◆ Human Brain Mapping 35:6067–6076 (2014) ◆

Opposing Effects of Oxytocin on Moral Judgment in Males and Females

Dirk Scheele,[1,2] Nadine Striepens,[1,2] Keith M. Kendrick,[3] Christine Schwering,[1,2] Janka Noelle,[1,2] Andrea Wille,[1,2] Thomas E. Schläpfer,[1,4,5] Wolfgang Maier,[1,6] and René Hurlemann[1,2]*

[1]Department of Psychiatry, University of Bonn, Bonn, Germany
[2]Department of Medical Psychology, University of Bonn, Bonn, Germany
[3]Key Laboratory for Neuroinformation, School of Life Science & Technology, University of Electronic Science & Technology of China (UESTC), Chengdu, People's Republic of China
[4]Departments of Psychiatry, Johns Hopkins University, Baltimore, Maryland
[5]Department of Mental Health, Johns Hopkins University, Baltimore, Maryland
[6]German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

Abstract: Current perspectives on the evolutionary roots of human morality suggest it arose to incentivize social cooperation by promoting feelings of disgust toward selfish behavior, although the underlying neural mechanisms remain unclear. To investigate whether the ancient mammalian neuropeptide oxytocin (OXT) influences self-referential processing in the domains of emotion evaluation and moral decision making, we conducted a pharmaco-functional magnetic resonance imaging (fMRI) and a behavioral experiment involving 157 healthy women and men who were treated with either OXT (24 IU) or placebo (PLC) intranasally. Our results show that OXT facilitated cortical midline responses during self-processing of disgust and selectively promoted self-interest moral judgments in men. In contrast, in women OXT increased the reaction time difference between accepted and rejected moral dilemmas and led them to suppress their self-interest and respond more altruistically for the benefit of others. Taken together, these findings suggest an OXT-related sexual dimorphism in human moral behavior which evolved adaptively to optimize both protection and nurturing of offspring by promoting selfish behavior in men and altruistic behavior in women. *Hum Brain Mapp 35:6067–6076, 2014.* © 2014 Wiley Periodicals, Inc.

Key words: disgust; functional magnetic resonance imaging; dilemma; neuropeptide; self-processing

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: Starting Independent Researcher Grant (NEMO—Neuromodulation of Emotion) jointly provided by the Ministry of Innovation, Science, Research & Technology of the German State of North Rhine-Westphalia (MIWFT) and the University of Bonn (R.H.); Contract grant sponsor: National Natural Science Foundation of China (K.M.K.); Contract grant number: 91132720.

Dirk Scheele, Nadine Striepens, and Keith M. Kendrick contributed equally to this work (shared first authorship)

*Correspondence to: René Hurlemann, Department of Medical Psychology, Department of Psychiatry, University of Bonn, Sigmund-Freud-Str. 25, 53105 Bonn, Germany. E-mail: renehurlemann@me.com

Received for publication 4 May 2014; Revised 21 July 2014; Accepted 28 July 2014.

DOI: 10.1002/hbm.22605

Published online 5 August 2014 in Wiley Online Library (wileyonlinelibrary.com).

