Reduced frontal cortical tracking of conflict between self-beneficial versus prosocial motives in Narcissistic Personality Disorder

David S. Stolz a,b,* , Aline Vater c, Björn H. Schott d,e, Stefan Roepke c,1 , Frieder M. Paulus a,b,1 , Sören Krach a,b

ARTICLE INFO

Keywords:
Narcissistic personality disorder
fMRI
Prosociality
Motivational conflict
Decision making

ABSTRACT

Narcissistic Personality Disorder (NPD) entails severe impairments in interpersonal functioning that are likely driven by self-beneficial and exploitative behavior. Here, we investigate the underlying motivational and neural mechanisms of prosocial decision-making by experimentally manipulating motivational conflict between self-beneficial and prosocial incentives. One group of patients diagnosed with NPD and a group of healthy controls (CTL) were scanned using functional magnetic resonance imaging while performing a prosocial decision-making task. In this task, we systematically varied the level of conflict between self-beneficial and prosocial options on each trial. We analyzed choice behavior, response times, and neural activity in regions associated with conflict monitoring to test how motivational conflict drives prosocial choice behavior. Participants in the NPD group behaved less prosocially than the CTL group overall. Varying degrees of motivational conflict between self-beneficial and prosocial options induced response variability in both groups, but more so in the CTL group. The NPD group responded faster than the CTL group, unless choosing prosocially, which slowed response times to a level comparable to the CTL group. Additionally, neural activity tracking motivational conflict in dorsomedial prefrontal cortex was reduced in the NPD group. Collectively, low generosity in NPD appears to arise from reduced consideration of prosocial motives, which obviates motivational conflict with self-beneficial motives and entails reduced activity in neural conflict monitoring systems. Yet, our data also indicate that NPD is not marked by an absolute indifference to others’ needs. This points to potentials for improving interpersonal relationships, effectively supporting the well-being of patients and their peers.

1. Introduction

From a clinical perspective, Narcissistic Personality Disorder (NPD) is characterized by a pervasive pattern of grandiosity, need for admiration, and lack of empathy (American Psychiatric Association, 2013). Accordingly, relationships in patients with NPD are described as largely superficial and serving as a means for self-esteem regulation, rather than being motivated by mutuality and genuine interest in the other person (American Psychiatric Association, 2013). Most of these descriptions derive from clinical observations, and it should be noted that they do not necessarily match the subjective experience of individuals diagnosed with NPD. Thus far, laboratory studies that aim to prove these concepts using experimental designs are lacking (Roepke and Vater, 2014). The goal of this study was to investigate self-beneficial behavior, a key interpersonal pattern in NPD, on a behavioral and neural level.

“Narcissism” can be conceptualized as a dimensional personality trait including grandiose and vulnerable features (Cain et al., 2008; Koepenik et al., 2021). Grandiose features are also reflected in the descriptive DSM-5 criteria set of NPD (American Psychiatric Association, 2013) and include an exaggerated sense of self-importance and...
need for admiration. Vulnerable features include traits such as anger, shame, and insecurity (Cain et al., 2008). A pervasive motif of grandiose narcissism is to maintain an exaggerated self-esteem, which is reflected in social behavior (Morf and Rhodewalt, 2001; Campbell et al., 2007). Non-clinically grandiose individuals overestimate their abilities (Farwell and Wohlwend-Lloyd, 1998), tend to be extraverted (Paulhus and Williams, 2002), and display interpersonal skills like charm (Back et al., 2013; Hart et al., 2017), which presumably facilitate making new acquaintances (Campbell et al., 2007; Campbell et al., 2006; Paulhus, 1998). In the long run, however, non-clinical grandiosity predicts more detrimental outcomes (Morf and Rhodewalt, 2001; Paulhus, 1998) reflected in antagonistic and defensive behavior (Back et al., 2013; Stucke and Sporer, 2002; Ferriday et al., 2011), vindictiveness (Dickinson and Pincus, 2003; Fatfouta et al., 2015), and dysfunctional relationships (Wurst et al., 2017; Sauls et al., 2019). Non-clinical narcissistic vulnerability similarly relates to defensiveness and self-enhancement (Hart et al., 2017), but also to emotional instability, low extraversion (Miller et al., 2011), low self-esteem (Crowe et al., 2018), and aggression (Pincus et al., 2009). Most behavioral studies on “narcissism” are performed in non-clinical samples using the Narcissism Personality Inventory (NPI) that mainly captures grandiose features of the construct (Pincus et al., 2009), making the transferability of these data into clinical samples disputable (Vater et al., 2013). Pathological narcissism likely represents a singular dimension, with NPD as an extreme (Aslinger et al., 2018) that is linked to reduced empathy (Ritter et al., 2011; Schulze et al., 2013), and distress of close others (Miller et al., 2007).

In healthy subjects, laboratory studies have shown that higher trait narcissism predicts higher individual acquisitiveness but diminishes the group’s total harvest in common goods games (Campbell et al., 2005). Similarly, higher levels of self-esteem vulnerability predict reduced generosity, lower sensitivity to potential punishment for low generosity, and increased retaliation (Böckler et al., 2017). This coheres with classic findings that non-clinical grandiosity predicted aggression, especially following self-esteem threats (Bushman and Baumeister, 1998). Yet, when prosocial motives gain salience through empathizing or perspective taking, both self-beneficial actions (Tusche et al., 2016; Piaciello et al., 2013; Gallo et al., 2018) and aggression (Paciello et al., 2013; Carlo et al., 2010) diminish in healthy individuals, and the arising arbitration between self-beneficial and prosocial motives entails slower reaction times (RTs) (Evans et al., 2015; Cornelissen et al., 2011; Yamagishi et al., 2017; Hutcherson et al., 2015). Habitually self-beneficial/prosocial individuals respond quickly when behaving self-beneficially/prosocially, whereas deviations from default behavior are slower (Yamagishi et al., 2017). Hence, to the extent that individuals inform their behavior by reconciling opposing interpersonal motives, experimentally induced conflict between such motives should slow down responses and potentially shift them away from default tendencies (Evans et al., 2015; Yamagishi et al., 2017).

