1 **Oxytocin facilitates self-serving rather than altruistic tendencies in competitive**

2 **social interactions via orbitofrontal cortex**

3

4 **Xiaolei Xu, Congcong Liu, Xinqi Zhou, Yuanshu Chen, Zhao Gao, Feng Zhou,**

5 **Juan Kou, Benjamin Becker, Keith M Kendrick**

6

7 The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Laboratory for

8 NeuroInformation, University of Electronic Science and Technology of China,

9 Chengdu, Sichuan 610054, China

10

11 **Correspondence to:**

12

13 Keith M Kendrick and Benjamin Becker, The Clinical Hospital of Chengdu Brain

14 Science Institute, MOE Key Laboratory of Neuroinformation, No.2006, Xiyuan Ave.,

15 West Hi-Tech Zone, Chengdu, Sichuan 611731, China

16

17 e-mail: k.kendrick.uestc@gmail.com and ben_becker@gmx.de

18

Abstract

While the neuropeptide oxytocin can facilitate empathy and altruistic behavior it may also promote self-serving tendencies in some contexts, and it remains unclear if it would increase altruistic or self-interest behaviors when they compete within a social situation. The current between-subject, double-blind, placebo-controlled fMRI study investigated the effect of intranasal oxytocin on empathy for social exclusion using a modified online ball-tossing game which incorporated monetary rewards and both the potential to display altruistic and self-interest behaviors. Results showed that when subjects in both oxytocin and placebo groups were observing a player being excluded (victim) by other players in the game there was activation in the mentalizing network. When subjects then played both with the victim and the players who had excluded them they threw more balls to the victim player, indicative of an altruistic response. However, subjects in the oxytocin group threw more balls to the excluder players indicative of greater self-interest, since the latter would be perceived as more likely to reciprocate to maximize financial gain. This behavioral effect of oxytocin was associated with greater medial orbitofrontal cortex activation when playing with the excluders and negatively correlated with trait-altruism scores. Overall, our findings suggest that in the context of competing motivations for exhibiting altruistic or self-interest behavior oxytocin enhanced self-interest and this was associated with greater activation in frontal reward areas.

Key words

Oxytocin, altruism, self-interest, social interaction, orbitofrontal cortex

Introduction

Empathy represents a core social function that allows individuals to recognize and understand the emotional states of others and respond to them accordingly (Eisenberg, & Eggum, 2009). Empathy has two main components, cognitive empathy which includes cognitive processes of perspective-taking allowing us to infer the mental states of others and emotional empathy reflecting a direct affective reaction involving understanding, sharing and responding appropriately to their feelings (Bernhardt, & Singer, 2012; Shamay-Tsoory, 2011; Shamay-Tsoory et al., 2009). Emotional empathy can further be sub-divided in direct and indirect components which have also been described as emotional contagion and a more general form of emotional arousal (Decety, 2011; Hurlemann et al., 2010). Although the cognitive and emotional components of empathy are partly dissociable (Shamay-Tsoory et al., 2009), it is proposed that our empathic experience involves a dynamic interplay between the two components with an explicit representation of another person's affective state being a prerequisite, thereby making initial cognitive empathy necessary for empathic empathy to occur (Decety, & Jackson, 2004; Hillis, 2014).

Numerous studies have focused specifically on empathy in response to physical harm and neural networks responding to empathy for others in pain overlap substantially with those engaged during actual experience of pain (Marsh, 2018). Similar to physical pain, being socially rejected or excluded by others causes painful experience in humans and both share common neural circuits (Eisenberger et al., 2003). Moreover, observing others being excluded evokes lower need satisfactions and negative affect in the same way as when being excluded personally (Wesselmann, Bagg, & Williams, 2009). Empathy for physical pain represents a prototypical example of emotional contagion driven primarily by automatic bottom-up processes such as described in the perception-action mechanistic model (PAM) (Preston, & de Waal, 2002). In contrast, empathy for social pain requires a greater cognitive component, particularly mentalizing and theory of mind to integrate the complex social information (Bruneau, Pluta, & Saxe, 2012). This difference is also reflected in the underlying neural networks, with empathic responses to physical pain engaging regions involved

73 in autonomic arousal and emotional reactivity such as the dorsal anterior cingulate

74 (dACC) and the anterior insula (AI) (Lamm, Decety, & Singer, 2011; Yao et al., 2016),

75 whereas empathy for social exclusion additionally engages regions involved in

76 mentalizing and theory of mind such as dorsomedial, medial and ventromedial

77 prefrontal cortex (DMPFC, MPFC and VMPFC) (Masten, Morelli, & Eisenberger,

78 2011).

79   Empathy for those in suffering leads to distressed feelings and a motivation for

80 helping behavior (FeldmanHall, Dalgleish, Evans, & Mobbs, 2015). Indeed, the

81 motivation for altruistic and prosocial behavior is an evolutionary outcome of empathy

82 (Rumble, Van Lange, & Parks, 2010; de Waal, 2008). In line with this strong

83 motivational component a previous study demonstrated that reciprocal altruism was

84 associated with increased activity in the reward system, suggesting that it may be

85 reinforcing (Rilling et al., 2002). However, in everyday life altruism is often exhibited

86 in the absence of an expected reciprocity and sometimes even occurs at the cost of an

87 individual's self-interest (de Waal, 2008). Batson proposed the empathy-altruism

88 hypothesis that empathic concern referring to other-oriented emotions produced

89 altruistic motivational states with the goal of increasing the welfare of others rather than

90 self (Batson, 2011). This costly altruism is predominantly exhibited towards genetically

91 and socially close others but might also extend to strangers (Vekaria, Brethel-Haurwitz,

92 Cardinale, Stoycos, & Marsh, 2017). Furthermore, costly altruistic behaviors and

93 associated activity in reward-related regions are driven by other-oriented empathy

94 rather than experienced personal distress (FeldmanHall, Dalgleish, & Mobbs, 2015). In

95 social contexts altruistic punishment represents another form of costly altruistic

96 behavior that aims both to help the victim and to enforce group norms and cooperation

97 (West, Griffin, & Gardner, 2007; Boyd, Gintis, Bowles, & Richerson, 2003). Both

98 forms of costly altruism have been associated with activity in core reward processing

99 regions including the ventral striatum (De Quervain, Fischbacher, Treyer, &

100 Schellhammer, 2004).

101   The hypothalamic neuropeptide oxytocin (OXT) has been found to influence a

102 number of different aspects of social cognition and to promote the formation and

103   maintenance of social bonds (Kendrick et al., 2017; Striepens, Kendrick, Maier, &

104   Hurlemann, 2011; Macdonald, & Macdonald, 2010). Studies administering intranasal

105   OXT have shown that it modulates core pain empathy regions, including the ACC and

106   insula, as well as mentalizing regions and reward-related striato-frontal circuits

107   although in a context-dependent manner (see Wigton et al., 2015; Bethlehem, van Honk,

108   Auyeung, & Baron-Cohen, 2013). A number of genetic studies have also reported links

109   between OXT-receptor polymorphisms and empathy in both Caucasian and Chinese

110   populations (Rodrigues, Saslow, Garcia, John, & Keltner, 2009; Smith, Porges, Norman,

111   Connelly, & Decety, 2014; Wu, Li, & Su, 2012). Intranasal OXT has been shown to

112   particularly enhance emotional - rather than cognitive - empathy in both Caucasian and

113   Chinese subjects (Geng et al., 2018; Hurlemann et al., 2010) and to concomitantly

114   reduce amygdala reactivity (Geng et al., 2018; Becker et al., 2018). On the other hand,

115   insula responses can be either enhanced (Riem et al., 2011; Striepens et al., 2012), or

116   decreased in the context of pain empathy (Bos, Montoya, Hermans, Kevsers, & van

117   Honk, 2015) and embarrassment (Becker et al., 2018).

118   In line with the emotional empathy-enhancing effects of intranasal OXT, a previous

119   study has reported that it also increased altruistic behavior towards an ostracized

120   individual (Riem et al., 2013). However, in this study exhibiting altruistic behavior did

121   not come at any cost to the participants whereas in real-life situations it often does, and

122   particularly in contexts where it might result in reduced financial gain or other aspects

123   that conflict with self-interest (Camerer, & Fehr, 2006; Miller D, 2001). Although many

124   prosocial effects of OXT have been reported it has also been found to promote self-

125   serving lying and group-serving dishonesty (Sindermann, 2018, Shalvi & De Dreu,

126   2014). In addition, OXT's detrimental effect on honesty was found in a competitive

127   environment and driven by conformity with the behavior of peers (Aydogan, Jobst,

128   D'Ardenne, Müller, & Kocher, 2017). On the other hand, intranasal OXT also tends to

129   increase both altruistic help as well as costly altruistic (altruistic punishment) behaviors

130   (Hu et al., 2016; Aydogan, Furtner, et al., 2017). Thus, although a number of studies

131   have demonstrated the effects of OXT on empathy, altruistic and self-serving behaviors

132   it is unclear what its functional role may be at both the behavioral and neural level when

133   these motivations are competing in a social situation.

134   The current placebo-controlled double-blind fMRI study therefore aimed at

135   determining the effects of intranasal OXT on competing behavioral tendencies between

136   empathy-motivated altruism and self-interest. To this end a modified Cyberball

137   paradigm was employed during which subjects initially observed a social exclusion

138   situation and subsequently engaged in the game. The Cyberball game is a classic and

139   widely used paradigm to induce social exclusion which leads to aversive and painful

140   feelings by threatening an individual's four fundamental needs including belonging,

141   control, self-esteem and meaningful existence (Williams, 2009). The negative

142   emotional responses to being socially excluded are found even when individuals know

143   that they were excluded by a computer (Zadro, Williams, & Richardson, 2004). Brain

144   imaging studies have demonstrated that the neural circuits that mediate exclusion-

145   induced psychological pain resemble those responding to physical pain with common

146   activations seen in the pain (e.g. anterior cingulate cortex, ACC) (Eisenberger et al.,

147   2003) and somatosensory processing (e.g. secondary somatosensory cortex, dorsal

148   posterior insula) systems (Kross, Berman, Mischel, Smith, & Wager, 2011). The

149   Cyberball paradigm has also successfully been employed to induce social pain empathy

150   by observing others being excluded in this game (Wesselmann, Bagg, & Williams, 2009)

151   and subsequent prosocial behavior towards the excluded individual (Masten et al.,

152   2011). Finally, previous studies have demonstrated a high sensitivity of this paradigm

153   to capture effects of OXT on social behavior in both healthy subjects (Xu et al., 2017)

154   and individuals with autism (Andari et al., 2010).

