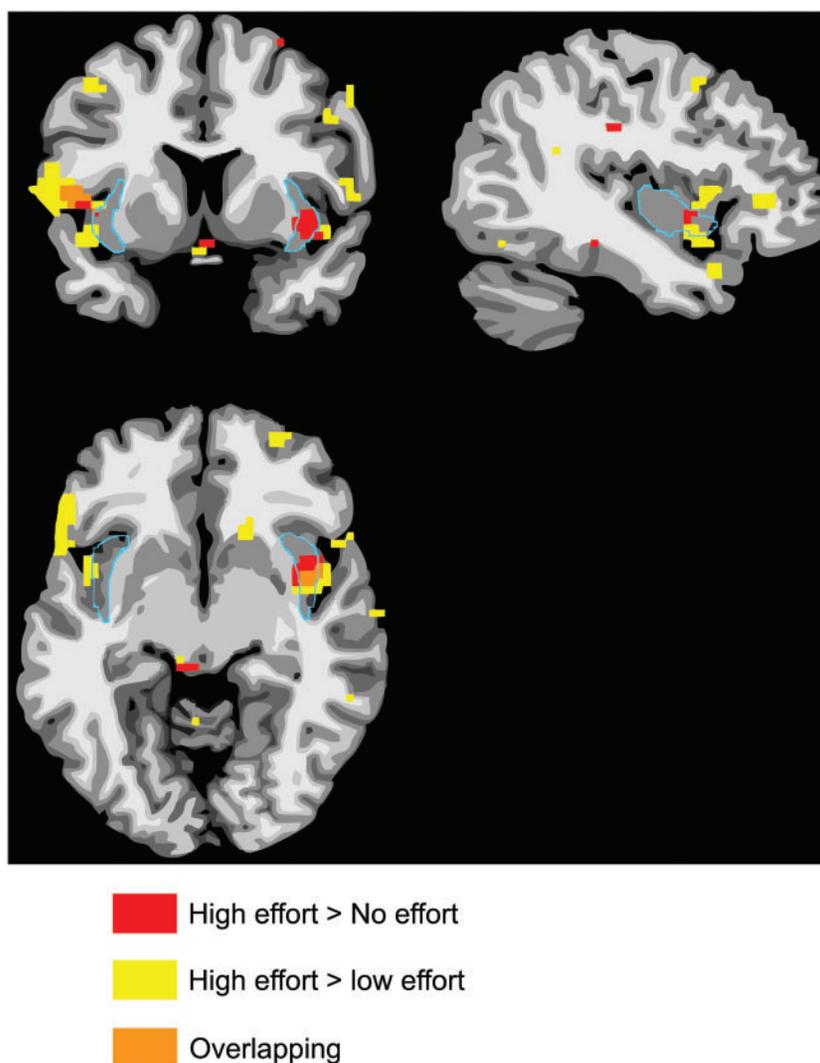
Fig. 4 AI. Overlapping effects (orange) of the contrasts high effort > no effort (red) and high effort > low effort (yellow). The insula is framed in blue. There was a stronger positive modulation of the BOLD signal by the relative loss after high effort (high effort > no effort in red, thresholded at $t > 2.67$, corresponding to $P < 0.005$, uncorrected; high effort > low effort in yellow, thresholded at $t > 2.49$, corresponding to $P < 0.008$, uncorrected). MNI coordinates: $X = -42$, $Y = 9$, $Z = -8$.

specific for the high effort condition. Neither the amygdala nor the dACC showed a significant loss-magnitude related activation.

To further illustrate the differences between the high effort and the other two conditions, we performed an additional analysis for the contrast high effort > low + no effort, showing significant (FWE corrected) differences between these conditions in the sgACC and NAcc for the endowment event and bilaterally in the insula for the loss event (see Figures S4 and S5).