Regarding NPD, this evidence from non-clinical samples stimulates divergent hypotheses: Either, pronounced self-beneficial behavior, as also frequently observed in non-clinical narcissism (Campbell et al., 2005; Böckler et al., 2017; Bushman and Baumeister, 1998), could go along with low motivational conflict if self-beneficial motives trump prosocial motives. Alternatively, if interpersonal deliberation in NPD entails motives like being perceived as likeable or attractive (Sack et al., 2013; Hart et al., 2017), achievable for instance through generous actions and generally self-beneficial behavior should demand to first resolve conflict with co-existing prosocial motives. In the first hypothesis, low motivational conflict should be reflected in fast RTs, whereas the second hypothesis would predict slower RTs (Evans et al., 2015; Yamagishi et al., 2017).

Neurally, motivational conflict should relate to activity in areas of medial prefrontal cortex, such as dorsomedial prefrontal cortex (dmPFC), anterior cingulate cortex (ACC), and the supplementary motor complex (SMC, including supplementary motor area (SMA), and preSMA) (Nachev et al., 2008). These regions are commonly associated with cognitive control (Botvinick et al., 2004; Bonini et al., 2014), which also holds in the social domain (Cutler and Campbell-Meiklejohn, 2019; Feng et al., 2015; Zaki et al., 2010). Further, the ventromedial prefrontal cortex (vmPFC) is broadly associated with self-relevant motivational processes, e.g. during subjective value computations regarding competing choice options (Bartra et al., 2013) or action outcomes (Stolz et al., 2020), and this also pertains to situations of self-beneficial vs. prosocial decision making (Sul et al., 2015). Hence, activity in this region may also be relevant during prosocial decision making in NPD. Moreover, the vmPFC tracks state self-esteem (Will et al., 2017; Eisenberger et al., 2011), social inequality assessment (Tricomi et al., 2010; Aoki et al., 2014), and guides prosocial behavior in concert with the temporoparietal junction (TPJ) (Hutcherson et al., 2015; Strombach et al., 2015; Park et al., 2017). The TPJ includes regions at the border of temporal and parietal lobes including superior temporal sulcus (STS) and angular gyrus (AG) (Schurz et al., 2017), implicated in higher-level social-cognitive processes (Decety and Lamm, 2007; Saxe and Kanwisher, 2003) such as moral judgement (Young et al., 2010) or perspective taking (Tusche et al., 2016). Last, the anterior insula (AI) is involved in interoceptive awareness (Craig, 2002), empathic processes (Tusche et al., 2016; Kanske et al., 2015), detection of social norm violations (Feng et al., 2015). Together with ACC, it serves the neural processing of interpersonal motives (Hein et al., 2016).

Here, patients diagnosed with NPD and a group of healthy control subjects (CTL) were scanned using functional magnetic resonance imaging (fMRI) during a prosocial decision-making task. On each trial, participants had two options to split a monetary endowment between themselves and an unknown recipient. The splits systematically varied to manipulate the level of conflict between self-beneficial and prosocial motives. Additionally, some trials allowed punishing the recipient by taking money from them. This allowed us to explore potential differences between groups in how their prosocial behavior and associated neural activity change when self-beneficial actions are only possible by actively harming others. Concluding from previous studies in non-clinical samples with narcissistic traits (Campbell et al., 2005; Böckler et al., 2017; Bushman and Baumeister, 1998), we expected lower generosity in the group of patients diagnosed with NPD than in CTL group, reflecting their decision making to be eventually governed by self-beneficial motives. Additionally, we expected activity in neural conflict monitoring systems (i.e. dmPFC, ACC, SMC) (Botvinick et al., 2004) and differential RT patterns (Evans et al., 2015; Yamagishi et al., 2017) to track the motivational conflict implied in the offers, potentially allowing to differentiate patients with NPD from CTL subjects.

2. Material and Methods

2.1. Sample

A total of N = 42 participants were included in the study. Participants belonging to the NPD group (n = 21) were recruited from Charité - Universitätsmedizin Berlin, Campus Benjamin Franklin, and were both inpatients and outpatients, but were included in the study only after being discharged from treatment. CTL group participants (n = 21) were recruited by announcements in the press and via publicly placed advertisements. The study conformed to the Declaration of Helsinki and was approved by the local ethics committee at Charité Berlin (EA4/092/10). Participants gave written informed consent and received monetary compensation for participating in the study. Structured Clinical Interviews for DSM IV axes I and II Disorders (SCID-I; SCID-II; First et al., 1996, 1997; German Versions: Fydrich et al., 1997; Wittchen et al., 1997) were administered to all participants by a trained clinical psychologist to validate NPD diagnosis and assess comorbidity (Table 1). Potential participants for the CTL group were excluded if completion of SCID-I or SCID-II indicated the presence of any mental disorder. As assessed in prior studies of our group, interrater reliability (κ = 0.80)
Table 1
Demographics and psychometric characteristics of study samples.