155   In the current adaptation of the Cyberball paradigm, participants were scanned

156   while first observing 3 unknown individuals playing the game and where one player

157   (victim) was gradually excluded by the two other players (excluders) so that they gained

158   more money than the victim. Immediately after the observe session, participants were

159   given an opportunity to play with the victim and one of the excluders as well as another

160   new player. To create a situation of competing empathy and self-interest motivated

161   behavioral tendencies participants were told that any player in this paradigm receiving

162   a ball would receive an additional monetary reward (0.3 RMB/ball). In this case, the

163   excluder player was manipulated to be the most attractive cooperator to maximize self-
164   interest while playing with the victim of exclusion would be rather motivated by
165   empathy-induced altruistic behavior.

166   Considering convergent evidence for OXT-enhanced emotional empathy, we
167   hypothesized that it would increase empathy for the victim of exclusion during the first
168   stage of the experiment. We additionally hypothesized that OXT-facilitated empathy
169   would be accompanied by increased activation in (social) pain (dACC and anterior
170   insula) and mentalizing networks (medial frontal, posterior parietal and temporal
171   regions, i.e. posterior superior temporal sulcus (pSTS), posterior cingulate cortex (PCC)
172   and precuneus (Eisenberger et al., 2003; Masten et al., 2011; Singer, 2006). In the
173   second stage of the experiment when the subjects were actively engaged, we
174   hypothesized that if OXT promotes costly altruistic behavior subjects should increase
175   the proportion of their throws to the victim and correspondingly throw less to the
176   excluder, whereas if it promotes increased self-serving behavior then this should result
177   in a greater proportion of throws to the excluder player who should be more likely to
178   reciprocate, thus increasing the financial gain. Given the engagement of striato-
179   orbitofrontal reward processing circuits (Singer, 2006; Kringelbach, 2005; Spicer et al.,
180   2007) in both monetary reward-anticipation (O'Doherty, Kringelbach, Rolls, Hornak,
181   & Andrews, 2001) as well as altruistic behavior (FeldmanHall, Dalgleish, Evans, &
182   Mobbs, 2015) we hypothesized that the OXT-induced behavioral preference for a player
183   would be mirrored by increased activity in this circuit. Given that cultural orientation,
184   i.e. horizontal independence modulated the effect of OXT following social exclusion
185   (Xu et al., 2017), and higher trait altruism has been associated with stronger empathic
186   brain responses (Haas et al., 2015) these traits were additionally assessed. Finally, in
187   view of our previous study demonstrating long-term effects on memory for and
188   preference for replaying with specific players (Xu et al., 2017) we also investigated
189   these same factors one week after the social exclusion experiment.
190

191   **Materials and Methods**
192   **Participants**

193   82 healthy Chinese male university students (right-handed, age = 18-27 years, Mean ±

194   SE = 21.36 ± 0.24 years) participated in the present study and were randomly assigned

195   to receive either placebo (PLC) (n = 41, age = 18-26 years, Mean ± SE = 21.68 ± 0.34

196   years) or OXT nasal spray (n = 41, age = 18-27 years, Mean ± SE = 21.05 ± 0.34 years;

197   PLC vs. OXT, t = 1.31, p = 0.195) in a double-blind between-subject pharmacological

198   fMRI experiment. Subjects reported being free from current or a history of psychiatric

199   or neurological disorders and did not use any medication in the four weeks before the

200   experiment and were asked to abstain from caffeine and alcohol in the 24 hours before

201   the experiment.

202   Study procedures were approved by the local ethics committee at the UESTC and

203   adhered to the latest revision of the Declaration of Helsinki. Each participant provided

204   written informed consent before the experiment and received monetary compensation

205   for participation. Study protocols were pre-registered at clinical trials.gov

206   (https://www.clinicaltrials.gov/ct2/show/NCT03122067, Trial ID: NCT03122067).

207

208   **Procedure**

209   To control for potential confounding effects of pre-treatment differences in affective

210   state and empathy-related domains, validated Chinese versions of established

211   questionnaires to assess mood (Positive and Negative Affect Schedule, PANAS)

212   (Watson, & Clark, 1988), anxiety (State-Trait Anxiety Inventory, STAI), (Barnes, Harp,

213   & Jung, 2002), (Liebowitz Social Anxiety Scale, LSAS) (Heimberg et al., 1999),

214   depression (Beck Depression Inventory, BDI) (Beck, Steer, & Brown, 1996), trait

215   autism (Adult Autism Spectrum Quotient, ASQ) (Baron-Cohen, Wheelwright, Skinner,

216   Martin, & Clubley, 2001), trait empathy (Interpersonal Reactivity Index-C, IRI-C) (Siu,

217   & Shek, 2005), and early life stress (Childhood Trauma Questionnaire, CTQ)

218   (Bernstein, & Fink, 1998) were administered. In line with our previous study on the

219   effects of OXT on social exclusion (Xu et al., 2017) we explored associations with the

220   'horizontal independence' (HI) scale of the Individualism and Collectivism Scale (ICS)

221   (Singelis, Triandis, Bhawuk, & Gelfand, 1995). In the context of previously reported

222   associations between trait altruism and empathic brain activity (Haas et al., 2015), as

223 well as a potential contribution of altruism-driven attempts to compensate the victim

224 after observing them being excluded we additionally explored associations between the

225 effects of OXT and the level of pre-treatment altruistic prosocial behavior as assessed

226 by the altruistic subscale of the Prosocial Tendency Measure (PTM). This altruistic

227 subscale in the PTM specifically assesses concerns about individuals in need of help

228 that incurs a cost to the helper (Carlo, & Randall, 2002) and thus appears of particular

229 relevance with respect to the present paradigm. Mood (PANAS) was additionally

230 assessed after the experiment to control for unspecific effect of treatment on these

231 domains.

232   After participants completed the questionnaires portrait photos with neutral facial

233 expressions were taken from each subject and served as stimuli depicting them during

234 the subsequent Cyberball game. Finally, subjects were asked to rate portrait photos with

235 neutral facial expressions of their subsequent fellow-players in the experiment with

236 respect to likeability, trustworthiness and valence (for details see paradigm description).

237 Subjects next self-administered 24 IU OXT (Oxytocin-spray, Sichuan Meike

238 Pharmaceutical Co., Ltd; 3 puffs of 4IU per nostril with 30s between each puff) or

239 placebo (PLC – identical sprays with the same ingredients other than OXT – i.e. sodium

240 chloride and glycerol) 45 minutes before the start of the experimental paradigm.

241 Administration adhered to a standardized protocol for the intranasal administration of

242 OXT (Guastella et al., 2013).

243

244 **Modified Cyberball paradigm**

245 Participants were told they would first observe and then participate in a ball-tossing

246 game ('Cyberball') with 4 other participants online while they were in the fMRI scanner

247 and that the other 4 players were sitting in separate compartments in a nearby behavioral

248 testing room to avoid direct personal interaction during the entire experiment. However,

249 in fact the 4 players were fictitious and preprogrammed in the experimental paradigm.

250 The revised Cyberball task in this study included two sessions acquired during separate

251 fMRI runs. The first session (OBSERVE condition) aimed to prime the subject's

252 attitude towards each of the players observed (victim or excluder, see Fig. 1a) in order

253   to influence their subsequent behavior when they participated in the game during the

254   second session (PLAY condition, see Fig. 1b). During the OBSERVE run, subjects

255   simply observed 3 individuals playing the game and were told that each player in the

256   game would receive 0.3RMB reward for every ball thrown to them. Subjects were

257   explicitly instructed to observe the game and to consider what each player might be

258   thinking or feeling during it (Masten et al., 2011). The first run (OBSERVE) contained

259   10 blocks with each block lasting for 30 seconds, starting with 4 'fair' blocks during

260   which the 3 players threw the ball equally often to each other, followed by two

261   'transition' blocks during which one of the players was gradually excluded and 4

262   'exclusion' blocks during which one player (victim) was totally excluded by the other

263   two players (excluder 1 and excluder 2). After each round the subject was shown how

264   much money each of the 3 players had gained (displayed for 12 s). In this way the

265   subject observing the game was expected to detect that one of the players (victim) was

266   eventually excluded by the two other players (excluders) and that the result of this was

267   the two excluders had learned that by throwing only to each other they ended up gaining

268   more money at the expense of the victim who gained less. The second run (PLAY)

269   involved the subject and 3 other players, including the victim, one excluder (1 or 2)

270   from the first run and a new player and comprised 6 rounds with a fixed number of 24

271   ball throws per round. All virtual players were programmed to throw the balls equally

272   to the other 3 players. Subjects were told that during the first four rounds they would

273   be informed how much each player would have gained (0.3RMB per ball received –

274   amount gained per player was displayed for 12 s) but that they would only actually

275   receive the monetary reward from the final 2 rounds. This was done to allow us to assess

276   whether having a strong motivational component of monetary rewards in every round

277   might have biased behavior towards greater self-interest behavior and reduce altruistic

278   behavior towards the victim from the first game. To confirm whether the pattern of

279   throwing to the other players was different when a monetary incentive was included,

280   we initially explored potential interaction effects between treatment and the monetary

281   condition (i.e. first 4 vs. last two blocks). In line with our expectation no main effect of

282   monetary condition or interaction effects with treatment were found (for details see

283   supplementary information) and consequently this factor was discarded from all further

284   analyses.