Behavioral study: effort changes active donation behavior

Consistent with the neural pattern of increased insular sensitivity to relative loss magnitude we observed that our experimental conditions influence active donation decisions in our additional behavioral experiment (see Additional Results and Figures S1 and S2 in Supplementary Material). Donation rates were lowest in the high effort condition. Comparing means between the high effort and the no effort condition revealed a strong trend for lower donations after high effort (dependent samples *t*-test, $P = 0.06$, one-sided). The number of subjects that donated more after high effort than after no effort was significantly lower than the number of subjects showing the opposite pattern (sign-test, $P = 0.022$, one-sided); see Figure S2.

The effect of effort on insular loss processing is modulated by personal attribution style

We extracted the parameter estimates (parametric modulators for reward and relative loss) for the peak voxels (+5 mm sphere) showing overlapping effects for high effort > no effort and high effort > low effort, and correlated the parameters with the subscales of the locus of control questionnaire and with a difference score between the scales internality and chance. This yields 16 comparisons (4 locations × 4 scales), demanding a significance level of $P < 0.003$, to test on an overall error probability of 0.05. On this significance level, there was a positive correlation ($r = 0.574$, $P = 0.001$, two-sided) of the loss-related activity in the left insula/frontal operculum and the scale chance (scale of the locus of control questionnaire, which measures the tendency of a person to attribute causes to chance or luck). Furthermore, the difference between the two scales internality and chance significantly correlated with the loss-related activity ($r = -0.593$, $P = 0.001$, two-sided).

DISCUSSION

Our study demonstrates effects of reward magnitude on reward-related brain activity in the sgACC and in the NAcc only for money gained with high effort but not for low- or non-effort gains. Conversely, we observed that a loss magnitude related signal in the AI was significantly increased after high-effort gains compared with low- or non-effort gains. These results show that reward and loss processing in the human brain critically depend on the history of a reward.

Both the sgACC and NAcc have been implicated in the processing of rewards. Rather than any measure of objective reward value, it has been shown that activity in these brain regions reflects subjective reward value, which depends on several contextual factors, e.g. previous (Elliott *et al.*, 2000) or alternative outcomes (Breiter *et al.*, 2001), the reward of others (Fliessbach, 2007) and personal assets (Tobler *et al.*, 2007). Only a limited number of studies have addressed the effect of the action preceding a reward on the processing of that reward. Zink *et al.* (2004) showed that an active choice before a reward is followed by a higher subsequent NAcc signal than for passively receiving a reward in a probabilistic reinforcement learning paradigm. This finding indicates that the *salience* of a reward, i.e. how much it is relevant for future actions, affects reward signals in the NAcc. Another recent study demonstrates higher reward signals in

the medial orbitofrontal cortex after a skill demanding task compared to a luck-dependent task (Vostroknutov *et al.*, 2012). As in our study, the magnitude of the monetary reward was varied in this study. In contrast to our study, however, the payoff depended on the subject's performance. Therefore, reward activity in this condition could also result from a feedback about performance. Our study demonstrates that effort prior to a reward increases the sensitivity to monetary rewards even though reward magnitude was not related to the amount of effort.

Effort enters the decision process as a cost (Walton *et al.*, 2007), and is consistently related to decreased reward related brain activity at the decision phase (Croxson *et al.*, 2009; Kurniawan *et al.*, 2010; Prévost *et al.*, 2010). Botvinick *et al.* (2009) showed that not only at the stage of decision making but also at the stage of a receipt of a reward there was a stronger BOLD signal for low effort trials compared with high effort trials. The authors interpret this as an effort discounting pattern in the NAcc. However, this study did not parametrically vary reward magnitude and thus can only address absolute levels of reward-related BOLD signals for a constant reward magnitude. Rather than such absolute signals, our results show that the BOLD signal is more strongly related to reward magnitude after effort. It is important to note that this does not automatically imply that monetary gains after effort generally induce higher signals as illustrated by additional analyses (see Additional Results and Figure S6 in Supplementary Material). This analysis reveals that reward signals are not generally lower after effortless gains. Rather, the modulation of the BOLD signal by reward magnitude is increased after high effort. Therefore, our results do not contradict studies showing decreased reward-related brain activity in effort-related decision tasks. We rather suggest that subjects become more sensitive to differences in reward magnitude after effort. One possible (but speculative) explanation for our result is that effort induces expectancies about the following outcome. The actually occurring outcomes are then related to this expectancy, i.e. they become reference dependent. After no or low effort, such expectancies might be missing or weaker and therefore, subjects might become relatively indifferent against outcome magnitude. Therefore, our findings are consistent with the idea of reference dependency of both utility (Tversky and Kahneman, 1991) and reward-related brain activity (Fliessbach *et al.*, 2007; Tobler *et al.*, 2007). Given that we did not find significant associations between BOLD signals and reward magnitude in the other conditions one might even speculate that a clear reference point for the evaluation of the resulting outcome only emerges if previous effort was undertaken.