|                        | CTL group (n = 19) | NPD group (n = 18) | M       | SD      | M       | SD      | t      | p (2-sided) |
|------------------------|-------------------|--------------------|---------|---------|---------|---------|--------|-------------|
| age                    | 38.84             | 38.94              | 2.14    | 0.03    | 2.14    | 0.03    | -0.03  | .975        |
| count                  | 5/14              | 5/13               | 0.01    | .920    |         |         |        |             |
| female/male            |                   |                    |         |         |         |         |        |             |
| Axis I diagnosis       |                   |                    |         |         |         |         |        |             |
| Depressive episode     | 0                 | 9                  |         |         |         |         |        |             |
| (mild)                 |                   |                    |         |         |         |         |        |             |
| Depressive episode     | 0                 | 3                  |         |         |         |         |        |             |
| (moderate)             |                   |                    |         |         |         |         |        |             |
| Recurrent depressive   | 0                 | 1                  |         |         |         |         |        |             |
| disorder               |                   |                    |         |         |         |         |        |             |
| (moderate)             |                   |                    |         |         |         |         |        |             |
| Recurrent depressive   | 0                 | 3                  |         |         |         |         |        |             |
| disorder (currently    |                   |                    |         |         |         |         |        |             |
| in remission)          |                   |                    |         |         |         |         |        |             |
| Axis II diagnosis      |                   |                    |         |         |         |         |        |             |
| NPD                    | 0                 | 18                 |         |         |         |         |        |             |
| BPD                    | 0                 | 6                  |         |         |         |         |        |             |
| APD                    | 0                 | 1                  |         |         |         |         |        |             |
| HPD                    | 0                 | 1                  |         |         |         |         |        |             |
| NPI total              | 2.37              | 0.81               | 3.97    | 0.70    | -6.47   | <0.001  |        |             |
| Contingent self-       | 2.34              | 1.12               | 4.07    | 0.94    | -5.10   | <0.001  |        |             |
| esteem                 |                   |                    |         |         |         |         |        |             |
| Exploitative           | 2.21              | 1.03               | 4.02    | 1.19    | -4.95   | <0.001  |        |             |
| Self-sacrificing Self- | 2.78              | 1.25               | 3.64    | 1.22    | -2.11   | .042    |        |             |
| enhancement            |                   |                    |         |         |         |         |        |             |
| Hiding the Self        | 2.71              | 1.08               | 4.37    | 0.89    | -5.06   | <0.001  |        |             |
| Grandiose Fantasy      | 2.15              | 0.83               | 3.83    | 1.05    | -5.42   | <0.001  |        |             |
| Devaluing               | 2.11              | 0.73               | 3.83    | 1.06    | -5.78   | <0.001  |        |             |
| Entitlement Rage       | 2.30              | 0.93               | 3.94    | 1.06    | -5.05   | <0.001  |        |             |
| NPI total              | 2.11              | 0.73               | 3.83    | 1.06    | -5.05   | <0.001  |        |             |
| Leadership             | 2.05              | 1.22               | 2.72    | 1.22    | -2.59   | <0.001  |        |             |
| Grandiosity/           | 2.00              | 1.63               | 4.39    | 1.91    | -4.09   | <0.001  |        |             |
| exhibitionism          |                   |                    |         |         |         |         |        |             |
| Entitlement/           | 1.00              | 0.88               | 2.83    | 1.10    | -5.51   | <0.001  |        |             |
| exploitative           |                   |                    |         |         |         |         |        |             |
| BDI                    | 2.95              | 4.62               | 20.78   | 9.93    | -6.94   | <0.001  |        |             |

Note: CTL = control group. NPD = Narcissistic Personality Disorder. BPD = Borderline Personality Disorder. APD = Anxious Personality Disorder. HPD = Histrionic Personality Disorder. a = Welch test was performed instead of student’s t-test due to unequal variances.

2.2. Procedure

Before the study, a set of paper–pencil questionnaires was sent to all participants (Table 1). Upon arrival at the MRI facility, all participants were briefed about the study protocol and gave written informed consent. They were informed that inside the MRI they were going to repeatedly decide how to share an endowment of 1000 points (equivalent to 10 €) between themselves and another person (henceforth: recipient) by choosing between two alternative splits of this endowment. To ensure interpersonal relevance of the task, they were informed that after task completion, a number of their choices would be selected randomly, defining how much money would be paid to them and to the next study participant (i.e. the recipient). Additionally, participants received money based on the choices of one other participant who was assigned randomly. All participants were tested alone and never met any other participant.

2.3. Experimental task

Stimulus presentation was performed using the Presentation software package (Neurobehavioral Systems, Albany, CA). On each trial, one option was displayed on the left and the other on the right side of the screen, with the amounts for the participant and the recipient on the top and the bottom, respectively (Fig. 1). Participants chose by pressing a button with either the index or the middle finger. In a baseline (BL) condition, with mean number of points assigned to each player and the preset maximum deviation from the mean assignment (max +/-). The split assigning most money to the participant was defined to be the self-beneficial option, whereas the prosocial option was the one splitting the endowment in a relatively more beneficial way for the recipient. Note that due to this definition, the prosocial option did not necessarily assign more money to the recipient than to the participant. There were 12 different, pseudorandomized trial sequences that were designed to optimize BOLD efficiency. Thus, the number of trials in each sequence varied between 223 and 229 (see section 2.3 for details).
conditions, both options assigned precisely 500 points to each person. In five
other conditions the level of conflict between self-beneficial and prosocial
motives was manipulated by systematically splitting the endowment, with one option favoring the participant more strongly (self-beneficial option) than the other option (prosocial option). In one condition, options were very similar, inducing low motivational conflict (low conflict, LC). On other trials, one option assigned a high payoff to the participant, leaving little to the recipient, whereas the other option proposed a fairer split (medium conflict, MC). Still other trials did not allow to resolve motivational conflict this way, because avoiding self-beneficial actions required to choose an extremely prosocial option (high conflict, HC).

