

Research Articles: Behavioral/Cognitive

Altruism under stress: cortisol negatively predicts charitable giving and neural value representations depending on mentalizing capacity

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

Cite as: J. Neurosci 2022; 10.1523/JNEUROSCI.1870-21.2022

Received: 15 September 2021

Revised: 9 February 2022

Accepted: 9 February 2022

This Early Release article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.jneurosci.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Copyright © 2022 the authors

1 **Altruism under stress: cortisol negatively predicts charitable**
 2 **giving and neural value representations depending on**
 3 **mentalizing capacity**

4 *Running title: Altruism under stress*

5 *Section: Behavioral/Cognitive*

6 Stefan Schulreich^{1*} (ORCID iD 0000-0001-9708-1545)

7 Anita Tusche^{2,3} (ORCID iD 0000-0003-4180-8447)

8 Philipp Kanske⁴ (ORCID iD 0000-0003-2027-8782)

9 Lars Schwabe¹ (ORCID iD 0000-0003-4429-4373)

10 ¹ Department of Cognitive Psychology, Faculty of Psychology and Human Movement
 11 Science, Universität Hamburg, 20146 Hamburg, Germany. ² Queen's Neuroeconomics
 12 Laboratory, Departments of Psychology and Economics, Queen's University, ON K7L 3N6
 13 Kingston, Canada. ³ California Institute of Technology, HSS, 1200 East California Boulevard
 14 Pasadena, California 91125. ⁴ Institute of Clinical Psychology and Behavioral Neuroscience,
 15 Faculty of Psychology, Technische Universität Dresden, 01187 Dresden, Germany.

16 *** Corresponding author:** Department of Cognitive Psychology, Universität Hamburg, Von-
 17 Melle-Park 5, 20146 Hamburg, Germany. Phone +49 40 42838-6523, fax +49 40 42838-
 18 4373, e-mail stefan.schulreich@uni-hamburg.de

20 **Number of pages:** 49; figures: 6; tables: 4

21 **Number of words (revised):** abstract: 250; introduction: 764; discussion: 1500

22 **Acknowledgments:** This work was supported by Universität Hamburg. We also thank
 23 Jehona Muslija and Gudrun Grätschus for their assistance during data collection and Carlo
 24 Hiller for programming the task.

25 **Conflict of interest:** The authors declare no competing financial interests.

Abstract

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 (fMRI). 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 (rDLPFC) were less predictive of post-manipulation 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.

51 Significance Statement

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

62

63

64

65

66

67

68

69

70

71

72 **Introduction**

73 Altruism involves other-regarding behavior at a cost to the self. It is a complex phenomenon
 74 that emerged in many species (Burkart et al., 2014). Humans often behave altruistically even
 75 in the absence of direct future benefits, such as in (anonymous) donations to charitable
 76 organizations (Hare et al., 2010; Tusche et al., 2016; Obeso et al., 2018). However, altruism
 77 varies across contexts. For instance, initial evidence suggests that stress might interfere with
 78 altruism. Acute psychosocial stress (Vinkers et al., 2013) and stress-induced increases in the
 79 glucocorticoid hormone cortisol (Starcke et al., 2011; but see Singer et al., 2017) have been
 80 linked to reduced altruistic decisions. To date, however, the neural mechanisms through
 81 which stress or cortisol may alter altruism are largely unknown.

82 Altruistic behavior is supported by networks of brain regions associated with social
 83 cognition and value-based decision-making (Tusche et al., 2016; Bellucci et al., 2020; Tusche
 84 and Bas, 2021). This includes prefrontal regions whose functions can be significantly
 85 impaired by acute stress and the stress hormone cortisol (for reviews, see Arnsten, 2009;
 86 Vogel et al., 2016) such as the dorsolateral prefrontal cortex (DLPFC) (Qin et al., 2009;
 87 Bogdanov and Schwabe, 2016) and dorsomedial prefrontal cortex (DMPFC) (Devilbiss et al.,
 88 2017). At the cognitive level, both regions critically support *mentalizing* (also referred to as
 89 Theory-of-Mind [ToM]) – the ability to attribute mental states (e.g., beliefs, desires,
 90 intentions) to others. For instance, the DMPFC is an important node in a well-described
 91 mentalizing network (Schurz et al., 2020), which also includes the temporoparietal junction
 92 (TPJ) and middle temporal gyrus (MTG). The DLPFC also critically contributes to
 93 mentalizing (Costa et al., 2008; Kalbe et al., 2010) as well as context-dependent cognitive
 94 control (Tusche and Hutcherson, 2018).

95 Mentalizing is a key contributor to altruism (Waytz et al., 2012; Tusche et al., 2016).
 96 Hence, a stress-induced impairment in this socio-cognitive process might be a plausible

97 mechanism through which stress impairs altruistic behavior. This notion is in line with initial
 98 behavioral evidence of stress-induced impairments of mentalizing (Smeets et al., 2009; Leder
 99 et al., 2013). Our main hypotheses are that 1) acute stress decreases altruism via altered
 100 prefrontal functioning, and 2) participants who strongly engage the mentalizing network
 101 (“high mentalizers”) are particularly prone to these stress-induced decrements. An alternative,
 102 more affective route via which stress might influence altruism is *empathy* (see, e.g., Tomova
 103 et al., 2017), which refers to the isomorphic representation of others’ affective states (e.g.,
 104 vicariously feeling others’ suffering) (Lockwood, 2016), or *compassion*, which involves
 105 caring feelings for others (Weng et al., 2013). Empathy is supported by a network comprising
 106 the anterior mid-cingulate cortex (amCC) and anterior insula (AI) (Lamm et al., 2011;
 107 Lockwood, 2016), while compassion has been associated with reward-related regions (e.g.,
 108 striatum; Kanske et al., 2015), and both are contributors to altruism (Weng et al., 2013;
 109 Tusche et al., 2016; Tomova et al., 2017).

110 We also hypothesize that stress-related effects on altruism are mediated through the
 111 action of the major stress hormone cortisol. Cortisol can exert its influence via earlier non-
 112 genomic actions (around <1h post-stressor) or later genomic processes in neurons (Hermans
 113 et al., 2014; Joëls et al., 2018). Non-genomic cortisol actions are particularly likely to play a
 114 role in altruism, given that cortisol elevations predicted or even statistically mediated
 115 decrements in mentalizing performance (Smeets et al., 2009; Leder et al., 2013) and DLPFC
 116 functioning (Qin et al., 2009), and given that these observations were made within the first
 117 hour following the stressor. Moreover, one study found altered altruistic choice only in an
 118 earlier but not later phase (Singer et al., 2021; but see Vinkers et al., 2013 for no difference).
 119 To note, none of those studies investigated stress effects in the very early phase dominated by
 120 autonomic stress-reactivity, which vanishes within minutes after stressor offset (Nater et al.,

2006). Consequently, we believe the phase of non-genomic 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 (Figure 1A). The post-phase was within the previously described window of non-genomic action of cortisol (<1h; 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 (EmpaToM; Kanske et al., 2015).

Materials and Methods

Participants

A total of 50 right-handed volunteers [24 women, 26 men; mean (M) age $\pm$ standard deviation (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 labs (e.g., Bogdanov and Schwabe, 2016; Nitschke et al., 2020), exclusion criteria included current physical or mental conditions, substantial under- or overweight [body mass index (BMI) <18.5 or >28.5], medication or drug intake, smoking, a lifetime history of any neurological or psychiatric disorder, and any MRI contraindications. We also excluded women using hormonal contraceptives due to possible alterations in the stress response (e.g., Lovallo et al., 2019) and those in pregnancy or lactation due to ethical reasons (i.e., to avoid any potential adverse

146 effects on mother or child). Participation of female participants was not restricted to a
 147 particular phase of their menstrual cycle. The final sample and both groups included women
 148 distributed across all phases (i.e., follicular vs. luteal). Furthermore, we asked participants to
 149 refrain from physical exercise, meals, and caffeine intake within the two hours before testing.

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

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

162 The final sample consisted of 35 participants (15 women, 20 men; 23.49 ± 4.14 years).
 163 To confirm that the sample size is sufficient to detect effects of interest, we implemented an
 164 a-priori power analysis using G*Power 3.1 (Faul et al., 2009). Given the mixed design with
 165 repeated measures, our final sample allowed for the detection of a small-to-medium effect of
 166 the stress manipulation on donations and neural responses from the pre- to post-treatment
 167 session, that is a group $\times$ time [between-within] interaction with Cohen's $f = 0.18$, with a
 168 statistical power of 90% and alpha $P = 0.05$, assuming correlation among repeated measures
 169 of 0.8, and nonsphericity correction $\epsilon = 1$. Wherever possible (i.e., when choice variability is
 170 not a prerequisite), we complemented our main analyses with analyses on the larger sample

171 (N = 44) that included those nine subjects with invariant decisions to check for the robustness
 172 of our results and observed here largely comparable results.

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

177 **Experimental Design**

178 All experimental sessions took place between 08:00 and 13:00 to mitigate the influence of the
 179 diurnal rhythm of cortisol (Edwards et al., 2001). The allocation of the two experimental
 180 conditions (stress vs. control group) to specific slots within this time window was
 181 randomized. Participants completed a series of tasks and measures in the lab (Figure 1A). To
 182 obtain a baseline measure of participants' mentalizing capacity, we administered the
 183 EmpaToM task (Kanske et al., 2015; see below for more details). To assess our primary
 184 measures of interest, altruistic behavior and the associated neural responses, we used a
 185 charitable donation task (adapted from Böckler et al., 2016; Tusche et al., 2016; see below for
 186 more details), while simultaneously collecting fMRI data. To test the impact of stress on
 187 charitable giving and its underlying neural processes, we used a mixed design that assessed
 188 donations and neural activity *before* and *after* (within-subject factor *time*) a standardized
 189 stress or control manipulation (TSST; Kirschbaum et al., 1993) (between-subjects factor
 190 *group*). This manipulation was accompanied with a series of stress-parameter measurements
 191 (see below). We randomly assigned participants to the stress or control group, with the only
 192 constraint of relatively balanced gender distribution. The final sample (N = 35) was roughly
 193 equally distributed across the stress (N = 18 [7 women, 11 men]; 23 ± 3.71 years) and the
 194 control group (N = 17 [8 women, 9 men]; 24 ± 4.61 years). At the end of the experimental

195 session, participants received their remuneration for participation and were fully debriefed.

196 The whole session lasted for about 150-180 min.

197 **Stress Manipulation**

198 In the stress condition, participants completed the Trier Social Stress Test (TSST;
199 Kirschbaum et al., 1993) (Figure 1A). The TSST is a standardized, well-validated laboratory
200 task to elicit subjective stress, a sympathetic stress response, and glucocorticoid secretion via
201 the hypothalamo-pituitary-adrenal (HPA) axis (Kirschbaum et al., 1993; Kudielka et al.,
202 2007). Participants first anticipated and prepared for a mock job interview (3 min). They
203 could take notes but could not use them in the following free speech (5 min), in which they
204 explained why they are the ideal candidate for the job. The free speech was followed by a
205 demanding arithmetic task of counting backward in steps of 17 from the number 2043 as fast
206 as possible (5 min). During both tasks, participants were videotaped and stood in front of a
207 panel of two rather cold and unresponsive experimenters (1 male, 1 female – different from
208 the experimenters that performed the rest of the experimental procedure), creating a social-
209 evaluative context. In the control condition, participants gave a 5-min speech about a topic of
210 their choice (e.g., last holiday) and performed a much simpler 5-min arithmetic task (i.e.,
211 counting forwards in steps of 5 starting from 0), following previous applications (e.g.,
212 Bogdanov and Schwabe, 2016; Vogel et al., 2018). During the control condition, participants
213 were neither videotaped nor monitored by a panel.