285

286
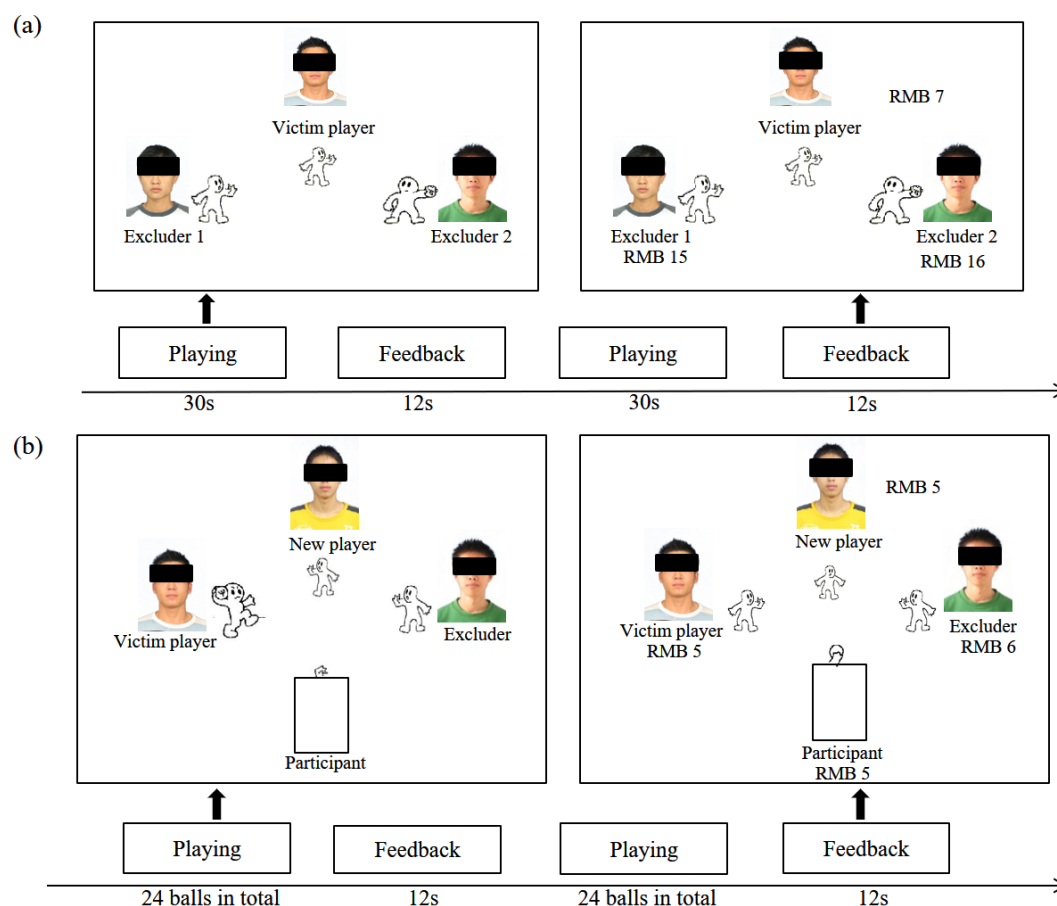
Figure 1. The Cyberball game paradigm employed. (a) Subjects first completed an "Observe" session which included a total of 10 blocks (4 blocks where players threw equally to each other followed by 2 blocks where two players (excluders) started to throw more often to each other and finally 4 blocks where the excluders threw exclusively to each other and did not throw to the other player (victim) at all). Blocks lasted for 30s during a 12s period between each block the subjects viewed how much money each of the players had won. Subjects were instructed during their observation of the game to consider what the individuals playing it were thinking and feeling. (b) Subjects next completed a "Play" session where they played the game together with the victim and one of the excluders and a novel player. This session was for 6 blocks with 24 ball throws in each block (After each block the subject could see how much money they and the other players had won and a 2 s maximum was allowed to throw the ball to avoid a monetary deduction).

Manipulation check and potential confounders

As a manipulation check subjects were asked to rate the other players in terms of their likeability, trustworthiness and valence using a visual analogue scale (VAS, 0-100) before treatment administration and following the OBSERVE and PLAY sessions respectively. To further evaluate the manipulation subjects were asked to report if any specific events had happened during the OBSERVE session (e.g. "Did all three players treat each other fairly?", "Was there anyone who was treated unfairly?"). They were also asked to rate how empathic they felt towards the excluded player (score between 0-100). As additional control variables the PANAS questionnaire was administered again to test whether OXT had altered participants' mood (PANAS) after the paradigm.

Long term effects of OXT – follow-up after one week

Subjects were asked to return to the laboratory to complete follow-up assessments one week after the Cyberball game. The assessment included ratings of the other players with respect to likeability, trustworthiness and valence. Moreover, a surprise memory test was employed to examine if the previous Cyberball game had an effect on social recognition memory and whether this was influenced by OXT. Finally, subjects were asked if they were willing to play with the previous players again and rated how much they would like to do so (1-9 scale).

Behavioral data analyses

Statistical analyses for the questionnaires and behavioral ratings were performed using SPSS 18.0 software (SPSS Inc., Chicago, Illinois, USA). Post-hoc analyses of interaction effects were performed employing Bonferroni correction for multiple comparisons. Associations between traits, behavior and neural indices were examined using Pearson correlation and differences in the correlations between the two groups were further examined using Fisher's Z test with Bonferroni correction.

Image acquisition

328   Imaging data were collected using a 3T GE Discovery MR750 system (General Electric,

329   Milwaukee, WI, USA) using the following sequence parameters: TR = 2000 ms; TE =

330   30 ms; flip angle = 90°; number of slices = 43; slice thickness = 3.2 mm; FOV = 220 ×

331   220 mm2; matrix = 64 × 64; slice orientation = axial. High-resolution whole-brain T1-

332   weighted images were additionally acquired using a spoiled gradient echo pulse

333   sequence to improve normalization of the functional data (TR = 6 ms; TE = 2 ms; flip

334   angle = 12°; number of slices = 156; slice thickness = 1 mm; FOV = 256 × 256 mm2;

335   matrix = 256 × 256).

336

337   **fMRI data analysis**

338   Statistical   Parametric   Mapping   as   implemented   in   SPM12

339   (http://www.fil.ion.ucl.ac.uk/spm/) was used to preprocess and analyze the

340   neuroimaging data. The first 6 volumes of each functional neuroimaging time-series

341   were removed to allow for T1 equilibration. Preprocessing included slice timing, image

342   realignment to correct for head motion, normalization into the Montreal Neurological

343   Institute (MNI) space resampled at 3×3×3 mm voxel size, and spatially smoothed using

344   an 8 mm FWHM Gaussian kernel. Generalized linear models (GLM) were built to

345   investigate the BOLD signal changes. A 128-second high-pass filter was applied to

346   further control for low-frequency noise artifacts.

347   The first-level design matrix for the OBSERVE run was modelled using a blocked-

348   design matrix including the first 4 blocks as the inclusion condition and the last 4 blocks

349   as the exclusion condition. The monetary reward feedback and the middle two blocks

350   with the transfer between inclusion and exclusion were additionally modelled and the

351   six head motion parameters included.

352   The first-level design matrix for the PLAY run was modelled using an event related

353   design matrix to specifically examine the throws made by the participant to the other

354   individual players. Ball-tosses towards the excluder, victim and new player were

355   implemented as experimental conditions and modelled as separate events. To

356   specifically model the expectation phase for a reciprocal action, independent from the

357   decision phase, the time between the other player receiving the ball form the participant

358   and the throw of that player was modelled as experimental event. Monetary feedback,

359   rating periods and head motion parameters were additionally included in the matrix.

360   Due to technical issues during the fMRI assessment and excessive motion (> 3mm),

361   data from 8 participants had to be excluded (OXT = 4, PLC = 4), leading to a final

362   sample size of OXT =37 and PLC = 37 for the imaging analysis. Effects of OXT during

363   the OBSERVE and PLAY sessions were assessed by employing independent t-tests. In

364   line with the main aim of the study the contrast [exclusion > inclusion] was used for the

365   OBSERVE condition, and for the PLAY condition player-specific contrasts were

366   examined [excluder, victim, new player]. The threshold p-value level was set at <0.05

367   cluster-level with family wise error (FWE) correction for multiple comparisons and an

368   initial cluster forming threshold at the voxel-level of $p<0.001$, uncorrected (in line with

369   recommendations by Eklund et al., 2016; Slotnick; 2017)

370

371   **Results**

372   **Potential confounders**

373   No significant differences between subjects in the PLC and OXT groups were found in

374   pre-treatment affective state and empathy-related domains (Independent t-tests, Table

375   1). Importantly, mood as assessed by the PANAS did not differ after the experiment,

376   arguing against unspecific confounding effects of treatment on mood.

377

378   **Table 1** Questionnaire scores in PLC and OXT group before and after treatment.

Measurements	Placebo	Oxytocin	t-value	p-value
Before treatment				
Positive and Negative Affect Schedule (PANAS)				
Positive	29.15±1.10	28.93±0.79	0.16	0.87
Negative	15.20±0.79	14.29±0.75	0.83	0.41
State-Trait Anxiety Inventory (STAI)				
State	37.68±1.44	36.93±1.32	0.39	0.70
Trait	37.63±1.20	39.56±1.22	-1.13	0.26

Liebowitz Social Anxiety Scale (LSAS)				
Avoid	19.85±1.40	21.56±1.31	-0.89	0.38
Fear	22.56±1.53	20.2±1.34	1.16	0.25
Beck Depression Inventory (BDI-II)	5.41±1.32	6.10±1.09	-0.40	0.69
Adult Autism Spectrum Quotient (ASQ)	20.51±0.70	18.76±0.80	1.64	0.10
Interpersonal Reactivity Index (IRI)	39.76±0.99	40.29±1.03	-0.38	0.71
Childhood Trauma Questionnaire (CTQ)	56.20±0.80	56.00±0.91	0.16	0.87
Individualism and Collectivism Scale (ICS)	5.31±0.11	5.36±0.12	-0.31	0.76
Prosocial Tendency Measure (PTM)-altruism	15.93±0.43	16.76±0.40	-1.41	0.16
After treatment				
PANAS-Positive	25.68±1.48	24.46±1.19	0.64	0.52
PANAS-Negative	12.85±0.57	13.60±0.86	-0.73	0.47

379

380 **Effects of perceived social exclusion and oxytocin on likeability, trustworthiness**

381 **and valence ratings**

382 A 2 (Treatment: OXT/PLC) × 3 (Player: Excluder/Victim/New) mixed ANOVA with

383 subjects' likeability, trustworthiness and valence rating scores as dependent variables

384 revealed no significant differences in these three dimensions between the PLC and OXT

385 groups before the experiment (Fig. 2a). However, after the OBSERVE session during

386 which subjects watched the victim player being excluded, their ratings of the victim

387 player for the three dimensions were significantly higher than for the excluders (Fig.