INTRODUCTION

The evolutionarily conserved neuropeptide oxytocin (OXT) influences a diverse repertoire of mammalian social behaviors [Eckstein and Hurlemann, 2013; Insel, 2010; Striepens et al., 2011]. For example, OXT can facilitate social-cognitive skills including trust, cooperation and empathy [Hurlemann et al., 2010; Kosfeld et al., 2005] in humans and these prosocial effects may be partially mediated by the peptide's anxiolytic action, particularly in the amygdala [Olff et al., 2013]. While anxiety reduction may enable approach-related behaviors, a reduced amygdala response to harm signals may also interfere with our ability to detect environmental threats on the basis of social cues. From an evolutionary perspective, a mechanism which leaves an organism unprepared for attacks by a predator appears maladaptive. However, we have recently shown that when OXT suppresses amygdala responses, other brain regions including the insula may become more influential in biasing cognitive and emotional processing towards increased preparedness for self-defense [Striepens et al., 2012]. The insula is also important for self-processing and interoceptive awareness [Ernst et al., 2014; Lamm and Singer, 2010] and so OXT may potentially be producing a heightened sense of self which can in turn lead to increased perspective-taking and empathic concern. Other regions implicated in self-processing include a number of cortical midline regions which form part of the default-mode network (DMN), most notably the medial prefrontal cortex, cingulate cortex, and precuneus [Qin and Northoff, 2011]. However, to date no study has systematically investigated whether both neural and behavioral aspects of self-processing are specifically affected by the peptide, although one study has reported enhanced self-rated scores on some positive personality traits [Cardoso et al., 2012]. In terms of other behavioral OXT effects, while many studies have demonstrated facilitation of prosocial behaviors [Hurlemann et al., 2010; Kosfeld et al., 2005; Olff et al., 2013], an increasing number have reported that it can promote more selfish ones such as envy and jealousy [Shamay-Tsoory et al., 2009] and diminish trust in others, adherence to fairness norms and even reduce social co-operation in some contexts [Bartz et al., 2011; Declerck et al., 2010; Radke and de Bruijn, 2012]. Furthermore, OXT has been shown to facilitate the initial sensation of psychosocial stress and to enhance stress-related responses in the anterior cingulate cortex (ACC) and precuneus [Eckstein et al., 2014]. Moreover, OXT promotes in-group favoritism and in some contexts aggression toward out-groups [De Dreu et al., 2010, 2011] which can also be interpreted in terms of increased self-interest.

In the current study, we have, therefore, first investigated whether OXT selectively enhances self- as opposed to other-referential emotional responses and impacts neural activation in the insula and/or cortical midline structures in the DMN involved in self-processing. Specifically, in a functional MRI (fMRI) experiment (Exp. 1), we measured behavioral and neural responses during performance of a self- versus other-referential paradigm using neutral and emotional expression face stimuli including both positive (happiness) and negative (disgust) valences. Disgust rather than fear-expressions were chosen because the insula is particularly involved in both discriminating and feeling this emotion [Wicker et al., 2003]. Additionally, not only is disgust a distinctive [Chapman and Anderson, 2012] and primitive emotion aiding self-preservation through avoidance or vomiting of potentially toxic substances, but also, importantly, it can be expressed in the context of moral judgments of behaviors performed either by ourselves or by others [Chapman et al., 2009; Eskine et al., 2011].

A plethora of findings from lesion, fMRI, and mood-induction studies support current concepts that emotional states play a key role in moral decision making [Greene et al., 2001; Koenigs et al., 2007; Moll et al., 2005]. Two recent reports have also suggested an involvement of OXT in moral decision-making with one finding that a specific receptor polymorphism is associated with increased attribution of culpability for inflicting accidental harm [Walter et al., 2012] and the other showing that intranasal treatment makes subjects less likely to sacrifice individual in-group members to save larger numbers of out-group ones using a few classical moral dilemma scenarios [De Dreu et al., 2011]. However, neither of these studies addressed the question of whether OXT influences moral decision making only in social contexts, which specifically involve self-benefit. We, therefore, performed a second behavioral experiment (Exp. 2) investigating OXT-induced changes in moral decision making using a large number of diverse moral dilemmas involving either self- or non-self-benefit scenarios or where there was no moral component.

Our hypothesis was that if OXT selectively enhances self-protection and self-interest then it should bias behavioral responses toward endorsement of moral dilemmas involving self-benefit and facilitate neural responses in cortical midline regions and the insula during self-referential processing. Conversely, if OXT is more generally involved in promoting generosity and altruism toward others in social contexts, we hypothesized that it would decrease behavioral endorsement of self-benefit scenarios in moral dilemmas, and either increase neural responses only during other-referential processing or perhaps produce an equivalent enhancement of both self- and other-referential processing.