214 To assess whether the stress manipulation was successful and determine the
215 individual stress reactivity, we measured subjective and physiological stress parameters at
216 several time points across the experiment. At the subjective level, participants rated the
217 perceived stressfulness, difficulty, and unpleasantness of the TSST or control procedure on a
218 scale from 0 (“not at all”) to 100 (“very much”) immediately after the procedure. As
219 indicators of sympathetic nervous system activity, blood pressure and pulse were measured

220 using an upper-arm monitor (Omron) at several time points: before (~ -70 min relative to
 221 TSST onset) and after the first fMRI session (-10 min), during the stress/control manipulation
 222 ($+8$ min), and before ($+20$ min) and after the second fMRI session ($\sim +70$ min). Blood
 223 pressure and pulse measurement was repeated twice and then averaged at a given time point
 224 to increase reliability. To quantify cortisol concentrations, saliva samples were collected from
 225 participants at several time points before and after the stress/control manipulation (about -70
 226 min, -10 min, $+20$ min, $+70$ min relative to TSST onset) using Salivette collection devices
 227 (Sarstedt, Nümbrecht, Germany). Saliva samples were stored at -18°C and analyzed for
 228 cortisol concentrations using a luminescence assay (IBL International, Hamburg, Germany).
 229 As an integrated measure of the cortisol response to the stress manipulation, we calculated
 230 the area under the curve with respect to the increase (AUCi) from before (-10 min) to $+70$
 231 min after the offset of the TSST/control procedure (Pruessner et al., 2003).

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

240 **Donation Task**

241 Altruistic behavior was measured using a charitable donation task (adapted from Böckler et
 242 al., 2016; Tusche et al., 2016) before and after the stress/control manipulation while
 243 simultaneously collecting fMRI data (Figure 1A). The post-session took place from
 244 approximately 30 to 60 minutes following treatment onset, overlapping with the phase of the

245 stress response mainly characterized by non-genomic cortisol action (Hermans et al., 2014;
 246 Joëls et al., 2018). The pre- and post-session consisted of 40 trials each (80 trials in total),
 247 arranged in four functional runs (blocks) à ten trials.

248 In each trial, participants were first presented with a short description of a real-world
 249 charitable organization (reading phase; terminated by button press with a maximum of up to
 250 25 s; for the complete set of charity descriptions [in German], see OSF project page:
 251 <https://osf.io/u46yj/>). Next, participants had to decide how much to donate to the charity
 252 (range of €0 to €20 in steps of €1) (decision phase; up to 8 s). Participants responded using a
 253 slider from a randomized starting position. After a variable inter-stimulus interval [ISI] from
 254 2 to 6 s, three rating questions were presented in a randomized order. Participants rated their
 255 experienced (1) *empathy* (“Felt with others?”, in the sense of sharing an affective state), (2)
 256 *compassion* (“Compassion for others?”, in the sense of warm, tender feelings towards others),
 257 and (3) *perspective taking* (“Took the perspective of others?” [i.e., of the beneficiaries of the
 258 charity]) (rating phase; up to 8 s per rating; for complete instructions, see
 259 <https://osf.io/u46yj/>). Participants responded using a slider on a 9-point scale (ranging from
 260 “not at all” to “very strong”). Trials were separated by another variable ISI (2 to 6 s). Across
 261 all blocks, 90 different charities were presented. The pre- and post-sessions consisted of a
 262 total of 80 trials. Participants completed one block of ten trials outside the scanner before the
 263 main experiment. This block did not contain rating questions and served both as a training
 264 block and as a control for the potential influence of the rating task on donation behavior.

265 Each participant was presented with each of the 90 charities once. The assignment of
 266 charities to a particular block was fixed and based on donations observed in a behavioral pre-
 267 test using an independent subject sample ($N = 27$ [20 women, 7 men], 25.25 ± 6.15 years).
 268 Charities were selected and grouped such that average donations were comparable across task
 269 blocks (pre-test: $F(8, 81) = 0.029$, $P = 0.99$, $\eta_p^2 = 0.003$; main experiment: $F(8, 81) = 0.555$,

270 $P = 0.81$, $\eta_p^2 = 0.052$) and to ensure coverage of a broad range of giving behavior across the
 271 ten charities in each block. The latter was crucial as sufficient variance is a prerequisite for
 272 the *Multivariate Pattern Analysis* described below. Figure 1B illustrates the variability of
 273 charity-wise donations within the task blocks of our main experiment. The order of charities
 274 within a task block and the order of the nine task blocks were randomized across participants.

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

287 **EmpaToM Task**

288 To assess participants' mentalizing capacity in complex social settings, we administered the
 289 well-established EmpaToM task outside of the scanner (Kanske et al., 2015; Tholen et al.,
 290 2020; Hildebrandt et al., 2021). This behavioral task was performed before the donation task
 291 and the stress/control manipulation (Figure 1A), providing an independent baseline (i.e., pre-
 292 manipulation) measure of individual differences in participants' mentalizing capacity. The
 293 task simultaneously assesses socio-affective responses (empathy, compassion) in social

294 settings, which allows us to examine the specificity of mentalizing-related effects on
 295 charitable giving.

296 We used a brief version of the EmpaToM consisting of 24 trials. Each trial started
 297 with a fixation cross (1–3 s), after which the name of a person (1 s) appeared, followed by a
 298 short video recounting an autobiographical episode (~ 15 s). The videos differed in
 299 emotionality (neutral vs. negative contents) and in whether their content is mentalizing-
 300 related (e.g., beliefs, deception) or not, later giving rise to mentalizing-related vs. factual
 301 questions, respectively (yielding a 2 x 2 factorial design). The videos showed six actors (3
 302 females, 3 males), each of whom recounted one story per condition (6 actors x 4 conditions =
 303 24 trials). After each video, participants rated their empathic affective response (“How do you
 304 feel?”; from “very negative” to “very positive” on a scale from -3 to 3) and their compassion
 305 for the person in the video (“How much compassion do you feel?”; from “none” to “very
 306 much” on a scale from 0 to 6) (4 s per rating, fixed order). Participants responded by moving
 307 a slider. A multiple-choice question with three response options was presented after a
 308 variable delay of 1–3 s. The question either demanded mentalizing (e.g., “Anna thinks that
 309 [...]” [12 trials]) or factual reasoning (e.g., “It is correct that [...]” [12 trials]) on the contents
 310 of the previous video. Participants responded by pressing one of three buttons assigned to the
 311 three choice options (up to 15 s). The rate of correct responses (accuracy) in the mentalizing-
 312 related questions served as our measure of *mentalizing capacity*. For a detailed description of
 313 the task validation, example stories, and questions for each experimental condition, see
 314 Kanske et al. (2015).

315 **Control Measures**

316 Prior to the experiment, participants completed an online survey at home (implemented via
 317 the SosciSurvey platform; Leiner 2020), which included demographic questions, the Beck
 318 Depression Inventory (BDI; Hautzinger et al., 2006) and the Trier Inventory of Chronic

319 Stress (TICS; Schulz and Schlotz, 1999). These measures ensured that experimental groups
 320 (stress vs. control) were matched in terms of age, depression scores, and chronic stress after
 321 randomization (all P s > 0.313).

322 Behavioral Data Analysis

323 We will focus in this section on our primary analyses of stress-parameter and choice data.

324 Supplemental analyses will be described in the course of the *Results* section.

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

337 *Impact of acute stress on altruism.* To investigate the effect of our stress manipulation
 338 on altruistic choice and whether this effect depends on baseline mentalizing capacity, we
 339 fitted a Generalized Linear Mixed Model (GLMM: Choice – Full Model 1) with donations as
 340 the dependent variable and the following predictors: *time* (pre vs. post) as repeated-measures
 341 factor, *group* (stress vs. control) as between-subjects factor and *mentalizing capacity* as
 342 captured in the EmpaToM as a covariate. In addition to main effects, we also modeled all
 343 two-way and three-way interactions and the intercept. The time- and group-factor were

344 effect-coded (i.e., using weights of -1 and +1) and the covariate was mean-centered so that
 345 the resulting main effects (and intercept) truly reflect average effects (and not effects for a
 346 single zero-coded category/for the covariate at zero). The model was estimated with a robust
 347 covariance matrix estimator.

348 *Cortisol-related effects on altruism.* Given that both altruism (Starcke et al., 2011)
 349 and mentalizing (Smeets et al., 2009; Leder et al., 2013) have been found to depend on stress-
 350 hormone dynamics, we fitted another GLMM (Choice – Full Model 2) to assess whether
 351 changes in cortisol ($\Delta\text{cortisol}$, captured in the AUCi), *mentalizing capacity* (EmpaToM), or
 352 an interaction of $\Delta\text{cortisol} \times \text{mentalizing}$ predicted changes in donations. The GLMM
 353 modeled *time* (pre vs. post) as a within-subject factor and used identical estimation
 354 procedures as the previous GLMM. We fitted this model on choice data across groups
 355 because cortisol varied strongly over time in both experimental groups. Nevertheless, we also
 356 formally compared this model with a more complex model including *group* as an additional
 357 factor (and another model including *gender*) in terms of model fit, assessed via the Bayesian
 358 Information Criterion (BIC) and the Akaike Information Criterion (AIC), and found support
 359 for the simpler model without these factors (see *Results*). Significant interaction effects were
 360 decomposed using appropriate follow-up models. Specifically, a three-way $\Delta\text{cortisol} \times$
 361 *mentalizing* $\times$ *time* interaction was decomposed by follow-up Generalized Linear Models
 362 with $\Delta\text{cortisol}$, *mentalizing* and their interaction as predictors, fitted for each phase (pre vs.
 363 post) separately (Decomposition PRE & POST) and change scores (Decomposition POST-
 364 PRE). This post-hoc decomposition are essential to test whether pre-to-post cortisol dynamics
 365 related to the manipulation rather than pre-existing differences drive the interaction effect in
 366 our full model. The emerging two-way interactions between the continuous predictors
 367 $\Delta\text{cortisol} \times \text{mentalizing}$ in the post-phase were decomposed using simple-slopes analysis
 368 (Preacher et al., 2006). This analysis assesses the relationship between a predictor ($\Delta\text{cortisol}$)

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 non-significant two-way interactions (i.e., in the pre-phase 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, Erlangen, Germany), equipped with a 64-channel head coil. We acquired gradient-echo T_2^* -weighted echo-planar-images (EPI). For each of the 8 functional runs of the donation task (4 pre-stress/control, 4 post-stress/control), we collected a series of volumes using a slice thickness of 2 mm and isotropic voxel size of 2 mm^2 , 60 slices aligned to the anterior commissure-posterior commissure (AC-PC) line and acquired in descending order, repetition time (TR) = 2000 ms, echo time (TE) = 30 ms, flip angle = 60%, and field of view (FOV) = 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 gradient-echo (MPRAGE) sequence was employed to acquire high-resolution ($0.8 \times 0.8 \times 0.9 \text{ mm}$) 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 T₁ saturation effects. The remaining functional images were spatially realigned and distortion-corrected using the field map, slice-time corrected, co-registered 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 half-maximum (FWHM) Gaussian kernel.

fMRI Analysis

For each subject and session (pre- vs. post-manipulation), we estimated two General Linear Models (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 *GLM1_{post}* were used to obtain *trial-wise* measures of blood-oxygenation-level-dependent (BOLD) responses during the donation task (separately for the pre- and post-phase). 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.,

418 dependent variables) for our primary multivariate analysis. Furthermore, two additional
 419 regressors modeled the reading phases (R41) and the rating phases (R42), and six motion
 420 regressors accounted for residual motion-related signal changes (R43-R48).