388 2b), confirming successful experimental manipulation. Moreover, the decreased ratings

389 for the excluder players remained stable after subjects played with them during the

390 PLAY run and were even maintained one week later (Figs. 2c, 2d).
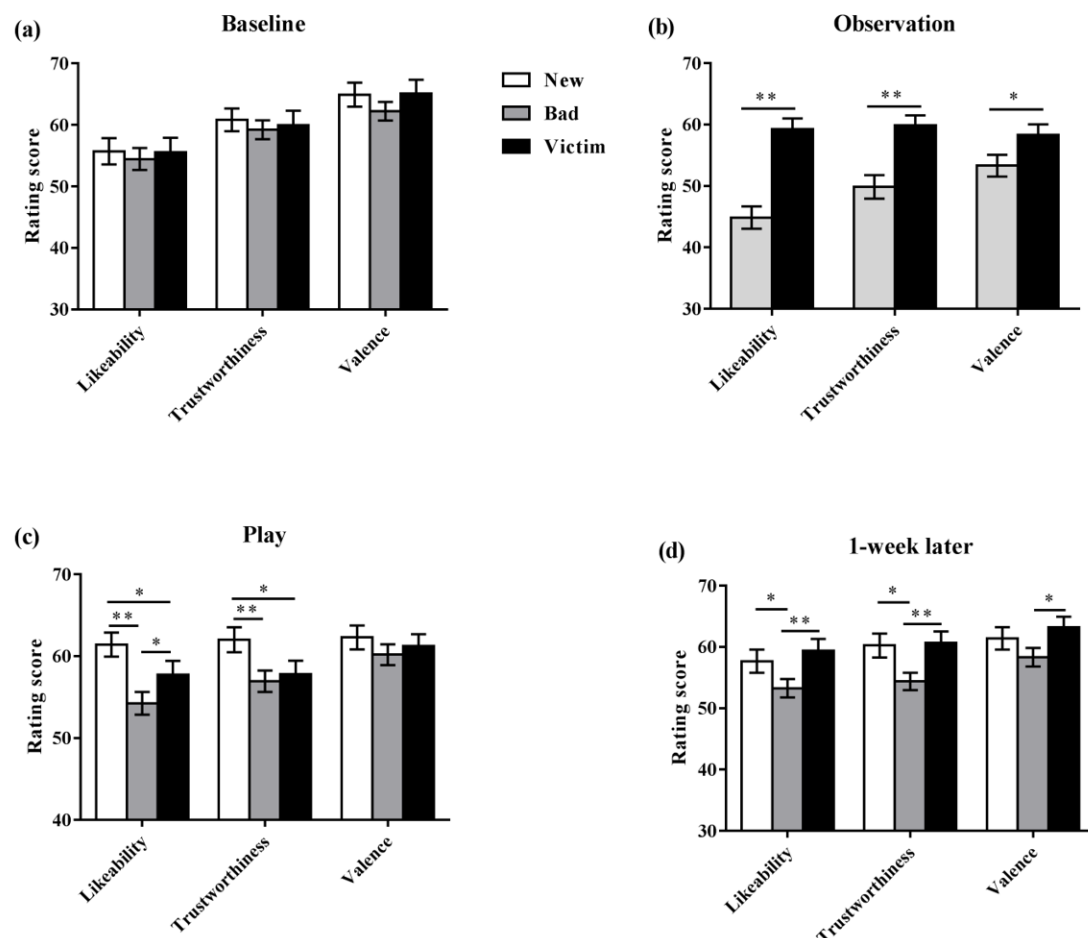
391

Figure 2. Rating scores of likeability, trustworthiness and valence in combined oxytocin and placebo-treated groups for (a) before treatment, (b) after treatment during the OBSERVE session, (c) after PLAY session, (d) 1-week after treatment and observing and playing the Cyberball game. *$p<0.05$, **$p<0.005$.

Manipulation check and empathic responses

At the end of the experiment all subjects reported that they realized one player was excluded during the OBSERVE session further confirming successful manipulation. To examine the effects of OXT on subjects' empathy scores for the victim player, an independent two sample t-test was performed but revealed no significant differences between the treatment groups (PLC = 63.62, OXT = 58.97, $t = 0.98$, $p = 0.33$, Cohen's $d = 0.224$).

406 **Behavior during the PLAY phase of the paradigm**

407 For the PLAY condition, the percentage of ball tosses to each of the other three players

408 served as the dependent variable and was subjected to a 2 (Treatment: PLC/OXT) × 3

409 (Players: Excluder/Victim/New) mixed ANOVA. Results revealed a main effect of

410 player ($F_{2,160}$ =7.05, $p = 0.001$, η^2_p=0.081), and a marginal significant interaction effect

411 between treatment and player ($F_{2,160} = 2.74$, $p = 0.068$, $\eta^2_p =0.033$). Post-hoc

412 Bonferroni-corrected paired comparisons showed that subjects in both groups threw

413 more balls to the victim than to the excluder players (victim = 36.7%, excluder =

414 30.28%, $p = 0.002$, Cohen's $d = 0.690$). Further exploratory analysis of the interaction

415 effect demonstrated that the OXT group threw significantly more balls to the excluder

416 player relative to the PLC group (to excluder player: PLC = 27.97%, OXT = 32.60%, t

417 $= -2.46$, $p = 0.016$, Cohen's $d = 0.543$, Fig. 3).
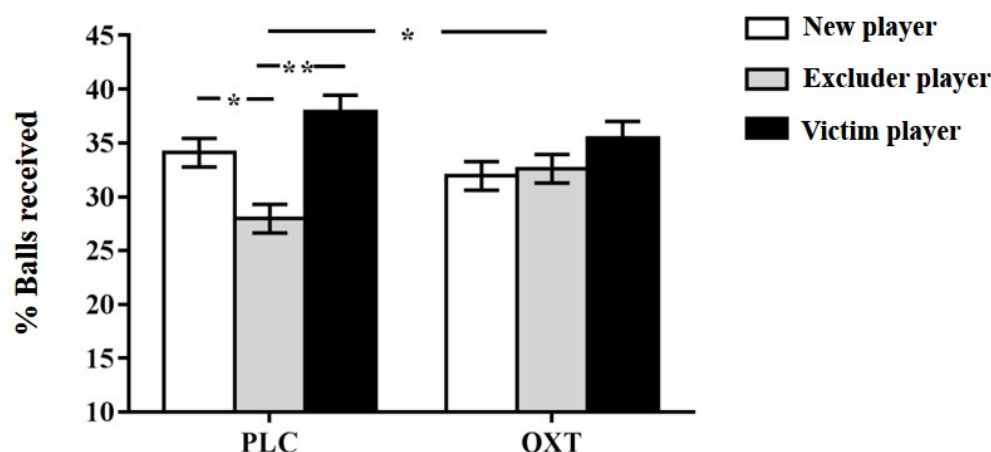
418

419

420 **Figure 3.** Percentage (mean ± sem) of balls that participants threw to the excluder, new, and

421 victim players respectively. *$p<0.05$, **$p<0.005$.

422

423 **Neural activation changes during the Cyberball task**

424 The task-related network engaged during the observation of social exclusion and effects

425 of OXT on the underlying neural activity [contrast, exclusion > inclusion] were firstly

426 investigated during the OBSERVE session. Analysis of neural changes during the

427 OBSERVE session in the combined PLC and OXT groups revealed significant

428 increased activity in the posterior cingulate cortex (PCC), left midcingulate cortex

429 (lMCC), precuneus, left inferior parietal lobule (lIPL) and right posterior superior

430 temporal sulcus (rpSTS) when subjects observed the victim being excluded (Fig. 4,

431 table 2). In line with the lack of OXT effects on the post OBSERVE behavioral ratings,

432 no significant neural differences were observed between the treatment groups. Thus,

433 any group differences during the subsequent PLAY session were unlikely to be driven

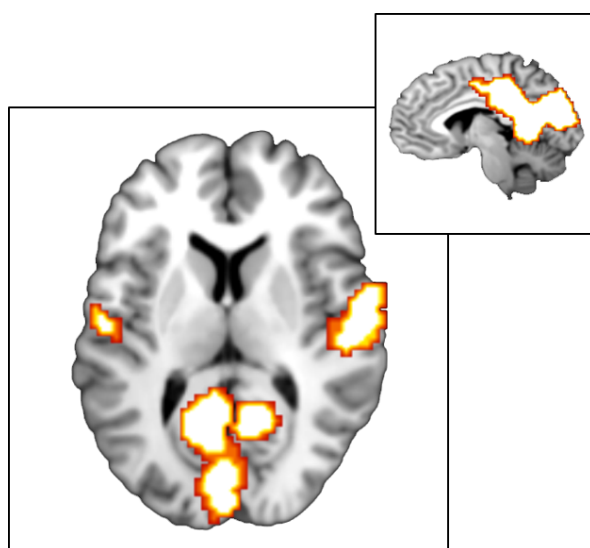434 by effects of OXT during the preceding encoding phase of the social interaction.

435

436

437 **Figure 4.** Neural responses to observed exclusion > inclusion in the entire sample (i.e. oxytocin and

438 placebo treated groups combined – $p < 0.05$; FWE corrected, cluster level).

439

440 **Table 2** Brain regions that positively activated under observed exclusion > inclusion. $p<0.05$ (FWE

441 corrected, cluster level)

Regions	Cluster k	x	y	z	t
Posterior cingulate cortex	2100	-3	-84	24	7.00
Midcingulate cortex, L		0	-36	39	6.70
Precuneus, L		-9	-51	3	6.57
Inferior parietal lobule, L	169	-42	-75	39	6.35
Superior temporal sulcus, R	257	60	0	3	5.82
		60	-18	9	3.60
Rolandic operculum, R		48	-15	15	4.23

442

443   For the PLAY session, a 2 (Treatment: PLC/OXT) × 3 (Player:

444   Excluder/Victim/New) mixed ANOVA using the flexible factorial model in SPM was

445   performed on the whole brain level but no significant results were found (FWE-

446   corrected $p < 0.05$). Based on the behavioral results indicating that OXT administration

447   specifically increased the proportion of throws made to the excluder player, an

448   independent t-test was performed to examine neural activity differences between the

449   treatment groups during this condition. Results indicated significantly stronger left

450   medial orbitofrontal cortex (mOFC) activity ($k = 106$, $t = 4.40$, $x/y/z$: -6/24/-6, $p = 0.047$,

451   Fig. 5) when playing with excluder players in the OXT group compared to the PLC

452   group (FWE-corrected $p < 0.05$). In line with the behavioral findings exploratory t-tests

453   revealed no significant treatment effects on neural activity when playing with the victim