METHODS

Subjects

We studied 23 healthy men (mean age ± SD: 25.57 ± 3.29 years) in Exp. 1, and 74 healthy men (24.76 ± 2.80 years) and 60 healthy women (23.48 ± 2.83 years) in Exp. 2. All subjects volunteered after giving written, informed consent. All subjects in Exps. 1 and 2 were free of current and past

TABLE I. Experiment 1: demographics and neuropsychological performance

	Mean (± SD)
Age (years)	25.57 (3.29)
Education (years)	17.74 (2.22)
RAVLT[a]	
Trial 1–5[b]	61.1 (7.85)
Trial 6 retention[c]	13.14 (2.03)
Trial 7 delayed recall[d]	13.32 (1.78)
LPS-4[e]	31.18 (3.84)
MWT-B[f]	31.36 (2.80)
TMT-A[g]	24.08 (6.60)
TMT-B[g]	59.60 (13.52)
Digit-span, forward	8.41 (1.40)
Digit-span, backwards	7.95 (2.08)
Trait anxiety[h]	32.50 (6.09)
BDI[i]	2.50 (2.78)

Verbal declarative memory performance was assessed using a German adaption of the
[a]RAVLT (Rey Auditory Verbal Learning Test) and included
[b]learning performance across five trials (maximum possible score 75)
[c]susceptibility to interference (maximum possible score 15), and
[d]delayed recall (maximum possible score 15). Nonverbal reasoning IQ was assessed by the
[e]LPS (Leistungsprüfsystem) subtest 4 (maximum possible score 40). Verbal IQ based on lexical decisions was assessed by the
[f]MWT-B (Mehrfachwahl–Wortschatz–Intelligenz–Test, Teil B) (maximum possible score 37), visual attention and task-switching was assessed using the
[g]TMT-A and TMT-B (Trail-making test A, B) (results displayed in seconds), and working memory performance was assessed using the digit-span forward and backward test (maximum possible score 14). Anxiety symptoms were assessed by the
[h]State Trait Anxiety Inventory and depressive symptoms by the self-report
[i]BDI (Beck Depression Inventory, Version II). See also Supporting Information Table S1.

physical or psychiatric illness, as assessed by medical history and a Structured Clinical Interview for DSM-IV axis I (SCID-I) and axis II disorders (SCID-II). A comprehensive neuropsychological test battery showed that all subjects were within a normal range of cognitive performance and that there were no significant differences between the OXT and PLC groups in Exp. 2 (cf. Tables I and II). The study was approved by the Institutional Review Board (IRB) of the Medical Faculty of the University of Bonn and was performed in compliance with the latest revision of the Declaration of Helsinki. A detailed synopsis of all experimental procedures is provided in the Supporting Information (SI).

Drug Application

Exp. 1 was a counter-balanced, double-blind, within-subject design, and we repeatedly scanned our subjects 30 min after they received either intranasal OXT (24IU; Syntocinon-Spray, Novartis; three puffs per nostril, each with 4IU OXT) or PLC (sodium chloride solution) treatment. The intranasal administration of OXT has been shown to increase central OXT concentration in the cerebrospinal fluid [Striepens et al., 2013]. In Exp. 2, we used a between-subject design and the behavioral tests started 30 min after the administration of OXT or PLC.

Experiment 1

For the fMRI task, two picture sets each with 30 disgusted, 30 neutral, and 30 happy faces, were taken from the validated "Karolinska Directed Emotional Face" database [Lundqvist et al., 1998]. The photographs were shown for 3 s in a quasi-randomized order (no more than two emotional stimuli were shown in a row). The subjects were instructed to rate the intensity of their own arousal and the arousal intensity of the depicted person while they viewed the picture using an 8-point scale [1 (minimum) to 8

TABLE II. Experiment 2: demographics, neuropsychological performance and personality traits