421

422 **Multivariate Pattern Analysis (MVPA)**

423 *Neural decoding of donations.* Our multivariate decoding analysis aimed to identify brain
 424 regions that encode trial-by-trial variations in donations in their multivoxel response patterns.
 425 Donations served here as an indicator of the value people place on specific charities. In a first
 426 step, we aimed to detect brain areas that allow decoding individuals' trial-wise donations
 427 before the stress manipulation (baseline/pre-phase, *GLM1pre*). Next, we examined whether
 428 the predictive information in these brain areas varies as a function of participants'
 429 mentalizing capacity. These two steps served two important functions. First, we tested
 430 whether we could replicate the previously observed neural decoding of donations (Tusche et
 431 al., 2016). Second, value coding associated with mentalizing capacity was hypothesized to be
 432 subject to cortisol-related alterations. In other words, the detected areas served as regions of
 433 interest (ROIs) to test for stress- and cortisol-related changes in value coding.

434 To this end, we applied a whole-brain searchlight decoding approach. This approach
 435 does not depend on a priori assumptions about informative brain regions and ensures
 436 unbiased information mapping throughout the whole brain (Kriegeskorte et al., 2006; Haynes
 437 et al., 2007). For each participant, we defined a sphere (radius = 5 voxels) around a given
 438 voxel v_i of the acquired brain volume (Libby et al., 2014; Solanas et al., 2020). For each of
 439 the N voxels within this sphere, we extracted trial-wise parameter estimates of *GLM1pre* for
 440 the neural responses during the decision phases (R1-R40) of the pre-stress donations (Figure
 441 1A). Extracted activation patterns were transformed into N -dimensional pattern vectors. This
 442 was done for each of the four runs (à 10 trials) separately. Pattern vectors of all runs but one

443 (“training data”) were used to train a support vector regression (SVR) model, as implemented
 444 in LIBSVM (<http://www.csie.ntu.edu.tw/~cjlin/libsvm>; Chang and Lin, 2011) using a linear
 445 kernel (nu-SVR) and a fixed regularization parameter ($c = 1$). This provided the basis of the
 446 following prediction of the donation amounts of the remaining run (“test data”) based on their
 447 neural response patterns. The procedure was repeated four times, always using a different run
 448 as a test dataset (leave-one-run-out cross-validation). Splitting the dataset into training and
 449 test datasets and run-wise cross-validation is a measure to control for potential problems of
 450 overfitting (Poldrack et al., 2020). The amount of predictive information on generosity was
 451 defined as the average Fisher’s Z-transformed correlation coefficient between the donations
 452 predicted by the SVR model and participant’s actual donations in these trials (Kahnt et al.,
 453 2014; Tusche et al., 2016). This predictive-accuracy value was then assigned to the central
 454 voxel of the searchlight cluster, and the procedure was repeated for every voxel of the
 455 acquired brain volume, resulting in a 3D map of average predictive accuracies for each
 456 participant.

457 These decoding maps were smoothed with a Gaussian kernel (8 mm FWHM) and
 458 submitted to a random-effects group analysis to identify brain regions that encode trial-wise
 459 donations across participants (simple t-test against baseline as implemented in SPM12). For
 460 this whole-brain analysis, we applied a cluster-forming threshold of $P \leq 0.001$, FWE-
 461 corrected for multiple comparisons at the cluster level ($P_{\text{FWE}} \leq 0.05$). These analysis steps
 462 were then repeated for post-stress donation blocks (GLM1_{post}) to examine the impact of
 463 stress on neural value representations (see below).

464 *Mentalizing and neural value representations.* Having identified activation patterns
 465 that decode donations, we proceeded to the next question. Does predictive information in
 466 (some of) these areas vary for people with high and low mentalizing capacity? To address
 467 this question, we identified high and low mentalizers based on the accuracy in mentalizing-

468 related questions in the independent EmpaToM task (median-split). Note that this task does
 469 not require prosocial decision-making and assesses the general capacity to mentalize. Next,
 470 we tested for a difference in predictive neural information on donations between high and
 471 low mentalizers ($MENT_{high} > MENT_{low}$) using a two-sample t-test. We restricted this test to
 472 brain areas previously shown to robustly predict donations on the group level (whole-brain
 473 decoding of donations, see above). The statistical test of decoding maps (of donations) for
 474 high vs. low mentalizers was performed at a more lenient statistical threshold of $P \leq 0.005$
 475 (and peak- $P \leq 0.001$) and an extent threshold of $k \geq 40$ voxels. Note that this analysis is
 476 designed to identify regions of interest (ROIs) for the subsequent test of stress on donations
 477 on the neural level. Thus, we opted against using more stringent statistical thresholds and
 478 corrections for multiple test comparisons that were used for the rest of our main analyses.

479 *Impact of stress on the mentalizing-valuation-relationship.* To understand the impact
 480 of stress on this interplay between mentalizing capacity and neural value coding, we
 481 examined the effect of acute stress on predictive neural information identified above.
 482 Specifically, we created spherical ROIs (5mm radius) around the peaks of predictive
 483 information that varied as a function of mentalizing capacity prior to the stress manipulation
 484 (see High MENT > Low MENT contrast in Table 2). These brain regions were identified
 485 irrespective of stress effects on donations. Thus, the resulting ROIs are fully independent,
 486 mitigating the risk of circular analysis and double-dipping (Kriegeskorte et al., 2009). ROI
 487 masks are provided on <https://osf.io/u46yj/>. For each spherical ROI, we estimated the average
 488 decoding accuracy on donations for the pre-stress donations (decoding based on $GLM1_{pre}$)
 489 and the post-stress donation sessions (decoding based on $GLM1_{post}$), respectively, which
 490 served as dependent variables in our statistical models on stress- and cortisol-related effects.
 491 Given an observed cortisol-related effect on donations (see Results), we employed a GLMM
 492 (MVPA – Full Model) to predict decoding accuracies on donations for each of our four ROIs

(i.e., as separate GLMMs with identical predictors). The predictors – Δ cortisol, mentalizing capacity, time, and their interactions – perfectly matched those of the behavioral GLMM (Choice – Full Model 2). We also used the same identical 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 testing for cortisol-related or group-related changes in the decoding maps of the pre- versus post-manipulation phase. To this end, we ran another whole-brain searchlight-analysis on the post-session ($GLM1_{post}$), identical to our initial analysis for the pre-session ($GLM1_{pre}$), and created difference maps (post minus pre) that served as dependent variables to GLMs with the (demeaned) predictors Δ cortisol (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 v. 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 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). $GLM2_{pre}$ and $GLM2_{post}$ included a regressor denoting the decision phases per session (R1) and a parametric regressor denoting donation amounts (R2). Furthermore, two additional regressors

518 modeled the reading phases (R3) and the rating phases (R4) as regressors of no interest. Six
 519 movement parameters were again included as nuisance regressors (R5-R10).

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

528

529 **Data and Code Availability**

530 Behavioral and (aggregated) fMRI data, group-level decoding maps, region-of-interest (ROI)
 531 masks, Matlab scripts for the SVR decoding analysis (whole-brain searchlight and ROI-based
 532 approach), and task material (instructions, charity descriptions) are publicly available on the
 533 project's Open Science Framework (OSF) page (<https://osf.io/u46yj/>). The SVR decoding
 534 analysis was implemented using the free LIBSVM toolbox
 535 (<http://www.csie.ntu.edu.tw/~cjlin/libsvm>; Chang and Lin, 2011) and Matlab R2019a
 536 (Mathworks). Further material will be available from the corresponding author upon request.

537

538

539

540

541

542

543 Results

544 **Manipulation Check I: Subjective and Physiological Stress Responses.** As a manipulation
545 check, we first examined the effectiveness of the experimental stress manipulation in terms of
546 subjective feelings, sympathetic and glucocorticoid (cortisol) reactivity. At the subjective
547 level, the TSST was experienced as significantly more stressful, unpleasant, and difficult than
548 the control condition (all P s < 0.001; see Table 1).

549 At the psychophysiological level, the TSST induced strong sympathetic arousal, as
550 indicated by a significant increase in systolic and diastolic blood pressure as well as pulse
551 compared to the control condition [group $\times$ time interaction for systolic blood pressure:
552 $F(2.902, 81.259) = 17.061, P < 0.001, \eta_p^2 = 0.379$; for diastolic blood pressure: $F(4, 112) =$
553 $20.391, P < 0.001, \eta_p^2 = 0.421$; for pulse: $F(2.816, 78.859) = 14.689, P < 0.001, \eta_p^2 = 0.344$].
554 Pairwise post-hoc comparisons revealed that blood pressure and pulse were significantly
555 elevated during the TSST relative to the control condition (all P s < 0.001), but not at other
556 time points of measurement (all P s ≥ 0.08 , in particular not in the pre-manipulation
557 measurements, all P s ≥ 0.47 ; Table 1), as expected for a transient sympathetic activation.

558 Furthermore, while there was a significant decrease in salivary cortisol in the control
559 group across the experimental session due to the circadian rhythm of cortisol [time: $F(2.239,$
560 $73.876) = 13.275, P < 0.001, \eta_p^2 = 0.29$], cortisol was significantly increased after the TSST,
561 relative to the control condition [group $\times$ time: $F(2.239, 73.876) = 10.424, P < 0.001, \eta_p^2 =$
562 0.24]. This increase peaked at the +20 min measurement right before the post-session of the
563 donation task (see Figure 2 and Table 1). Post-hoc comparisons revealed significant group
564 differences right before ($t(33) = 3.117, P = 0.004$) and after the post-session ($t(33) = 4.2, P <$
565 0.001), indicating higher glucocorticoid activity in the stress than the control group
566 throughout the whole donation task in the post-session. In contrast, the stress and control
567 group did not differ in cortisol levels at both time points before the experimental

manipulation (both P s > 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 ($r = -0.16$, $P = 0.36$) and is not related to mentalizing capacity ($r = -0.03$, $P = 0.88$). The cortisol responder rate (i.e., percentage of participants with a cortisol increase of >2 nmol/l from the pre-manipulation 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: min-max: -1.26 to 11.48 nmol/l, range: 12.74 nmol/l; control: min-max: -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- 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 prior to any experimental stress manipulation (i.e., in the pre-phase), we observed substantial variability in average donations across participants (minimum: €2.80; maximum: €17.45; mean $\pm$ SD: €12.07 $\pm$ €3.63) as well as in contributions to different charities within individuals (mean of participants' 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 (see Figure 1B), making it unlikely that the 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

593 mentalizing and prosocial behaviors such as charitable giving (Waytz et al., 2012; Tusche et
 594 al., 2016; Bellucci et al., 2020). Before turning to our main analyses, we therefore checked
 595 whether this relationship is also reflected in the present data. In a first step, we examined data
 596 in the independent EmpaToM task to assess the variance in mentalizing capacity across
 597 participants. On average, participants performed well in the mentalizing-related questions of
 598 the EmpaToM (mean accuracy $\pm$ SD: 66.67% $\pm$ 16.17%; for the distribution of scores, see
 599 Figure 3). Despite random group assignment, the stress group displayed a higher baseline
 600 mentalizing performance than the control group (72.22% vs. 60.78%, $P = 0.034$). However,
 601 our main models explicitly accounted for this variable as a covariate/predictor (see below).
 602 Next, in line with previous research, baseline mentalizing capacity in the EmpaToM
 603 correlated positively with average generosity (subject-wise mean donations) across sessions
 604 ($r(33) = 0.351$, $P = 0.039$) and in both sessions separately (pre: $r(33) = 0.313$, $P = 0.034$ [one-
 605 tailed]; post: $r(33) = 0.351$, $P = 0.025$; not significantly different from each other, $P = 0.168$).
 606 Moreover, mentalizing capacity in the EmpaToM was positively related to average self-
 607 reported mentalizing (i.e., perspective-taking ratings) in the donation task ($r(33) = 0.325$, $P =$
 608 0.029 [one-tailed]). Thus, participants with higher mentalizing performance in the
 609 independent task tended to recruit mentalizing more strongly in the donation task as well.
 610 Subjects also displayed considerable degrees of empathic reactivity (1.82 ± 0.85 [difference
 611 score from -6 to 6 for emotional > neutral videos]) and compassion (3.13 ± 0.64 [6-point
 612 scale]) in the EmpaToM (see also Figure 3) with no significant differences between groups
 613 ($P_s > 0.202$). However, variance in these affective measures was not significantly associated
 614 with overall generosity in the donation task (compassion: $r(33) = 0.234$, $P = 0.176$; empathy:
 615 $r(33) = 0.2$, $P = 0.249$). Hence, mentalizing capacity was the only robust EmpaToM predictor
 616 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 prior to any stress manipulation, participants reported varying degrees of perceived mentalizing (5.71 ± 1.10 [9-point scale]), empathy (mean $\pm$ 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 socio-cognitive and -affective processes 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 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 s > 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 4A illustrates average donations in both groups over time. 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 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 *Methods*). Contrary to our hypothesis, we did not observe a significant change in donations after the stress manipulation ($group \times time$ interaction: $F(1, 62) = 0.996, P = 0.322$), nor a moderation of this effect by mentalizing capacity ($group \times time \times mentalizing$ interaction: $F(1, 62) = 0.414, P = 0.522$). There was also no significant overall group difference in donations ($M_{\text{stress}} = €13.18; M_{\text{control}} = €10.8$; main effect of *group*: $F(1, 62) = 2.001, P = 0.162$), no significant main effect of *time* across groups ($F(1, 62) = 0.035, P = 0.853$), nor any other significant effects (all $P_s \geq 0.114$).

