Behavioral/Cognitive

Altruism under Stress: Cortisol Negatively Predicts Charitable Giving and Neural Value Representations Depending on Mentalizing Capacity

Stefan Schulreich,[1] **Anita Tusche,**[2,3] **Philipp Kanske,**[4] **and Lars Schwabe**[1]

[1]Department of Cognitive Psychology, Faculty of Psychology and Human Movement Science, Universität Hamburg, 20146 Hamburg, Germany, [2]Queen's Neuroeconomics Laboratory, Departments of Psychology and Economics, Queen's University, Kingston, Ontario K7L 3N6, Canada, [3]Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, California 91125, and [4]Institute of Clinical Psychology and Behavioral Neuroscience, Faculty of Psychology, Technische Universität Dresden, 01187 Dresden, Germany

Altruism, defined as costly other-regarding behavior, varies considerably across people and contexts. One prominent context in which people frequently must decide on how to socially act is under stress. How does stress affect altruistic decision-making and through which neurocognitive mechanisms? To address these questions, we assessed neural activity associated with charitable giving under stress. Human participants (males and females) completed a charitable donation task before and after they underwent either a psychosocial stressor or a control manipulation, while their brain activity was measured using functional magnetic resonance imaging. As the ability to infer other people's mental states (i.e., mentalizing) predicts prosocial giving and may be susceptible to stress, we examined whether stress effects on altruism depend on participants' general capacity to mentalize, as assessed in an independent task. Although our stress manipulation per se had no influence on charitable giving, increases in the stress hormone cortisol were associated with reductions in donations in participants with high mentalizing capacity, but not in low mentalizers. Multivariate neural response patterns in the right dorsolateral prefrontal cortex (DLPFC) were less predictive of postmanipulation donations in high mentalizers with increased cortisol, indicating decreased value coding, and this effect mediated the (moderated) association between cortisol increases and reduced donations. Our findings provide novel insights into the modulation of altruistic decision-making by suggesting an impact of the stress hormone cortisol on mentalizing-related neurocognitive processes, which in turn results in decreased altruism. The DLPFC appears to play a key role in mediating this cortisol-related shift in altruism.

Key words: altruism; cortisol; decision-making; fMRI; social cognition; stress

Significance Statement

Altruism is a fundamental building block of our society. Emerging evidence indicates a major role of acute stress and stress-related neuromodulators in social behavior and decision-making. How and through which mechanisms stress may impact altruism remains elusive. We observed that the stress hormone cortisol was linked to diminished altruistic behavior. This effect was mediated by reduced value representations in the right dorsolateral prefrontal cortex and critically depended on the individual capacity to infer mental states of others. Our findings provide novel insights into the modulation of human altruism linked to stress-hormone dynamics and into the involved sociocognitive and neural mechanisms, with important implications for future developments of more targeted interventions for stress-related decrements in social behavior and social cognition.

Introduction

Altruism involves other-regarding behavior at a cost to the self. It is a complex phenomenon that emerged in many species (Burkart et al., 2014). Humans often behave altruistically even in the absence of direct future benefits, such as in (anonymous) donations to charitable organizations (Hare et al., 2010; Tusche et al., 2016; Obeso et al., 2018). However, altruism varies across contexts. For instance, initial evidence suggests that stress might interfere with altruism. Acute psychosocial stress (Vinkers et al.,

Received Sep. 15, 2021; revised Feb. 9, 2022; accepted Feb. 9, 2022.

Author contributions: S.S., A.T., P.K., and L.S. designed research; S.S. performed research; S.S. analyzed data; S.S., A.T., P.K., and L.S. wrote the paper.

This work was supported by Universität Hamburg (S.S.) and NIMH Conte Center 2P50 MH094258 (A.T.). We thank Jehona Muslija and Gudrun Grätschus for assistance during data collection, and Carlo Hiller for programming the task.

The authors declare no competing financial interests.

Correspondence should be addressed to Stefan Schulreich at stefan.schulreich@uni-hamburg.de.

https://doi.org/10.1523/JNEUROSCI.1870-21.2022

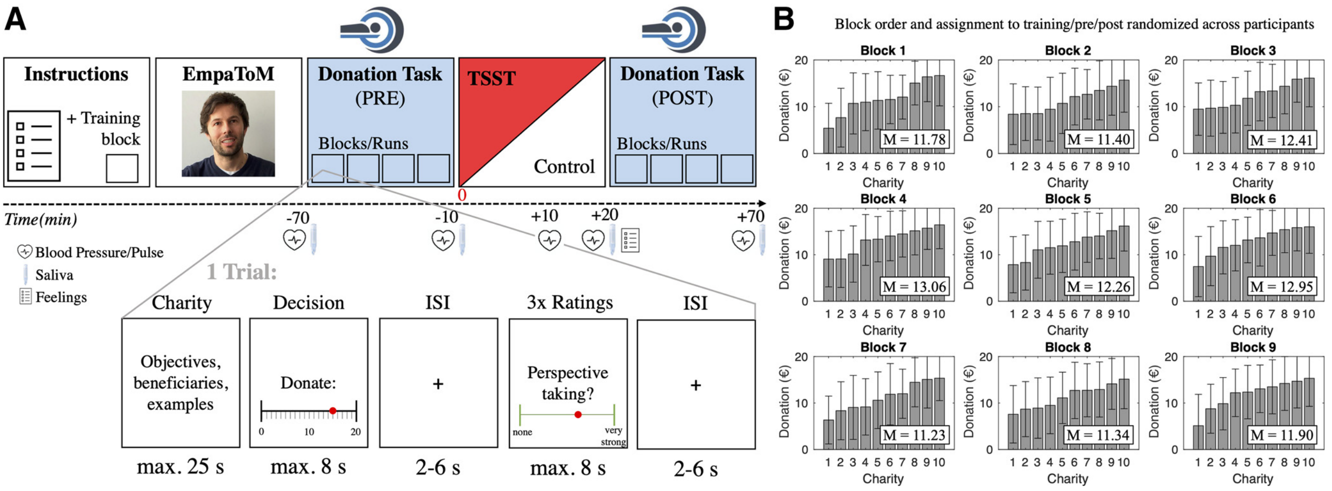

Figure 1. ***A***, Experimental sequence. Before the main fMRI experiment, participants completed one training block of the donation task as well as the EmpaToM as a behavioral baseline measure of mentalizing capacity and socioaffective processes. Afterward, two sessions (PRE and POST) of the donation task were performed in the scanner (within-subject factor), with a stress (TSST) or control manipulation in between (between-subjects factor, implemented outside the scanner). ***B***, Variability of donations across charities within the nine task blocks. Donation blocks were randomly distributed per participant across the prephase and postphase (and the one rating-free block before the EmpaToM and fMRI experiment). Based on pretest data, each block was constructed in a way that charities elicited a broad range of donations in the main experiment (i.e., from low to high). Donations are depicted in ascending order in each block (i.e., the order does not reflect the actual sequence within a block). Mean (M) donations did not differ significantly between blocks ($p = 0.81$; see Materials and Methods). Error bars represent ±1 SD.

2013) and stress-induced increases in the glucocorticoid hormone cortisol (Starcke et al., 2011; but see Singer et al., 2017) have been linked to reduced altruistic decisions. To date, however, the neural mechanisms through which stress or cortisol may alter altruism are largely unknown.

Altruistic behavior is supported by networks of brain regions associated with social cognition and value-based decision-making (Tusche et al., 2016; Bellucci et al., 2020; Tusche and Bas, 2021). This includes prefrontal regions whose functions can be significantly impaired by acute stress and the stress hormone cortisol (for reviews, see Arnsten, 2009; Vogel et al., 2016), such as the dorsolateral prefrontal cortex (DLPFC; Qin et al., 2009; Bogdanov and Schwabe, 2016) and dorsomedial prefrontal cortex (DMPFC; Devilbiss et al., 2017). At the cognitive level, both regions critically support mentalizing [also referred to as Theory-of-Mind (ToM)]—the ability to attribute mental states (e.g., beliefs, desires, intentions) to others. For instance, the DMPFC is an important node in a well described mentalizing network (Schurz et al., 2020), which also includes the temporoparietal junction (TPJ) and middle temporal gyrus (MTG). The DLPFC also critically contributes to mentalizing (Costa et al., 2008; Kalbe et al., 2010) as well as context-dependent cognitive control (Tusche and Hutcherson, 2018).

Mentalizing is a key contributor to altruism (Waytz et al., 2012; Tusche et al., 2016). Hence, a stress-induced impairment in this sociocognitive process might be a plausible mechanism through which stress impairs altruistic behavior. This notion is in line with initial behavioral evidence of stress-induced impairments of mentalizing (Smeets et al., 2009; Leder et al., 2013). Our main hypotheses are that (1) acute stress decreases altruism via altered prefrontal functioning and (2) participants who strongly engage the mentalizing network ("high mentalizers") are particularly prone to these stress-induced decrements. An alternative, more affective route via which stress might influence altruism is empathy (Tomova et al., 2017), which refers to the isomorphic representation of others' affective states (e.g., vicariously feeling others' suffering; Lockwood, 2016), or compassion, which involves caring feelings for others (Weng et al., 2013). Empathy is supported by a network comprising the anterior mid-cingulate cortex (aMCC) and anterior insula (AI; Lamm et al., 2011; Lockwood, 2016), while compassion has been associated with reward-related regions (e.g., striatum; Kanske et al., 2015), and both are contributors to altruism (Weng et al., 2013; Tusche et al., 2016; Tomova et al., 2017).

We also hypothesize that stress-related effects on altruism are mediated through the action of the major stress hormone cortisol. Cortisol can exert its influence via earlier nongenomic actions (approximately <1 h poststressor) or later genomic processes in neurons (Hermans et al., 2014; Joëls et al., 2018). Nongenomic cortisol actions are particularly likely to play a role in altruism, given that cortisol elevations predicted or even statistically mediated decrements in mentalizing performance (Smeets et al., 2009; Leder et al., 2013) and DLPFC functioning (Qin et al., 2009), and given that these observations were made within the first hour following the stressor. Moreover, one study found altered altruistic choice only in an earlier phase, but not in a later phase (Singer et al., 2021; but for no difference see Vinkers et al., 2013). To note, none of those studies investigated stress effects in the very early phase dominated by autonomic stress reactivity, which vanishes within minutes after stressor offset (Nater et al., 2006). Consequently, we believe the phase of nongenomic cortisol action to be particularly relevant for detecting stress-related effects on altruism and associated neural activity.

To test whether stress and cortisol in particular negatively (or positively) affect altruism and via which neurocognitive mechanisms, participants made charitable donation decisions before and after undergoing a standardized psychosocial stress protocol [Trier Social Stress Test (TSST); Kirschbaum et al., 1993] or a control manipulation (Fig. 1*A*). The postphase was within the previously described window of nongenomic action of cortisol (<1 h; Hermans et al., 2014; Joëls et al., 2018), but well after autonomic activity returned to baseline. While participants made donation decisions their brain activity was measured with functional magnetic resonance imaging (fMRI). Participants also completed an independent, well validated task to assess their general tendency to mentalize and empathize (the EmpaToM; Kanske et al., 2015).

Materials and Methods

Participants

A total of 50 right-handed volunteers (24 women, 26 men; mean age ± SD, 23.90 ± 4.03 years) with normal or corrected-to-normal vision participated in this experiment. Before the experiment, we checked exclusion criteria in a standardized phone interview. Following previous studies in our and other laboratories (Bogdanov and Schwabe, 2016; Nitschke et al., 2020), exclusion criteria included current physical or mental conditions, substantial underweight or overweight (body mass index <18.5 or >28.5), medication or drug intake, smoking, a lifetime history of any neurologic or psychiatric disorder, and any MRI contraindications. We also excluded women using hormonal contraceptives because of possible alterations in the stress response (Lovallo et al., 2019) and those in pregnancy or lactation because of ethical reasons (i.e., to avoid any potential adverse effects on mother or child). Participation of female participants was not restricted to a particular phase of their menstrual cycle. The final sample and both groups included women distributed across all phases (i.e., follicular vs luteal). Furthermore, we asked participants to refrain from physical exercise, meals, and caffeine intake within the 2 h before testing.

Six participants were excluded from all analyses for the following reasons: repeatedly exceeding the maximum reading duration in the charity description phase, indicating potentially incomplete task processing (N = 1); clinically relevant depression scores [N = 1; Beck Depression Inventory (BDI) score >30]; self-reported claustrophobic feelings in the MRI session, which might have induced a stressful situation in this control group subject (N = 1); having experienced the TSST before (N = 1); and for being outliers in mentalizing capacity (N = 1, z = −2.87) and self-reported compassion (N = 1, z = −3.47) in the EmpaToM.

Nine participants had to be excluded from the main analyses because of a lack of variability in donations (a prerequisite to examine value coding during altruistic decision-making on the neural level). Four of these participants chose the maximum donation amount in every single trial (i.e., ceiling effect), and five participants chose the maximum amount over at least one block and displayed very low variability in the remaining blocks.