Further, one condition allowed participants to increase their gain by subtracting money from the recipient, or to resort to a fairer split (medium conflict punishment, MCP). Finally, one condition also allowed participants to increase their payoff at the recipient’s cost, but the alternative would be to choose an extremely prosocial option (high conflict punishment, HCP). On each trial, following 500 ms fixation, a choice had to be made within the 3000 ms display of the options, otherwise the trial was coded as invalid. Each trial onset was triggered by an MRI pulse, and the inter-trial-interval (ITI) was jittered between 500 and 6500 ms in steps of 2000 ms (further information in Supplementary Methods and Supplementary Tables S1 & S2).

2.4. Statistical analyses of behavioral data

Statistical analyses of behavioral data were performed using JASP (JASP Team, 2019, 2019), jamovi (The jamovi project, 2020), and the p.adjust() function in R (R Core Team, 2013). We analyzed percentages of prosocial choices and median RTs for the different experimental conditions using repeated measures analysis of variance (ANOVA) and nonparametric tests, as appropriate. Since both choice options in the BL condition offered the same amount of 500 points to the player and the recipient, no self-beneficial or prosocial option was defined and BL choice data were not analyzed. However, RTs from the BL condition were analyzed and it served as an explicit baseline for the analysis of fMRI data. To test the effects of increasing motivational conflict on behavior and neural activity, we compared HC trials against the average of LC and MC trials (LC\&MC), because the latter both allowed to resolve the conflict with the same pro-social option in which the participant still received more payoff but amounts were rather similar (i.e. 550 and 450, respectively). This was not the case in the HC condition, in which the pro-social option implied a relative loss for the participant and gain for the recipient (200 and 800, respectively). To test the main effect of possible punishment, we averaged MCP and HCP trials (PUN) and compared this to the average of MC and HC trials (noPUN). Last, to test for automatic expressions of habitual response tendencies (Cornelissen et al., 2011; Yamagishi et al., 2017), we compared RTs in dependence of the decision made (i.e., self-beneficial vs. prosocial) and between groups in a mixed two-way ANOVA. One participant in the CTL group and three participants in the NPD group never chose the prosocial option and were therefore excluded from these specific analyses. Therefore, the sample sizes in this analysis were n = 18 (CTL) and n = 15 (NPD).

2.5. FMRI data acquisition and preprocessing

Echo-planar images (EPIs) were acquired on a Magnetom TrioTim Syngo (Siemens, Munich, Germany). The entire run consisted of 600 EPIs (37 ascending slices; 20% gap; 3°×3°×3 mm voxels; TR = 2000 ms; TE = 30 ms; FA = 70°; FoV = 192; GRAPPA = 2). Additionally, a high-resolution T1-weighted MPRADE image (176 ascending slices; 50% gap; 1°×1°×1 mm voxels; TR = 1900 ms; TE = 2.52 ms; FA = 9°; FoV = 256; GRAPPA = 2) was acquired to improve spatial normalization. All fMRI data were preprocessed and analyzed using SPM 12 (Ashburner et al., 2014) in Matlab R2019b (The Mathworks, Natick, Massachusetts, USA). EPIs were corrected for acquisition delay (slice timing), with the middle slice as reference, and for head motion (realignment). The MPRADE image was coregistered to the mean EPI obtained from realignment and segmented using the unified segmentation algorithm implemented in SPM. EPIs were normalized to a common stereotactic reference frame (MNI) space using the transformation parameters obtained from segmentation, and smoothed using an 8 mm full-width-at-half-maximum isotropic Gaussian kernel.

2.6. Statistical analysis of fMRI data

All fMRI data were analyzed using SPM12 (Ashburner et al., 2014), using a two-stage mixed effects model. The first-level general linear model (GLM) contained six regressors of interest, one for the BL condition and one for each of the Choice conditions, plus the six rigid-body movement parameters obtained from realignment as regressors of no interest, to account for variance related to head motion, and a single constant. At the second level, we constructed a full-factorial ANOVA model, which included the contrast images of each Choice condition against the BL condition (e.g. HC > BL), separated by group, resulting in the within-subjects factor condition with five levels, and the between-subjects factor group (CTL, NPD).

Last, we aimed to test whether within-subject variations in neural activity during prosocial decision making can be plausibly interpreted as reflective of decision conflict, which has been previously shown to be reflected in response times (Evans et al., 2015; Yamagishi et al., 2017). Thus, we set up an additional GLM for each participant, in which we modelled BL trials separately from Choice trials, and parametrically modulated, in this order, the latter with dummy predictors that indicated whether medium (MC, MCP) or high (HC, HCP) conflict trials were presented, and whether or not a given trial was a PUN trial (i.e., MCP, HCP). Last, we included a parametric modulator indicating the RT on each trial (log-transformed and z-scored). Parametric modulators were orthogonalized and additional regressors were included that modelled invalid trials and head motion.

2.7. Statistical inference for fMRI data

Statistical inference for all fMRI analyses was based on a significance threshold of p < .05 (family-wise error (FWE) corrected at peak level) in combination with a cluster-extent threshold of k = 10.