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 ($\Delta\text{cortisol}$, captured in the AUCi), *mentalizing capacity* (EmpaToM), or an interaction of $\Delta\text{cortisol} \times \text{mentalizing}$ 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 varied strongly over time in both experimental groups (see Figure 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

667 accounts for potential differential effects of cortisol or mentalizing capacity across groups, it
 668 showed an inferior model fit (BIC: 410.79 vs. 402.56 for the simpler model; AIC: 407.05 vs.
 669 398.51), suggesting a general effect across groups. Accounting for potential gender
 670 differences by including gender as an additional predictor also resulted in an inferior model
 671 fit (hence, there is no evidence in favor of systematic gender differences in our [limited]
 672 sample). Based on these model comparisons, we report the results of the simpler model in the
 673 following and in Table 2 (Choice – Full Model 2).

674 We hypothesized that increases in cortisol would be associated with reduced altruism
 675 over time (for a simple bivariate relationship, also see Figure 4B), and that this association
 676 might be moderated by mentalizing capacity. In line with this notion, the GLMM revealed a
 677 significant $\Delta\text{cortisol} \times \text{mentalizing} \times \text{time}$ interaction ($F(1, 62) = 5.174, P = 0.026$). To
 678 decompose this 3-way interaction, we fitted two follow-up Generalized Linear Models for the
 679 pre- and post-phase separately (see Table 2, Decomposition PRE & POST) and for change
 680 scores (Decomposition POST-PRE). These decomposition models also included $\Delta\text{cortisol}$,
 681 *mentalizing capacity*, and their *interaction* as predictors. In the pre-phase decomposition,
 682 there were no significant predictors (all $P_s \geq 0.083$), although by tendency mentalizing
 683 capacity positively predicted charitable giving ($B = 5.870, SE = 3.527, P = 0.096$). In contrast,
 684 for post-phase donations, we observed a significant positive effect of mentalizing capacity (B
 685 $= 7.834, SE = 3.747, P = 0.037$). More importantly, we also found a significant $\Delta\text{cortisol} \times$
 686 *mentalizing* interaction ($B = -0.059, SE = 0.025, P = 0.019$). The latter finding aligns with our
 687 hypothesis of a cortisol-related effect on altruistic choice that depends on individuals'
 688 mentalizing capacity.

689 This two-way interaction was decomposed further using a simple-slopes analysis
 690 (Preacher et al., 2006), which assesses the relationship between $\Delta\text{cortisol}$ and post-phase
 691 donations at different levels of mentalizing capacity (± 1 SD). Figure 4C illustrates the simple

692 slopes of the significant post-phase interaction. For comparison, it also illustrates the simple
 693 slopes for the non-significant interaction in the pre-phase and donation-change scores. While
 694 we observed a significant negative association between changes in cortisol and post-phase
 695 donations for high mentalizing capacity ($B_{\text{highMENT}(+1\text{SD})} = -0.011$, $\text{SE} = 0.005$, $P = 0.028$),
 696 there was no significant (and a numerically positive) cortisol-related association for low
 697 mentalizing capacity ($B_{\text{lowMENT}(-1\text{SD})} = 0.008$, $\text{SE} = 0.005$, $P = 0.101$; see Figure 4C [POST]).
 698 In other words, only for individuals with higher mentalizing capacity, we observed that
 699 increases in cortisol were associated with relative decreases in charitable giving. When
 700 examining pre-post changes directly (Figure 4C [POST-PRE]), donations even increased over
 701 time in high mentalizers under low $\Delta\text{cortisol}$. Yet, under high $\Delta\text{cortisol}$ this relationship
 702 reversed to decreased donations in these individuals.

703 For exploratory purposes, we also fitted identical group-based and cortisol-based
 704 GLMMs with the other two social capacity measures (*empathy* or *compassion*) and with the
 705 factual-reasoning measure obtained in the EmpaToM task as a predictor instead of
 706 mentalizing capacity (separate models). We did not observe any significant main effects or
 707 interactions regarding those predictors in these supplemental models (all $P_s \geq 0.107$). In
 708 particular, none of the $\Delta\text{cortisol} \times \text{empathy/compassion/factual-reasoning} \times \text{time}$ interactions
 709 reached significance (all $P_s \geq 0.156$). These exploratory findings support the notion of
 710 specificity of our main results: the association between cortisol and altruism was moderated
 711 uniquely by mentalizing capacity. We also did not observe an interaction effect for response
 712 times ($P = 0.734$; for response times adjusted for the initial position of the choice slider
 713 through a regression model: $P = 0.111$). Thus, any brain responses associated with the
 714 choice-related effect are unlikely merely due to differences in decision speed.

715 Finally, we also fitted two exploratory regression models using either a composite
 716 score of autonomic reactivity (average of Z-scored base-to-peak increases in blood pressure

717 and pulse) or a composite score of subjective stress ratings, together with mentalizing
 718 capacity and interaction terms as predictors to investigate the potential role of other stress
 719 parameters. We observed no significant effects related to these autonomic and subjective
 720 stress indices (all P s ≥ 0.114). In particular, the interaction of *time* $\times$ Δ *autonomic/subjective-*
 721 *stress* $\times$ *mentalizing* interactions (all P s ≥ 0.131) did not reach significance. These
 722 exploratory results might indicate a unique role of cortisol in the observed effects on altruism.

723

724 **Neural decoding of trial-by-trial variations in donations.** As a first step in our fMRI
 725 analysis, we used a multivariate whole-brain searchlight analysis to identify brain regions that
 726 encode trial-by-trial variations in donations before any stress manipulation (i.e., pre-phase).
 727 In line with previous research (Tusche et al., 2016; Bellucci et al., 2020), we observed a
 728 range of brain areas that reliably decoded donations (Figure 5 and Table 3). Notably, this
 729 included regions previously associated with mentalizing such as the bilateral middle temporal
 730 gyrus (MTG)/temporoparietal junction (TPJ), precuneus and dorsomedial prefrontal cortex
 731 (DMPFC) (Kanske et al., 2015; Schurz et al., 2020).

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

739 We then turned to investigate potential stress-induced or cortisol-related effects on
 740 neural value coding (see next section). To this end, we created ROIs (also see Methods)
 741 based on the results reported above. Specifically, we created spherical ROIs (5mm-radius)

centered at the activation peaks of brain areas in which information predictive of donations varied for high and low mentalizers (for peak coordinates, see Table 3, High MENT > Low MENT). Note that the construction of the ROIs was based on *pre*-phase data only (i.e., before any stress manipulation) and thus 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 employed GLMMs to predict decoding accuracies on donations for each of our four ROIs (separate GLMMs) based on time, Δ cortisol, mentalizing capacity, and their interaction (matching the predictors of the behavioral GLMM). Mirroring our behavioral results, this model revealed a significant Δ cortisol $\times$ mentalizing $\times$ time interaction ($F(1, 62) = 9.347, P = 0.003$) for decoding accuracies in the rDLPFC (MVPA – Full Model in Table 4; for an illustration of pre- and post SVR decoding accuracies, see Figure 6A). 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 3-way interaction, we again fitted two follow-up Generalized Linear Models for the pre- and post-phase separately (see Decomposition PRE & POST in Table 4) and for change scores (Decomposition POST-PRE). These again included Δ cortisol, mentalizing capacity, and their interaction as predictors. Matching our behavioral findings, we did not observe a significant Δ cortisol $\times$ mentalizing interaction for pre-phase decoding accuracies in the rDLPFC ($B = 0.002, SE = 0.001, P = 0.096$). However, this interaction was significant for the post-phase ($B = -0.003, SE = 0.001, P = 0.011$). Importantly, the interaction effect for the post-phase is also significantly stronger

766 than for the pre-phase in an identical model on the change scores ($B = -0.004$, $SE = 0.001$, P
767 $= 0.002$), in line with the interaction with time in our full GLMM above.

768 The significant $\Delta cortisol \times mentalizing$ interaction for post-phase decoding accuracies
769 was again decomposed using a simple-slopes analysis (Figure 6B [POST]). While we
770 observed a significant negative association between changes in cortisol and post-phase
771 decoding accuracies in the rDLPFC for high mentalizing capacity ($B_{highMENT(+1SD)} = -0.001$,
772 $SE = 0.0003$, $P = 0.007$), there was no significant cortisol-related association for low
773 mentalizing capacity ($B_{lowMENT(-1SD)} = 0.0002$, $SE = 0.0002$, $P = 0.304$). In other words, only
774 for high mentalizers, we observed that increased cortisol was associated with decreased
775 neural representations of donations in the rDLPFC (also see Figure 6B [POST-PRE]).

776 No further significant $\Delta cortisol \times mentalizing \times time$ interactions (all $P_s > 0.276$) or
777 other cortisol-related effects (all $P_s > 0.096$) were detected in the GLMMs for any of the
778 other ROIs. Likewise, no additional brain region was identified in our exploratory set of
779 group-based ROI and whole-brain analyses. Furthermore, there were no significant cortisol-
780 or group-related effects in our supplemental univariate fMRI analysis.

781 As a sanity check, we used a permutation test to assess whether the rDLPFC reliably
782 encoded donations in both the pre- and post-phase and to ensure that the results did not
783 emerge by chance. Specifically, for each participant and per phase (i.e. pre vs. post),
784 permutation distributions were created by breaking up the mapping of donations and neural
785 response patterns (10000-fold). We then compared the average “real” decoding accuracies
786 (i.e., ROI-wise mean across participants) to the sampled permutation distributions. We
787 observed that the rDLPFC reliably encoded donations in both the pre-phase ($r = 0.14$, $P_{perm} <$
788 0.0001) and the post-phase ($r = 0.09$, $P_{perm} = 0.0015$).