	OXT group Mean (±SD)	PLC group Mean (±SD)	*t*	*P*
Males				
Age (years)	24.35 (2.97)	25.16 (2.58)	−1.25	0.22
Education (years)	16.61 (2.58)	17.11 (1.71)	−0.98	0.33
MWT-A[a]	30.33 (2.85)	31.33 (2.85)	−1.45	0.15
TMT-A[b]	25.23 (7.11)	25.25 (7.66)	−0.02	0.99
TMT-B[b]	62.88 (18.98)	60.90 (19.77)	0.38	0.71
BDI[c]	4.73 (4.28)	3.35 (3.98)	1.44	0.16
Trait anxiety[d]	36.32 (7.88)	34.46 (8.83)	0.96	0.34
EPQ realism[e]	57.68 (13.53)	52.89 (12.69)	1.57	0.12
EPQ idealism[e]	63.54 (11.03)	62.35 (10.25)	0.48	0.63
Females				
Age (years)	23.23 (2.91)	23.73 (2.77)	−0.68	0.50
Education (years)	16.41 (2.09)	16.25 (4.19)	0.19	0.85
MWT-A[a]	30.18 (2.60)	31.38 (2.73)	1.70	0.09
TMT-A[b]	23.83 (8.81)	22.29 (8.03)	0.68	0.50
TMT-B[b]	58.84 (19.80)	58.89 (17.53)	−0.01	0.99
BDI[c]	2.58 (2.59)	3.81 (4.29)	−1.25	0.22
Trait anxiety[d]	39.17 (6.92)	38.81 (7.39)	0.19	0.85
EPQ realism[e]	53.37 (13.82)	57.70 (11.26)	−1.33	0.19
EPQ idealism[e]	71.00 (7.70)	67.77 (8.77)	1.52	0.14

Verbal IQ based on lexical decisions was assessed by the
[a]MWT-A (Mehrfachwahl-Wortschatz-Intelligenz-Test Teil A) (maximum possible score 37) and visual attention and task-switching was assessed using the
[b]TMT-A and TMT-B (Trail-making test A, B) (results displayed in seconds). Depressive symptoms were assessed by the self-report
[c]BDI (Beck Depression Inventory, Version II), and anxiety symptoms by the
[d]STAI (State Trait Anxiety Inventory). Moral thoughts were measured using Forsyth's
[e]EPQ (Ethics Position Questionnaire).
OXT, oxytocin; PLC, placebo. See also Supporting Information Table S2.

(maximum)] presented for 4 s immediately after the photograph. Between trials, a fixation cross (mean duration: 4 s, range 3–7 s) served as a low-level baseline.

Experiment 2

The moral dilemma scenarios in Exp. 2 were taken from Greene et al. [2001] and consisted of 20 non-moral, 25 personal moral, and 17 impersonal examples. Based on the ratings of our pilot study, the scenarios were subdivided into non-self-benefit and self-benefit dilemmas (cf. Supporting Information Table S3). The scenarios were presented as text on a computer screen and by pressing a button the participants could advance to a question related to the scenario ("Is it appropriate to…?"). All the moral and nonmoral dilemmas scenarios are self-referential in that they ask what "you" would decide to do given a certain set of circumstances. The possible answers were always either "no" or "yes," with the latter indicating endorsement of the proposed action. There was no time limit for reading the scenario or answering the question.

RESULTS

All subjects were within a normal range of cognitive performance and there were no significant differences between the OXT and PLC group in Experiment 2 (all P values > 0.05; cf. Tables I and II).