The final sample consisted of 35 participants (15 women, 20 men; mean age, 23.49 ± 4.14 years). To confirm that the sample size is sufficient to detect the effects of interest, we implemented an a priori power analysis using G*Power 3.1 (Faul et al., 2009). Given the mixed design with repeated measures, our final sample allowed for the detection of a small-to-medium effect of the stress manipulation on donations and neural responses from the pretreatment to the posttreatment session, that is a group × time (between–within) interaction with Cohen's f = 0.18, with a statistical power of 90% and α at p = 0.05, assuming a correlation among repeated measures of 0.8, and a nonsphericity correction of $\in$= 1. Wherever possible (i.e., when choice variability is not a prerequisite), we complemented our main analyses with analyses on the larger sample (N = 44) that included those nine subjects with invariant decisions to check for the robustness of our results and observed here largely comparable results.

All participants gave written informed consent before the experiment and received a compensation of €30 plus a possible bonus in the donation task (see below). The study protocol was in line with the Declaration of Helsinki and approved by the ethics committee of the Faculty of Psychology and Human Movement Science at the Universität Hamburg.

Experimental design

All experimental sessions took place between 8:00 A.M. and 1:00 P.M. to mitigate the influence of the diurnal rhythm of cortisol (Edwards et al., 2001). The allocation of the two experimental conditions (stress vs control group) to specific slots within this time window was randomized. Participants completed a series of tasks and measures in the laboratory (Fig. 1*A*). To obtain a baseline measure of participants' mentalizing capacity, we administered the EmpaToM task (Kanske et al., 2015; see below for more details). To assess our primary measures of interest, altruistic behavior and the associated neural responses, we used a charitable donation task (adapted from Böckler et al., 2016; Tusche et al., 2016; see below for more details), while simultaneously collecting fMRI data. To test the impact of stress on charitable giving and its underlying neural processes, we used a mixed design that assessed donations and neural activity before and after (within-subject factor time) a standardized stress or control manipulation (TSST; Kirschbaum et al., 1993; the between-subjects factor group). This manipulation was accompanied by a series of stress parameter measurements (see below). We randomly assigned participants to the stress or control group, with the only constraint of relatively balanced gender distribution. The final sample (N = 35) was approximately equally distributed across the stress group [N = 18 (7 women, 11 men); mean age, 23 ± 3.71 years] and the control group (N = 17 [8 women, 9 men]; mean age, 24 ± 4.61 years). At the end of the experimental session, participants received their remuneration for participation and were fully debriefed. The whole session lasted for ∼150–180 min.

Stress manipulation

In the stress condition, participants completed the TSST (Kirschbaum et al., 1993; Fig. 1*A*). The TSST is a standardized, well validated laboratory task to elicit subjective stress, a sympathetic stress response, and glucocorticoid secretion via the hypothalamo–pituitary–adrenal axis (Kirschbaum et al., 1993; Kudielka et al., 2007). Participants first anticipated and prepared for a mock job interview (3 min). They could take notes but could not use them in the following free speech (5 min), in which they explained why they are the ideal candidate for the job. The free speech was followed by a demanding arithmetic task of counting backward in steps of 17 from the number 2043 as fast as possible (5 min). During both tasks, participants were videotaped and stood in front of a panel of two rather cold and unresponsive experimenters (1 male, 1 female, different from the experimenters that performed the rest of the experimental procedure), creating a social-evaluative context. In the control condition, participants gave a 5 min speech about a topic of their choice (e.g., last holiday) and performed a much simpler 5 min arithmetic task (i.e., counting forward in steps of 5 starting from 0), following previous applications (Bogdanov and Schwabe, 2016; Vogel et al., 2018). During the control condition, participants were neither videotaped nor monitored by a panel.

To assess whether the stress manipulation was successful and to determine the individual stress reactivity, we measured subjective and physiological stress parameters at several time points across the experiment. At the subjective level, participants rated the perceived stressfulness, difficulty, and unpleasantness of the TSST or control procedure on a scale from 0 ("not at all") to 100 ("very much") immediately after the procedure. As indicators of sympathetic nervous system activity, blood pressure and pulse were measured using an upper arm monitor (Omron) at several time points: before (approximately −70 min relative to TSST onset) and after the first fMRI session (−10 min), during the stress/control manipulation (+8 min), and before (+20 min) and after the second fMRI session (approximately +70 min). Blood pressure and pulse measurement was repeated twice and then averaged at a given time point to increase reliability. To quantify cortisol concentrations, saliva samples were collected from participants at several time points before and after the stress/control manipulation (approximately −70, −10, +20, and +70 min relative to TSST onset) using Salivette collection devices (Sarstedt). Saliva samples were stored at −18°C and analyzed for cortisol concentrations using a luminescence assay (IBL International). As an integrated measure of the cortisol response to the stress manipulation, we calculated the area under the curve with respect to the increase (AUCi) from before (−10 min) to +70 min after the onset of the TSST/control procedure (Pruessner et al., 2003).

One participant in the control group provided only three of the four cortisol values. In line with previous recommendations (Tabachnik and Fidell, 2013), we imputed the single missing value in the following ways. First, we performed a multiple regression that predicts the respective data point from the cortisol values of the other time points for subjects in the control group (excluding the participant in question). Second, we used the regression coefficients to estimate the missing cortisol value in the respective participant. Our main behavioral and multivariate pattern analysis (MVPA) findings remained significant when we excluded this participant for a robustness check.

Donation task

Altruistic behavior was measured using a charitable donation task (adapted from Böckler et al., 2016; Tusche et al., 2016) before and after the stress/control manipulation while simultaneously collecting fMRI data (Fig. 1*A*). The postsession took place from ~30 to 60 min following treatment onset, overlapping with the phase of the stress response mainly characterized by nongenomic cortisol action (Hermans et al., 2014; Joëls et al., 2018). The presession and postsession consisted of 40 trials each (80 trials in total), arranged in four functional runs (blocks) of 10 trials.

In each trial, participants were first presented with a short description of a real-world charitable organization [reading phase; terminated by a button press with a maximum of up to 25 s; for the complete set of charity descriptions (in German); see Open Science Framework (OSF) project page: https://osf.io/u46yj/]. Next, participants had to decide how much to donate to the charity (range, €0 to €20 in steps of €1; decision phase, up to 8 s). Participants responded using a slider from a randomized starting position. After a variable interstimulus interval (ISI) from 2 to 6 s, three rating questions were presented in a randomized order. Participants rated their experienced (1) empathy ("Felt with others?," in the sense of sharing an affective state), (2) compassion ("Compassion for others?," in the sense of warm, tender feelings toward others), and (3) perspective taking ["Took the perspective of others?" (i.e., of the beneficiaries of the charity); rating phase; up to 8 s per rating; for complete instructions, see https://osf.io/u46yj/]. Participants responded using a slider on a 9 point scale (ranging from "not at all" to "very strong"). Trials were separated by another variable ISI (2–6 s). Across all blocks, 90 different charities were presented. Together, the presession and postsession consisted of a total of 80 trials. Participants completed one block of 10 trials outside the scanner before the main experiment. This block did not contain rating questions and served both as a training block and as a control for the potential influence of the rating task on donation behavior.

Each participant was presented with each of the 90 charities once. The assignment of charities to a particular block was fixed and based on donations observed in a behavioral pretest using an independent-subject sample [*N* = 27 (20 women, 7 men); mean age, 25.25 ± 6.15 years). Charities were selected and grouped such that average donations were comparable across task blocks (pretest: $F_{(8,81)} = 0.029$, $p = 0.99$, η_p^2 = 0.003; main experiment: $F_{(8,81)} = 0.555$, $p = 0.81$, η_p^2 = 0.052) and to ensure coverage of a broad range of giving behavior across the 10 charities in each block. The latter was crucial as sufficient variance is a prerequisite for the multivariate pattern analysis described below. Figure 1*B* illustrates the variability of charity-wise donations within the task blocks of our main experiment. The order of charities within a task block and the order of the nine task blocks were randomized across participants.

Before the task, participants were informed that one donation trial would be randomly selected at the end of the experiment and implemented. This procedure ensured that participants treated each trial independently (instead of dividing their endowment among different charities, which would reflect a portfolio effect). The charity would receive the total amount donated in that trial, and participants could keep 25% of the amount not donated. For instance, if a participant donated €12 of their €20 endowment, then €12 was transferred to the charity after the experiment, and €2 [25% of the amount not donated (€8)] was added to the participant's remuneration. Thus, choices in the donation task were costly (as they reduce personal gains) and had real consequences, which ensures that donations are consistent with participants' actual preferences. A partial (instead of full) payout of the nondonated amount was implemented to not override other-regarding preferences and to provide a moderate donation incentive (Tusche et al., 2016).

EmpaToM task

To assess participants' mentalizing capacity in complex social settings, we administered the well established EmpaToM task outside of the scanner (Kanske et al., 2015; Tholen et al., 2020; Hildebrandt et al., 2021). This behavioral task was performed before the donation task and the stress/control manipulation (Fig. 1*A*), providing an independent baseline (i.e., premanipulation) measure of individual differences in participants' mentalizing capacity. The task simultaneously assesses socioaffective responses (empathy, compassion) in social settings, which allows us to examine the specificity of mentalizing-related effects on charitable giving.

We used a brief version of the EmpaToM consisting of 24 trials. Each trial started with a fixation cross (1–3 s), after which the name of a person (1 s) appeared, followed by a short video recounting an autobiographical episode (~15 s). The videos differed in emotionality (neutral vs negative contents) and in whether their content is mentalizing related (e.g., beliefs, deception) or not, later giving rise to mentalizing-related versus factual questions, respectively (yielding a 2 × 2 factorial design). The videos showed six actors (three females, three males), each of whom recounted one story per condition (6 actors × 4 conditions = 24 trials). After each video, participants rated their empathic affective response ("How do you feel?"; from "very negative" to "very positive" on a scale from −3 to 3) and their compassion for the person in the video ("How much compassion do you feel?"; from "none" to "very much" on a scale from 0 to 6; 4 s per rating, fixed order). Participants responded by moving a slider. A multiple-choice question with three response options was presented after a variable delay of 1–3 s. The question either demanded mentalizing [e.g., "Anna thinks that (...)" (12 trials)] or factual reasoning (e.g., "It is correct that (...)" (12 trials)] on the contents of the previous video. Participants responded by pressing one of three buttons assigned to the three choice options (up to 15 s). The percentage of correct responses (accuracy) in the mentalizing-related questions served as our measure of mentalizing capacity. For a detailed description of the task validation, example stories, and questions for each experimental condition, see the study by Kanske et al. (2015).

Control measures

Before the experiment, participants completed an online survey at home [implemented via the SoSciSurvey platform (https://www.soscisurvey.de)], which included demographic questions, the BDI (Hautzinger et al., 2006), and the Trier Inventory of Chronic Stress (Schulz and Schlotz, 1999). These measures ensured that experimental groups (stress vs control) were matched in terms of age, depression scores, and chronic stress after randomization (all *p* values > 0.313).

Behavioral data analysis

We will focus in this section on our primary analyses of stress parameters and choice data. Supplemental analyses will be described in the course of the Results section.

Stress reactivity. As a manipulation check, we first examined the effectiveness of the experimental stress manipulation in terms of subjective feelings, sympathetic and glucocorticoid (cortisol) reactivity. At the subjective level, we analyzed whether the stress and control groups differ in their self-reported ratings of stressfulness, unpleasantness, and difficulty immediately after the TSST procedure, using two-sample *t* tests. Physiologic parameters (i.e., systolic and diastolic blood pressure, pulse, and salivary cortisol levels) were subjected to a general linear model (GLM) with the within-subject factor time (denoting time points of measurement across the experiment) and the between-subjects factor group (stress vs control condition). A differential response to the experimental manipulation would be reflected by a significant group × time interaction effect. To decompose this interaction and to assess at which time points groups differ from each other, we used *post hoc* pairwise comparisons.

Impact of acute stress on altruism. To investigate the effect of our stress manipulation on altruistic choice and whether this effect depends on baseline mentalizing capacity, we fitted a generalized linear mixed model (GLMM; choice – full model 1) with donations as the dependent variable and the following predictors: time (pre vs post) as a repeated-measures factor; group (stress vs control) as a between-subjects factor; and mentalizing capacity as captured in the EmpaToM as a covariate. In addition to main effects, we also modeled all two-way and three-way interactions and the intercept. The time and group factor were effect coded (i.e., using weights of −1 and +1), and the covariate was mean centered so that the resulting main effects (and intercept) truly reflect average effects (and not effects for a single zero-coded category/for the covariate at zero). The model was estimated with a robust covariance matrix estimator.