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

792

793 **The negative cortisol-altruism association is mediated by reduced value coding in the**
 794 **DLPFC in high mentalizers.** So far, we observed that increases in cortisol were associated
 795 with (i) decreases in charitable giving, and (ii) decreases in donation-decoding accuracies in
 796 the rDLPFC, but only for individuals with higher mentalizing capacity. Furthermore, using
 797 robust regression to establish a brain-behavior link, we show that pre-post changes in
 798 decoding accuracies in the DLPFC positively predicted pre-post changes in charitable giving
 799 ($B = 1.617, P = 0.014$). In other words, decreases in decoding accuracies from the pre- to
 800 post-session were associated with decreases in charitable giving (Figure 6C).

801 This raises the question of whether these observations are directly linked. Can the
 802 association between changes in cortisol and donations be explained (i.e., was it mediated) by
 803 changes in neural donation-value decoding? Second, is this mediation moderated by
 804 mentalizing capacity (i.e., present for high mentalizers only)? A moderated mediation model
 805 provided evidence in favor of these conjectures (Figure 6D). We observed that pre-post
 806 changes in SVR-decoding accuracy in the rDLPFC mediated the positive association between
 807 pre-post changes in cortisol and donations for participants with high ($\beta_{\text{highMENT}(+1\text{SD})} = -0.27$,
 808 $SE = 0.19$, 90% CI [-0.62, -0.01]) and medium mentalizing capacity ($\beta_{\text{mediumMENT}(\text{mean})} = -0.1$,
 809 $SE = 0.09$, 90% CI [-0.27, -0.0002]), but not for participants with lower baseline mentalizing
 810 capacity ($\beta_{\text{lowMENT}(-1\text{SD})} = 0.08$, $SE = 0.1$, 90% CI [-0.09, 0.22]) (90% CIs excluding 0 reflect
 811 $P < 0.05$ [one-tailed]). In other words, only in high mentalizers, we observed a significant
 812 decline in donations over time following increases in cortisol, mediated by decreases in
 813 multivariate neural value representations for donations in the rDLPFC.

814 **Discussion**

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

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

836 The identified negative association between increasing cortisol and charitable giving
 837 in high (but not low) mentalizers might indicate a cortisol-related disruption of mentalizing-
 838 related cognitive processes that otherwise would contribute to altruistic choice. This notion is

839 consistent with evidence suggesting that acute stress and cortisol in particular can impair
 840 mentalizing (Smeets et al., 2009; Leder et al., 2013), but these observations have not yet been
 841 linked to similar disruptions of altruism (Starcke et al., 2011). Our observation that cortisol-
 842 related decrements in altruism depend on mentalizing capacity indicates such a link. This also
 843 demonstrates that (independent) task-based measures of individuals' general capacity to
 844 mentalize present a unique angle to study the role of stress (hormones) in altruism.

845 On the neural level, we established a link between a cortisol-related disruption of
 846 value coding in the rDLPFC and reduced altruism in high mentalizers. In theory, there are
 847 two possibilities how DLPFC functioning could be linked to mentalizing. First, DLPFC
 848 activity might directly reflect *core* mentalizing processes. This notion is in line with our
 849 observation that – before any stress manipulation – high mentalizers displayed higher
 850 rDLPFC donation-decoding accuracies. Moreover, brain-stimulation studies suggest that the
 851 DLPFC causally contributes to mentalizing (Costa et al., 2008; Kalbe et al., 2010). Second,
 852 the DLPFC, among other regions, represents a “*co-opted*” system relevant to mentalizing
 853 (Siegal and Varley, 2002). This might also explain why DLPFC activity has not been
 854 consistently observed in fMRI meta-analyses of mentalizing tasks (Kogler et al., 2020;
 855 Schurz et al., 2020; but see Molenberghs et al., 2016). However, the DLPFC is frequently
 856 reported in neuroimaging studies on prosocial decision-making (Waytz et al., 2012; Tusche et
 857 al., 2016; Bellucci et al., 2020). The DLPFC has been suggested to contribute to altruistic
 858 choice via general-purpose and context-dependent cognitive control. Across social and non-
 859 social domains, the rDLPFC flexibly encodes values of choice options consistent with the
 860 current regulatory focus and goals. During altruistic choice, rDLPFC activation patterns
 861 reflect reduced inputs of self-related motives when individuals focus on others' thoughts and
 862 feelings (Tusche and Hutcherson, 2018). The rDLPFC is also involved in the controlled shift
 863 from a self-centered to an other-centered perspective (Thirioux et al., 2014). Based on our

864 data alone, we cannot be sure whether altered rDLPFC activity reflects changes in
 865 mentalizing or other decision-relevant processes. To note, the underlying process appears not
 866 to contribute to factual reasoning, given that we did not find similar effects for factual-
 867 reasoning capacity. Future studies might leverage a neurocomputational approach (Hampton
 868 et al., 2008; Tusche and Bas, 2021) to further delineate the mechanistic role of the rDLPFC in
 869 mentalizing and mediating cortisol-related effects on altruism.

870 Mentalizing-related processes – whether core or co-opted – are not the only
 871 contributors to altruism. Empathy and compassion are potent affective drivers (Batson et al.,
 872 2015; Tusche et al., 2016; Böckler et al., 2018). Acute stress can *increase* empathy and
 873 prosociality. For instance, one study found stress-enhanced activity in the empathy network
 874 (AI and aMCC), which predicted altruistic choices in an independent dictator game (Tomova
 875 et al., 2017). This is consistent with other reports of increased altruism under stress or
 876 elevated cortisol (von Dawans et al., 2012; Singer et al., 2017; Margittai et al., 2018), but it is
 877 unclear whether altered empathy mediated this effect. Notably, we did not find any evidence
 878 for stress- or cortisol-associated *increases* in altruism. Likewise, there were no interaction
 879 effects with baseline empathy and compassion in the EmpaToM.

880 The seemingly inconsistent effects of acute stress or glucocorticoids on altruism
 881 might be explained by the influence of context, individual factors, and their interaction. For
 882 instance, acute stress is thought to enhance altruism particularly when the target's need is
 883 salient (Buchanan and Preston, 2014), consistent with enhanced empathy when actually
 884 seeing others receiving painful treatment (Tomova et al., 2017). Differences in salience might
 885 also explain why empathy and compassion ratings in the EmpaToM (salient videos) did not
 886 significantly – though descriptively positively – predict donations (less salient charity-texts).
 887 Social closeness might also play a role as it can increase empathy (Engert et al., 2014) and
 888 moderate the stress-altruism relationship (Singer et al., 2021). However, perceived social

889 closeness did not emerge as a significant predictor of donations and neural responses in the
 890 donation task (Tusche et al., 2016). Text-based stimuli might generally induce less perceived
 891 closeness than other more salient stimuli for which it may play a more critical role. Moreover,
 892 whereas some individuals display a high general propensity to empathize, others strongly
 893 engage in mentalizing. These two capacities are independent of each other on a behavioral
 894 and neural level (Kanske et al., 2016). Hence, while our data indicate cortisol-related
 895 decrements in altruism in high mentalizers, stronger empathizers might show opposite effects,
 896 particularly when the need of others is salient. Compassionate individuals might display still
 897 other context-dependent effects, given that compassion has unique neural correlates (e.g., in
 898 reward-related regions; Klimecki et al., 2014; Kanske et al., 2015). Future studies might
 899 benefit from advancing this situation-person-interaction perspective by comparing different
 900 contexts in relation to individual traits and states. A similar perspective might ultimately
 901 inform target-specific interventions to alleviate stress-related social disruptions in clinical,
 902 economic, and other settings. For instance, stress-prone mentalizers may benefit from stress-
 903 reduction treatments and a (compensatory) training of their mentalizing abilities to avoid
 904 stress translating to deficits in prosociality. It would also be interesting to investigate how
 905 altruistic choice in the laboratory (though incentivized) relates to real-world charity and other
 906 forms of altruism, or vice versa, whether they are important determinants of behavior in the
 907 laboratory.

908 The post-phase of our donation task matched a phase of the stress response
 909 characterized by non-genomic cortisol action (<1h following stressor onset; Hermans et al.,
 910 2014; Joëls et al., 2018). Hence, the observed cortisol-related effects might be explainable via
 911 this mode of action. In contrast, sympathetic activity returned to baseline before post-phase
 912 donations and was unrelated to altruistic choice. To control for potential influences of other
 913 (co-varying) stress-related factors and to provide evidence that enhanced cortisol *causally*

914 decreases neural value coding and altruism in high mentalizers, future studies could use
 915 pharmacological manipulations of cortisol (and noradrenergic) activity (Metz et al., 2020).
 916 Furthermore, other stress components may exert (differential) effects on altruistic choice at
 917 different time scales. While the influence of catecholamines (e.g., on prefrontal functioning;
 918 Arnsten, 2009) might be stronger in an earlier phase, genomic (vs. non-genomic) effects of
 919 cortisol come into play only later (Singer et al., 2021). Future experimental designs might
 920 leverage different time scales to assess different phases of the stress response. Interestingly,
 921 the observed cortisol-related effect on altruism was not specific to the TSST condition.
 922 Instead, variability in stress-hormone changes across both groups explained variability in
 923 donations. This effect, however, did not translate into a group difference in altruism, despite
 924 elevated cortisol in the stress group. This might be explained by a considerable overlap in the
 925 group distributions of cortisol changes and the moderation of the cortisol effect by
 926 mentalizing capacity, which itself varies across participants of both groups. We argue,
 927 however, that our findings are still relevant to stress contexts, given cortisol elevations after
 928 the TSST.

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

936
 937
 938

939 **References**

- 940 Arnsten AFT (2009) Stress signalling pathways that impair prefrontal cortex structure and
 941 function. *Nat Rev Neurosci* 10:410–422.
- 942 Batson CD, Lishner DA, Stocks EL (2015) The empathy-altruism hypothesis. In: *The Oxford*
 943 *handbook of prosocial behavior* (Schroeder DA, Graziano WG, eds), pp 259–281.
 944 Oxford University Press.
- 945 Bellucci G, Camilleri JA, Eickhoff SB, Krueger F (2020) Neural signatures of prosocial
 946 behaviors. *Neurosci Biobehav Rev* 118:186–195.
- 947 Böckler A, Tusche A, Schmidt P, Singer T (2018) Distinct mental trainings differentially
 948 affect altruistically motivated, norm motivated, and self-reported prosocial behaviour.
 949 *Sci Rep* 8:1–14.
- 950 Böckler A, Tusche A, Singer T (2016) The Structure of Human Prosociality: Differentiating
 951 Altruistically Motivated, Norm Motivated, Strategically Motivated, and Self-Reported
 952 Prosocial Behavior. *Soc Psychol Personal Sci* 7:530–541.
- 953 Bogdanov M, Schwabe L (2016) Transcranial stimulation of the dorsolateral prefrontal
 954 cortex prevents stress-induced working memory deficits. *J Neurosci* 36:1429–1437.
- 955 Buchanan TW, Preston SD (2014) Stress leads to prosocial action in immediate need
 956 situations. *Front Behav Neurosci* 8:1–6.
- 957 Burkart JM, Allon O, Amici F, Fichtel C, Finkenwirth C, Heschl A, Huber J, Isler K,
 958 Kosonen ZK, Martins E, Meulman EJ, Richiger R, Rueth K, Spillmann B,
 959 Wiesendanger S, Van Schaik CP (2014) The evolutionary origin of human hyper-
 960 cooperation. *Nat Commun* 5:1–9.
- 961 Chang CC, Lin CJ (2011) LIBSVM: A Library for support vector machines. *ACM Trans*
 962 *Intell Syst Technol* 2.
- 963 Costa A, Torriero S, Oliveri M, Caltagirone C (2008) Prefrontal and temporo-parietal
 964 involvement in taking others' perspective: TMS evidence. *Behav Neurol* 19:71–74.
- 965 Devilbiss DM, Spencer RC, Berridge CW (2017) Stress Degrades Prefrontal Cortex Neuronal
 966 Coding of Goal-Directed Behavior. *Cereb Cortex* 27:2970–2983.
- 967 Edwards S, Clow A, Evans P, Hucklebridge F (2001) Exploration of the awakening cortisol
 968 response in relation to diurnal cortisol secretory activity. *Life Sci* 68:2093–2103.
- 969 Engert V, Plessow F, Miller R, Kirschbaum C, Singer T (2014) Cortisol increase in empathic
 970 stress is modulated by emotional closeness and observation modality.
 971 *Psychoneuroendocrinology* 45:192–201.
- 972 Faul F, Erdfelder E, Buchner A, Lang A-G (2009) Statistical power analyses using G*Power
 973 3.1: Tests for correlation and regression analyses. *Behav Res Methods* 41:1149–1160.
- 974 Hampton AN, Bossaerts P, O'Doherty JP (2008) Neural correlates of mentalizing-related
 975 computations during strategic interactions in humans. *Proc Natl Acad Sci* 105:6741–
 976 6746.
- 977 Hare TA, Camerer CF, Knoepfle DT, Rangel A (2010) Value computations in ventral medial
 978 prefrontal cortex during charitable decision making incorporate input from regions
 979 involved in social cognition. *J Neurosci* 30:583–590.
- 980 Hautzinger M, Keller F, Kühner C (2006) Beck Depressions-Inventar (BDI-II). Revision.
 981 Frankfurt/Main: Harcourt Test Services.
- 982 Hayes AF (2018) Introduction to mediation, moderation, and conditional process analysis.
- 983 Haynes J-D, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE (2007) Reading hidden
 984 intentions in the human brain. *Curr Biol* 17:323–328.
- 985 Hermans EJ, Henckens MJAG, Joëls M, Fernández G (2014) Dynamic adaptation of large-
 986 scale brain networks in response to acute stressors. *Trends Neurosci* 37:304–314.