Experiment 1

Behavioral results

For the behavior during the fMRI task, a repeated measures analysis of variance (ANOVA) was computed with "treatment" (OXT vs. PLC), "perspective" (self vs. other), and "valence" (negative, neutral, and positive) as within-subject factors and the ratings as dependent variable. We found significant main effects of perspective ($F_{(1,22)} = 8.94$, $P < 0.01$, $\eta^2 = 0.29$) and valence ($F_{(2,22)} = 124.81$, $P < 0.01$, $\eta^2 = 0.85$) as well as an interaction of perspective and valence ($F_{(2,22)} = 67.83$, $P < 0.01$, $\eta^2 = 0.76$, cf. Supporting Information Fig. S1). There was neither a main nor interaction effect of treatment (all Ps > 0.21).

fMRI results

At the whole-brain level, there was no non-specific effect of OXT across all conditions. To examine valence- and perspective-dependent effects of OXT, we computed the following contrasts $[\text{disgusted–neutral}]_{OXT} > [\text{disgusted–neutral}]_{PLC}$ and $[\text{happy–neutral}]_{OXT} > [\text{happy–neutral}]_{PLC}$ separately for "self" and "other." Interestingly, we identified a significant cluster extending from the middle to anterior parts of the cingulate cortex and another cluster in the precuneus after OXT treatment specifically for disgusted faces in the "self" condition. An inspection of the extracted parameter estimates for these clusters confirmed that this effect was due to an increased response to the evaluation of disgusted faces after OXT treatment (Fig. 1A,B). In a next step, we compared this disgust-specific OXT effect in the "self" and "other" conditions and found that the precuneus activation was significantly stronger during the self-assessment (MNI x, y, z: -18, -58, 42, $Z = 4.19$, $k = 499$).

A region-of-interest analysis restricted to the insula revealed an OXT effect again only for the evaluation of disgusted faces, but this occurred both in the "self" (Fig. 1C) and "other" assessment conditions (MNI x, y, z: 36, 18, 6, $Z = 2.67$). Again, the extracted parameter estimates demonstrated that this effect was driven by an increased response to the rating of disgust after OXT administration.

Experiment 2

A mixed model ANOVA with the percentage endorsement of the dilemmas as dependent variable, the dilemma category (self-benefit moral, non-self-benefit moral and non-moral) as within-subject factor, as well as treatment (OXT vs. PLC) and sex as between-subject factors yielded a main effect of category ($F_{(1.59,206.31)} = 272.97$, $P < 0.01$, $\eta^2 = 0.68$), an interaction between category and sex ($F_{(1.59,206.31)} = 3.64$, $P = 0.03$, $\eta^2 = 0.03$), and an interaction between treatment, sex and the dilemma type ($F_{(1.59,206.31)} = 9.28$, $P < 0.01$, $\eta^2 = 0.07$). The category effect consisted of a linear trend ($F_{(1,130)} = 215.05$, $P < 0.01$, $\eta^2 = 0.62$), with self-benefit scenarios being least and non-moral dilemmas being most endorsed. The interaction between category and sex was related to a higher endorsement in men, particularly for self-benefit ($t_{(132)} = 2.44$, $P = 0.02$, $d = 0.42$) and non-self-benefit ($t_{(110.46)} = 2.60$, $P = 0.01$, $d = 0.49$) moral dilemmas. There was no main treatment effect ($P = 0.57$), but post hoc independent t-tests revealed that OXT compared to PLC selectively increased approval of self-benefit items in men ($t_{(72)} = 2.45$, $P = 0.02$, $d = 0.58$; Fig. 2A) and decreased the approval in women ($t_{(58)} = -2.49$, $P = 0.02$, $d = 0.65$; Fig. 2B, see also Supporting Information Fig. S2). In an analysis based on Greene's taxonomy of personal and impersonal moral dilemmas, no significant OXT effects emerged (all Ps > 0.69). These results argue against a general OXT effect on utilitarian or deontological morality.