Cortisol-related effects on altruism. Given that both altruism (Starcke et al., 2011) and mentalizing (Smeets et al., 2009; Leder et al., 2013) have been found to depend on stress hormone dynamics, we fitted another GLMM (choice – full model 2) to assess whether changes in cortisol (Δcortisol; captured in the AUCi), mentalizing capacity (EmpaToM), or an interaction of Δcortisol × mentalizing predicted changes in donations. The GLMM modeled time (pre vs post) as a within-subject factor and used identical estimation procedures as the previous GLMM. We fitted this model on choice data across groups because cortisol dynamics varied strongly across participants of both groups. Nevertheless, we also formally compared this model with a more complex model including group as an additional factor (and another model including gender) in terms of model fit, assessed via the Bayesian information criterion (BIC) and the Akaike information criterion (AIC), and found support for the simpler model without these factors (see Results). Significant interaction effects were decomposed using appropriate follow-up models. Specifically, a three-way Δcortisol × mentalizing × time interaction was decomposed by follow-up generalized linear models with Δcortisol, mentalizing, and their interaction as predictors, fitted for each phase (pre vs post) separately (Decomposition PRE and POST) and change scores (Decomposition POST-PRE). This *post hoc* decomposition is essential to test whether pre-to-post cortisol dynamics related to the manipulation rather than pre-existing differences drive the interaction effect in our full model. The emerging two-way interactions between the continuous predictors Δcortisol × mentalizing in the postphase were decomposed using simple-slopes analysis (Preacher et al., 2006). This analysis assesses the relationship between a predictor (Δcortisol) and the dependent variable (donations) at different levels of the other predictor (±1 SD in mentalizing capacity). Again, the *post hoc* decomposition of interactions is essential for interpreting the source and direction of an effect (e.g., whether specific levels of predictors drive effects). For comparison, the simple slopes are also reported for nonsignificant two-way interactions (i.e., in the prephase of the donation task) and for pre–post change scores. Please note that all simple-slopes analyses are (second-order) decompositions of a significant three-way interaction.

Behavioral data were analyzed using MATLAB R2019a (MathWorks) and SPSS 25 (IBM). The significance level was set at $p \leq 0.05$. All reported *p*-values are two tailed, if not explicitly indicated otherwise. In the case of violations of sphericity, Greenhouse–Geisser correction was applied.

MRI acquisition and preprocessing

Functional imaging was conducted using a 3 T Magnetom Prisma MRI scanner (Siemens), equipped with a 64-channel head coil. We acquired gradient-echo T_2^*-weighted echoplanar images (EPIs). For each of the eight functional runs of the donation task (four premanipulation, four postmanipulation), we collected a series of volumes using a slice thickness of 2 mm and an isotropic voxel size of 2 mm^2, with 60 slices aligned to the anterior commissure–posterior commissure line and acquired in descending order, repetition time (TR) = 2000 ms, echo time (TE) = 30 ms, flip angle = 60%, and field of view = 224 × 224. After the four functional runs in each session, we obtained a static field map for off-line image distortion correction of the EPI scans. After the donation task, an additional magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence was used to acquire high-resolution (0.8 × 0.8 × 0.9 mm) T_1-weighted structural images for each participant (TR = 2.5 s, TE = 2.12 ms, 256 slices).

Preprocessing of functional images was performed using SPM12 (http://www.fil.ion.ucl.ac.uk/spm/) implemented in MATLAB (MathWorks). For each run, the first five functional images were discarded from the analysis to avoid T_1 saturation effects. The remaining functional images were spatially realigned and distortion corrected using the field map, slice time corrected, coregistered to the structural image, followed by spatial normalization to the Montreal Neurological Institute (MNI) stereotaxic standard space. The resulting (unsmoothed) images were used as inputs to our multivariate decoding analysis (the decoding maps were later smoothed for a whole-brain analysis, see below). Only for the complementary univariate analysis, preprocessing also included spatial smoothing using an 8 mm full-width at half-maximum (FWHM) Gaussian kernel.

fMRI analysis

For each subject and session (premanipulation vs postmanipulation), we estimated two GLMs of the neural responses. Task-related regressors of both GLMs were modeled as boxcar functions with a duration of the associated trial phase (e.g., decision phase) and convolved with a canonical hemodynamic response function. We applied a 128 s high-pass cutoff filter to eliminate low-frequency drifts in the data. GLM1 served to generate the inputs for our multivariate analysis, whereas GLM2 was part of our complementary univariate analysis. We will start by describing GLM1 and our decoding analysis in detail, after which we will continue with the univariate analysis.

GLM1_{pre} and $\text{GLM1}_{\text{post}}$ were used to obtain trial-wise measures of blood oxygenation level-dependent responses during the donation task (separately for the prephase and postphase). In line with a previous fMRI implementation of the task (Tusche et al., 2016), the models included a regressor for each of the 40 decision phases per session (R1–R40 for the 40 decisions), and the associated hemodynamic response estimates served as inputs (i.e., predictor variables) for our primary multivariate analysis. Furthermore, two additional regressors modeled the reading phases (R41) and the rating phases (R42), and six motion regressors accounted for residual motion-related signal changes (R43–R48).

MVPA

Neural decoding of donations. Our multivariate decoding analysis aimed to identify brain regions that encode trial-by-trial variations in donations in their multivoxel response patterns. Donations served here as an indicator of the value people place on specific charities. In a first step, we aimed to detect brain areas that allow decoding individuals' trial-wise donations before the stress manipulation [baseline/prephase (GLM1_{pre})]. Next, we examined whether the predictive information in these brain areas varies as a function of participants' mentalizing capacity. These two steps served two important functions. First, we tested whether we could replicate the previously observed neural decoding of donations (Tusche et al., 2016). Second, value coding associated with mentalizing capacity was hypothesized to be subject to cortisol-related alterations. In other words, the detected areas served as regions of interest (ROIs) to test for stress- and cortisol-related changes in value coding.

To this end, we applied a whole-brain searchlight decoding approach. This approach does not depend on a priori assumptions about informative brain regions and ensures unbiased information mapping throughout the whole brain (Kriegeskorte et al., 2006; Haynes et al., 2007). For each participant, we defined a sphere (radius = 5 voxels) around a given voxel, v_i, of the acquired brain volume (Libby et al., 2014; Solanas et al., 2020). For each of the *N* voxels within this sphere, we extracted trial-wise parameter estimates of GLM1_{pre} for the neural responses during the decision phases (R1–R40) of the prephase donations (Fig. 1*A*). Extracted activation patterns were transformed into *N*-dimensional pattern vectors. This was done for each of the four runs (10 trials per run) separately. Pattern vectors of all runs but one ("training data") were used to train a support vector regression (SVR) model, as implemented in LIBSVM (http://www.csie.ntu.edu.tw/∼cjlin/libsvm; Chang and Lin, 2011) using a linear kernel (ν-SVR) and a fixed regularization parameter ($c = 1$). This provided the basis of the following prediction of the donation amounts of the remaining run ("test data") based on their neural response patterns. The procedure was repeated four times, always using a different run as a test dataset (leave-one-run-out cross-validation). Splitting the dataset into training and test datasets and run-wise cross-validation is a measure to control for potential problems of overfitting (Poldrack et al., 2020). The amount of predictive

Table 1. Subjective and physiological stress parameters at different time points (in minutes relative to stress manipulation onset)

	Stress condition		Control condition	
	Mean	SD	Mean	SD
Subjective feelings (+5)				
Stressfulness	62.94*	24.94	30.00	28.50
Unpleasantness	74.12*	25.26	34.12	32.61
Difficulty	72.35*	15.62	30.00	23.45
Systolic blood pressure (mmHg)				
Pre-1st fMRI (−70)	115.27	9.60	116.73	14.93
Post-1st fMRI (−10)	116.23	13.14	114.17	14.07
TSST/control (+8)	135.70*	14.05	114.87	11.15
Pre-2nd fMRI (+20)	122.27	12.06	115.67	11.66
Post-2nd fMRI (+70)	119.67	13.05	121.63	14.75
Diastolic blood pressure (mmHg)				
Pre-1st fMRI (−70)	78.57	9.16	78.00	6.18
Post-1st fMRI (−10)	80.47	10.44	77.67	6.99
TSST/control (+8)	97.80*	7.87	79.10	5.17
Pre-2nd fMRI (+20)	85.73	10.04	79.60	8.07
Post-2nd fMRI (+70)	81.50	9.71	81.33	9.96
Pulse (bpm)				
Pre-1st fMRI (−70)	81.50	12.19	74.97	10.15
Post-1st fMRI (−10)	76.80	13.34	73.07	9.11
TSST/control (+8)	96.67*	16.32	72.97	10.22
Pre-2nd fMRI (+20)	79.17	14.50	74.57	9.38
Post-2nd fMRI (+70)	81.00	11.23	75.33	9.98

*Significant group difference at $p < 0.001$.

information on generosity was defined as the average Fisher's z-transformed correlation coefficient between the donations predicted by the SVR model and participant's actual donations in these trials (Kahnt et al., 2014; Tusche et al., 2016). This predictive accuracy value was then assigned to the central voxel of the searchlight cluster, and the procedure was repeated for every voxel of the acquired brain volume, resulting in a 3D map of average predictive accuracies for each participant.

These decoding maps were smoothed with a Gaussian kernel (8 mm FWHM) and submitted to a random-effects group analysis to identify brain regions that encode trial-wise donations across participants (simple t test against baseline as implemented in SPM12). For this whole-brain analysis, we applied a cluster-forming threshold of $p \leq 0.001$, familywise error (FWE) corrected for multiple comparisons at the cluster level ($p_{\text{fwe}} \leq 0.05$). These analysis steps were then repeated for postmanipulation donation blocks (GLM1$_{\text{post}}$) to examine the impact of stress on neural value representations (see below).

Mentalizing and neural value representations. Having identified activation patterns that decode donations, we proceeded to the next question. Does predictive information in (some of) these areas vary for people with high and low mentalizing capacity? To address this question, we identified high and low mentalizers based on the accuracy in mentalizing-related questions in the independent EmpaToM task (median split). Note that this task does not require prosocial decision-making and assesses the general capacity to mentalize. Next, we tested for a difference in predictive neural information on donations between high mentalizers and low mentalizers (MENT$_{\text{high}}$ > MENT$_{\text{low}}$) using a two-sample t test. We restricted this test to brain areas previously shown to robustly predict donations on the group level (whole-brain decoding of donations; see above). The statistical test of decoding maps (of donations) for high versus low mentalizers was performed at a more lenient statistical threshold of $p \leq 0.005$ (and peak $p \leq 0.001$) and an extent threshold of $k \geq 40$ voxels. Note that this analysis is designed to identify ROIs for the subsequent test of stress on donations on the neural level. Thus, we opted against using more stringent statistical thresholds and corrections for multiple test comparisons that were used for the rest of our main analyses.

Impact of stress on the mentalizing–valuation relationship. To understand the impact of stress on this interplay between mentalizing capacity and neural value coding, we examined the effect of acute stress on predictive neural information identified above. Specifically, we created spherical ROIs (5 mm radius) around the peaks of predictive information that varied as a function of mentalizing capacity before the stress manipulation (Table 2, MENT$_{\text{high}}$ > MENT$_{\text{low}}$ contrast). These brain regions were identified regardless of stress effects on donations. Thus, the resulting ROIs are fully independent, mitigating the risk of circular analysis and double-dipping (Kriegeskorte et al., 2009). ROI masks are provided at https://osf.io/u46yj/. For each spherical ROI, we estimated the average decoding accuracy for the prephase donations (decoding based on GLM1$_{\text{pre}}$) and for postphase donations (decoding based on GLM1$_{\text{post}}$), respectively, which served as dependent variables in our statistical models on stress- and cortisol-related effects. Given an observed cortisol-related effect on donations (see Results), we used a GLMM (MVPA – full model) to predict decoding accuracies on donations for each of our four ROIs (i.e., as separate GLMMs with identical predictors), whose predictors—Δcortisol, mentalizing capacity, time, and their interactions—perfectly matched those of the behavioral GLMM (choice – full model 2). We also used the same estimation procedures and an identical follow-up decomposition of interaction effects. As an exploratory analysis, we fitted group-based GLMMs to examine potential group differences across time.

Finally, we complemented this ROI-based approach with an exploratory whole-brain approach by testing for cortisol-related or group-related changes in the decoding maps of the premanipulation versus postmanipulation phase. To this end, we ran another whole-brain searchlight analysis on the postsession (GLM1$_{\text{post}}$), identical to our initial analysis for the pre-session (GLM1$_{\text{pre}}$), and created difference maps (post – pre) that served as dependent variables to GLMs with the (demeaned) predictors Δcortisol (or group), mentalizing capacity and their interaction.

Moderated mediation analysis. In a final step, we also examined whether observed cortisol-related changes in neural value coding mediated the observed cortisol-related changes in altruistic choice in high (but not low) mentalizers (i.e., moderated mediation), thereby providing a direct brain–behavioral link. Specifically, we used the PROCESS toolbox version 3.4.1. (Hayes, 2018) to set up a model ("Model 7" within the toolbox) that tests whether the cortisol–donation association can be explained via changes in right DLPFC (rDLPFC) activity patterns. One outlier subject had to be excluded from this analysis (Cook's distance = 0.5; z = 2.63).