- 987 Hildebrandt MK, Jauk E, Lehmann K, Maliske L, Kanske P (2021) Brain activation during
 988 social cognition predicts everyday perspective-taking: A combined fMRI and ecological
 989 momentary assessment study of the social brain. *Neuroimage* 227:117624.
- 990 Joëls M, Karst H, Sarabdjitsingh RA (2018) The stressed brain of humans and rodents. *Acta*
 991 *Physiol* 223:1–10.
- 992 Kahnt T, Park SQ, Haynes JD, Tobler PN (2014) Disentangling neural representations of
 993 value and salience in the human brain. *Proc Natl Acad Sci U S A* 111:5000–5005.
- 994 Kalbe E, Schlegel M, Sack AT, Nowak DA, Dafotakis M, Bangard C, Brand M, Shamay-
 995 Tsoory S, Onur OA, Kessler J (2010) Dissociating cognitive from affective theory of
 996 mind: a TMS study. *Cortex* 46:769–780.
- 997 Kanske P, Böckler A, Trautwein F-M, Singer T (2015) Dissecting the social brain:
 998 Introducing the EmpaToM to reveal distinct neural networks and brain-behavior
 999 relations for empathy and Theory of Mind. *Neuroimage* 122:6–19.
- 1000 Kanske P, Böckler A, Trautwein FM, Leemann FHP, Singer T (2016) Are strong
 1001 empathizers better mentalizers? Evidence for independence and interaction between the
 1002 routes of social cognition. *Soc Cogn Affect Neurosci* 11:1383–1392.
- 1003 Kirschbaum C, Pirke KM, Hellhammer DH (1993) The “Trier Social Stress Test” - a tool for
 1004 investigating psychobiological stress responses in a laboratory setting.
 1005 *Neuropsychobiology* 28:76–81.
- 1006 Klimecki OM, Leiberg S, Ricard M, Singer T (2014) Differential pattern of functional brain
 1007 plasticity after compassion and empathy training. *Soc Cogn Affect Neurosci* 9:873–879.
- 1008 Kogler L, Müller VI, Werninghausen E, Eickhoff SB, Derntl B (2020) Do I feel or do I
 1009 know? Neuroimaging meta-analyses on the multiple facets of empathy. *Cortex* 129:341–
 1010 355.
- 1011 Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping.
 1012 *Proc Natl Acad Sci* 103:3863–3868.
- 1013 Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems
 1014 neuroscience: The dangers of double dipping. *Nat Neurosci* 12:535–540.
- 1015 Kudielka BM, Hellhammer DH, Kirschbaum C (2007) Ten Years of Research with the Trier
 1016 Social Stress Test - Revisited. In: *Social Neuroscience: Integrating biological and*
 1017 *psychological explanations of social behavior* (Harmon-Jones E, Winkielman P, eds), pp
 1018 56–83. The Guilford Press, New York.
- 1019 Lamm C, Decety J, Singer T (2011) Meta-analytic evidence for common and distinct neural
 1020 networks associated with directly experienced pain and empathy for pain. *Neuroimage*
 1021 54:2492–2502.
- 1022 Leder J, Häusser JA, Mojzisch A (2013) Stress and strategic decision-making in the beauty
 1023 contest game. *Psychoneuroendocrinology* 38:1503–1511.
- 1024 Leiner DJ (2020) SoSci Survey (Version 3.2.05-i) [Computer software]. Available at
 1025 <https://www.sosicurvey.de>.
- 1026 Libby LA, Hannula DE, Ranganath C (2014) Medial temporal lobe coding of item and spatial
 1027 information during relational binding in working memory. *J Neurosci* 34:14233–14242.
- 1028 Lockwood PL (2016) The anatomy of empathy: Vicarious experience and disorders of social
 1029 cognition. *Behav Brain Res* 311:255–266.
- 1030 Lovallo WR, Cohoon AJ, Acheson A, Vincent AS, Sorocco KH (2019) Cortisol stress
 1031 reactivity in women, diurnal variations, and hormonal contraceptives: studies from the
 1032 Family Health Patterns Project. *Stress* 22:421–427.
- 1033 Margittai Z, van Wingerden M, Schnitzler A, Joëls M, Kalenscher T (2018) Dissociable roles
 1034 of glucocorticoid and noradrenergic activation on social discounting.
 1035 *Psychoneuroendocrinology* 90:22–28.
- 1036 Metz S, Waiblinger-Grigull T, Schulreich S, Chae WR, Otte C, Heckerlen HR, Wingenfeld K

- 1037 (2020) Effects of hydrocortisone and yohimbine on decision-making under risk.
 1038 Psychoneuroendocrinology 114:8–11.
- 1039 Molenberghs P, Johnson H, Henry JD, Mattingley JB (2016) Understanding the minds of
 1040 others: A neuroimaging meta-analysis. *Neurosci Biobehav Rev* 65:276–291.
- 1041 Nater UM, La Marca R, Florin L, Moses A, Langhans W, Koller MM, Ehlert U (2006)
 1042 Stress-induced changes in human salivary alpha-amylase activity - Associations with
 1043 adrenergic activity. *Psychoneuroendocrinology* 31:49–58.
- 1044 Nitschke JP, Sunahara CS, Carr EW, Winkelman P, Pruessner JC, Bartz JA (2020) Stressed
 1045 connections: cortisol levels following acute psychosocial stress disrupt affiliative
 1046 mimicry in humans. *Proc R Soc B Biol Sci* 287.
- 1047 Obeso I, Moisa M, Ruff CC, Dreher JC (2018) A causal role for right temporo-parietal
 1048 junction in signaling moral conflict. *Elife* 7:1–16.
- 1049 Poldrack RA, Huckins G, Varoquaux G (2020) Establishment of Best Practices for Evidence
 1050 for Prediction: A Review. *JAMA Psychiatry* 77:534–540.
- 1051 Preacher KJ, Curran PJ, Bauer DJ (2006) Computational tools for probing interactions in
 1052 multiple linear regression, multilevel modeling, and latent curve analysis. *J Educ Behav*
 1053 *Stat* 31:437–448.
- 1054 Pruessner JC, Kirschbaum C, Meinlschmid G, Hellhammer DH (2003) Two formulas for
 1055 computation of the area under the curve represent measures of total hormone
 1056 concentration versus time-dependent change. *Psychoneuroendocrinology* 28:916–931.
- 1057 Qin S, Hermans EJ, van Marle HJF, Luo J, Fernández G (2009) Acute Psychological Stress
 1058 Reduces Working Memory-Related Activity in the Dorsolateral Prefrontal Cortex. *Biol*
 1059 *Psychiatry* 66:25–32.
- 1060 Sandi C, Haller J (2015) Stress and the social brain: Behavioural effects and neurobiological
 1061 mechanisms. *Nat Rev Neurosci* 16:290–304.
- 1062 Schulreich S, Schwabe L (2021) Causal Role of the Dorsolateral Prefrontal Cortex in Belief
 1063 Updating under Uncertainty. *Cereb Cortex* 31:184–200.
- 1064 Schulz P, Schlotz W (1999) The Trier Inventory for the Assessment of Chronic Stress
 1065 (TICS): Scale construction, statistical testing, and validation of the scale work overload.
 1066 *Diagnostica* 45:8–19.
- 1067 Schurz M, Radua J, Tholen MG, Maliske L, Margulies DS, Mars RB, Sallet J, Kanske P
 1068 (2020) Toward a hierarchical model of social cognition: A neuroimaging meta-analysis
 1069 and integrative review of empathy and theory of mind. *Psychol Bull*.
- 1070 Schwabe L, Haddad L, Schachinger H (2008) HPA axis activation by a socially evaluated
 1071 cold-pressor test. *Psychoneuroendocrinology* 33:890–895.
- 1072 Siegal M, Varley R (2002) Neural systems involved in “theory of mind.” *Nat Rev Neurosci*
 1073 3:463–471.
- 1074 Singer N, Binapfl J, Sommer M, Wüst S, Kudielka BM (2021) Everyday moral decision-
 1075 making after acute stress: do social closeness and timing matter? *Stress* 24:468–473.
- 1076 Singer N, Sommer M, Döhnelt K, Zänkert S, Wüst S, Kudielka BM (2017) Acute
 1077 psychosocial stress and everyday moral decision-making in young healthy men: The
 1078 impact of cortisol. *Horm Behav* 93:72–81.
- 1079 Smeets T, Dziobek I, Wolf OT (2009) Social cognition under stress: Differential effects of
 1080 stress-induced cortisol elevations in healthy young men and women. *Horm Behav*
 1081 55:507–513.
- 1082 Solanas MP, Vaessen M, De Gelder B (2020) Computation-based feature representation of
 1083 body expressions in the human brain. *Cereb Cortex* 30:6376–6390.
- 1084 Starcke K, Polzer C, Wolf OT, Brand M (2011) Does stress alter everyday moral decision-
 1085 making? *Psychoneuroendocrinology* 36:210–219.
- 1086 Tabachnik BB, Fidell LS (2013) Using multivariate statistics, 6. ed. Boston, MA: Pearson.