454   or the new player (both ps > 0.348).
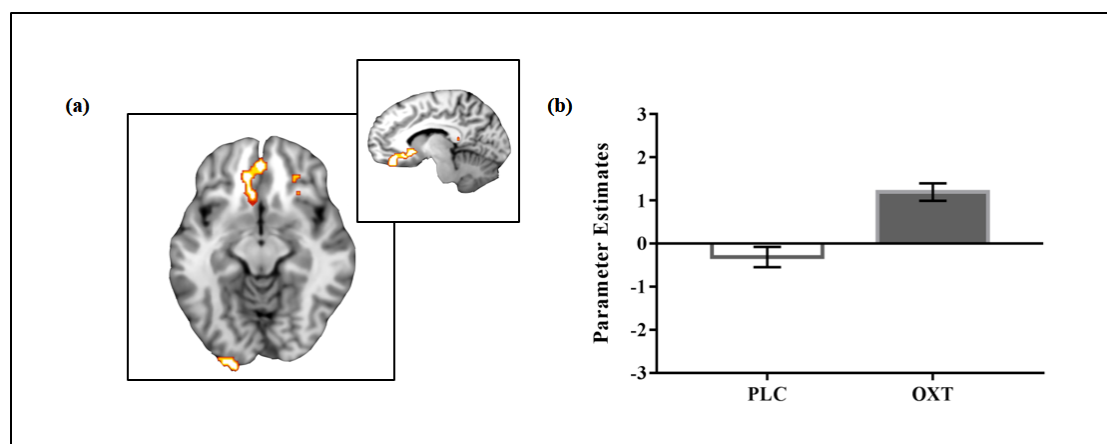
455

456

457   **Figure 5.** (a) Neural response to excluder player under oxytocin (OXT) > placebo (PLC). $p < 0.05$

458   (FWE corrected, cluster level), (b) Corresponding beta parameters of the left mOFC in the PLC and

459   OXT group.

460

461   **Associations between trait altruism behavior ratings and neural activity**

462   A potential modulatory influence of pre-treatment variations in trait altruism (PTM

463   scores) on behavioral and neural responses was investigated using a correlation analysis.

464   No significant associations between the proportion of balls thrown to the three different

465   players and PTM scores were found in either the PLC (Excluder – r = -0.172, p = 0.283;

466   Victim – r = 0.091, p = 0.57; New – r = 0.075, p = 0.64) or OXT (Excluder – r = -0.131,

467   p = 0.413; Victim – r = 0.227, p = 0.154; New – r = -0.114, p = 0.478) groups, or

468   differences between the groups (all ps>0.41). Correlations between left mOFC activity

469   and PTM scores during play with the excluder were also examined in the PLC and OXT

470   groups. Results revealed that PTM scores were not associated with mOFC activity

471   following PLC, but that OXT treatment established a significant negative association

472   between PTM trait altruism and activity in this region (PLC: r = 0.294, p = 0.094; OXT:

473   r = -0.329, p = 0.047; significant different correlations confirmed by Fisher's z = 2.59,

474   p = 0.040, Cohen's q = 0.28, Bonferroni corrected, Fig. 6). This shows that OXT

475   particularly increased left mOFC activity in individuals with lower trait altruism.

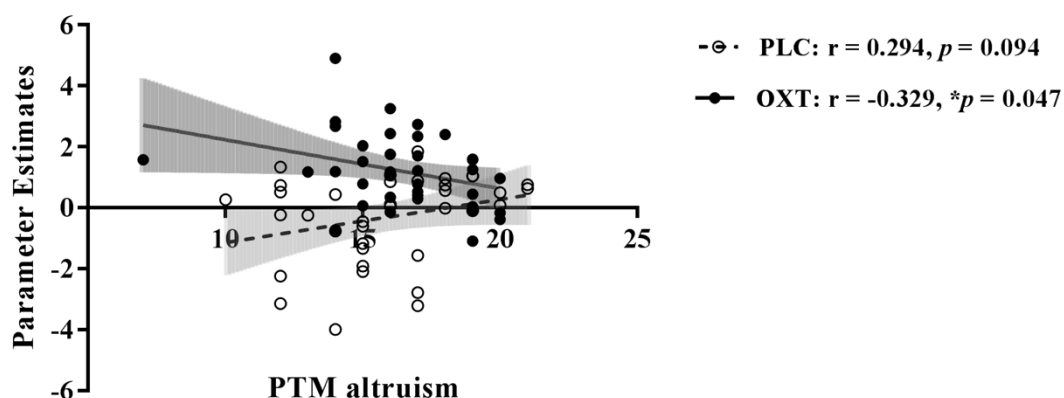476   Horizontal Independence was not associated with either behavior or mOFC activity.

477

478
479   **Figure 6.** Correlations between pre-treatment trait altruism score (PTM) and neural activity when

480   interacting with the excluder player in the placebo (PLC) and oxytocin (OXT) treatment groups.

481   Best-fit line with 95% confidence bands, *p < 0.05

482

483   **Follow-up assessment after one week**

484   Accuracy on memory for the faces of the different players was assessed in the two

485   treatment groups during a surprise recognition memory test one week after the initial

486   Cyberball games, however both groups achieved a very high accuracy (all OXT treated

487   subjects scored 100% accuracy as did 38/41 of the PLC treated ones) and so no

488   meaningful statistical comparison could be made. A 2 (Treatment: PLC/OXT) × 3

489   (Player: Excluder/Victim/New) mixed ANOVA using the willingness to replay rating

490   scores as a dependent variable revealed a main effect of player ($F_{2,160} = 10.46, p < 0.001$,

491    $\eta^2_p = 0.116$), with subjects reporting a stronger preference to play with the victim and

492   new players again as compared to the excluder players (excluder = 3.36±0.27 vs. victim

493   = 4.89±0.32, $p = 0.002$, Cohen's $d = 0.562$; excluder vs. new = 4.93±0.29, $p < 0.001$,

494   Cohen's $d = 0.611$). A treatment × player interaction ($F_{2,160} = 3.55, p = 0.031, \eta^2_p = $

495   0.042,) showed additionally that whereas subjects in the PLC group were significantly

496   less willing to play again with excluders compared to the victim or new players

497   (excluder vs. victim, excluder vs. new, both post hoc ps < .005) this was not the case in

498   the OXT group (both ps > 0.097).

499

500   **Discussion**

501   The current study aimed firstly to establish whether OXT treatment enhanced empathic

502   behavior and neural responses towards observing that a person (victim) is being socially

503   excluded, and secondly whether it promoted altruistic or self-serving behaviors and

504   associated neural responses under circumstances when these two behaviors are in

505   competition with one another. Overall, on the behavioral level following observation of

506   a modified Cyberball game both groups showed strong empathic responses towards the

507   victim players and also greater likeability and trustworthiness ratings for them

508   compared with excluder players, although OXT did not potentiate this. On the neural

509   level, observation of social exclusion was accompanied by increased activity in the

510   mentalizing network, including core regions such as PCC, pSTS, IPL and precuneus,

511   however, in line with the lack of behavioral OXT effects, neural activity patterns were

512   not influenced by OXT treatment. During the subsequent play phase of the paradigm

513   subjects in both treatment groups threw more balls to the victim player but this effect

514   was not enhanced by OXT, suggesting that it did not promote greater altruistic behavior.

515   However, the OXT group threw significantly more balls to the excluder player

516   suggesting that OXT promoted self-serving decisions since playing with the excluder

517   should lead to a higher monetary pay-off. In line with our hypothesis, increased self-

518   serving behavior following OXT was associated with stronger activation in the mOFC

519   reward system when subjects were playing with the excluder player. Furthermore, OXT

520   established a negative relationship between mOFC activity and trait altruism, an

521   association that was absent following PLC. One week after the Cyberball game while

522   the PLC group expressed a greater preference to play again with the victim and novel

523   players compared to the excluder, in the OXT group there was no such difference

524   indicating that they maintained their greater interest in playing with individuals who

525   might potentially help them gain larger rewards. Thus overall, our results demonstrate

526   that when altruistic and self-serving motivations are in competition OXT rather than

527   promoting altruism actually enhances selfish decision making.

528   Our hypothesis that OXT would enhance empathic responses towards the victims

529   of exclusion in the Cyberball game was not supported at either the behavioral or the

530   neural level. Previous research combined OXT treatment with the multifaceted empathy

531   task to demonstrate that it enhanced emotional but not cognitive empathy towards

532   individuals in both positive and negative valence contexts (Hurlemann et al., 2010;

533   Geng et al., 2018), an effect that was associated with suppressed amygdala responses

534   (Geng et al., 2018). Oxytocin also increased empathic embarrassment in male and

535   female subjects and this was associated with decreased amygdala and insula cortex

536   responses, but with no effect on mentalizing networks (Becker et al., 2018). Empathic

537   embarrassment can be considered as an example of social pain and another study has

538   also reported that OXT decreased insula responses to viewing people in pain (Bos et al.,

539   2015). In the current study in both the PLC and OXT groups there was evidence at the

540   whole brain level for increased activation in core mentalizing regions (for convergent

541   findings see also (Masten et al., 2011)) but no responses in the pain network (notably

542   the insula).

543   In general, empathic ratings by subjects for the victim were not that high and since

544   altered activation was only observed in the mentalizing network, which we have shown

545   in the context of empathic embarrassment is not influenced by OXT (Becker et al.,

546   2018), it seems possible that the empathy experienced in the current context was more

547   cognitive than emotional. Indeed, given that the participants could not directly observe

548   the other players and static neutral faces were used, the social pain had to be inferred

549   by mentalizing how the victim may emotionally experience exclusion from the game.

550   Together with the fact that mentalizing system is a core component of the cognitive

551   empathy system (Shamay-Tsoory, 2011), this may explain why OXT failed to have any

552   impact on empathic ratings or associated likeability and trustworthiness ratings for the

553   victim in the current study since it has a greater influence on emotional rather than

554   cognitive empathy.

555   Our original hypothesis that in a situation where altruistic and self-serving

556   motivations were in competition OXT would enhance altruistic responses was also not

557   supported in the current experiment. Overall, participants in both groups threw more

558   balls to the victim player indicative of an altruistic response and validating the

559   experimental manipulation, however OXT had no effect on this. Similarly, both groups

560   rated the likeability and trustworthiness of excluder players lower than that of victims

561   after observing the game and also one week later after playing it, but again this was not

562   influenced by OXT. This general finding is in agreement with previous studies which

563   also found that subjects exhibit greater prosocial behavior towards individuals who

564   have been observed to be socially excluded (Masten et al., 2011; Van Der Meulen, Van

565   IJzendoorn, & Crone, 2016). On the other hand, the OXT group threw a significantly

566   greater proportion of balls to the excluder player indicative of an enhanced self-serving

567   motivation since the excluder player would have been perceived as being more likely

568   to reciprocate and potentially result in greater financial gain for the participant.