Univariate fMRI analysis

We complemented our main multivariate analyses with a univariate analysis by estimating two further GLMs (one for each session) based on smoothed data (8 mm FWHM). GLM2$_{\text{pre}}$ and GLM2$_{\text{post}}$ included a regressor denoting the decision phases per session (R1) and a parametric regressor denoting donation amounts (R2). Furthermore, two additional regressors modeled the reading phases (R3) and the rating phases (R4) as regressors of no interest. Six movement parameters were again included as nuisance regressors (R5–R10).

We extracted the following two kinds of parameter estimates: (1) of R1 to assess (differences in) average decision-related brain activity; and (2) of R2 to assess (differences in) brain activity that linearly predicted donation amounts. Similar to our multivariate analysis, we performed an ROI-based analysis on the extracted decision-related or donation-encoding parameter estimates (average over all voxels within the 5-mm-radius spheres). Likewise, these estimates served as dependent variables in matching GLMMs with Δcortisol, mentalizing capacity, time, and their interactions as predictors. We also fitted complementary group-based GLMMs and performed identical whole-brain analyses on difference maps (post minus pre).

Data availability

Behavioral and (aggregated) fMRI data, group-level decoding maps, ROI masks, MATLAB scripts for the SVR decoding analysis (whole-brain searchlight and ROI-based approach), and task material (instructions, charity descriptions) are publicly available on the OSF page of the project (https://osf.io/u46yj/). The SVR decoding analysis was implemented

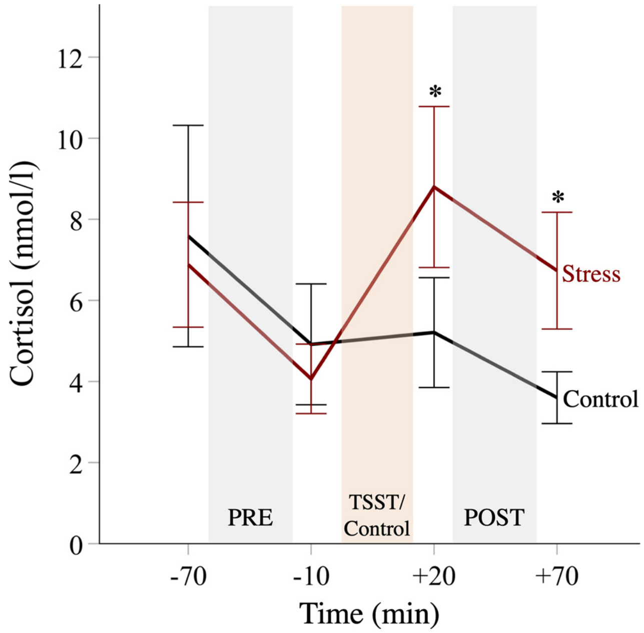

Figure 2. Time courses of salivary cortisol levels. Following the TSST, the stress group displayed elevated cortisol levels relative to the control group. Error bars represent 95% CIs.

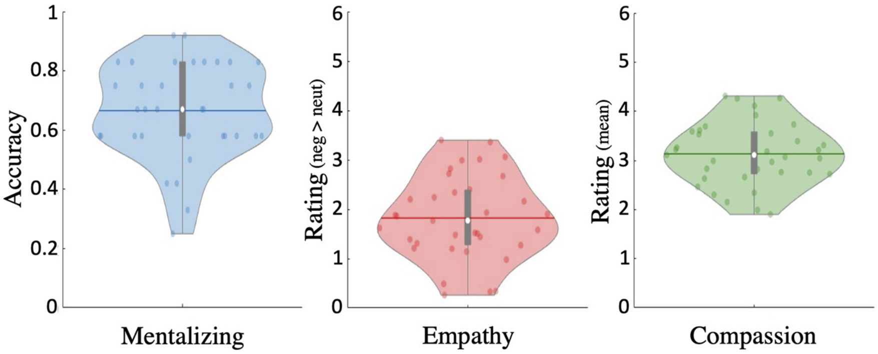

Figure 3. Violin distribution plots of EmpaToM scores. Mentalizing capacity was assessed as the rate of correct responses (accuracy) in mentalizing-related questions in the EmpaToM task (Kanske et al., 2015). Empathy was assessed with valence ratings of participants' current affective state after watching the videos and by creating a difference score for (negative) valence following negative > neutral videos. More positive scores reflect more negative affect and thus more empathy after watching negative relative to neutral videos. Compassion was assessed with ratings of felt compassion (mean across all videos). Horizontal colored lines, mean scores across subjects; white data point, median; thick gray vertical lines, boxplots.

using the free LIBSVM toolbox (http://www.csie.ntu.edu.tw/~cjlin/libsvm; Chang and Lin, 2011) and MATLAB R2019a (MathWorks). Further material will be available from the corresponding author on request.

Results

Manipulation check I: subjective and physiological stress responses

As a manipulation check, we first examined the effectiveness of the experimental stress manipulation in terms of subjective feelings, and sympathetic and glucocorticoid (cortisol) reactivity. At the subjective level, the TSST was experienced as significantly more stressful, unpleasant, and difficult than the control condition (all p values < 0.001; Table 1).

At the psychophysiological level, the TSST induced strong sympathetic arousal, as indicated by a significant increase in systolic and diastolic blood pressure as well as pulse compared with the control condition (group × time interaction for systolic blood pressure: $F_{(2.902,81.259)} = 17.061$, $p < 0.001$, $\eta_p{}^2 = 0.379$; for diastolic blood pressure: $F_{(4,112)} = 20.391$, $p < 0.001$, $\eta_p{}^2 = 0.421$; for pulse: $F_{(2.816,78.859)} = 14.689$, $p < 0.001$, $\eta_p{}^2 = 0.344$). Pairwise *post hoc* comparisons revealed that blood pressure and pulse were significantly elevated during the TSST relative to the control condition (all p values < 0.001), but not at other time points of measurement (all p values ≥ 0.08; in particular not in the premanipulation measurements, all p values ≥ 0.47; Table 1), as expected for a transient sympathetic activation.

Furthermore, while there was a significant decrease in salivary cortisol in the control group across the experimental session because of the circadian rhythm of cortisol (time: $F_{(2.239,73.876)} = 13.275$, $p < 0.001$, $\eta_p{}^2 = 0.29$), cortisol was significantly increased after the TSST, relative to the control condition (group × time: $F_{(2.239,73.876)} = 10.424$, $p < 0.001$, $\eta_p{}^2 = 0.24$). This increase peaked at the +20 min measurement right before the postsession of the donation task (Fig. 2). *Post hoc* comparisons revealed significant group differences right before ($t_{(33)} = 3.117$, $p = 0.004$) and after the postsession ($t_{(33)} = 4.2$, $p < 0.001$), indicating higher glucocorticoid activity in the stress group than in the control group throughout the whole donation task in the postsession. In contrast, the stress and control groups did not differ in cortisol levels at both time points before the experimental manipulation (both p values > 0.3). Moreover, the degree of cortisol reactivity to the manipulation, as assessed via the AUCi (Pruessner et al., 2003), did not significantly depend on baseline cortisol levels ($r_{(33)} = -0.16$, $p = 0.36$) and is not related to mentalizing capacity ($r_{(33)} = -0.03$, $p = 0.88$). The cortisol responder rate [i.e., percentage of participants with a cortisol increase of >2 nmol/l from the premanipulation baseline (−10 min) to peak; Schwabe et al., 2008] was 88.9% in the stress group, and this rate was larger than in the control group (23.5%; $\chi(1) = 15.251$, $p < 0.001$). As also indicated in Figure 2, there was considerable interindividual variability in (base-to-peak) cortisol reactivity in both groups (stress group: minimum to maximum, −1.26 to 11.48 nmol/L; range, 12.74 nmol/L; control group: minimum to maximum, −3.23 to 3.61 nmol/L; range, 6.84 nmol/L).

Manipulation check II: variability in donation behavior

As another crucial manipulation check, we assessed the subject-wise and charity-wise variability in donations. The latter was a desired consequence of the construction of the donation task and was subsequently exploited in subject-wise regressions of trial-by-trial donations in our multivariate decoding analysis. Even before any experimental stress manipulation (i.e., in the prephase), we observed substantial variability in average donations across participants (minimum, €2.80; maximum, €17.45; mean donation ± SD, €12.07 ± €3.63) as well as in contributions to different charities within individuals (mean of participant SDs, €4.91; range, €1.65 to €8.68). Moreover, there was a substantial variation in donations to specific charities within task blocks and across subjects (Fig. 1*B*), making it unlikely that the

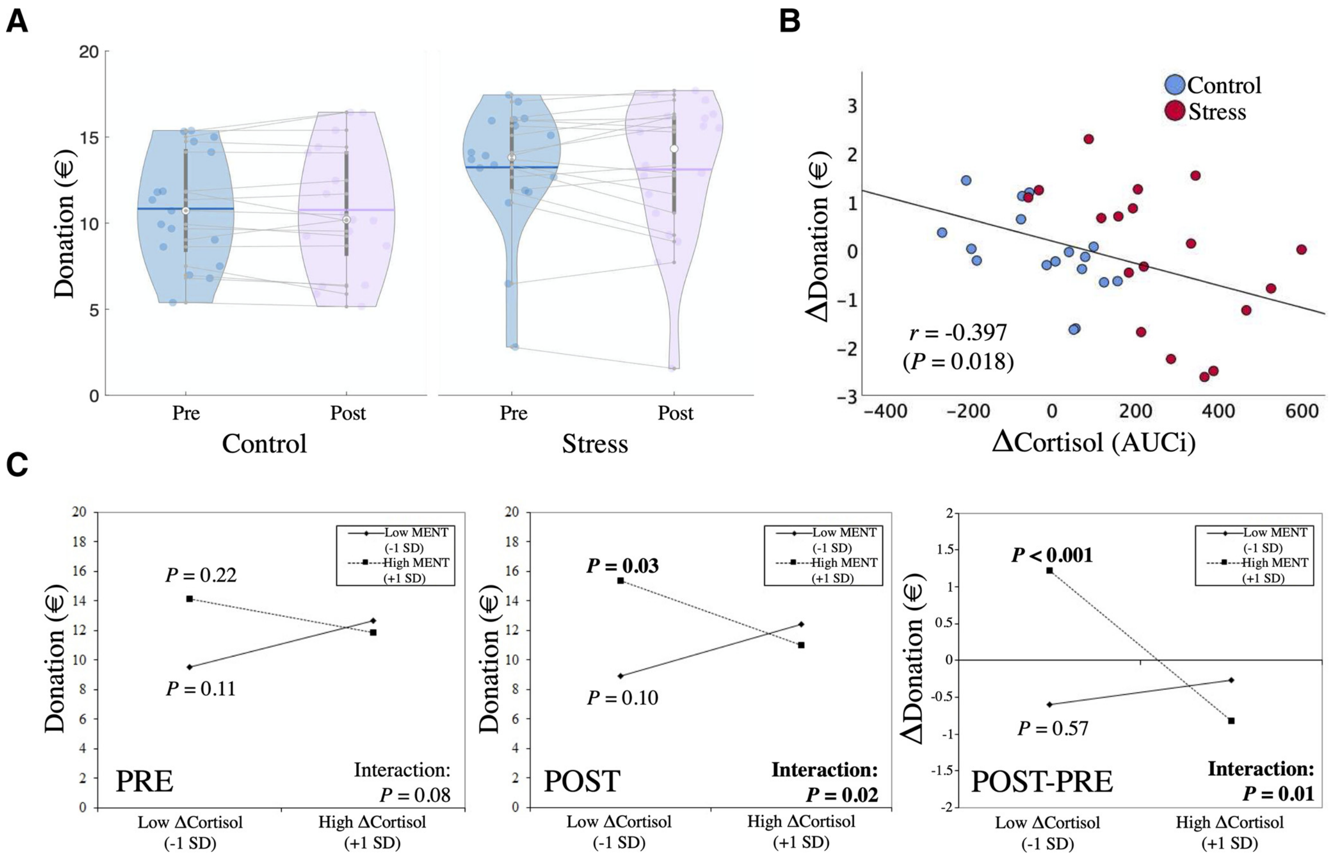

Figure 4. ***A***, Violin distribution plots of mean donations across groups and sessions (horizontal colored lines, mean donations across subjects; white data point, median; thick gray vertical lines, box-plots). ***B***, Increases in cortisol were associated with decreases in charitable giving across participants and groups. This negative association could also be observed in both groups separately (stress group: $r = 0.51$, $p = 0.031$; control group: $r = 0.56$, $p = 0.021$). ***C***, ΔCortisol × mentalizing interaction plots of the simple slopes at +1 SD above and –1 SD below the mean of the moderator (mentalizing capacity) and Δcortisol for donations in the prephase, postphase, and the change over time (POST – PRE). Only for postphase donations, we observed a significant negative association between cortisol changes and donations for high mentalizers, but not for low mentalizers. This moderated negative association was also significantly stronger compared with the prephase (as indicated by the significant Δcortisol × mentalizing × time interaction in the GLMM and in the simple-slopes analysis that directly tests for POST-PRE changes).

multivariate neural decoding of donations was driven merely by unique properties of particular charity stimuli.