2.7.1. Region-of-Interest analyses to examine motivational conflict during prosocial decision making

Since our primary focus was to investigate how motivational conflict engages regions associated with conflict monitoring, we used Neurosynth (Yarkoni et al., 2011) to perform an automated meta-analysis of functional neuroimaging studies by searching for the term “conflict” and obtaining the association test map thresholded at p < .01 (false-discovery-rate-corrected; 337 studies). This map was then limited to include only clusters of 5 or more contiguous voxels and used as a region-of-interest (ROI), largely covering dorsal ACC, SMC, and aspects of bilateral AI (henceforth: conflict mask). We used this ROI to test hypotheses regarding (group differences in) the level of motivational conflict during prosocial decision making.

2.7.2. Whole brain analyses to explore effects of punishment

Since we did not have any precise a priori expectation regarding the localization of neural responses to the possibility of exerting punishment during the task, the corresponding analyses were performed on the whole-brain level.
3. Results

3.1. Participants in NPD group acted more self-beneficially, but avoided punishing recipients

Across all Choice conditions, the NPD group chose the prosocial option less often than the CTL group (Mann-Whitney-U test: $W = 260.5$, one-sided $p = .003$; Fig. 2 & Table 2).

When motivational conflict was low, the CTL group behaved more prosocially than the NPD group (LC&MC: $W = 266.5$, one-sided $p = .002$). When conflict increased, prosocial choices diminished in both groups (difference: LC&MC minus HC; CTL group: $W = 171.00$, one-sided $p < .001$; NPD group: $W = 100.5$, one-sided $p = .003$), and this effect was larger in the CTL group ($W = 288.50$, two-sided $p < .001$). However, under HC, the CTL group still behaved more prosocially than the NPD group ($W = 229$, one-sided, Holm-corrected $p = 0.046$). Thus, the difference in prosociality between the groups was reduced, but still present, when prosocial behavior was very disadvantageous for the participant.

Across groups, prosociality increased when self-beneficial actions required to punish the recipient (difference PUN minus noPUN, CTL: $W = 158$, one-sided $p < .001$; NPD: $W = 87$, one-sided $p = .016$). This increase was descriptively, but not significantly larger in the CTL than in the NPD group ($W = 219$, one-sided $p = .074$). Hence, having to punish others for one’s own benefit led to more prosocial behavior.

Fig. 2. Behavioral data. a. Percentages of prosocial choices for the CTL group and NPD group overall (left), for trials with low and medium (LC&MC) vs high motivational conflict (HC; middle), and separately for trials without (noPUN: MC, HC), and with punishment (PUN: MCP, HCP). b. Median reaction times in seconds for BL and Choice trials (left), for LC&MC trials and HC trials (middle), and separately for PUN and noPUN trials. Where shown, small crossed bars represent interaction between group (CTL, NPD) and factor on x-axis. Central bars of boxplots show median, lower and upper box borders show 25th and 75th percentile and whiskers end at last data point within 1.5 times the interquartile range from lower or upper box border.
3.2. Responses in NPD group were fast for self-beneficial decisions and slow for prosocial decisions

RTs did not generally differ between groups (ANOVA with factors trial type [BL, Choice] and group [CTL, NPD]; test for effect of group: F(1,35) = 3.39, p = .074, $\eta^2_p=0.09$). However, there was a significant main effect of trial type (F(1,35) = 37.15, p < .001, $\eta^2_p=0.52$), that was modulated by group (trial type*group interaction: F(1,35) = 11.09, $p = .002$, $\eta^2_p=0.24$; Fig. 2). Post-hoc tests showed that the CTL group responded slower during Choice trials ($t(54.2)=3.21$, two-sided $p_{Holm}<0.009$), but there was no difference during BL trials ($t(54.2)=0.02$, two-sided $p_{Holm}=0.987$). Moreover, motivational conflict slowed RTs in the CTL group (Choice > BL: $t(35)=6.76$, $p_{Holm}<0.001$), but not in the NPD group (t(35) = 1.93, $p_{Holm}=0.186$). Thus, both groups responded comparably fast unless interpersonal motives needed to be reconciled, which slowed responding only in the CTL group.

Median RTs did not differ between self-beneficial and prosocial decisions across groups (ANOVA with factors decision [self-beneficial, prosocial] and group [CTL, NPD]: no main effect of decision: F(1,31) = 4.03, $p = .054$, $\eta^2_p=0.12$; factor group: F(1,31) = 2.91, $p = .098$, $\eta^2_p=0.09$). However, the effect of decision differed between groups (decision*group: F(1,32) = 8.56, $p = .006$, $\eta^2_p=0.22$; Fig. 3), showing that the CTL group was slower than the NPD group during self-beneficial choices (CTL > NPD: $t(49.2)=2.93$, $p_{Holm}=0.026$; Fig. 3), but not during prosocial choices (CTL > NPD: $t(49.2)=0.03$, $p_{Holm}=1.000$). Moreover, RTs within the CTL group did not differ between decisions ($t(31)=0.68$, $p_{Holm}=1.000$), but the NPD group was faster for self-beneficial than for prosocial decisions ($t(31)=-3.34$, $p_{Holm}=0.013$), which was not due to group-specific effects of the experimental conditions (Supplementary Table S3).

RTs were insensitive to level of motivational conflict across groups (ANOVA with factors level of conflict [LC, MC, HC] and group [CTL, NPD]: F(1,35) = 1.86, $p = .181$, $\eta^2_p=0.05$). Yet, we found a main effect of group (F(1,35) = 8.53, $p = .006$, $\eta^2_p=0.20$), that was driven by the HC condition (CTL > NPD: $t(35)=3.21$, $p_{Holm}=0.017$). The interaction of both factors was not significant (F(1,35) = 3.50, $p = .070$, $\eta^2_p=0.09$).