We also computed the reaction time (RT) difference between scenarios in which a behavior was accepted or rejected and found a main effect of scenario type ($F_{(1.78,225.48)} = 20.04$, $P < 0.01$, $\eta^2 = 0.14$), with the RT difference for self-benefit dilemmas being significantly longer compared to non-self-benefit ($t_{(130)} = -3.68$, $P < 0.01$, $d = 0.65$) or non-moral ($t_{(130)} = 5.55$, $P < 0.01$, $d = 0.97$) ones. This RT difference may indicate a conflict between the intuitive and automatic rejection of the described behavior, on the one hand and the deliberate self-benefit-driven

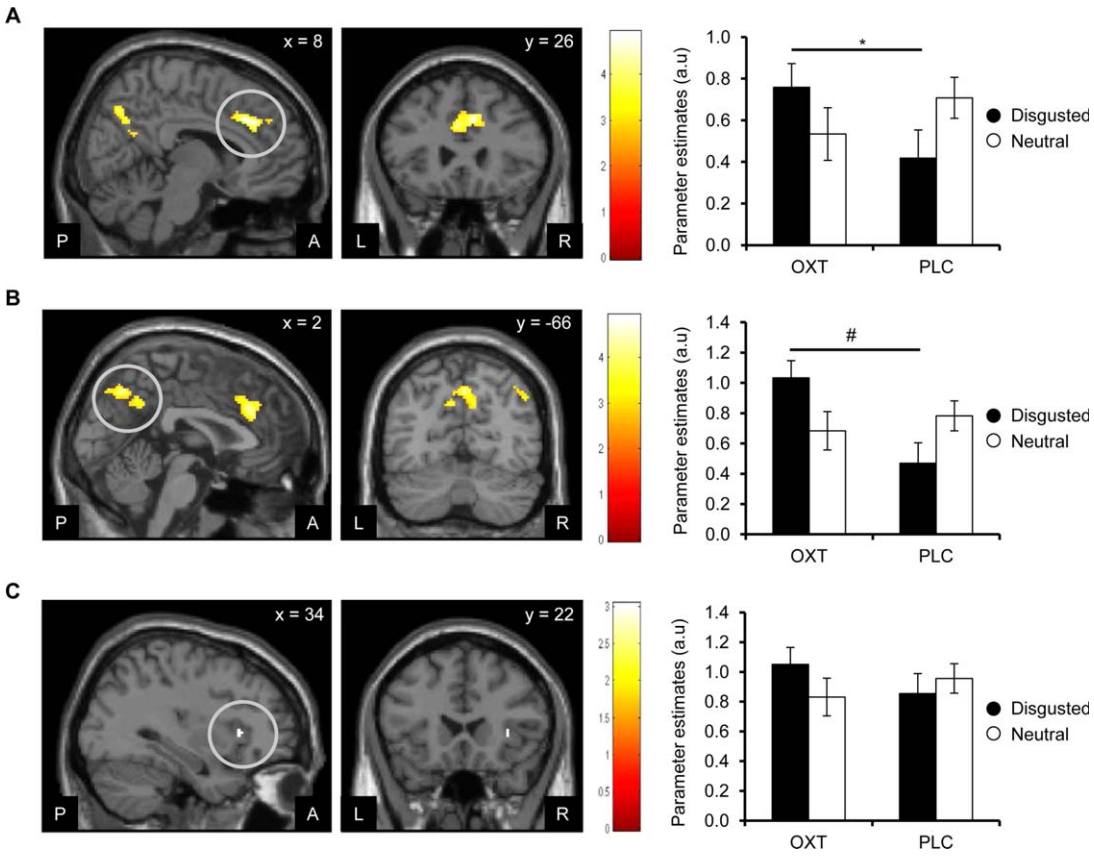

Figure 1.

OXT's influence on self-referential processing of social signals of disgust (Experiment 1; $n = 23$). OXT enhanced neural activity in the anterior and middle cingulate cortex ($OXT_{(Disgusted>Neutral)} > PLC_{(Disgusted>Neutral)}$; display threshold $P < 0.005$ uncorrected; MNI peak coordinates *x*, *y*, *z*: 8, 26, 34, $Z = 4.00$, $k = 648$) (**A**), in the precuneus (MNI peak coordinates *x*, *y*, *z*: 2, −66, 44, $Z = 3.67$, $k = 333$) (**B**), and in the anterior insula (MNI peak coordinates *x*, *y*, *z*: 34, 22, 4, $Z = 2.75$) (**C**). Error bars indicate the standard error of the mean (SEM). See also Supporting Information Figure S1. Abbreviations: A, anterior; L, left hemisphere; OXT, oxytocin; P, posterior; PLC, placebo; R, right hemisphere; $^{*}P < 0.05$; $^{\#}P < 0.10$.