569   While a number of previous studies have demonstrated that OXT can facilitate

570   altruistic behaviors in terms of cooperation, generosity, trait-judgements and valuing

571   other's possessions, these have mainly involveed contexts where personal costs to

572   individuals were absent or low (Declerck, Boone, & Kiyonari, 2010; Andari et al., 2010;

573   Riem et al., 2011; Zhao et al., 2016; 2017). The finding in the current study that under

574   circumstances where there is perceived to be a potential cost of altruism in terms of

575   reduced personal gain argues for the primary function of OXT as enhancing the

576   motivation for resource acquisition. Where individuals do exhibit costly altruistic

577   behavior, this is paralleled by increased empathic concern and altered activation in the

578   ventral tegmental area, caudate and subgenual anterior cingulate which are important

579   for promoting social attachment and caregiving (FeldmanHall, Dalgleish, Evans, &

580   Mobbs, 2015). Although OXT has been show to modulate neural processing in these

581   regions in social and non-social contexts (e.g. Scheele et al., 2013; Mickey et al., 2016;

582   Zhao et al., 2019) it did not affect activity in this circuitry during the present paradigm,

583   further indicating its lack of effect on promoting altruism in the current context.

584   Previous studies have reported that OXT can promote lying for the benefit of in-

585   group members, including participants themselves, although not lying purely for self-

586   gain (Shalvi, & De Dreu, 2014) and that it reduced honesty for personal gain only in a

587   competitive environment (Aydogan et al., 2017). However, OXT can also in some

588   circumstances promote pure self-serving lying to increase personal gain in men when

589   there is no risk of discovery (Sindermann et al., 2018) and can increase acceptance of

590   self-benefit moral dilemmas, but importantly not acceptance of other types of moral

591   dilemma (Scheele et al., 2014). Interestingly, OXT effects on self-serving lying for

592   financial gain are modulated by OXT receptor genotype (Sindermann et al., 2018) and

593   thus it is possible that the effects of OXT on altruistic compared with self-benefit

594   behaviors when they are in conflict might be to some extent genotype dependent. Taken

595   together, and in line with our current results, accumulating evidence therefore suggests

596   that OXT can promote personal self-interest in some contexts.

597   On the neural level the present study revealed that increased self-serving behavior

598   following OXT was neurally underpinned by regional- and player-specific increased

599   mOFC activation when participants interacted with excluder players. The mOFC is

600   involved in monitoring associations between previous stimuli with reward and tracking

601   response-outcome probabilities during changing reward contingencies (Elliott, Dolan,

602   & Frith, 2000; Kringelbach & Rolls, 2004). Moreover, the mOFC codes the value of

603   different behavioral options (Padoa-Schioppa, & Assad, 2006), including the value of

604   expected monetary gains (Breiter, Aharon, Kaheman, Dale, & Shizgal, 2001), and

605   activity in this region increases with monetary reward magnitude (O'Doherty,

606   Kringelbach, Rolls, Hornak, & Andrews, 2001). Thus, in the current context increased

607   mOFC responses may reflect an enhanced value of the expected greater monetary

608   reward when cooperating with the excluder player in subjects in the OXT group.

609   Additionally, OXT treatment produced a negative association between PTM trait

610   altruism and mOFC activation that was absent during PLC treatment. This suggests that

611   at the neural level OXT particularly increased the value of the potential monetary gain

612   in subjects with low baseline altruistic tendencies. Possibly individuals with high trait

613   altruism might be less likely to experience a greater anticipation of gaining a greater

614   monetary reward by playing with the excluder under OXT, as a result of greater feelings

615   of guilt evoked by having to exclude the other players, and notably the victim.

616   While both behavioral and neural effects of OXT observed in the current paradigm

617   do indicate a shift towards a self-serving rather than altruistic motivation it is notable

618   that the pattern of altered bias is quite subtle. Under OXT, participants do not actually

619   play more with the excluders than with either the victim or the novel player and

620   effectively exhibit an egalitarian playing pattern which is therefore unlikely to create

621   any feelings of exclusion in any of the other players. This is in contrast to participants

622   in the PLC group who show a clear pattern of excluding the previous excluder by

623   comparison with both the victim and novel players. As such the effect of OXT could be

624   viewed as promoting self-benefit behavior but only if it does not damage others and

625   therefore possibly generate significant feelings of guilt. Indeed, this is in accordance

626   with the findings that OXT increases lying for self-gain when individuals know that

627   there is no chance their lies will be discovered or that this comes at the cost of reducing

628   the financial gain of others (Sinderman et al., 2018). Alternatively, it might be argued

629   that subjects in the PLC group are exhibiting altruistic punishment towards the

630   excluders and OXT is reducing the desire to inflict such punishment. However, altruistic

631   punishment is strongly associated with altered amygdala function (Scheele et al., 2012)

632   and there was no evidence for differential amygdala responses in the PLC and OXT

633   groups. While a previous study has reported that OXT can promote altruistic

634   punishment of defectors, and feelings of anger and disappointment towards them, in an

635   economic game context, it also increases co-operation with them thereby increasing

636   self-gain (Prisoner's dilemma game: Aydogan et al., 2017). Thus, on balance, it is likely

637   that in the current context OXT primarily biases individuals towards an optimal self-

638 gain strategy, although without simultaneously doing so by overtly damaging others

639 emotionally. Clearly, to further establish this it would be necessary to investigate the

640 effects of OXT under circumstances where increasing self-gain would also significantly

641 damage others emotionally.

642 There are several limitations in the present study which should be noted. Firstly,

643 the victim observed being excluded in the Cyberball paradigm in the current study was

644 a stranger to the participant and therefore it is possible that if they had been a partner,

645 relative or other in-group member then OXT may have had the opposite effect by

646 enhancing empathic and altruistic behaviors rather than self-serving ones. Further

647 studies would be required to establish this. Secondly, and relatedly, only male

648 participants were included in the current study and previous studies reporting OXT-

649 induced enhancement of self-serving behaviors have also only included male subjects.

650 A number of studies have reported opposite neural and behavioral effects of OXT in

651 males and females (Gao et al., 2016; Luo et al., 2017) and one study found that while it

652 enhanced acceptance of self-benefit moral dilemmas in men it had the opposite effect

653 in women (Scheele et al., 2014). Thus, it remains possible that in a similar competing

654 motivation situation OXT may have facilitated altruistic rather than self-serving

655 behavior in women.

656 In summary, our current findings demonstrate that under circumstances where self-

657 serving and altruistic behaviors are in competition OXT promotes an increased

658 tendency in men towards self-benefit behavior, and this is associated with increased

659 activation in the mOFC indicative of greater reward anticipation. Furthermore, the

660 effects of OXT on mOFC are strongest in individuals with lower trait altruism. Thus,

661 contrary to expectation, OXT did not promote altruistic behavior although equally it did

662 not reduce it to the point where it might potentially have produced negative feelings in

663 others. It seems likely therefore that in men OXT tends to bias individuals towards

664 acquisition of resources for self-benefit rather than helping others. However, this self-

665 serving bias may not extend to the point where it generates strong negative feelings in

666 others which could result in significant feelings of guilt and risk of punishment for

667 social norm violations.

668

Acknowledgments

672

Conflict of interest

The authors declare no financial interests or potential conflict of interest.

675

Author contributions

X.X., B.B, and K.M.K designed the study and wrote the paper. X.X., C.L., Y.C. carried out the study. X.X., X.Z., Z.G., F.Z., and J.K. analyzed the data. All authors contributed to and have approved the final version of the manuscript.

680

References

Andari, E., Duhamel, J.-R., Zalla, T., Herbrecht, E., Leboyer, M., & Sirigu, A. (2010). Promoting social behavior with oxytocin in high-functioning autism spectrum disorders. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 4389–4394. https://doi.org/10.1073/pnas.0910249107

Aydogan, G., Furtner, N. C., Kern, B., Jobst, A., Müller, N., & Kocher, M. G. (2017). Oxytocin promotes altruistic punishment. *Social Cognitive and Affective Neuroscience, 12,* 1740–1747. https://doi.org/10.1093/scan/nsx101

Aydogan, G., Jobst, A., D'Ardenne, K., Müller, N., & Kocher, M. G. (2017). The Detrimental Effects of Oxytocin-Induced Conformity on Dishonesty in Competition. *Psychological Science, 28,* 751–759. https://doi.org/10.1177/0956797617695100

Barnes, L. L., Harp, D., & Jung, W. S. (2002). Reliability generalization of scores on the Spielberger state-trait anxiety inventory. *Educational and psychological measurement, 62,* 603-618. https://doi.org/10.1177/0013164402062004005

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism Spectrum Quotient : Evidence from Asperger syndrome/high functioning

autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31,* 5–17. https://doi.org/10.1023/A:1005653411471

Batson, C. D. (2011). Altruism in humans. Oxford University Press, USA.

Beck, A. T., Steer, R. A., Brown, G. K. (1996). Manual for Beck depression inventory. San Antonio, TX: The Psychological Corporation.

Becker, B., GENG, Y., Zhao, W., Zhou, F., Ma, X., Yao, S., & Kendrick, K. M. (2018). Oxytocin facilitates empathic-and self-embarrassment ratings by attenuating amygdala and anterior insula responses. *Frontiers in endocrinology*, *9*, 572. https://doi.org/10.3389/fendo.2018.00572

Bernhardt, B. C., & Singer, T. (2012). The Neural Basis of Empathy. *Annual Review of Neuroscience, 35,* 1–23. https://doi.org/10.1146/annurev-neuro-062111-150536

Bernstein, D. P., & Fink, L. (1998). Childhood trauma questionnaire: A retrospective self-report: Manual. San Antonio, TX: The Psychological Corporation.