![Fig. 3. Reaction times separated by decision. Small crossed bars represent interaction between group (CTL, NPD) and decision made (self-beneficial, prosocial). Central bars of boxplots show median, lower and upper box borders show 25th and 75th percentile and whiskers end at last data point within 1.5 times the interquartile range from lower or upper box border.](image)

Thus, when prosociality was disadvantageous for the participants,
responses in the CTL group were slower than in the NPD group.

Last, the option to punish did not impact reaction times across groups (ANOVA with factors punishment [noPUN, PUN] and group [CTL, NPD]: \(F(1,35) = 0.00, p = .983, \eta^2_p = 0.00\)). Again, the CTL group responded slower than the NPD group (main effect of group: \(F(1,35) = 8.30, p = .007, \eta^2_p = 0.19\)), which was due to noPUN trials (CTL > NPD: \(t(43.2) = 3.06, p_{\text{Holm}} = 0.023\)). No significant interaction was found (punishment*group: \(F(1,35) = 1.06, p = .311, \eta^2_p = 0.03\)).

3.3. Conflict-related neural activity during prosocial decision making was reduced in NPD group

Across groups, presence of motivational conflict (Choice > BL) elicited activity in right dorsal midcingulate cortex. This effect was stronger for the CTL than the NPD group in left dorsomedial prefrontal cortex and left preSMA (Fig. 4; see Table 3 for an overview of all fMRI analyses and Supplementary Tables S4-S11 for full activation coordinates), while the opposite comparison (i.e. Choice > BL * NPD > CTL) did not yield any significant results. To explore the group difference, we then tested CTL > NPD in each condition separately (see Table 4). This indicated that the overall group difference for Choice > BL was driven by conditions in which motivational conflict was high (HC) and in which, additionally, punishment was possible (HCP). Together, these findings suggest that the consideration of highly conflicting motivations and potential interpersonal consequences of one’s behavior is linked to conflict-related neural activity particularly in the CTL group.

Contrasting HC against the average of LC and MC (HC > LC&M) to test for effects of increasing motivational conflict on neural activity did not yield any significant effects or interactions.

Across groups, when self-beneficial behavior required punishing others (PUN > noPUN), activity increased in right AG, right inferior temporal gyrus (ITG). To understand this effect in more detail, we tested PUN > noPUN in each group separately. This demonstrated that the overall effect in right ITG was driven by the CTL group, which also showed significantly stronger activations for PUN vs noPUN trials in left AI, and right middle temporal gyrus. In the NPD group, no significant effects of PUN > noPUN were found. However, despite the presence of an effect for PUN > noPUN in one but not the other group, no interaction of group and PUN > noPUN was found. In sum, when self-beneficial behavior was possible only when simultaneously punishing others, this induced stronger activity in brain regions like right AG and ITG across groups, and this effect was driven by the CTL group.

3.4. Interindividual variability in neural task responses is not predicted by variability in prosocial behavior

Furthermore, we tested whether interindividual variability in task-related neural activity relates to between-subject differences in prosociality. We thus regressed the first-level Choice > BL contrast images on the participants’ mean percentages of prosocial choices, while controlling for group differences, which did not yield any significant effects. Regressing the contrast HC > LC&M on the corresponding difference in prosociality (i.e., HC minus LC&M) did not yield any significant effects. Similarly, regressing the contrast PUN > noPUN on the corresponding difference in prosociality (i.e., PUN minus noPUN) did not yield any significant effects.

Fig. 4. Neural activity during the experimental task. a. Across groups, neural activity in regions associated with the term conflict (green outline; a priori defined "conflict" region-of-interest, see section 2.7.1 for details on definition of the mask) including dorsomedial prefrontal cortex (dmPFC) and supplementary motor area (SMA) was higher for the contrast Choice > BL (displayed at \(p < .0001\), uncorrected, at whole-brain level). This effect was larger in the CTL group than in the NPD group, visualized in the bar plots. Neural activation data in bar plots are based on mean parameter estimate across all voxels in dmPFC and SMA, respectively, in which the interaction of Choice > BL * CTL > NPD was significant at \(p < .05\), small-volume FWE-corrected inside the conflict mask). b. Neural activity in right angular gyrus (AG) and right inferior temporal gyrus (ITG) was higher for PUN than for noPUN trials, but did not differ between groups. Bar plots show mean parameter estimates across all voxels in right AG and right ITG in which the contrast PUN > noPUN was significant at \(p < .05\), whole-brain FWE-corrected (displayed at \(p < .005\), uncorrected, at whole-brain level). Inset in bottom right corner depicts activation in right AG, rotated around the y axis by 45°. Errorbars are +/- 1 standard error of the mean.
8

3.5. Reaction times predict conflict-related neural activity across participants

Last, under motivational conflict, slower participants (contrast Choice > BL regressed on difference in median reaction times for Choice trials minus BL trials) showed higher activity in right midcingulate gyrus, left AI, and left SMA, while controlling for group differences. Similarly, slower RTs due to increasing conflict were associated with higher neural activity in SMA, and left AI. In contrast, differences in median reaction times between PUN and noPUN did not predict differences in neural activity between these conditions. Finally, slower trial-by-trial RTs predicted neural activity in ACC and bilateral AI. Collectively, these findings provide further support for our interpretation that observed neural responses in regions covered by the conflict mask play a role in the neural processing of motivational conflict.