reasoning, on the other. We also observed an interaction of category and treatment ($F_{(1.78,225.48)} = 7.38$, $P < 0.01$, $\eta^2 = 0.06$) and no further interactions (all Ps > 0.79). However, an exploratory analysis revealed that OXT enhanced the RT differences only for female participants ($t_{(42.97)} = 2.69$, $P = 0.02$, $d = 0.82$; Fig. 2D), but not for men ($P = 0.16$; Fig. 2C), suggesting that OXT enhanced the underlying tension between these competing mechanisms only in women.

OXT did not affect cognitive abilities in general as indicated by the absence of any influence of OXT on the RT for reading or responding to the dilemmas (all Ps > 0.14). In view of the fact that non-self-benefit and non-moral dilemmas had both a higher overall endorsement rate and shorter accept versus reject RT compared with self-benefit ones, it was possible that OXT effects in the latter might have only been revealed because in general self-benefit dilemmas were less likely to be endorsed. We, therefore, performed a separate analysis using 10 non-self-benefit and non-moral dilemmas, which yielded the lowest endorsement rates. This still revealed no significant influence of OXT treatment for either the non-self-benefit (mean % endorsed, ♀ OXT = 2.67 ± 5.21, ♀ PLC = 4.67 ± 6.29, $P = 0.19$; ♂ OXT = 6.76 ± 7.84, ♂ PLC = 5.95 ± 7.25, $P = 0.65$) or non-moral dilemma categories (♀ OXT = 25.33 ± 15.25, ♀ PLC = 23.67 ± 13.26, $P = 0.65$; ♂ OXT = 23.24 ± 12.03, ♂ PLC = 26.76 ± 14.15, $P = 0.25$). Thus, we can exclude item difficulty as a possible confound. In both Exps. 1 and 2, OXT had no modulatory influence on state anxiety or positive/negative affect (cf. Supporting Information Tables S1 and S2).

DISCUSSION

In the present study, we aimed at elucidating potential neural and behavioral effects of OXT on self-referential