Bethlehem, R. A. I., van Honk, J., Auyeung, B., & Baron-Cohen, S. (2013). Oxytocin, brain physiology, and functional connectivity: A review of intranasal oxytocin fMRI studies. *Psychoneuroendocrinology, 38,* 962–974. https://doi.org/10.1016/j.psyneuen.2012.10.011

Bos, P. A., Montoya, E. R., Hermans, E. J., Keysers, C., & van Honk, J. (2015). Oxytocin reduces neural activity in the pain circuitry when seeing pain in others. *NeuroImage, 113,* 217–224. https://doi.org/10.1016/j.neuroimage.2015.03.049

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences, 100,* 3531–3535. https://doi.org/10.1073/pnas.0630443100

Breiter, H. C., Aharon, I., Kahneman, D., Dale, A., & Shizgal, P. (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron, 30,* 619-639. https://doi.org/10.1016/S0896-6273(01)00303-8

Bruneau, E. G., Pluta, A., & Saxe, R. (2012). Distinct roles of the "Shared Pain" and "Theory of Mind" networks in processing others' emotional suffering. *Neuropsychologia, 50,* 219–231.

728   https://doi.org/10.1016/j.neuropsychologia.2011.11.008

729   Carlo, G., & Randall, B. A. (2002). The development of a measure of prosocial

730   behaviors for late adolescents. *Journal of youth and adolescence, 31,* 31-44.

731   Camerer, C. F., & Fehr, E. (2006). When does" economic man" dominate social

732   behavior?. *Science, 311,* 47-52. https://doi.org/10.1126/science.1110600

733   de Waal, F. B. M. (2008). Putting the Altruism Back into Altruism: The Evolution of

734   Empathy.   *Annual   Review   of   Psychology,   59,*   279–300.

735   https://doi.org/10.1146/annurev.psych.59.103006.093625

736   Decety, J. (2011). The neuroevolution of empathy. *Annals of the New York Academy of*

737   *Sciences, 1231,* 35–45. https://doi.org/10.1111/j.1749-6632.2011.06027.x

738   Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy.

739   *Behavioral   and   cognitive   neuroscience   reviews,   3,*   71-100.

740   https://doi.org/10.1177/1534582304267187

741   Declerck, C. H., Boone, C., & Kiyonari, T. (2010). Oxytocin and cooperation under

742   conditions of uncertainty: The modulating role of incentives and social

743   information.   *Hormones   and   Behavior,   57,*   368–374.

744   https://doi.org/10.1016/j.yhbeh.2010.01.006

745   De Quervain, D. J., Fischbacher, U., Treyer, V., & Schellhammer, M. (2004). The neural

746   basis   of   altruistic   punishment.   *Science,   305,*   1254-1258.   doi:

747   10.1126/science.1100735

748   Eisenberg, N., & Eggum, N. D. (2009). Empathic responding: Sympathy and personal

749   distress.   *The   social   neuroscience   of   empathy,   6,*   71-83.

750   https://doi.org/10.7551/mitpress/9780262012973.003.0007

751   Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt?

752   An   fMRI   study   of   social   exclusion.   *Science,   302,*   290–292.

753   https://doi.org/10.1126/science.1089134

754   Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: why fMRI inferences

755   for spatial extent have inflated false-positive rates. *Proceedings of the National*

756   *Academy of Sciences U S A,* 201602413. https://doi.org/10.1073/pnas.1602413113

757   Elliott, R., Dolan, R. J., & Frith, C. D. (2000). Dissociable functions in the medial and

lateral orbitofrontal cortex: Evidence from human neuroimaging studies. *Cerebral Cortex, 10,* 308–317. https://doi.org/10.1093/cercor/10.3.308

FeldmanHall, O., Dalgleish, T., Evans, D., & Mobbs, D. (2015). Empathic concern drives costly altruism. *NeuroImage, 105,* 347–356. https://doi.org/10.1016/j.neuroimage.2014.10.043

Gao S, Becker B, Luo L, Geng Y, Zhao W, Yin Y, Hu J, Gao Z, Gong Q, Hurlemann R, Yao D and Kendrick KM (2016) Oxytocin, the peptide that bonds the sexes also divides them. *Proceedings of the National Academy of Sciences U S A, 113*, 7650-7654. https://doi.org/10.1073/pnas.1602620113

Geng, Y., Zhao, W., Zhou, F., Ma, X., Yao, S., Hurlemann, R., ... & Kendrick, K. M. (2018). Oxytocin enhancement of emotional empathy: generalization across cultures and effects on amygdala activity. *Frontiers in neuroscience, 12.* doi: [10.3389/fnins.2018.00512]

Guastella, A. J., Hickie, I. B., McGuinness, M. M., Otis, M., Woods, E. A., Disinger, H. M., … Banati, R. B. (2013). Recommendations for the standardisation of oxytocin nasal administration and guidelines for its reporting in human research. *Psychoneuroendocrinology, 38,* 612–625. https://doi.org/10.1016/j.biopsych.2009.09.020

Haas, B. W., Brook, M., Remillard, L., Ishak, A., Anderson, I. W., & Filkowski, M. M. (2015). I know how you feel: The warm-altruistic personality profile and the empathic brain. *PloS one, 10*, e0120639. https://doi.org/10.1371/journal.pone.0120639

Heimberg, R. G., Horner, K. J., Juster, H. R., Safren, S. A., Brown, E. J., Schneier, F. R., & Liebowitz, M. R. (1999). Psychometric properties of the Liebowitz social anxiety scale. *Psychological medicine, 29,* 199-212. https://doi.org/10.1017/S0033291798007879

Hillis, A.E. (2014). Inability to empathize: brain lesions that disrupt sharing and understanding another's emotions. *Brain 137,* 981-997. https://doi.org/10.1093/brain/awt317

Hu, Y., Scheele, D., Becker, B., Voos, G., David, B., Hurlemann, R., & Weber, B. (2016).

788   The effect of oxytocin on third-party altruistic decisions in unfair situations: An

789   fMRI study. *Scientific Reports, 6,* 1–10. https://doi.org/10.1038/srep20236

790 Hurlemann, R., Patin, A., Onur, O.A., Cohen, M.X., Baumgartner, T., Metzler, S., et al.

791   (2010). Oxytocin Enhances Amygdala-Dependent, Socially Reinforced Learning

792   and Emotional Empathy in Humans. *Journal of Neuroscience, 30,* 4999-5007.

793   https://doi.org/10.1523/JNEUROSCI.5538-09.2010

794 Kendrick K.M., Guastella A.J., Becker B. (2017) Overview of Human Oxytocin

795   Research. In: Hurlemann R., Grinevich V. (eds) Behavioral Pharmacology of

796   Neuropeptides: Oxytocin. Current Topics in Behavioral Neurosciences, vol 35.

797   Springer, Cham

798 Kringelbach, M. L., & Rolls, E. T. (2004). The functional neuroanatomy of the human

799   orbitofrontal cortex: Evidence from neuroimaging and neuropsychology. *Progress*

800   *in Neurobiology, 72,* 341–372. https://doi.org/10.1016/j.pneurobio.2004.03.006

801 Kringelbach, M. L. (2005). The human orbitofrontal cortex: linking reward to hedonic

802   experience. *Nature Reviews Neuroscience, 6,* 691-702. doi: 10.1038/nrn1747

803 Kross, E., Berman, M. G., Mischel, W., Smith, E. E., & Wager, T. D. (2011). Social

804   rejection shares somatosensory representations with physical pain. *Proceedings of*

805   *the National Academy of Sciences U S A, 108,* 6270–6275.

806   https://doi.org/10.1073/pnas.1102693108

807 Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and

808   distinct neural networks associated with directly experienced pain and empathy

809   for pain. *NeuroImage, 54,* 2492–2502.

810   https://doi.org/10.1016/j.neuroimage.2010.10.014

811 Luo L, Becker B, Geng Y, Zhao Z, Gao S, Zhao W, Yao S, Zheng X, Ma X, Zhao G,

812   Hu J, Kendrick KM (2017) The sex-dependent effect of oxytocin to masked

813   negative faces: associations with anxiety. *Neuroimage, 162,* 127-137.

814   https://doi.org/10.1016/j.neuroimage.2017.08.079

815 MacDonald, K., & MacDonald, T. M. (2010). The peptide that binds: a systematic

816   review of oxytocin and its prosocial effects in humans. *Harvard review of*

817   *psychiatry, 18,* 1-21. https://doi.org/10.1016/j.neuroimage.2017.08.079

818 Marsh, A. A. (2018). The neuroscience of empathy. *Current Opinion in Behavioral*
819   *Sciences, 19,* 110–115. https://doi.org/10.1016/j.cobeha.2017.12.016

820 Masten, C. L., Morelli, S. A., & Eisenberger, N. I. (2011). An fMRI investigation of
821   empathy for "social pain" and subsequent prosocial behavior. *NeuroImage, 55,*
822   381–388. https://doi.org/10.1016/j.neuroimage.2010.11.060

823 Mickey, B. J., Heffernan, J., Heisel, C., Peciña, M., Hsu, D. T., Zubieta, J. K., & Love,
824   T. M. (2016). Oxytocin modulates hemodynamic responses to monetary incentives
825   in   humans.   *Psychopharmacology,   233,   3905-3919.*
826   https://doi.org/10.1007/s00213-016-4423-6

827 Miller, D. T. (2001). The norm of self-interest. The next phase of business ethics:
828   Integrating psychology and ethics (pp. 193-210). Emerald Group Publishing
829   Limited.

830 O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001).
831   Abstract reward and punishment representations in the human orbitofrontal cortex.
832   *Nature Neuroscience, 4,* 95–102. https://doi.org/10.1038/82959

833 Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode
834   economic value. *Nature, 441,* 223-226. https://doi.org/10.1038/nature04676

835 Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases.
836   *Behavioral   and   Brain   Sciences,   25,   1–20.*
837   https://doi.org/10.1017/S0140525X02000018

838 Riem, M. M. E., Bakermans-Kranenburg, M. J., Pieper, S., Tops, M., Boksem, M. A.
839   S., Vermeiren, R. R. J. M., … Rombouts, S. A. R. B. (2011). Oxytocin modulates
840   amygdala, insula, and inferior frontal gyrus responses to infant crying: A
841   randomized   controlled   trial. *Biological   Psychiatry,   70,   291–297.*
842   https://doi.org/10.1016/j.biopsych.2011.02.006

843 Riem, M. M., Bakermans-Kranenburg, M. J., Huffmeijer, R., & van IJzendoorn, M. H.
844   (2013). Does intranasal oxytocin promote prosocial behavior to an excluded fellow
845   player? A randomized-controlled trial with Cyberball. *Psychoneuroendocrinology*,
846   *38*, 1418-1425. https://doi.org/10.1016/j.psyneuen.2012.12.023

847 Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D.