Mentalizing capacity predicts charitable giving

We hypothesized that altruism would be particularly susceptible to stress- or cortisol-related influences in individuals with high mentalizing capacities ("high mentalizers"). This hypothesis rests on prior evidence linking mentalizing and prosocial behaviors such as charitable giving (Waytz et al., 2012; Tusche et al., 2016; Bellucci et al., 2020). Before turning to our main analyses, we therefore checked whether this relationship is also reflected in the present data. In a first step, we examined data in the independent EmpaToM task to assess the variance in mentalizing capacity across participants. On average, participants performed well in the mentalizing-related questions of the EmpaToM (mean accuracy ± SD, 66.67 ± 16.17%; Fig. 3, distribution of scores). Despite random group assignment, the stress group displayed a higher baseline mentalizing performance than the control group (72.22% vs 60.78%, $p = 0.034$). However, our main models explicitly accounted for this variable as a covariate/predictor (see below). Next, in line with previous research, baseline mentalizing capacity in the EmpaToM correlated positively with average generosity (subject-wise mean donations) across sessions ($r_{(33)} = 0.351$, $p = 0.039$) and in both sessions separately (pre: $r_{(33)} = 0.313$, $p = 0.034$, one-tailed test; post: $r_{(33)} = 0.378$, $p = 0.025$; not significantly different from each other, $p = 0.168$). Moreover, mentalizing capacity in the EmpaToM was positively related to average self-reported mentalizing (i.e., perspective-taking ratings) in the donation task ($r_{(33)} = 0.325$, $p = 0.029$, one-tailed test). Thus, participants with higher mentalizing performance in the independent task tended to recruit mentalizing more strongly in the donation task as well. Subjects also displayed considerable degrees of empathic reactivity (1.82 ± 0.85; difference score from −6 to 6 for emotional > neutral videos) and compassion (3.13 ± 0.64; 6 point scale) in the EmpaToM (but see Fig. 3) with no significant differences between groups (p values > 0.202). However, variance in these affective measures was not significantly associated with overall generosity in the donation task (compassion: $r_{(33)} = 0.234$, $p = 0.176$; empathy: $r_{(33)} = 0.2$, $p = 0.249$). Hence, mentalizing capacity was the only robust EmpaToM predictor of generosity in the donation task.

For completeness, we also checked whether the decision-related ratings in the donation task are linked to generosity. Even before any stress manipulation, participants reported varying degrees of perceived mentalizing [5.71 ± 1.10 (9 point scale)], empathy [mean ± SD: 4.78 ± 1.24 (9-point scale)], and compassion [5.78 ± 1.16 (9 point scale)] regarding the beneficiaries of the charities. We estimated a linear mixed regression model to assess whether these sociocognitive and socioaffective ratings were also associated with altruistic behavior. This model included trial-wise donations as the dependent variable and the trial-wise ratings as independent variables (modeled as fixed

Table 2. Statistical models assessing cortisol- and mentalizing-related effects on donations

Predictor	β	SE	Test statistic*	*p*-value
Choice – full model 2 (GLMM)				
Constant (intercept)	11.983	0.565	21.223	<0.001
Time	−0.057	0.077	−0.744	0.460
ΔCortisol	<−0.001	0.003	−0.002	0.998
Mentalizing capacity	6.852	2.320	2.953	0.004
ΔCortisol × time	−0.001	<0.001	−2.924	0.005
Mentalizing × time	0.982	0.514	1.910	0.061
ΔCortisol × mentalizing	−0.050	0.017	−2.883	0.005
ΔCortisol × mentalizing × time	−0.009	0.004	−2.275	0.026
Decomposition (PRE)				
Constant (intercept)	12.041	0.552	476.620	<0.001
ΔCortisol	0.001	0.003	0.146	0.702
Mentalizing capacity	5.870	3.527	2.771	0.096
ΔCortisol × mentalizing	−0.041	0.024	3.0004	0.083
Decomposition (POST)				
Constant (intercept)	11.926	0.586	414.252	<0.001
ΔCortisol	−0.001	0.003	0.134	0.715
Mentalizing capacity	7.834	3.747	4.372	0.037
ΔCortisol × mentalizing	−0.059	0.025	5.462	0.019
Decomposition (POST-PRE)				
Constant (intercept)	−0.116	0.155	0.560	0.454
ΔCortisol	−0.002	<0.001	7.536	0.006
Mentalizing capacity	1.962	0.990	3.925	0.048
ΔCortisol × mentalizing	−0.018	0.007	7.144	0.008

*Test statistic for full model (GLMM): *t* value; decomposition models (GLMs): Wald-χ^2 score (default SPSS outputs).

effects) and participants (random effects). Consistent with previous findings (Tusche et al., 2016) and the idea that mentalizing is a driving force of altruistic behavior, trial-by-trial ratings of participants' engagement in mentalizing predicted generosity (self-reported perspective-taking: $B = 1.34$, $p < 0.001$). Compassion also emerged as significant positive predictors of generosity ($B = 0.95$, $p < 0.001$), whereas self-reported empathy was not significantly associated with charitable giving ($B = 0.04$, $p = 0.45$). We also checked whether the inclusion of the rating questions had a general effect on donations. However, donations in the rating-free control block did not significantly differ from later blocks with ratings (all pair-wise comparisons with p values > 0.21), indicating that the inclusion of the ratings generally did not alter donation decisions.

Together, especially mentalizing emerged as a robust predictor of generosity across independent tasks and might hence function as a particularly plausible moderator of stress- or cortisol-related effects on altruistic choice and neural activity.

Cortisol increases are linked to reduced charitable giving

Our main set of behavioral analyses examined stress-related effects on charitable giving. First, we tested whether the stress group displayed altered charitable giving after the TSST, relative to the control group. Figure 4*A* illustrates average donations in both groups over time. We fitted a GLMM (choice – full model 1) with donations as the dependent variable and the following predictors: time (pre vs post) as repeated-measures factor, group (stress vs control) as between-subjects factor, and mentalizing capacity (as captured in the EmpaToM) as a covariate (for more details, see Materials and Methods). Contrary to our hypothesis, we did not observe a significant change in donations after the stress manipulation (group × time interaction: $F_{(1,62)} = 0.996$, $p = 0.322$) or a moderation of this effect by mentalizing capacity (group × time × mentalizing interaction: $F_{(1,62)} = 0.414$, $p = 0.522$). There was also no significant overall group difference in donations ($\text{mean}_{\text{stress}}$ = €13.18; $\text{mean}_{\text{control}}$ = €10.8; main effect of group: $F_{(1,62)} = 2.001$, $p = 0.162$), no significant main effect of time across groups ($F_{(1,62)} = 0.035$, $p = 0.853$), or any other significant effects (all p values ≥ 0.114).

Although our initial analysis did not reveal a significant effect of the stressor per se on altruism on the group level, the above model ignores potentially important variability in stress-related parameters. Participants differed in their stress response, as captured in individuals' changes in cortisol. Given that prosocial behavior might specifically depend on cortisol activity (Starcke et al., 2011), and mentalizing might be particularly sensitive to fluctuations in cortisol (Smeets et al., 2009; Leder et al., 2013), we hypothesized that cortisol-related effects on altruism would be moderated by mentalizing capacity. In a second analysis that is more sensitive to variability in cortisol, we therefore fit a GLMM (choice – full model 2) to assess whether changes in cortisol (Δcortisol, captured in the AUCi) or mentalizing capacity (EmpaToM), or an interaction of Δcortisol × mentalizing predicted changes in donations. Time (pre vs post) was included as a within-subject factor. We fitted this model on choice data across groups because cortisol dynamics varied stronlgy in both experimental groups (Fig. 2). Notably, we also fitted another, more complex model that also included group as a predictor (including all main effects and interactions). While this model explicitly accounts for potential differential effects of cortisol or mentalizing capacity across groups, it showed an inferior model fit (BIC: 410.79 vs 402.56 for the simpler model; AIC: 407.05 vs 398.51), suggesting a general effect across groups. Accounting for potential gender differences by including gender as an additional predictor also resulted in an inferior model fit [hence, there is no evidence in favor of systematic gender differences in our (limited) sample]. Based on these model comparisons, we report the results of the simpler model in the following and in Table 2 (choice – full model 2).

We hypothesized that increases in cortisol would be associated with reduced altruism over time (for a simple bivariate relationship; see Fig. 4*B*), and that this association might be moderated by mentalizing capacity. In line with this notion, the GLMM revealed a significant Δcortisol × mentalizing × time interaction ($F_{(1,62)} = 5.174$, $p = 0.026$). To decompose this three-way interaction, we fitted two follow-up generalized linear models for the prephase and postphase separately (Table 2, Decomposition PRE and POST) and for change scores (Decomposition POST-PRE). These decomposition models also included Δcortisol, mentalizing capacity, and their interaction as predictors. In the prephase decomposition, there were no significant predictors (all p values ≥ 0.083), although by tendency mentalizing capacity positively predicted charitable giving ($B = 5.870$, $SE = 3.527$, $p = 0.096$). In contrast, for postphase donations, we observed a significant positive effect of mentalizing capacity ($B = 7.834$, $SE = 3.747$, $p = 0.037$). More importantly, we also found a significant Δcortisol × mentalizing interaction ($B = -0.059$, $SE = 0.025$, $p = 0.019$). The latter finding aligns with our hypothesis of a cortisol-related effect on altruistic choice that depends on individuals' mentalizing capacity.

This two-way interaction was decomposed further using a simple-slopes analysis (Preacher et al., 2006), which assesses the relationship between Δcortisol and postphase donations at different levels of mentalizing capacity (±1 SD). Figure 4*C* illustrates the simple slopes of the significant postphase interaction. For comparison, it also illustrates the simple slopes for the nonsignificant interaction in the prephase and donation change scores. While we observed a significant negative association between changes in cortisol and postphase donations for high mentalizing capacity ($B_{\mathrm{highMENT}(+1\mathrm{SD})} = -0.011$, SE = 0.005, $p = 0.028$), there was no significant (and a numerically positive) cortisol-related association for low mentalizing capacity ($B_{\mathrm{lowMENT}(-1\mathrm{SD})} = 0.008$, SE = 0.005, $p = 0.101$; Fig. 4*C*, POST). In other words, only for individuals with higher mentalizing capacity, we observed that increases in cortisol were associated with relative decreases in charitable giving. When examining pre-post changes directly (Fig. 4*C*, POST-PRE), donations even increased over time in high mentalizers under low Δcortisol. Yet, under high Δcortisol this relationship reversed to decreased donations in these individuals.

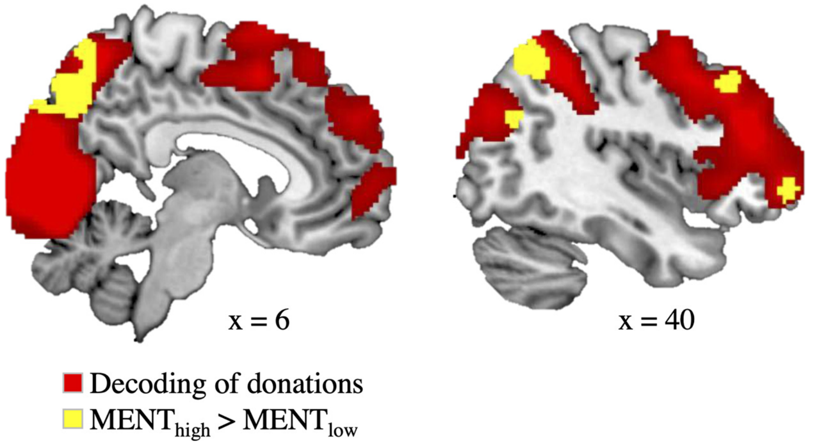

Figure 5. Neural decoding of trial-by-trial donations (red) in a whole-brain searchlight analysis using an SVR approach (cluster-forming threshold, $p \leq 0.001$; FWE-corrected at the cluster level, $p \leq 0.05$). Decoding accuracies in a subset of these brain areas (yellow) were increased with higher mentalizing capacity [thresholded with uncorrected $p \leq 0.005$ (peak, $p \leq 0.001$) and with a cluster extent threshold of $k \geq 40$ voxels]. Brain areas with differential value coding for high versus low mentalizers included the rDLPFC, the rMTG/rTPJ, a more ventrolateral part of the rMFG, and the precuneus. We observed no brain area that was more predictive of donations for low relative to high mentalizers. Group-level decoding maps are provided at https://osf.io/u46yj/.