- 1087 Thirioux B, Mercier MR, Blanke O, Berthoz A (2014) The cognitive and neural time course
 1088 of empathy and sympathy: An electrical neuroimaging study on self-other interaction.
 1089 Neuroscience 267:286–306.
- 1090 Tholen MG, Trautwein FM, Böckler A, Singer T, Kanske P (2020) Functional magnetic
 1091 resonance imaging (fMRI) item analysis of empathy and theory of mind. Hum Brain
 1092 Mapp 41:2611–2628.
- 1093 Tomova L, Majdandžić J, Hummer A, Windischberger C, Heinrichs M, Lamm C (2017)
 1094 Increased neural responses to empathy for pain might explain how acute stress increases
 1095 prosociality. Soc Cogn Affect Neurosci 12:401–408.
- 1096 Tusche A, Bas LM (2021) Neurocomputational models of altruistic decision-making and
 1097 social motives: Advances, pitfalls, and future directions. WIREs Cogn Sci:1–29.
- 1098 Tusche A, Bockler A, Kanske P, Trautwein F-M, Singer T (2016) Decoding the Charitable
 1099 Brain: Empathy, Perspective Taking, and Attention Shifts Differentially Predict
 1100 Altruistic Giving. J Neurosci 36:4719–4732.
- 1101 Tusche A, Hutcherson CA (2018) Cognitive regulation alters social and dietary choice by
 1102 changing attribute representations in domain-general and domain-specific brain circuits.
 1103 Elife 7:1–35.
- 1104 Vinkers CH, Zorn J V., Cornelisse S, Koot S, Houtepen LC, Olivier B, Verster JC, Kahn RS,
 1105 Boks MPM, Kalenscher T, Joëls M (2013) Time-dependent changes in altruistic
 1106 punishment following stress. Psychoneuroendocrinology 38:1467–1475.
- 1107 Vogel S, Fernández G, Joëls M, Schwabe L (2016) Cognitive Adaptation under Stress: A
 1108 Case for the Mineralocorticoid Receptor. Trends Cogn Sci 20:192–203.
- 1109 Vogel S, Klun LM, Fernández G, Schwabe L (2018) Stress affects the neural ensemble for
 1110 integrating new information and prior knowledge. Neuroimage 173:176–187.
- 1111 von Dawans B, Fischbacher U, Kirschbaum C, Fehr E, Heinrichs M (2012) The Social
 1112 Dimension of Stress Reactivity: Acute Stress Increases Prosocial Behavior in Humans.
 1113 Psychol Sci 23:651–660.
- 1114 Waytz A, Zaki J, Mitchell JP (2012) Response of dorsomedial prefrontal cortex predicts
 1115 altruistic behavior. J Neurosci 32:7646–7650.
- 1116 Weng HY, Fox AS, Shackman AJ, Stodola DE, Caldwell JZK, Olson MC, Rogers GM,
 1117 Davidson RJ (2013) Compassion Training Alters Altruism and Neural Responses to
 1118 Suffering. Psychol Sci 24:1171–1180.
- 1119
 1120
 1121
 1122
 1123
 1124
 1125
 1126
 1127
 1128
 1129
 1130
 1131
 1132
 1133
 1134
 1135

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 socio-affective processes. Afterwards, 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 9 task blocks. Donation blocks were randomly distributed per participant across the pre- and post-phase (and the one rating-free block prior to the EmpaToM and fMRI experiment). Based on pre-test data, each block was constructed in a way that charities elicited a broader 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) and mean donations did not differ significantly between blocks ($P = 0.81$, see *Methods*). Error bars represent ± 1 SD.

1147

1148

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% confidence intervals (CIs).

1151

1152

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). *Note:* Horizontal colored lines = mean scores across subjects; white data point = median; thick grey vertical lines = boxplots.

1160

1161

1162

1163

1164 Figure 4. **A)** Violin distribution plots of mean donations across groups and sessions (Horizontal colored lines =
 1165 mean donations across subjects; white data point = median; thick grey vertical lines = box-plots). **B)** Increases
 1166 in cortisol were associated with decreases in charitable giving across participants and groups. This negative
 1167 association could also be observed in both groups separately (stress group: $r = 0.51$, $P = 0.031$; control group: r
 1168 $= 0.56$, $P = 0.021$). **C)** Δ Cortisol $\times$ mentalizing interaction plots of the simple slopes at +1 SD above and -1 SD
 1169 below the mean of the moderator (mentalizing capacity) and Δ cortisol for donations in the pre-phase, post-phase,
 1170 and the change over time (post minus pre). Only for post-phase donations, we observed a significant negative
 1171 association between cortisol changes and donations for high mentalizers, but not for low mentalizers. This
 1172 moderated negative association was also significantly stronger compared to the pre-phase (as indicated by the
 1173 significant Δ cortisol $\times$ mentalizing $\times$ time interaction in the GLMM and in the simple-slopes analysis that
 1174 directly tests for POST-PRE changes).

1175

1176

1177 Figure 5. Neural decoding of trial-by-trial donations (red) in a whole-brain searchlight analysis using a support
 1178 vector regression (SVR) approach (cluster-forming threshold of $P \leq 0.001$, FWE-corrected at the cluster level at
 1179 $P \leq 0.05$). Decoding accuracies in a subset of these brain areas (yellow) were increased with higher mentalizing
 1180 capacity (thresholded with uncorrected $P \leq 0.005$ [peak- $P \leq 0.001$] and with cluster extent threshold of $k \geq 40$
 1181 voxels). Brain areas with differential value coding for high vs. low mentalizers included the right dorsolateral
 1182 prefrontal cortex (rDLPFC), the right middle temporal gyrus (rMTG)/temporoparietal junction (rTPJ), a more
 1183 ventrolateral part of the right middle frontal gyrus (rMFG) and the precuneus. We observed no brain area that
 1184 was more predictive of donations for low relative to high mentalizers. Group-level decoding maps are provided
 1185 on <https://osf.io/u46yj/>.

1186

1187

1188

1189

Figure 6. **A)** Value coding (i.e., SVR donation-decoding accuracies [Fisher's Z]) in the rDLPFC (5-mm sphere around peak [40 28 42]) in the pre- and post-session (changes illustrated through grey lines). **B)** Δ Cortisol $\times$ mentalizing interaction plots of the simple slopes at +1 SD above and -1 SD below the mean of the moderator (mentalizing capacity) and Δ cortisol for SVR decoding accuracies. Results are shown separately for the pre-phase, post-phase, and their changes over time (post minus pre). Only for post-phase decoding accuracies of donations, we observed a significant negative association between Δ cortisol and neural value coding in high mentalizers, but not low mentalizers. This moderated negative association was also significantly stronger compared to the pre-phase. **C)** Brain-behavior correlation: decreases in decoding accuracies in the rDLPFC predicted decreases in donations from the pre- to post-session (robust regression). **D)** Moderated mediation model: Cortisol increases were associated with reduced donations from the pre- to post-session in high mentalizers. This effect was mediated by reductions in value coding in the right dorsolateral prefrontal cortex (DLPFC). β coefficients represent standardized regression coefficients. β_{direct} is the direct association between Δ cortisol and Δ donations after the mediator (i.e., 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: -1SD 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 (5,000 bootstrap samples) provided a 90% confidence interval that did not span 0, indicating a significant mediation (one-tailed $P \leq 0.05$). ** two-tailed $P \leq 0.05$.

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				
Pre-1 st fMRI (~ -70)	115.27	9.60	116.73	14.93
Post-1 st fMRI (-10)	116.23	13.14	114.17	14.07
TSST/Control (+8)	135.70*	14.05	114.87	11.15
Pre-2 nd fMRI (+20)	122.27	12.06	115.67	11.66
Post-2 nd fMRI (~ +70)	119.67	13.05	121.63	14.75
Diastolic blood pressure				
Pre-1 st fMRI (~ -70)	78.57	9.16	78.00	6.18
Post-1 st fMRI (-10)	80.47	10.44	77.67	6.99
TSST/Control (+8)	97.80*	7.87	79.10	5.17
Pre-2 nd fMRI (+20)	85.73	10.04	79.60	8.07
Post-2 nd fMRI (~ +70)	81.50	9.71	81.33	9.96
Pulse				
Pre-1 st fMRI (~ -70)	81.50	12.19	74.97	10.15
Post-1 st fMRI (-10)	76.80	13.34	73.07	9.11
TSST/Control (+8)	96.67*	16.32	72.97	10.22
Pre-2 nd fMRI (+20)	79.17	14.50	74.57	9.38
Post-2 nd fMRI (~ +70)	81.00	11.23	75.33	9.98

Note: Units of measurement: *blood pressure*: mmHG; *pulse*: beats per minute (bpm); *cortisol*: nmol/l; *significant group difference with $P < 0.001$

1218
1219
1220
1221
1222
1223
1224
1225

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

Predictor	Beta	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-Chi²* score (default SPSS-outputs).

1226
1227
1228
1229
1230

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

Brain region	Side	BA	T	k	MNI (peaks)		
					x	y	z
Average effect ^a							
Multiregional cluster	R/L		12.36	52714	2	-86	2
- local peaks ^b							
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 & Inferior Parietal Lobule	R/L	7	5.38	3201	2	-72	44
Middle Frontal Gyrus (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

Note: L, Left hemisphere; R, right hemisphere; BA, Brodmann area; k, cluster size in voxels; DLPFC, dorsolateral prefrontal cortex; DMPFC, dorsomedial prefrontal cortex; TPJ, temporoparietal junction.

^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 sub-clusters 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.

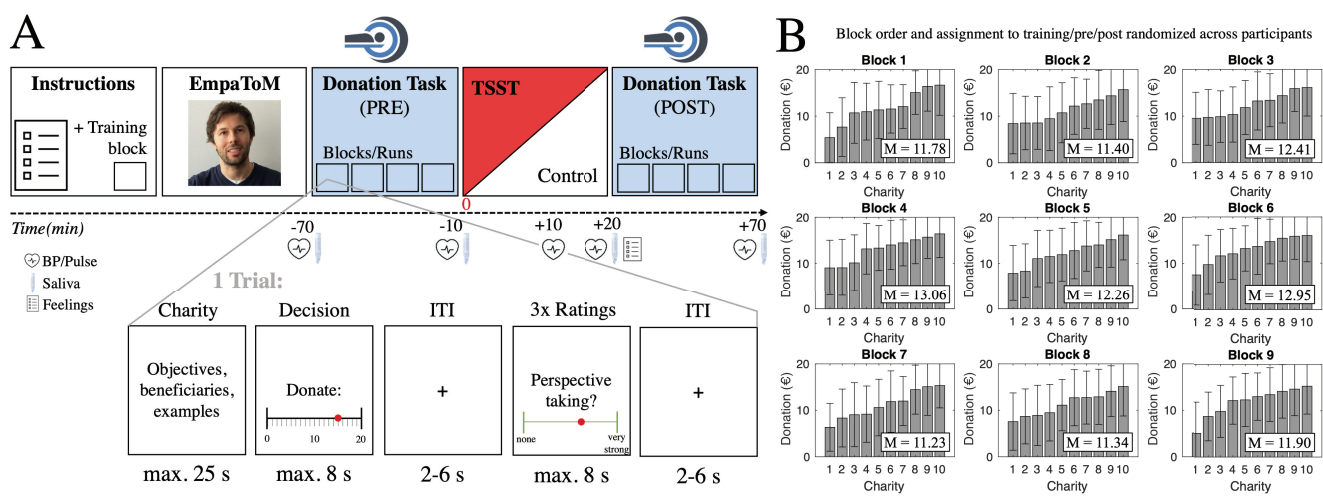
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244