848   (2002). A neural basis for social cooperation. *Neuron, 35,* 395–405.
849   https://doi.org/10.1016/S0896-6273(02)00755-9

850   Rodrigues, S. M., Saslow, L. R., Garcia, N., John, O. P., & Keltner, D. (2009). Oxytocin
851   receptor genetic variation relates to empathy and stress reactivity in humans.
852   *Proceedings of the National Academy of Sciences U S A, 106,* 21437–21441.
853   https://doi.org/10.1038/ncpneuro

854   Rumble, A. C., Van Lange, P. A., & Parks, C. D. (2010). The benefits of empathy: When
855   empathy may sustain cooperation in social dilemmas. *European Journal of Social*
856   *Psychology, 40,* 856-866. https://doi.org/10.1002/ejsp.659

857   Scheele D, Mihov Y, Kendrick KM, Maier W and Hurlemann R. (2012) Amygdala
858   lesion profoundly alters altruistic punishment. *Biological Psychiatry, 72,* e5-7.
859   https://doi.org/10.1016/j.biopsych.2012.01.028

860   Scheele D, Striepens N, Kendrick KM, Schwering C, Noelle J, Wille A, Schlapfer TE,
861   Maier W and Hurlemann R (2014) Opposing effects of oxytocin on moral
862   judgment in males and females. *Human Brain Mapping, 35,* 6067-6076
863   https://doi.org/10.1002/hbm.22605

864   Scheele, D., Wille, A., Kendrick, K. M., Stoffel-Wagner, B., Becker, B., Güntürkün,
865   O., ... & Hurlemann, R. (2013). Oxytocin enhances brain reward system responses
866   in men viewing the face of their female partner. *Proceedings of the National*
867   *Academy   of   Sciences   U   S   A,   110,*   20308-20313.
868   https://doi.org/10.1073/pnas.1314190110

869   Shalvi, S., & De Dreu, C. K. W. (2014). Oxytocin promotes group-serving dishonesty.
870   *Proceedings of the National Academy of Sciences U S A, 111,* 5503–5507.
871   https://doi.org/10.1073/pnas.1400724111

872   Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy:
873   a double dissociation between emotional and cognitive empathy in inferior frontal
874   gyrus   versus   ventromedial   prefrontal   lesions.   *Brain*,   *132*,   617-627.
875   https://doi.org/10.1093/brain/awn279

876   Shamay-Tsoory, S. G. (2011). The neural bases for empathy. *Neuroscientist, 17,* 18–24.
877   https://doi.org/10.1177/1073858410379268

878   Shamay-Tsoory, S. G., Abu-Akel, A., Palgi, S., Sulieman, R., Fischer-Shofty, M.,
879   Levkovitz, Y., & Decety, J. (2013). Giving peace a chance: Oxytocin increases
880   empathy to pain in the context of the Israeli-Palestinian conflict.
881   *Psychoneuroendocrinology,   38,*   3139–3144.
882   https://doi.org/10.1016/j.psyneuen.2013.09.015

883   Sindermann, C., Luo, R., Becker, B., Kendrick, K. M., & Montag, C. (2018). Oxytocin
884   promotes lying for personal gain in a genotype-dependent manner. *bioRxiv,*
885   361212. https://doi.org/10.1101/361212

886   Singelis, T. M., Triandis, H. C., Bhawuk, D. P., & Gelfand, M. J. (1995). Horizontal
887   and vertical dimensions of individualism and collectivism: A theoretical and
888   measurement   refinement.   *Cross-cultural   research,   29,*   240-275.
889   https://doi.org/10.1177/106939719502900302

890   Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading:
891   Review of literature and implications for future research. *Neuroscience and*
892   *Biobehavioral   Reviews,   30,*   855–863.
893   https://doi.org/10.1016/j.neubiorev.2006.06.011

894   Siu, A. M., & Shek, D. T. (2005). Validation of the Interpersonal Reactivity Index in a
895   Chinese   context.   *Research   on   Social   Work   Practice,   15,*   118-126.
896   https://doi.org/10.1177/1049731504270384

897   Slotnick, S. D. (2017). Cluster success: fMRI inferences for spatial extent have
898   acceptable   false-positive   rates.   *Cognitive   neuroscience,   8,*   150-155.
899   https://doi.org/10.1080/17588928.2017.1319350

900   Smith, K. E., Porges, E. C., Norman, G. J., Connelly, J. J., & Decety, J. (2014). Oxytocin
901   receptor gene variation predicts empathic concern and autonomic arousal while
902   perceiving   harm   to   others.   *Social   Neuroscience,   9,*   1–9.
903   https://doi.org/10.1080/17470919.2013.863223

904   Spicer, J., Galvan, A., Hare, T. A., Voss, H., Glover, G., & Casey, B. J. (2007).
905   Sensitivity of the nucleus accumbens to violations in expectation of reward.
906   *NeuroImage, 34,* 455-461. https://doi.org/10.1016/j.neuroimage.2006.09.012

907   Striepens, N., Kendrick, K. M., Maier, W., & Hurlemann, R. (2011). Prosocial effects

of oxytocin and clinical evidence for its therapeutic potential. *Frontiers in Neuroendocrinology, 32,* 426–450. https://doi.org/10.1016/j.yfrne.2011.07.001

Striepens, N., Scheele, D., Kendrick, K. M., Becker, B., Schäfer, L., Schwalba, K., … Hurlemann, R. (2012). Oxytocin facilitates protective responses to aversive social stimuli in males. *Proceedings of the National Academy of Sciences U S A, 109,* 18144–18149. https://doi.org/10.1073/pnas.1208852109

Van Der Meulen, M., Van IJzendoorn, M. H., & Crone, E. A. (2016). Neural correlates of prosocial behavior: Compensating social exclusion in a four-player cyberball game. *PLoS ONE, 11,* 1–13. https://doi.org/10.1371/journal.pone.0159045

Vekaria, K. M., Brethel-Haurwitz, K. M., Cardinale, E. M., Stoycos, S. A., & Marsh, A. A. (2017). Social discounting and distance perceptions in costly altruism. *Nature Human Behaviour, 1,* 0100. https://doi.org/10.1038/s41562-017-0100

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology, 54,* 1063.

Wesselmann, E. D., Bagg, D., & Williams, K. D. (2009). "I Feel Your Pain": The effects of observing ostracism on the ostracism detection system. *Journal of Experimental Social Psychology, 45,* 1308–1311. https://doi.org/10.1016/j.jesp.2009.08.003

West, S. A., Griffin, A. S., & Gardner, A. (2007). Social semantics: Altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology, 20,* 415–432. https://doi.org/10.1111/j.1420-9101.2006.01258.x

Wigton, R., Jocham Radua, P. A., Averbeck, B., Meyer-Lindenberg, A., McGuire, P., Shergill, S. S., & Fusar-Poli, P. (2015). Neurophysiological effects of acute oxytocin administration: systematic review and meta-analysis of placebo-controlled imaging studies. *Journal of psychiatry & neuroscience: JPN, 40,* E1. https://doi.org/10.1503/jpn.130289

Williams, K. D. (2009). Chapter 6 Ostracism: A Temporal Need-Threat Model. *Advances in Experimental Social Psychology, 41,* 275–314. https://doi.org/10.1016/S0065-2601(08)00406-1

Wu, N., Li, Z., & Su, Y. (2012). The association between oxytocin receptor gene polymorphism (OXTR) and trait empathy. *Journal of Affective Disorders, 138,* 468–472. https://doi.org/10.1016/j.jad.2012.01.009

Xu, X., Yao, S., Xu, L., Geng, Y., Zhao, W., Ma, X., ... & Kendrick, K. M. (2017). Oxytocin biases men but not women to restore social connections with individuals who socially exclude them. *Scientific reports, 7,* 40589. https://doi.org/10.1038/srep40589

Yao, S., Becker, B., Geng, Y., Zhao, Z., Xu, X., Zhao, W., … Kendrick, K. M. (2016). Voluntary control of anterior insula and its functional connections is feedback-independent and increases pain empathy. *NeuroImage, 130,* 230–240. https://doi.org/10.1016/j.neuroimage.2016.02.035

Zadro, L., Williams, K. D., & Richardson, R. (2004). How low can you go? Ostracism by a computer is sufficient to lower self-reported levels of belonging, control, self-esteem, and meaningful existence. *Journal of Experimental Social Psychology, 40,* 560-567. https://doi.org/10.1016/j.jesp.2003.11.006

Zhao W, Yao S, Li Q, Geng Y, Ma X, Luo L, Xu L, Kendrick KM (2016) Oxytocin blurs the self-other distinction during trait judgments and reduces medial prefrontal cortex responses. *Human Brain Mapping, 37,* 2512-2527. https://doi.org/10.1002/hbm.23190

Zhao W, Geng Y, Luo L, Zhao Z, Ma X, Xu L, Yao S, Kendrick KM (2017). Oxytocin increases the perceived value of both self- and other-owned items and alters medial prefrontal cortex activity in an endowment task. *Frontiers in Human Neuroscience. 11,* 272. https://doi.org/10.3389/fnhum.2017.00272

Zhao, Z., Ma, X., Geng, Y., Zhao, W., Zhou, F., Wang, J., ... & Becker, B. (2019). Oxytocin differentially modulates specific dorsal and ventral striatal functional connections with frontal and cerebellar regions. *NeuroImage, 184,* 781-789. https://doi.org/10.1016/j.neuroimage.2018.09.067

968   **Supplementary information**

969

970   In order to examine if there was any differential effect in the PLAY session of the

971   Cyberball game between the first 4 blocks where participants did not receive a monetary

972   reward and last 2 blocks where they did, we included the monetary condition as an

973   independent variable in the analysis. Given the high correlation between ball tosses to

974   excluder and victim players in the two stages (excluder player r = 0.477, $p < 0.001$;

975   victim player r = 0.340, $p = 0.002$), we used the ball tosses to new player as a baseline

976   and the dependent variable was used by subtracting ball tosses to new player from that

977   of the excluder and victim players. A 2 (Treatment: PLC/OXT) × 2 (Player: Bad/Victim)

978   × 2 (Monetary condition: First 4/Last 2) repeated ANOVA was performed and results

979   showed there was no significant effect of monetary condition (main effect: $p = 0.154$;

980   interaction: monetary condition × drug $p = 0.420$; player × monetary condition $p =$

981   0.409; player × monetary condition × drug $p = 0.146$) which indicated that subjects

982   were consistent in their pattern of play during the entire PLAY session.

983