This interpretation is further supported by the finding of a loss magnitude effect in the AI that was specific for the high effort condition. A wealth of empirical findings demonstrates the role of the anterior insular cortex in the processing of negative emotions (Seymour *et al.*, 2005; Pessiglione *et al.*, 2006). In human studies of decision making, insular activity has also been shown in the context of outcome evaluation (Diekhof *et al.*, 2012). At the anticipatory stage, AI activity negatively scales with expected value (Rolls *et al.*, 2008) and positively with the expectation of an aversive stimulus (Seymour *et al.*, 2004). At the stage of outcome evaluation, insular activity signals both regret and disappointment (Chua *et al.*, 2009), and is predictive for behavioral changes occurring after the experience of negative outcomes (O'Doherty *et al.*, 2003). Importantly, insular BOLD signal changes scale with the magnitude of a loss (relative to an alternative outcome) in the same way as NAcc and mPFC activity reflects reward magnitude (Kuhnen and Knutson, 2005). In line with those findings, we found an association between BOLD signal magnitude and the relative loss occurring at the forced donation. No such association was found with the absolute amount of money that was lost. Based on standard utility theory, we expected that the relative loss would predominantly

determine the subjective loss in utility, as the initial endowment sets a reference-point ('how much can be lost?') to which the magnitude of the loss is related. Our results are therefore consistent with previous reports which support the view that neural activity codes subjective, reference-dependent values rather than absolute values (Rangel and Hare, 2010). Based on these findings we interpret the AI activity occurring at the stage of the forced donation as reflecting the averseness associated with the loss of the subject's personal endowment. Notably, the forced donation event also implies a gain of money for the charity. This means that the overall utility of the outcome presented at the loss event might not completely be reflected by the own personal loss. Especially in subjects with strong prosocial preferences the gain for the charity might carry its own utility, as suggested by a recent study (Kuss et al., 2011). Since we did not formally test for prosocial preferences in the fMRI group, we cannot control for this factor. However, from the donation behavior in our behavioral study and from the mentioned study it appears that the majority of subjects prefer their own gains over gains for the charity. Therefore, we assumed that the utility loss occurring at the loss event is predominantly reflected by the own personal loss. Note that we did not find any NAcc significant activation related to the forced donation event. This is consistent with findings showing higher NAcc activation for voluntary acts of giving in contrast to forced donations (Harbaugh et al., 2007). The finding that insular activity is modulated by the magnitude of the loss in the high effort condition but not in the two other conditions therefore complements the finding from the endowment event. We infer that the sensitivity of subjects to loss magnitude is increased for effortfully gained money.

This effect in the AI was most pronounced in subjects with high external control perception. Individuals with this personality trait tend to attribute outcomes to external causes like chance. One might speculate that these subjects are more prone to a manipulation that explicitly induces an internal attribution like the high effort condition in our study. More generally, this personality trait is related to reduced cognitive control of emotions and to an increased sensitivity for punishment (Declerck et al., 2006). Therefore, the increased insular effect in externally attributing subjects could reflect a higher sensitivity of an aversive event, i.e. the loss of effortfully earned money.