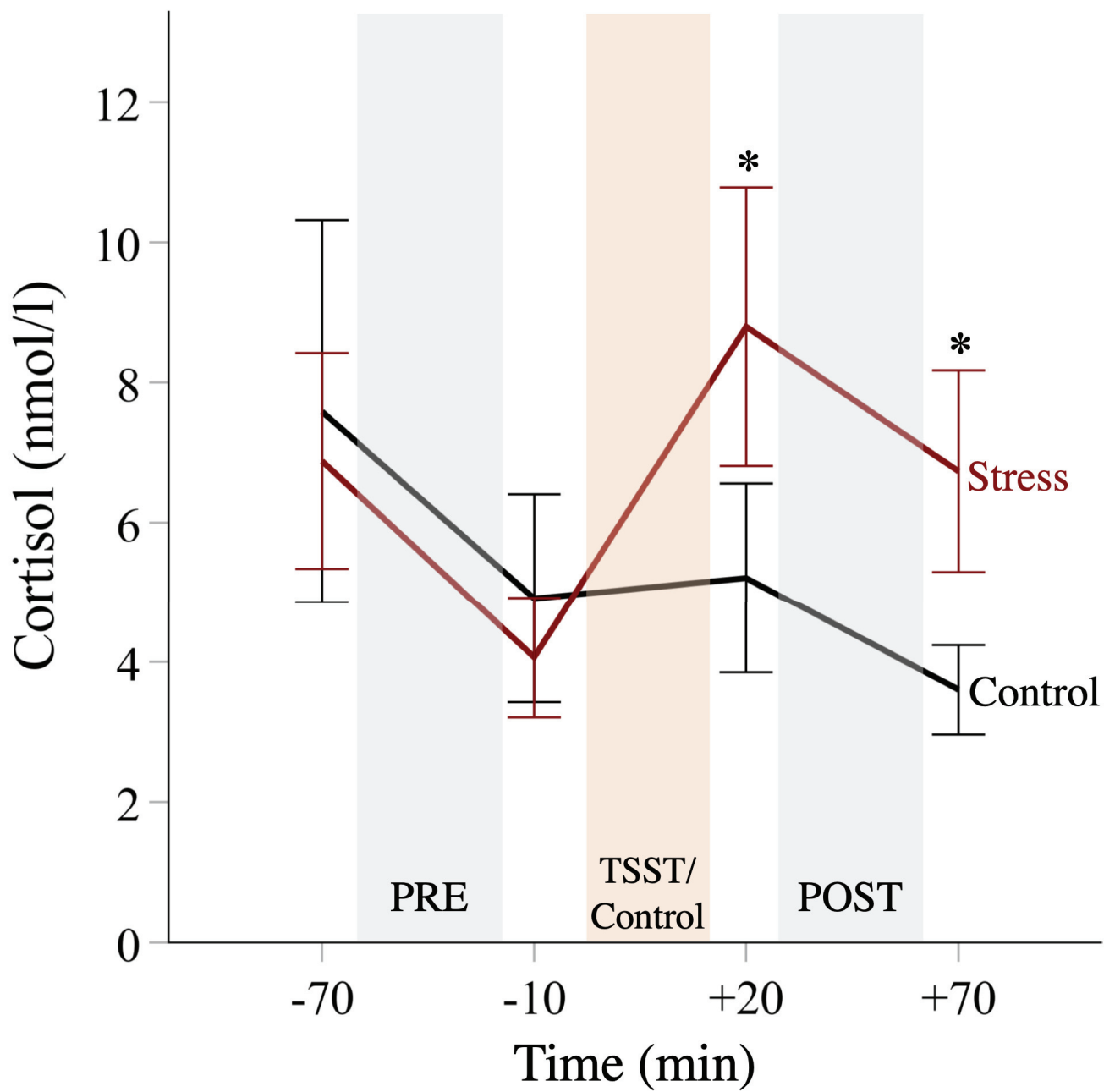
Table 4. Statistical models assessing cortisol- and mentalizing-related effects on value coding (SVR decoding accuracies) in the rDLPFC.

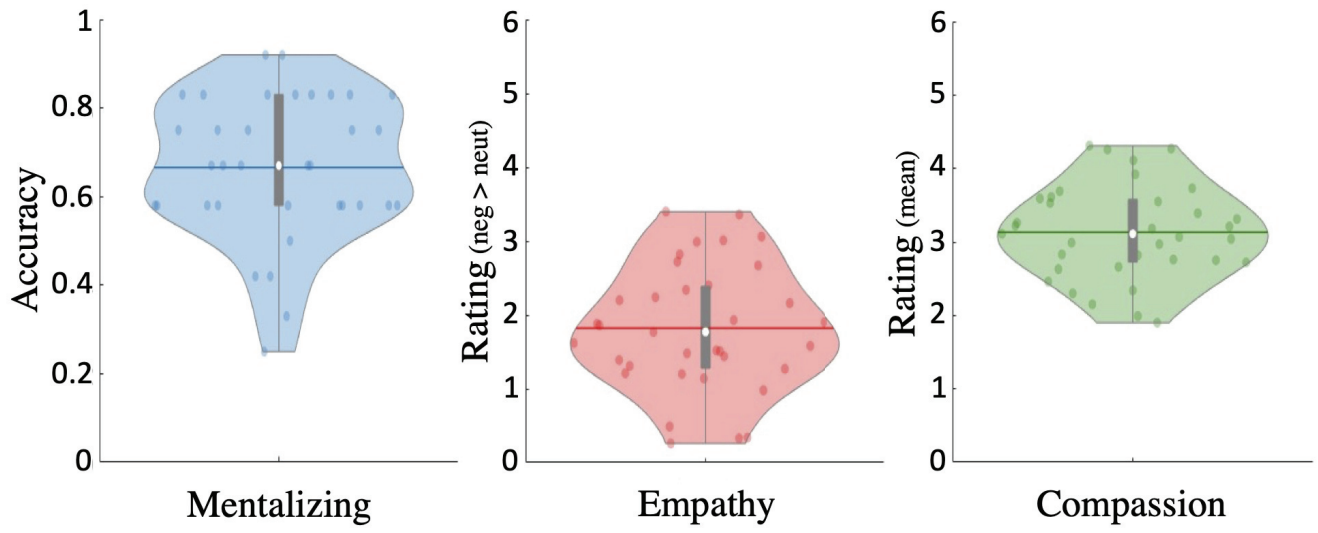
Predictor	Beta	SE	Test statistic*	P-value
MVPA – Full Model (GLMM)				
Constant (Intercept)	0.118	0.021	5.741	<0.001
Time	-0.024	0.024	-0.989	0.326
ΔCortisol	<-0.001	<0.001	-0.615	0.541
Mentalizing Capacity	0.352	0.126	2.798	0.007
ΔCortisol × Time	<-0.001	<0.001	-2.000	0.050
Mentalizing × Time	-0.268	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

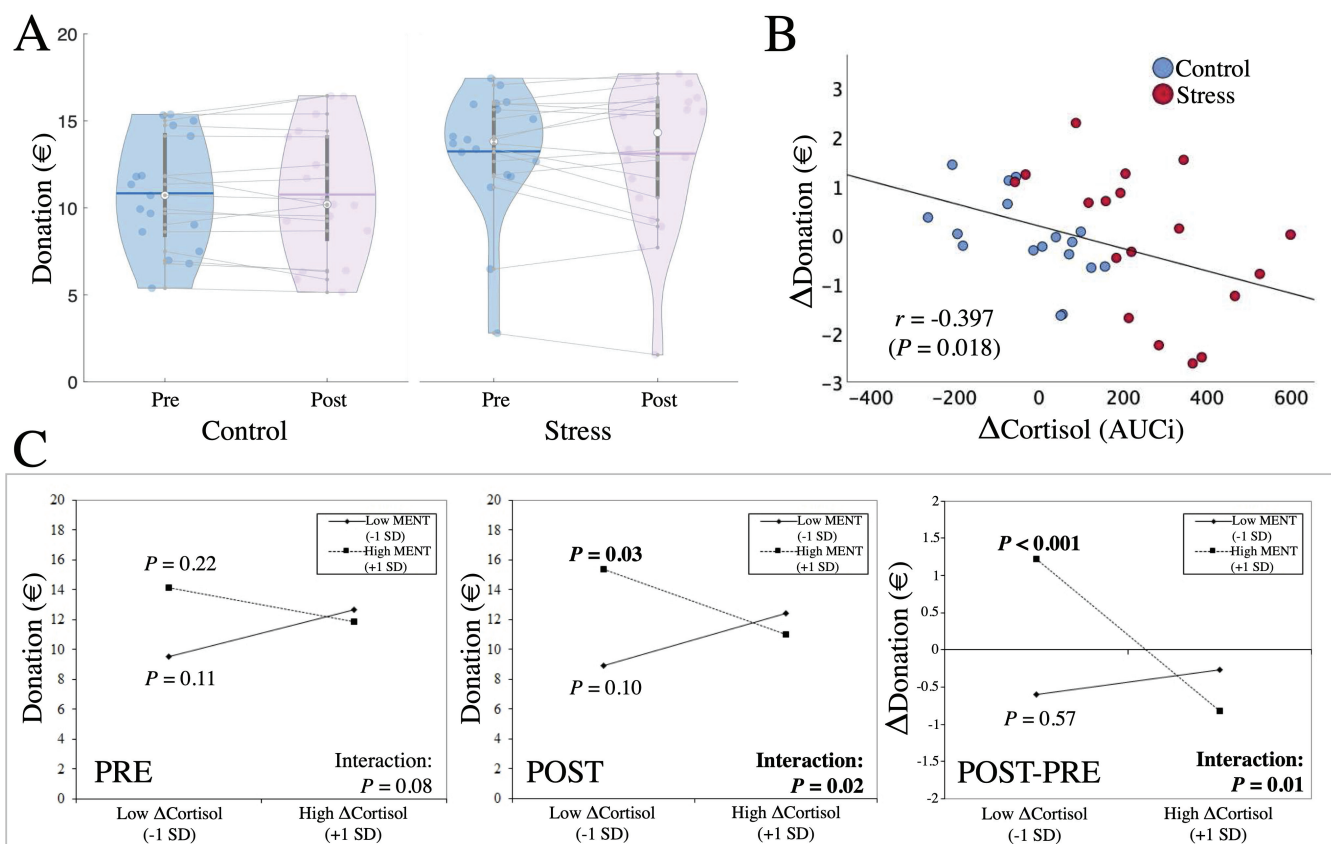
*Test statistic for full model (GLMM): *T*-value; decomposition models (GLMs): *Wald-Chi*² score (default SPSS-outputs).

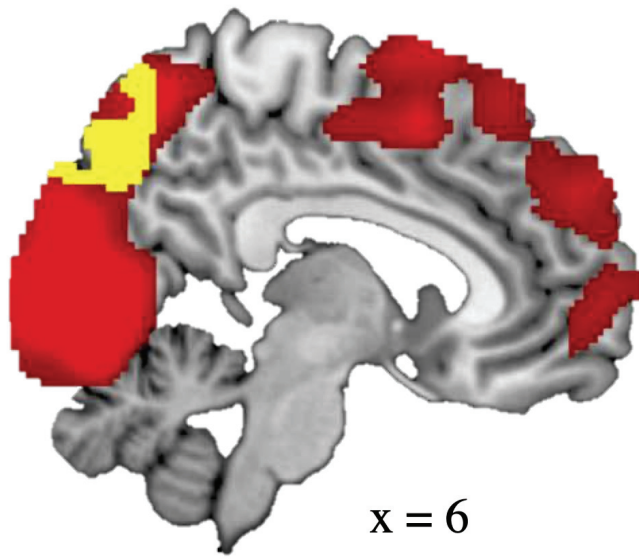
1245
1246
1247



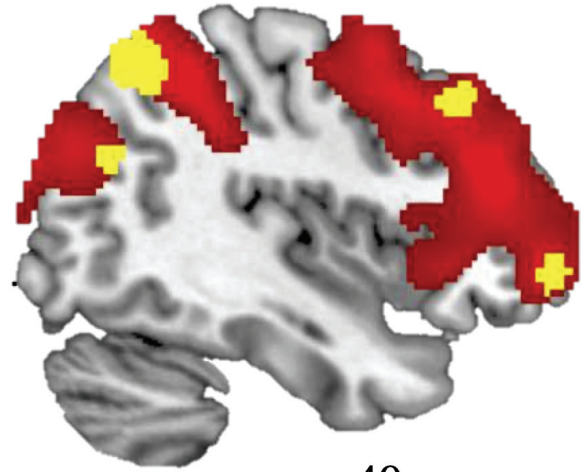








x = 6



x = 40

■ Decoding of donations
 ■ $MENT_{high} > MENT_{low}$