4. Discussion

In the present study we investigated how prosocial decision making...
and underlying neural processes in patients diagnosed with NPD are shaped by conflicting interpersonal motives. Compared to control subjects, the NPD group behaved less prosocially, responded faster when choice options implied high motivational conflict, and showed less neural activity in regions associated with conflict monitoring such as the SMA and cingulate cortex (Nachev et al., 2008; Botvinick et al., 2004). When prosociality was disadvantageous for the participants, it diminished in both groups, although control subjects still behaved more prosocially than the NPD group. Both groups were similarly affected by the requirement to punish others for their own benefit, which shifted behavior away from self-beneficial actions, and in the CTL group increased neural activity in right AG and ITG. Last, when participants in the NPD group acted prosocially, thus deviating from their self-beneficial mode of behavior, this slowed responding to a level comparable to the CTL group.

Our finding of low generosity in participants diagnosed with NPD is well in line with existing literature from non-clinical narcissistic samples (Campbell et al., 2005; Böckler et al., 2017). We hypothesized that this may either result from more readily available tendencies towards self-beneficial actions (Cornelissen et al., 2011) or, alternatively, from cognitively effortful arbitration between conflicting interpersonal motives; i.e. either to increase one’s own gain or to behave prosocially in order to be seen in a good light by others (Back et al., 2013; Hart et al., 2017). Our findings speak to the former hypothesis, as the mainly self-beneficial responses in the NPD group were generally fast, which was paralleled by comparably low conflict-related neural activity (Nachev et al., 2008; Botvinick et al., 2004), and rare prosocial choices were considerably slower (Evans et al., 2015; Yamagishi et al., 2017). Participants in the CTL group more readily utilized the fairness information provided in each trial to guide their decisions, causing motivational conflict and the need to engage in cognitive control (Evans et al., 2015; Botvinick et al., 2004). In sum, our data nicely coincide with earlier research showing that deviations from self-beneficial or prosocial habits are marked by slowing RTs (Yamagishi et al., 2017). Thus, using response times one may reveal individual differences in social value orientations (Cornelissen et al., 2011), and our data indicate that participants in the NPD were more strongly oriented toward self-beneficial actions.

Notably, behavior in the NPD group was not fully insensitive to its potential social consequences. Similar to the CTL group, participants in the NPD group acted more self-beneficially when prosociality was believed to be highly detrimental, and acted more prosocially when potential social consequences. Similar to the CTL group, participants in the NPD group acted prosocially, thus deviating from their self-beneficial mode of behavior, this slowed responding to a level comparable to the CTL group.

Regarding the fMRI data, neural activity in regions linked to cognitive control of behavior (Nachev et al., 2008), including SMC and ACC, was found to be higher for the CTL group than the NPD group. A post-hoc exploration of the individual conditions indicated that this was driven by trials in which prosocial and self-beneficial motives were highly conflicting (HC, HCP). Linking these findings with the behavioral data suggests the following interpretation: in the CTL group, decisions in the HC condition slowed down and turned less prosocial, suggesting that an increase in motivational conflict was finally resolved in self-beneficial ways. In the HCP condition, response times also slowed down, thus motivational conflict also seemed to increase, but was resolved in favor of the recipient. Thereby, participants in the CTL group avoided punishing the other person. In the NPD group, this behavioral pattern was roughly similar, though less pronounced, and neural responses to motivational conflict were attenuated. Collectively, this suggests that the requirement to resolve motivational conflict during prosocial behavior engages activity of brain regions more broadly associated with executive control (Botvinick et al., 2004), at least in the CTL group, although the precise ways in which this conflict is resolved depends on the particular context. Concerning NPD, future studies could aim to create experimental contexts that diminish potential floor effects regarding prosocial behavior and associated conflict processing, thus allowing to characterize these dynamics in greater detail.

When it was necessary to punish the recipient for one’s own benefit, activity in right AG increased when tested across groups, which, speculatively, could reflect social-cognitive processes like moral reasoning (Young et al., 2010). Alternatively, these effects could relate to domain-general cognitive processes like attentional reorienting (Seghier, 2013) that may also be engaged during social behavior. Similarly, ITG was activated when punishment was possible, which has previously been associated with fairness considerations (Feng et al., 2015). Here, future studies should attempt to employ experimental designs targeting specific (social-) cognitive processes to better understand their role in NPD and the (avoidance of) punishing others during social decision making. Moreover, it should be highlighted that our data are inconclusive regarding potential group differences regarding these particular effects: while increased activity in AG and ITG was found across groups, which was driven by the CTL group, this was not the case in the NPD group. However, no significant difference between groups was observed. Therefore, further research is required to draw more robust conclusions regarding social-cognitive processes and associated neural activity during social decision making in NPD.

Two potential limitations to our study are the relatively small sample size and the presence of comorbidity in the NPD group. The limited sample size arises in part due to the fact that the prevalence of individuals diagnosed with NPD is relatively low (Coid et al., 2006; Samuels et al., 2002), and therefore, it is difficult to recruit large samples. Yet, to our knowledge, no other study thus far has employed a combination of fMRI and an experimental paradigm to investigate NPD, making our study a relevant contribution to the field. Regarding the presence of comorbidity, individuals will mainly be diagnosed with NPD once they seek psychiatric treatment. This is part of a more general problem in research on NPD: since study participants tend to be contacted via psychiatric clinics, as was the case in the present study, samples tend to overemphasize the part of the narcissism spectrum (Krivan and Herlache, 2018) that is marked by elevated levels of self-esteem vulnerability, whereas individuals with grandiose self-esteem are comparably more difficult to recruit. Relatedly, increased levels of depression in our NPD group, while this was not the case in the CTL group. Yet, despite this difference between groups, we would argue that depression is unlikely to contribute to reduced prosociality and the associated findings on the neural level. This is because reduced prosociality is not a key characteristic of depression. Rather, depression is linked to self-conscious emotions like guilt (American Psychiatric Association, 2013) that are understood to promote prosocial behavior (Tangney et al., 2007; Vaish, 2018). However, to disentangle
effects of depression and NPD on prosocial behavior more clearly, future studies should directly compare samples diagnosed with either NPD or major depression.