Our results bear important implications. First, they suggest that the attribution of rewards to internal or external cause critically influences their evaluation as suggested by attribution theory (Wittig et al., 1981; Weiner, 2000). This is further supported by the influence of the locus of control-personality trait which specifically features the predominant attribution style of persons. Second, together with previous behavioral studies, our results inform behavioral economic studies on social preferences and preferences in general about the fact, that non-earned (windfall) money is differently evaluated than earned money. Third, our results are linked to the psychological concept of deservingness, defined as the balance between outcome and action that led to it (Feather et al., 2011). We speculate that the increased sensitivity to the reward and loss magnitude observed after previous effort contributes to the mechanism through which deservingness emerges. Finally, the results imply that outcomes become especially important when a preceding effort had to be taken. This assumption makes sense from an ecological point of view: if a person invests energy to obtain a reward, it should matter more whether the effort actually 'paid off'. The fact that neural correlates of this effect can be induced instantaneously by a simple experimental manipulation suggests that this is a very fundamental mechanism which might have strong biological foundations.

SUPPLEMENTARY DATA

Supplementary data are available at SCAN online.

REFERENCES

Alessandri, J., Darcheville, J.-C., Delevoye-Turrell, Y., Zentall, T.R. (2008). Preference for rewards that follow greater effort and greater delay. *Learning & Behavior*, 36(4), 352–8.

Botvinick, M.M., Huffstetler, S., McGuire, J.T. (2009). Effort discounting in human nucleus accumbens. *Cognitive, Affective & Behavioral Neuroscience*, 9(1), 16–27.

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(2), 619–39.

Chua, H.F., Gonzalez, R., Taylor, S.F., Welsh, R.C., Liberzon, I. (2009). Decision-related loss: regret and disappointment. *NeuroImage*, 47(4), 2031–40.

Clement, T.S., Feltus, J.R., Kaiser, D.H., Zentall, T.R. (2000). "Work ethic" in pigeons: reward value is directly related to the effort or time required to obtain the reward. *Psychonomic Bulletin & Review*, 7(1), 100–6.

Croxson, P.L., Walton, M.E., O'Reilly, J.X., Behrens, T.E.J., Rushworth, M.F.S. (2009). Effort-based cost-benefit valuation and the human brain. *The Journal of Neuroscience*, 29(14), 4531–41.

Declerck, C.H., Boone, C., De Brabander, B. (2006). On feeling in control: a biological theory for individual differences in control perception. *Brain and Cognition*, 62(2), 143–76.

Dehaene, S., Molko, N., Cohen, L., Wilson, A.J. (2004). Arithmetic and the brain. *Current Opinion in Neurobiology*, 14(2), 218–24.

Diekhof, E.K., Kaps, L., Falkai, P., Gruber, O. (2012). The role of the human ventral striatum and the medial orbitofrontal cortex in the representation of reward magnitude—an activation likelihood estimation meta-analysis of neuroimaging studies of passive reward expectancy and outcome processing. *Neuropsychologia*, 50(7), 1252–66.

Elliott, R., Friston, K.J., Dolan, R.J. (2000). Dissociable neural responses in human reward systems. *The Journal of Neuroscience*, 20(16), 6159–65.

Feather, N.T., McKee, I.R., Bekker, N. (2011). Deservingness and emotions: testing a structural model that relates discrete emotions to the perceived deservingness of positive or negative outcomes. *Motivation and Emotion*, 35(1), 1–13.

Fliessbach, K., Weber, B., Trautner, P., et al. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. *Science*, 318(5854), 1305–8.

Harbaugh, W.T., Mayr, U., Burghart, D.R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science*, 316(5831), 1622–5.

Johnson, A.W., Gallagher, M. (2011). Greater effort boosts the affective taste properties of food. *Proceedings of the Biological sciences/The Royal Society*, 278(1711), 1450–6.

Kolling, N., Behrens, T.E.J., Mars, R.B., Rushworth, M.F.S. (2012). Neural mechanisms of foraging. *Science*, 336(6077), 95–8.

Krampen, G. (1981). IPC-Fragebogen zu Kontrollüberzeugungen (German version of Locus of Control Questionnaire by Hanna Levenson, 1972) Göttingen: Hogrefe.