For exploratory purposes, we also fitted identical group-based and cortisol-based GLMMs with the other two social capacity measures (empathy or compassion) and with the factual-reasoning measure obtained in the EmpaToM task as a predictor instead of mentalizing capacity (separate models). We did not observe any significant main effects or interactions regarding those predictors in these supplemental models (all p values ≥ 0.107). In particular, none of the Δcortisol × empathy/compassion/factual-reasoning × time interactions reached significance (all p values ≥ 0.156). These exploratory findings support the notion of specificity of our main results: the association between cortisol and altruism was moderated uniquely by mentalizing capacity. We also did not observe an interaction effect for response times ($p = 0.734$; for response times adjusted for the initial position of the choice slider through a regression model, $p = 0.111$). Thus, any brain responses associated with the choice-related effect are unlikely merely because of differences in decision speed.

Finally, we also fitted two exploratory regression models using either a composite score of autonomic reactivity (average of *z*-scored base-to-peak increases in blood pressure and pulse) or a composite score of subjective stress ratings, together with mentalizing capacity and interaction terms as predictors to investigate the potential role of other stress parameters. We observed no significant effects related to these autonomic and subjective stress indices (all p values ≥ 0.114). In particular, the interaction of time × Δautonomic/subjective-stress × mentalizing interactions (all p values ≥ 0.131) did not reach significance. These exploratory results might indicate a unique role of cortisol in the observed effects on altruism.

Neural decoding of trial-by-trial variations in donations

As a first step in our fMRI analysis, we used a multivariate whole-brain searchlight analysis to identify brain regions that encode trial-by-trial variations in donations before any stress manipulation (i.e., prephase). In line with previous research (Tusche et al., 2016; Bellucci et al., 2020), we observed a range of brain areas that reliably decoded donations (Fig. 5, Table 3). Notably, this included regions previously associated with mentalizing such as the bilateral MTG/TPJ, precuneus, and DMPFC (Kanske et al., 2015; Schurz et al., 2020).

Next, we examined whether the predictive information on donations varied as a function of mentalizing capacity (as captured in the independent EmaToM task). We found that the rDLPFC, right MTG/TPJ, right middle frontal gyrus (rMFG), and the precuneus displayed significantly higher decoding accuracies in participants with high relative to low mentalizing capacity (Fig. 5, Table 3), indicating a stronger value representation in high mentalizers. Group-level decoding maps are available on https://osf.io/u46yj/.

We then turned to investigate potential stress-induced or cortisol-related effects on neural value coding (see next section). To this end, we created ROIs (also see Materials and Methods) based on the results reported above. Specifically, we created spherical ROIs (radius, 5 mm) centered at the activation peaks of brain areas in which information predictive of donations varied for high and low mentalizers (Table 3, peak coordinates; High MENT > Low MENT). Note that the construction of the ROIs was based on prephase data only (i.e., before any stress manipulation) and thus was independent of the changes in neural activity following the stress manipulation.

Cortisol elevations predict decreased neural representations of donations in the rDLPFC in high mentalizers

Our main set of fMRI analyses examined the neural basis of the functional link between increased cortisol, decreased charitable giving, and its moderation by mentalizing capacity. To this end, we used GLMMs to predict decoding accuracies on donations for each of our four ROIs (separate GLMMs) based on time, Δcortisol, mentalizing capacity, and their interaction (matching the predictors of the behavioral GLMM). Mirroring our behavioral results, this model revealed a significant Δcortisol × mentalizing × time interaction ($F_{(1,62)} = 9.347$, $p = 0.003$) for decoding accuracies in the rDLPFC (Fig. 6*A*, illustration of pre-SVR and post-SVR decoding accuracies; Table 4, MVPA – full model). This effect also survives a correction for the number of tests (i.e., for four ROIs) with an adjusted $p = 0.012$. To decompose this three-way interaction, we again fitted two follow-up generalized linear models for the prephase and postphase separately (Table 4, Decomposition PRE and POST) and for change scores (Decomposition POST-PRE). These again included Δcortisol, mentalizing capacity, and their interaction as predictors. Matching our behavioral findings, we did

Table 3. Whole-brain searchlight regression (SVR) of donations

					MNI (peaks)		
Brain region	Side	BA	*T*	*k*	*x*	*y*	*z*
Average effect[a]							
Multiregional cluster—local peaks[b]	R/L		12.36	52,714	2	−86	2
Lingual gyrus/calcarine sulcus	R/L	18/17	12.36		2	−86	2
Lingual gyrus	L	18	10.45		−10	−84	−6
Lingual gyrus	R	18	10.26		10	−76	−4
Precentral gyrus	R	6	8.44		52	2	48
Superior frontal gyrus (DLPFC)	R	8	7.62		20	48	48
Middle frontal gyrus (DLPFC)	R	8/9	6.77		34	28	40
Middle frontal gyrus	L	10	5.05		−36	54	6
Superior temporal gyrus (TPJ)	L	39	4.47		−52	−52	12
Middle temporal gyrus (TPJ)	L	39	4.40		−50	−58	2
Inferior frontal gyrus	L	44	4.46		−56	8	20
Superior frontal gyrus (DLPFC)	R/L	8	4.40		2	34	58
Inferior frontal gyrus/insula	R	47/13	4.39		40	28	−2
Inferior parietal lobule (TPJ)	L	40	4.34		−52	−38	32
Middle frontal gyrus	L	10	4.33		−14	50	12
Inferior parietal lobule/angular gyrus	L	40/39	4.24		−44	−64	48
Middle frontal gyrus (DMPFC)	R/L	10	4.23		0	64	22
Inferior parietal lobule (TPJ)	R	40	4.18		42	−38	38
High MENT > Low MENT[c]							
Precuneus, superior and inferior parietal lobule	R/L	7	5.38	3201	2	−72	44
DLPFC	R	8/9	3.80	100	40	28	42
Middle temporal gyrus (TPJ)	R	39	3.78	49	40	−66	26
Middle frontal gyrus	R	10	3.58	314	32	56	−2

L, Left hemisphere; R, right hemisphere; BA, Brodmann area; *k*, cluster size in voxels.

[a] Results are reported with a cluster-defining uncorrected threshold of $p \leq 0.001$, FWE-corrected for multiple comparisons at the cluster level ($p \leq 0.05$).

[b] To derive meaningful local peaks within the large cluster, we created subclusters using an uncorrected threshold of $p \leq 0.0001$ and report peak coordinates.

[c] Within the donation-coding brain areas (i.e., inclusive mask), we used an uncorrected threshold of $p \leq 0.005$ (together with an uncorrected $p_{\text{peak}} \leq 0.001$ and an extent threshold of $k \geq 40$) for the contrast high MENT > low MENT.

not observe a significant Δcortisol × mentalizing interaction for prephase decoding accuracies in the rDLPFC (B = 0.002, SE = 0.001, $p = 0.096$). However, this interaction was significant for the postphase (B = −0.003, SE = 0.001, $p = 0.011$). Importantly, the interaction effect for the postphase is also significantly stronger than for the prephase in an identical model on the change scores (B = −0.004, SE = 0.001, $p = 0.002$), in line with the interaction with time in our full GLMM above.

The significant Δcortisol × mentalizing interaction for postphase decoding accuracies was again decomposed using a simple-slopes analysis (Fig. 6*B*, POST). While we observed a significant negative association between changes in cortisol and postphase decoding accuracies in the rDLPFC for high mentalizing capacity ($B_{\text{highMENT}(+1\text{SD})} = -0.001$, SE = 0.0003, $p = 0.007$), there was no significant cortisol-related association for low mentalizing capacity ($B_{\text{lowMENT}(-1\text{SD})} = 0.0002$, SE = 0.0002, $p = 0.304$). In other words, only for high mentalizers, we observed that increased cortisol was associated with decreased neural representations of donations in the rDLPFC (Fig. 6*B*, POST-PRE).

No further significant Δcortisol × mentalizing × time interactions (all p values > 0.276) or other cortisol-related effects (all p values > 0.096) were detected in the GLMMs for any of the other ROIs. Likewise, no additional brain region was identified in our exploratory set of group-based ROI and whole-brain analyses. Furthermore, there were no significant cortisol- or group-related effects in our supplemental univariate fMRI analysis.

As a sanity check, we used a permutation test to assess whether the rDLPFC reliably encoded donations in both the prephase and postphase and to ensure that the results did not emerge by chance. Specifically, for each participant and per phase (i.e., pre vs post), permutation distributions were created by breaking up the mapping of donations and neural response patterns (10,000-fold). We then compared the average "real" decoding accuracies (i.e., ROI-wise mean across participants) to the sampled permutation distributions. We observed that the rDLPFC reliably encoded donations in both the prephase ($r = 0.14$, $p_{\text{perm}} < 0.0001$) and the postphase ($r = 0.09$, $p_{\text{perm}} = 0.0015$).

Together, we observed statistically reliable donation value coding in the rDLPFC across time at the group level. Importantly, value coding in the rDLPFC was reduced in the face of increased cortisol concentrations in high mentalizers.

The negative cortisol–altruism association is mediated by reduced value coding in the DLPFC in high mentalizers

So far, we observed that increases in cortisol were associated with (1) decreases in charitable giving and (2) decreases in donation decoding accuracies in the rDLPFC, but only for individuals with higher mentalizing capacity. Furthermore, using robust regression to establish a brain–behavior link, we show that pre-post changes in decoding accuracies in the DLPFC positively predicted pre-post changes in charitable giving (B = 1.617, $p = 0.014$). In other words, decreases in decoding accuracies from the presession to postsession were associated with decreases in charitable giving (Fig. 6*C*).

This raises the question of whether these observations are directly linked. Can the association between changes in cortisol and donations be explained (i.e., was it mediated) by changes in neural donation value decoding? Second, is this mediation moderated by mentalizing capacity (i.e., present for high mentalizers only)? A moderated mediation model provided evidence in favor of these conjectures (Fig. 6*D*). We observed that pre-post changes in SVR-decoding accuracy in the rDLPFC mediated the negative association between pre-post changes in cortisol and