5. Conclusion

So far, few studies have investigated the neural correlates of interpersonal behavior in NPD using experimental paradigms in combination with fMRI. By doing so, our data contribute to a more detailed understanding of NPD in terms of impairments in interpersonal functioning, as introduced by DSM-5, section 3 (American Psychiatric Association, 2013). Collectively, our data indicate that self-beneficial behavior in NPD relies on diminished consideration of prosocial motives, which obviates the need to resolve conflict with self-beneficial tendencies. However, this may reflect a reduced propensity, not ability, to act prosocially (Hepper et al., 2014; Melfert et al., 2015). Potentially, this reflects differences between NPD and, for instance, antisocial personality disorder, which could be tested explicitly in future studies. In terms of treatment implications, our results point to possibilities for enhancing prosocial behavior and thereby improving social relationships and well-being in NPD. Regarding social cognition and interpersonal behavior more broadly, our findings support the notion that these functions rely on domain-general control processes (Nachev et al., 2008; Botvinick et al., 2004; Bonini et al., 2014), that draw on contextual social information to decide between generous or self-beneficial behavior (Zaki et al., 2010).

Funding

Research leading to this publication has been funded by the Deutsche Forschungsgemeinschaft (Sören Krach; KR 3803/9–1; Aline Vater and Stefan Roepke; EXC 302; Björn Schott; SFB 779, TP A08).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nicl.2021.102800.

References

American Psychiatric Association (2013): Diagnostic and Statistical Manual of Mental Disorders, 5th ed. Washington, DC: American Psychiatric Association. https://doi.org/10.1176/appi.books.9870890425596.

Aoki, R., Matsumoto, M., Yonogida, Y., Iizima, K., Murayama, K., Sugiyama, A., Camerer, C.F., Adolphs, R., Matsumoto, K., 2014. Social equality in the number of choice options is represented in the ventromedial prefrontal cortex. J. Neurosci. 34 (18), 6413–6421.

Ashburner, J., Barnes, G., Chen, C., Daunizeau, J., Moran, R, Henson, R, et al. https://doi.org/10.1111/j.1365-2964.2006.02813.x.

Asplinger, E.N., Manuck, S.B., Pilkonis, P.A., Simms, L.J., Wright, A.G.C., 2018. Narcissist or narcissistic? Evaluation of the latent structure of narcissistic personality disorder. J. Abnorm. Psychol. 127 (5), 496–502.
Hepper, E.G., Hart, C.M., Sedikides, C., 2014. Moving Narcissus: Can Narcissists Be Empathic? Personal Soc Psychol Bull 40 (9), 1079–1091.

Hutchenron, C., Bushong, B., Rangel, A., 2015. A Neurocomputational Model of Altruistic Choice and Its Implications. Neuron 87 (2), 451–462.

JASP Team (2019): JASP (Version 0.11.1).

Kanske, P., Bockler, A., Trautwein, F.-M., Singer, T., 2015. Dissecting the social brain: Introducing the EmpaTom to reveal distinct neural networks and brain-behavior relations for empathy and Theory of Mind. Neuroimage 122, 6–19.

Koepenik, T., Jauk, E., Kanske, P., 2021. Lay theories of grandiose and vulnerable narcissism. Curr Psychol. https://doi.org/10.1007/s11214-020-01296-w.

Krzan, Z., Herlache, A.D., 2018. The Narcissism Spectrum Model: A Synthetic View of Narcissistic Personality. Personal Soc Psychol Rev 22 (1), 3–31.

Leary, M.R., 2005. Socioemotional theory and the pursuit of relational value: Getting to the root of self-esteem. Eur Rev Soc Psychol 16 (1), 75–111.

Meffert, H., Gazzola, V., Den Boer, J.A., Bartels, A.A.J., Keysers, C., 2013. Reduced spontaneous but relatively normal deliberate vicarious representations in psychopathy. Brain 136, 2550–2562.

Miller, J.D., Campbell, W.K., Pilkonis, P.A., 2007. Narcissistic personality disorder: A mixed blessing? J. Pers. Soc. Psychol. 74 (5), 1197–1206.

Paciello, M., Fida, R., Cerniglia, L., Tramontano, C., Cole, E., 2013. High cost helping: altruistic or selfish? J. Pers. Soc. Psychol. 104 (2), 308–320.

Paulhus, D.L., Williams, K.M., 2002. The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. J. Res. Pers. 36 (6), 556–587.

Roepke, S., 2014. Shame in patients with narcissistic personality disorder. Psychiatry Res. 187 (1-2), 241–247.

Saxe, R., Kanwisher, N., 2003. People thinking about thinking people: The role of the temporo-parietal junction in “theory of mind”. Neuroimage 19 (4), 1835–1842.

Stolz, D.S., Muller-Findler, L., Krach, S., Paulus, F.M., 2020. Internal control beliefs shape positive affect and associated neural dynamics during outcome valuation. Nat. Commun. 11, 1230.

Tangney, J.P., Stuewig, J., Mashek, D.J., 2007. Moral Emotions and Moral Behavior. Annu. Rev. Psychol. 58 (1), 345–372.

The jamovi project (2021). jamovi. (Version 1.8) [Computer Software]. Retrieved from https://www.jamovi.org.