Kuhnen, C.M., Knutson, B. (2005). The neural basis of financial risk taking. *Neuron*, 47(5), 763–70.

Kurniawan, I.T., Seymour, B., Talmi, D., Yoshida, W., Chater, N., Dolan, R.J. (2010). Choosing to make an effort: the role of striatum in signaling physical effort of a chosen action. *Journal of Neurophysiology*, 104(1), 313–21.

Kuss, K., Falk, A., Trautner, P., Elger, C.E., Weber, B., Fliessbach, K. (2011). A reward prediction error for charitable donations reveals outcome orientation of donators. *Social Cognitive and Affective Neuroscience*, 1–17, doi:10.1093/scan/nsr088.

Lewis, M. (1964). Some nondecremental effects of effort. *Journal of Comparative and Physiological Psychology*, 57(3), 367–72.

Liu, X., Hairston, J., Schrier, M., Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, 35(5), 1219–36.

Muehlbacher, S., Kirchler, E. (2009). Origin of endowments in public good games: the impact of effort on contributions. *Journal of Neuroscience, Psychology, and Economics*, 2(1), 59–67.

O'Doherty, J., Critchley, H., Deichmann, R., Dolan, R.J. (2003). Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *The Journal of Neuroscience*, 23(21), 7931–9.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., Frith, C.D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, 442(7106), 1042–5.

Prévost, C., Pessiglione, M., Météreau, E., Cléry-Melin, M.-L., Dreher, J.-C. (2010). Separate valuation subsystems for delay and effort decision costs. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 30(42), 14080–90.

Rangel, A., Hare, T. (2010). Neural computations associated with goal-directed choice. *Current Opinion in Neurobiology*, 20(2), 262–70.

Rolls, E.T., Grabenhorst, F., Parris, B.A. (2008). Warm pleasant feelings in the brain. *NeuroImage*, 41(4), 1504–13.

Rudebeck, P.H., Walton, M.E., Smyth, A.N., Bannerman, D.M., Rushworth, M.F.S. (2006). Separate neural pathways process different decision costs. *Nature Neuroscience*, 9(9), 1161–8.

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology*, 57, 87–115.

Seymour, B., Doherty, J.P.O., Dayan, P., et al. (2004). *Temporal difference models describe higher-order learning in humans*, Vol. 429(June), 664–7.

Seymour, B., O'Doherty, J.P., Koltzenburg, M., et al. (2005). Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nature Neuroscience*, *8*(9), 1234–40.

Stephens, D.W., Anderson, D. (2001). The adaptive value of preference for immediacy: when shortsighted rules have farsighted consequences. *Behavioral Ecology*, *12*(3), 330–9.

Tobler, P.N., Fletcher, P.C., Bullmore, E.T., Schultz, W. (2007). Learning-related human brain activations reflecting individual finances. *Neuron*, *54*(1), 167–75.

Tversky, A., Kahneman, D. (1991). Loss aversion in riskless choice: a reference-dependent model. *The Quarterly Journal of Economics*, *106*(4), 1039–61.

Vostroknutov, A., Tobler, P.N., Rustichini, A. (2012). Causes of social reward differences encoded in human brain. *Journal of Neurophysiology*, *107*(5), 1403–12.

Walton, M.E., Rudebeck, P.H., Bannerman, D.M., Rushworth, M.F.S. (2007). Calculating the cost of acting in frontal cortex. *Annals of the New York Academy of Sciences*, *1104*, 340–56.

Weiner, B. (2000). Intrapersonal and interpersonal theories of motivation from an attributional perspective. *Educational Psychology Review*, *12*(1), 1–14.

Wittig, M.A., Marks, G., Jones, G.A. (1981). Luck versus effort attributions: effect on reward allocations to self and other. *Personality and Social Psychology Bulletin*, *7*(1), 71–8.

Zink, C.F., Pagnoni, G., Martin-Skurski, M.E., Chappelow, J.C., Berns, G.S. (2004). Human striatal responses to monetary reward depend on saliency. *Neuron*, *42*(3), 509–17.