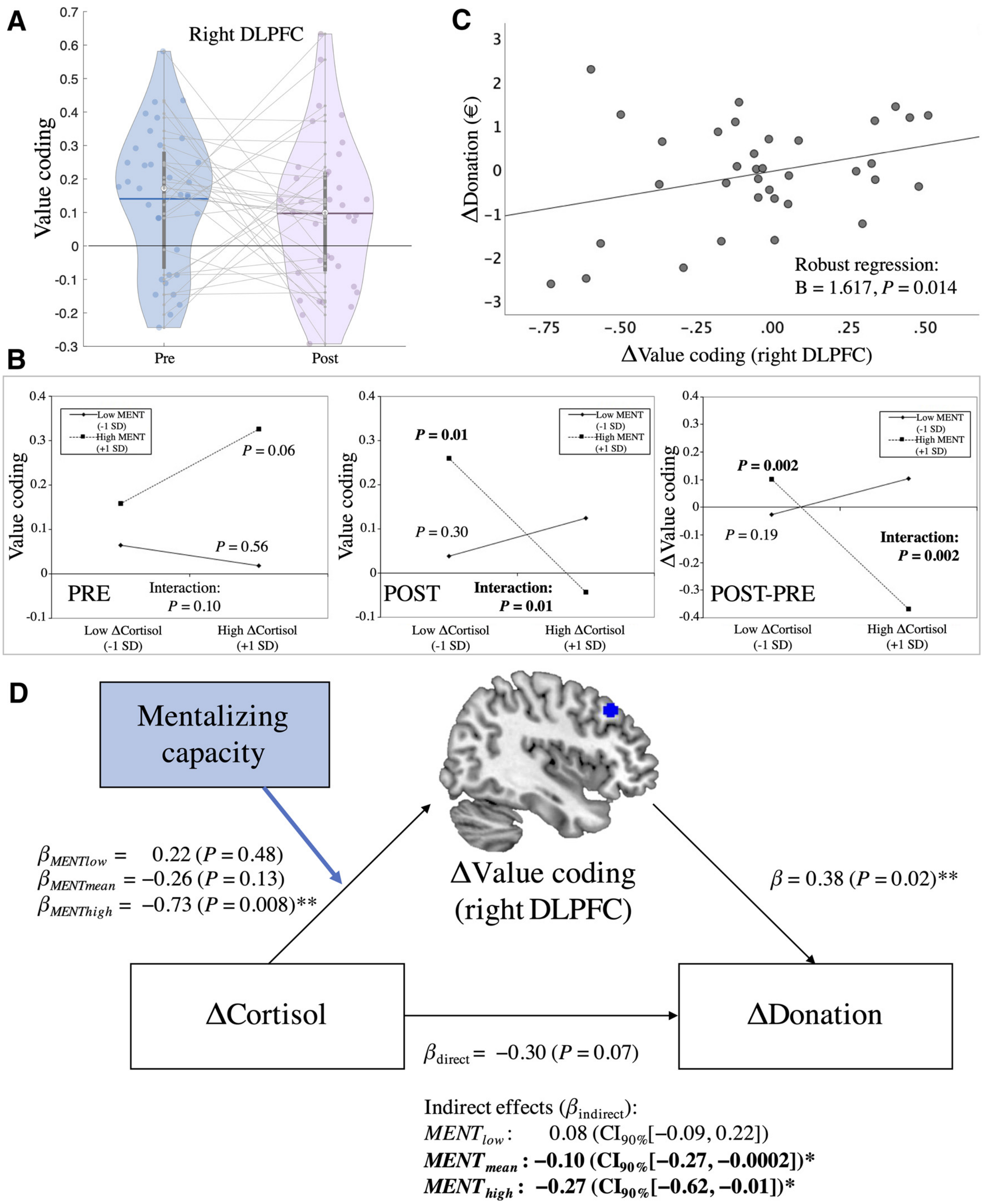

Figure 6. ***A***, Value coding [i.e., SVR donation-decoding accuracies (Fisher's *z*) in the rDLPFC (5 mm sphere around peak (MNI coordinates: 40, 28, 42)] in the presession and postsession (changes illustrated through gray lines). ***B***, ΔCortisol × mentalizing interaction plots of the simple slopes at +1 SD above and −1 SD below the mean of the moderator (mentalizing capacity) and Δcortisol for SVR decoding accuracies. Results are shown separately for the prephase, postphase, and their changes over time (POST − PRE). Only for postphase decoding accuracies of donations, we observed a significant negative association between Δcortisol and neural value coding in high mentalizers, but not low mentalizers. This moderated negative association was also significantly stronger compared with the prephase. ***C***, Brain–behavior correlation: decreases in decoding accuracies in the rDLPFC predicted decreases in donations from the presession to postsession (robust regression). ***D***, Moderated mediation model: cortisol increases were associated with reduced donations from the presession to postsession in high mentalizers. This effect was mediated by reductions in value coding in the rDLPFC. β-coefficients represent standardized regression coefficients. β_{direct} is the direct association between Δcortisol and Δdonations after the mediator (i.e., a change in SVR decoding accuracies in the rDLPFC) had been taken into account; β_{indirect} refers to the indirect effects that could be explained through altered neural value coding for each level of mentalizing capacity (low: −1 SD below the mean; high: +1SD above the mean). Here, the cortisol-related reductions in donations were mediated by reductions in value coding in the rDLPFC in average and high mentalizers. *For the significant indirect effect, bias-corrected bootstrapping (5000 bootstrap samples) provided a 90% confidence interval that did not span 0, indicating a significant mediation (one-tailed $p \leq 0.05$). **Two-tailed $p \leq 0.05$.

donations for participants with high [$\beta_{\text{highMENT}(+1\text{SD})} = -0.27$; SE = 0.19; 90% confidence interval (CI), −0.62, −0.01] and medium mentalizing capacity ($\beta_{\text{mediumMENT(mean)}} = -0.1$; SE = 0.09; 90% CI, −0.27, −0.0002), but not for participants with lower baseline mentalizing capacity ($\beta_{\text{lowMENT}(-1\text{SD})} = 0.08$; SE = 0.1; 90% CI, −0.09, 0.22; 90% CIs excluding 0 reflect $p < 0.05$, one-tailed test). In other words, only in high mentalizers, we observed a significant decline in donations over time

Table 4. Statistical models assessing cortisol- and mentalizing-related effects on value coding (SVR decoding accuracies) in the rDLPFC

Predictor	β	SE	Test statistic*	*p*-value
MVPA–full model (GLMM)				
Constant (intercept)	0.118	0.021	5.741	<0.001
Time	−0.024	0.024	−0.989	0.326
ΔCortisol	<−0.001	<0.001	−0.615	0.541
Mentalizing capacity	0.352	0.126	2.798	0.007
ΔCortisol × time	<−0.001	<0.001	−2.000	0.050
Mentalizing × time	−0.268	0.121	−2.221	0.030
ΔCortisol × mentalizing	−0.001	0.001	−0.852	0.397
ΔCortisol × mentalizing × time	−0.002	0.001	−3.057	0.003
Decomposition (PRE)				
Constant (intercept)	0.142	0.030	22.728	<0.001
ΔCortisol	<0.001	<0.001	1.349	0.245
Mentalizing capacity	0.619	0.156	15.677	<0.001
ΔCortisol × mentalizing	0.002	<0.001	2.769	0.096
Decomposition (POST)				
Constant (intercept)	0.095	0.033	8.098	0.004
ΔCortisol	<−0.001	<0.001	3.064	0.080
Mentalizing capacity	0.084	0.190	0.195	0.659
ΔCortisol × mentalizing	−0.003	0.001	6.416	0.011
Decomposition (POST-PRE)				
Constant (intercept)	−0.047	0.048	0.979	0.323
ΔCortisol	<−0.001	<0.001	3.998	0.046
Mentalizing capacity	−0.535	0.241	4.932	0.026
ΔCortisol × mentalizing	−0.004	0.001	9.347	0.002

*Test statistic for full model (GLMM), *t* value; for decomposition models (GLMs), Wald-χ^2 score (default SPSS outputs).

following increases in cortisol, mediated by decreases in multivariate neural value representations for donations in the rDLPFC.

Discussion

Stress and the involved glucocorticoids are important modulators of social behaviors and their neurobiological underpinnings (Sandi and Haller, 2015). Here we provide behavioral and neuroscientific evidence suggesting a cortisol-related decline in human altruism. Specifically, while we did not observe an effect of our stress/group manipulation per se, cortisol elevations were associated with reduced charitable giving from the presession to the postsession across groups. Notably, only participants with higher baseline mentalizing capacity—measured in an independent task—displayed that effect, but not low mentalizers. At the neural level, we found a similar interaction for value coding in the rDLPFC. Postphase activity patterns were less predictive of donations following cortisol increases in high mentalizers only. Crucially, reduced value coding in the rDLPFC mediated the negative association between cortisol and charitable giving in medium-to-high mentalizers, but not low mentalizers. This moderated mediation thereby provides a direct brain–behavior link. Our findings point to a critical role of the rDLPFC in altruism and its sensitivity to glucocorticoid influence, particularly in individuals who naturally strongly engage mentalizing to guide social behaviors.

Our findings are consistent with previous reports of cortisol-related decrements in altruistic choice (Starcke et al., 2011) and of antagonistic responses under stress (Sandi and Haller, 2015). Our study extends this line of research in two critical ways. First, we tested whether cortisol-related effects depended on mentalizing capacity—an important contributor to prosociality (Waytz et al., 2012; Tusche et al., 2016; Bellucci et al., 2020). Second, we provide evidence of a neural mechanism mediating cortisol-related effects on altruism.

The identified negative association between increasing cortisol and charitable giving in high (but not low) mentalizers might indicate a cortisol-related disruption of mentalizing-related cognitive processes that otherwise would contribute to altruistic choice. This notion is consistent with evidence suggesting that acute stress and cortisol in particular can impair mentalizing (Smeets et al., 2009; Leder et al., 2013), but these observations have not yet been linked to similar disruptions of altruism (Starcke et al., 2011). Our observation that cortisol-related decrements in altruism depend on mentalizing capacity indicates such a link. This also demonstrates that (independent) task-based measures of individuals' general capacity to mentalize present a unique angle to study the role of stress (hormones) in altruism.

On the neural level, we established a link between a cortisol-related disruption of value coding in the rDLPFC and reduced altruism in high mentalizers. In theory, there are two possibilities how DLPFC functioning could be linked to mentalizing. First, DLPFC activity might directly reflect core mentalizing processes. This notion is in line with our observation that, before any stress manipulation, high mentalizers displayed higher rDLPFC donation-decoding accuracies. Moreover, brain stimulation studies suggest that the DLPFC causally contributes to mentalizing (Costa et al., 2008; Kalbe et al., 2010). Second, the DLPFC, among other regions, represents a "co-opted" system relevant to mentalizing (Siegal and Varley, 2002). This might also explain why DLPFC activity has not been consistently observed in fMRI meta-analyses of mentalizing tasks (Kogler et al., 2020; Schurz et al., 2020; but see Molenberghs et al., 2016). However, the DLPFC is frequently reported in neuroimaging studies on prosocial decision-making (Waytz et al., 2012; Tusche et al., 2016; Bellucci et al., 2020). The DLPFC has been suggested to contribute to altruistic choice via general purpose and context-dependent cognitive control. Across social and nonsocial domains, the rDLPFC flexibly encodes values of choice options consistent with the current regulatory focus and goals. During altruistic choice, rDLPFC activation patterns reflect reduced inputs of self-related motives when individuals focus on others' thoughts and feelings (Tusche and Hutcherson, 2018). The rDLPFC is also involved in the controlled shift from a self-centered to an other-centered perspective (Thirioux et al., 2014). Based on our data alone, we cannot be sure whether altered rDLPFC activity reflects changes in mentalizing or other decision-relevant processes. To note, the underlying process appears not to contribute to factual reasoning, given that we did not find similar effects for factual-reasoning capacity. Future studies might leverage a neurocomputational approach (Hampton et al., 2008; Tusche and Bas, 2021) to further delineate the mechanistic role of the rDLPFC in mentalizing and mediating cortisol-related effects on altruism.

Mentalizing-related processes, whether core or co-opted, are not the only contributors to altruism. Empathy and compassion are potent affective drivers (Batson et al., 2015; Tusche et al., 2016; Böckler et al., 2018). Acute stress can increase empathy and prosociality. For instance, one study found stress-enhanced activity in the empathy network (AI and aMCC), which predicted altruistic choices in an independent dictator game (Tomova et al., 2017). This is consistent with other reports of increased altruism under stress or elevated cortisol levels (von Dawans et al., 2012; Singer et al., 2017; Margittai et al., 2018), but it is unclear whether altered empathy mediated this effect. Notably, we did not find any evidence for stress- or cortisol-associated increases in altruism. Likewise, there were no interaction effects with baseline empathy and compassion in the EmpaToM.

The seemingly inconsistent effects of acute stress or glucocorticoids on altruism might be explained by the influence of context, individual factors, and their interaction. For instance, acute stress is thought to enhance altruism, particularly when the need of the target is salient (Buchanan and Preston, 2014), consistent with enhanced empathy when actually seeing others receiving painful treatment (Tomova et al., 2017). Differences in salience might also explain why empathy and compassion ratings in the EmpaToM (salient videos) did not significantly, though descriptively positively, predict donations (less salient charity texts). Social closeness might also play a role as it can increase empathy (Engert et al., 2014) and moderate the stress–altruism relationship (Singer et al., 2021). However, perceived social closeness did not emerge as a significant predictor of donations and neural responses in the donation task (Tusche et al., 2016). Text-based stimuli might generally induce less perceived closeness than other more salient stimuli for which it may play a more critical role. Moreover, whereas some individuals display a high general propensity to empathize, others strongly engage in mentalizing. These two capacities are independent of each other on a behavioral and neural level (Kanske et al., 2016). Hence, while our data indicate cortisol-related decrements in altruism in high mentalizers, stronger empathizers might show opposite effects, particularly when the need of others is salient. Compassionate individuals might display still other context-dependent effects, given that compassion has unique neural correlates (e.g., in reward-related regions; Klimecki et al., 2014; Kanske et al., 2015). Future studies might benefit from advancing this situation–person interaction perspective by comparing different contexts in relation to individual traits and states. A similar perspective might ultimately inform target-specific interventions to alleviate stress-related social disruptions in clinical, economic, and other settings. For instance, stress-prone mentalizers may benefit from stress-reduction treatments and a (compensatory) training of their mentalizing abilities to avoid stress translating to deficits in prosociality. It would also be interesting to investigate how altruistic choice in the laboratory (though incentivized) relates to real-world charity and other forms of altruism, or vice versa, whether they are important determinants of behavior in the laboratory.

The postphase of our donation task matched a phase of the stress response characterized by nongenomic cortisol action (<1 h following stressor onset; Hermans et al., 2014; Joëls et al., 2018). Hence, the observed cortisol-related effects might be explainable via this mode of action. In contrast, sympathetic activity returned to baseline before postphase donations and was unrelated to altruistic choice. To control for potential influences of other (covarying) stress-related factors and to provide evidence that enhanced cortisol causally decreases neural value coding and altruism in high mentalizers, future studies could use pharmacological manipulations of cortisol (and noradrenergic) activity (Metz et al., 2020). Furthermore, other stress components may exert (differential) effects on altruistic choice at different timescales. While the influence of catecholamines (e.g., on prefrontal functioning; Arnsten, 2009) might be stronger in an earlier phase, genomic (vs nongenomic) effects of cortisol come into play only later (Singer et al., 2021). Future experimental designs might leverage different timescales to assess different phases of the stress response. Interestingly, the observed cortisol-related effect on altruism was not specific to the TSST condition. Instead, variability in stress–hormone changes across both groups explained variability in donations. This effect, however, did not translate into a group difference in altruism, despite elevated cortisol in the stress group. This might be explained by a considerable overlap in the group distributions of cortisol changes and the moderation of the cortisol effect by mentalizing capacity, which itself varies across participants of both groups. We argue, however, that our findings are still relevant to stress contexts, given cortisol elevations after the TSST.

In sum, the present study suggests that detrimental influences of acute stress hormone elevations on altruistic choice in high mentalizers are mediated by the rDLPFC. Our results thereby underline the potential susceptibility of mentalizing-related DLPFC functioning to cortisol. Future research might benefit from powerful neurocomputational models of choice and mentalizing, combined with causal manipulations of cortisol levels or neural activity (e.g., of the rDLPFC; Schulreich and Schwabe, 2021), to further elucidate the mechanisms underlying the modulation of altruism through stress hormone dynamics.

References

Arnsten AFT (2009) Stress signalling pathways that impair prefrontal cortex structure and function. Nat Rev Neurosci 10:410–422.

Batson CD, Lishner DA, Stocks EL (2015) The empathy-altruism hypothesis. In: The Oxford handbook of prosocial behavior (Schroeder DA, Graziano WG, eds), pp 259–281. Oxford, UK: Oxford UP.

Bellucci G, Camilleri JA, Eickhoff SB, Krueger F (2020) Neural signatures of prosocial behaviors. Neurosci Biobehav Rev 118:186–195.

Böckler A, Tusche A, Singer T (2016) The structure of human prosociality: differentiating altruistically motivated, norm motivated, strategically motivated, and self-reported prosocial behavior. Soc Psychol Personal Sci 7:530–541.

Böckler A, Tusche A, Schmidt P, Singer T (2018) Distinct mental trainings differentially affect altruistically motivated, norm motivated, and self-reported prosocial behaviour. Sci Rep 8:13560.

Bogdanov M, Schwabe L (2016) Transcranial stimulation of the dorsolateral prefrontal cortex prevents stress-induced working memory deficits. J Neurosci 36:1429–1437.

Buchanan TW, Preston SD (2014) Stress leads to prosocial action in immediate need situations. Front Behav Neurosci 8:5.

Burkart JM, Allon O, Amici F, Fichtel C, Finkenwirth C, Heschl A, Huber J, Isler K, Kosonen ZK, Martins E, Meulman EJ, Richiger R, Rueth K, Spillmann B, Wiesendanger S, Van Schaik CP (2014) The evolutionary origin of human hyper-cooperation. Nat Commun 5:4747.

Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2:1–27.

Costa A, Torriero S, Oliveri M, Caltagirone C (2008) Prefrontal and temporo-parietal involvement in taking others' perspective: TMS evidence. Behav Neurol 19:71–74.

Devilbiss DM, Spencer RC, Berridge CW (2017) Stress degrades prefrontal cortex neuronal coding of goal-directed behavior. Cereb Cortex 27:2970–2983.

Edwards S, Clow A, Evans P, Hucklebridge F (2001) Exploration of the awakening cortisol response in relation to diurnal cortisol secretory activity. Life Sci 68:2093–2103.

Engert V, Plessow F, Miller R, Kirschbaum C, Singer T (2014) Cortisol increase in empathic stress is modulated by emotional closeness and observation modality. Psychoneuroendocrinology 45:192–201.

Faul F, Erdfelder E, Buchner A, Lang A-G (2009) Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 41:1149–1160.

Hampton AN, Bossaerts P, O'Doherty JP (2008) Neural correlates of mentalizing-related computations during strategic interactions in humans. Proc Natl Acad Sci U|S|A 105:6741–6746.

Hare TA, Camerer CF, Knoepfle DT, Rangel A (2010) Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. J Neurosci 30:583–590.

Hautzinger M, Keller F, Kühner C (2006) Beck depression inventory (BDI-II), revision. Frankfurt, Germany: Harcourt Test Services.

Hayes AF (2018) Introduction to mediation, moderation, and conditional process analysis: a regression-based approach. New York: Guilford Publications.

Haynes J-D, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE (2007) Reading hidden intentions in the human brain. Curr Biol 17:323–328.

Hermans EJ, Henckens MJAG, Joëls M, Fernández G (2014) Dynamic adaptation of large-scale brain networks in response to acute stressors. Trends Neurosci 37:304–314.

Hildebrandt MK, Jauk E, Lehmann K, Maliske L, Kanske P (2021) Brain activation during social cognition predicts everyday perspective-taking: a combined fMRI and ecological momentary assessment study of the social brain. Neuroimage 227:117624.

Joëls M, Karst H, Sarabdjitsingh RA (2018) The stressed brain of humans and rodents. Acta Physiol (Oxf) 223:e13066.

Kahnt T, Park SQ, Haynes JD, Tobler PN (2014) Disentangling neural representations of value and salience in the human brain. Proc Natl Acad Sci U|S|A 111:5000–5005.

Kalbe E, Schlegel M, Sack AT, Nowak DA, Dafotakis M, Bangard C, Brand M, Shamay-Tsoory S, Onur OA, Kessler J (2010) Dissociating cognitive from affective theory of mind: a TMS study. Cortex 46:769–780.

Kanske P, Böckler 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.

Kanske P, Böckler A, Trautwein FM, Lesemann FHP, Singer T (2016) Are strong empathizers better mentalizers? Evidence for independence and interaction between the routes of social cognition. Soc Cogn Affect Neurosci 11:1383–1392.

Kirschbaum C, Pirke KM, Hellhammer DH (1993) The "Trier Social Stress Test"—a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology 28:76–81.

Klimecki OM, Leiberg S, Ricard M, Singer T (2014) Differential pattern of functional brain plasticity after compassion and empathy training. Soc Cogn Affect Neurosci 9:873–879.

Kogler L, Müller VI, Werminghausen E, Eickhoff SB, Derntl B (2020) Do I feel or do I know? Neuroimaging meta-analyses on the multiple facets of empathy. Cortex 129:341–355.

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci U|S|A 103:3863–3868.

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci 12:535–540.

Kudielka BM, Hellhammer DH, Kirschbaum C (2007) Ten years of research with the Trier Social Stress Test—revisited. In: Social neuroscience: integrating biological and psychological explanations of social behavior (Harmon-Jones E, Winkielman P, eds), pp 56–83. New York: Guilford Press.

Lamm C, Decety J, Singer T (2011) Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. Neuroimage 54:2492–2502.

Leder J, Häusser JA, Mojzisch A (2013) Stress and strategic decision-making in the beauty contest game. Psychoneuroendocrinology 38:1503–1511.

Libby LA, Hannula DE, Ranganath C (2014) Medial temporal lobe coding of item and spatial information during relational binding in working memory. J Neurosci 34:14233–14242.

Lockwood PL (2016) The anatomy of empathy: vicarious experience and disorders of social cognition. Behav Brain Res 311:255–266.

Lovallo WR, Cohoon AJ, Acheson A, Vincent AS, Sorocco KH (2019) Cortisol stress reactivity in women, diurnal variations, and hormonal contraceptives: studies from the Family Health Patterns Project. Stress 22:421–427.

Margittai Z, van Wingerden M, Schnitzler A, Joëls M, Kalenscher T (2018) Dissociable roles of glucocorticoid and noradrenergic activation on social discounting. Psychoneuroendocrinology 90:22–28.

Metz S, Waiblinger-Grigull T, Schulreich S, Chae WR, Otte C, Heekeren HR, Wingenfeld K (2020) Effects of hydrocortisone and yohimbine on decision-making under risk. Psychoneuroendocrinology 114:104589.

Molenberghs P, Johnson H, Henry JD, Mattingley JB (2016) Understanding the minds of others: a neuroimaging meta-analysis. Neurosci Biobehav Rev 65:276–291.

Nater UM, La Marca R, Florin L, Moses A, Langhans W, Koller MM, Ehlert U (2006) Stress-induced changes in human salivary alpha-amylase activity—associations with adrenergic activity. Psychoneuroendocrinology 31:49–58.

Nitschke JP, Sunahara CS, Carr EW, Winkielman P, Pruessner JC, Bartz JA (2020) Stressed connections: cortisol levels following acute psychosocial stress disrupt affiliative mimicry in humans. Proc R Soc B Biol Sci 287:20192941.

Obeso I, Moisa M, Ruff CC, Dreher JC (2018) A causal role for right temporo-parietal junction in signaling moral conflict. Elife 7:e40671.

Poldrack RA, Huckins G, Varoquaux G (2020) Establishment of best practices for evidence for prediction: a review. JAMA Psychiatry 77:534–540.

Preacher KJ, Curran PJ, Bauer DJ (2006) Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. J Educ Behav Stat 31:437–448.

Pruessner JC, Kirschbaum C, Meinlschmid G, Hellhammer DH (2003) Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. Psychoneuroendocrinology 28:916–931.

Qin S, Hermans EJ, van Marle HJF, Luo J, Fernández G (2009) Acute psychological stress reduces working memory-related activity in the dorsolateral prefrontal cortex. Biol Psychiatry 66:25–32.

Sandi C, Haller J (2015) Stress and the social brain: behavioural effects and neurobiological mechanisms. Nat Rev Neurosci 16:290–304.

Schulreich S, Schwabe L (2021) Causal role of the dorsolateral prefrontal cortex in belief updating under uncertainty. Cereb Cortex 31:184–200.

Schulz P, Schlotz W (1999) The Trier Inventory for the Assessment of Chronic Stress (TICS): scale construction, statistical testing, and validation of the scale work overload. Diagnostica 45:8–19.

Schurz M, Radua J, Tholen MG, Maliske L, Margulies DS, Mars RB, Sallet J, Kanske P (2020) Toward a hierarchical model of social cognition: a neuroimaging meta-analysis and integrative review of empathy and theory of mind. Psychol Bull 147:293–327.

Schwabe L, Haddad L, Schachinger H (2008) HPA axis activation by a socially evaluated cold-pressor test. Psychoneuroendocrinology 33:890–895.

Siegal M, Varley R (2002) Neural systems involved in "theory of mind". Nat Rev Neurosci 3:463–471.

Singer N, Sommer M, Döhnel K, Zänkert S, Wüst S, Kudielka BM (2017) Acute psychosocial stress and everyday moral decision-making in young healthy men: the impact of cortisol. Horm Behav 93:72–81.

Singer N, Binapfl J, Sommer M, Wüst S, Kudielka BM (2021) Everyday moral decision-making after acute stress: do social closeness and timing matter? Stress 24:468–473.

Smeets T, Dziobek I, Wolf OT (2009) Social cognition under stress: differential effects of stress-induced cortisol elevations in healthy young men and women. Horm Behav 55:507–513.

Solanas MP, Vaessen M, De Gelder B (2020) Computation-based feature representation of body expressions in the human brain. Cereb Cortex 30:6376–6390.

Starcke K, Polzer C, Wolf OT, Brand M (2011) Does stress alter everyday moral decision-making? Psychoneuroendocrinology 36:210–219.

Tabachnik BB, Fidell LS (2013) Using multivariate statistics, Ed 6. Boston: Pearson.

Thirioux B, Mercier MR, Blanke O, Berthoz A (2014) The cognitive and neural time course of empathy and sympathy: an electrical neuroimaging study on self-other interaction. Neuroscience 267:286–306.

Tholen MG, Trautwein FM, Böckler A, Singer T, Kanske P (2020) Functional magnetic resonance imaging (fMRI) item analysis of empathy and theory of mind. Hum Brain Mapp 41:2611–2628.

Tomova L, Majdandžõić J, Hummer A, Windischberger C, Heinrichs M, Lamm C (2017) Increased neural responses to empathy for pain might explain how acute stress increases prosociality. Soc Cogn Affect Neurosci 12:401–408.

Tusche A, Bas LM (2021) Neurocomputational models of altruistic decision-making and social motives: advances, pitfalls, and future directions. Cogn Sci 12:e1571.

Tusche A, Hutcherson CA (2018) Cognitive regulation alters social and dietary choice by changing attribute representations in domain-general and domain-specific brain circuits. Elife 7:e31185.

Tusche A, Böckler A, Kanske P, Trautwein F-M, Singer T (2016) Decoding the charitable brain: empathy, perspective taking, and attention shifts differentially predict altruistic giving. J Neurosci 36:4719–4732.

Vinkers CH, Zorn JV, Cornelisse S, Koot S, Houtepen LC, Olivier B, Verster JC, Kahn RS, Boks MPM, Kalenscher T, Joëls M (2013) Time-dependent changes in altruistic punishment following stress. Psychoneuroendocrinology 38:1467–1475.

Vogel S, Fernández G, Joëls M, Schwabe L (2016) Cognitive adaptation under stress: a case for the mineralocorticoid receptor. Trends Cogn Sci 20:192–203.

Vogel S, Kluen LM, Fernández G, Schwabe L (2018) Stress affects the neural ensemble for integrating new information and prior knowledge. Neuroimage 173:176–187.

von Dawans B, Fischbacher U, Kirschbaum C, Fehr E, Heinrichs M (2012) The social dimension of stress reactivity: acute stress increases prosocial behavior in humans. Psychol Sci 23:651–660.

Waytz A, Zaki J, Mitchell JP (2012) Response of dorsomedial prefrontal cortex predicts altruistic behavior. J Neurosci 32:7646–7650.

Weng HY, Fox AS, Shackman AJ, Stodola DE, Caldwell JZK, Olson MC, Rogers GM, Davidson RJ (2013) Compassion training alters altruism and neural responses to suffering. Psychol Sci 24:1171–1180.