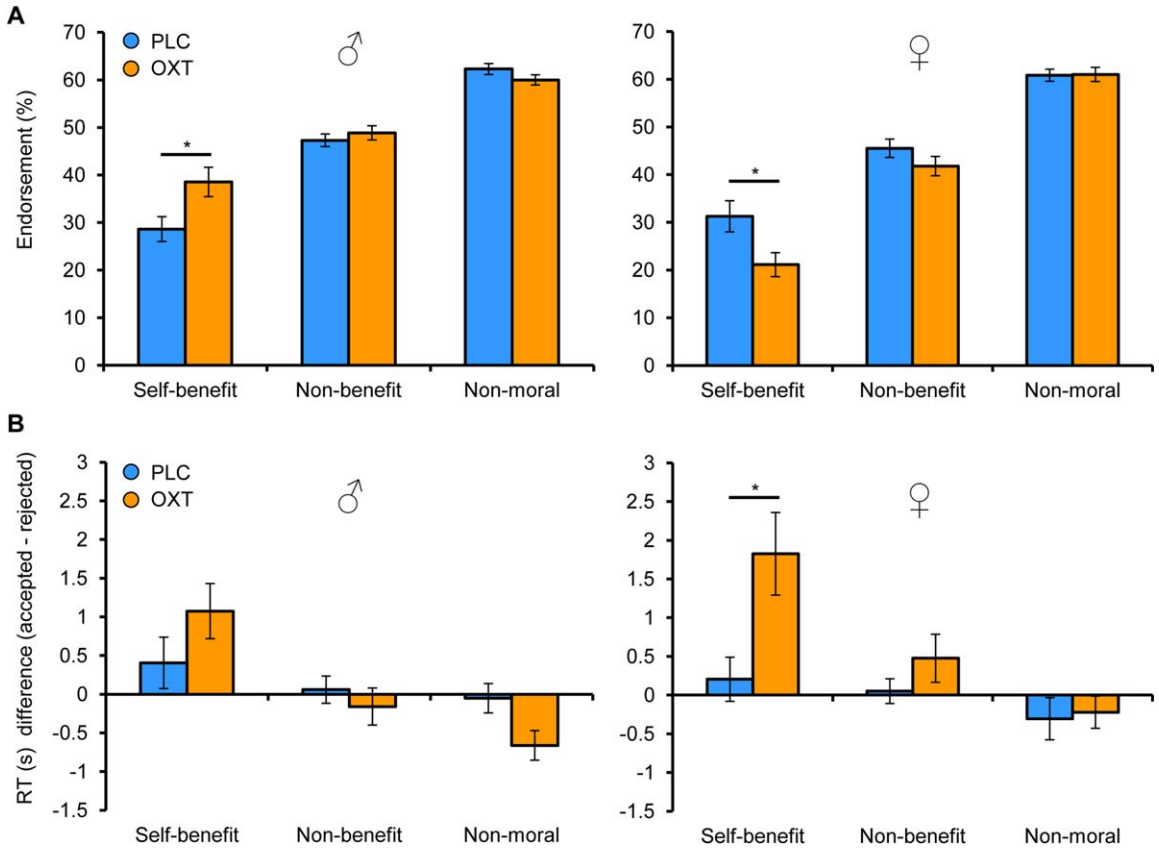

Figure 2.

OXT's effect on moral decision making (Experiment 2). OXT specifically increased the endorsement of self-benefit moral dilemmas in men ($n = 74$) and decreased the endorsement of these dilemmas in women ($n = 60$; (**A**)). Higher difference scores between the RT for the acceptance and rejection of self-benefit moral dilemmas compared to non-benefit or non-moral scenarios indicate that the automatic, default response in the former dilemma type is a rejection. OXT did not significantly alter this RT difference in men, but it increased the difference in women (**B**). Error bars indicate the standard error of the mean (SEM). See also Supporting Information Figure S2 and Table S3. Abbreviations: OXT, oxytocin; PLC, placebo; RT, reaction time; *$P < 0.05$.

processing using both self- and other-evaluation of arousal by positive (happy) and negative (disgust) emotional stimuli and endorsement of moral dilemmas with either a potential self- or non-self-benefit outcome. We report the first evidence that OXT increases neural activity in two cortical midline regions widely implicated in self-processing, that is, the precuneus and ACC, during self-referential processing of disgusted faces. Increased activation of the insula following OXT-treatment occurred during both self- and other-referential evaluation of disgust faces. Behavioral support for an impact of OXT on self-processing was also found since it biased male subjects to more strongly endorse self-benefit outcomes in moral dilemmas but did not influence decision making in ones without a self-benefit outcome or involving nonmoral contexts. By contrast, in female participants OXT directed the behavior toward other-regarding preferences, with self-benefit dilemmas being less endorsed and the RT difference between accepted and rejected self-benefit dilemmas being increased.

On the neural level, OXT facilitated self-referential neural responses in the ACC and precuneus and to both "self" and "other" in the insula, but there was no corresponding impact on arousal. Thus, enhanced self-related activation in the precuneus and ACC and of both self- and other-related activation in the insula are unlikely to reflect a facilitated emotional response but increased attention to identification and self-referencing of social cues most relevant for promoting a self-protection response. The ACC, insula, and precuneus all share functional connections [Cauda et al., 2011] and have been strongly implicated in interoceptive awareness and self-referential processing

[Cabanis et al., 2013; Caseras et al., 2013; Damasio and Carvalho, 2013]. The anterior insula also mediates convergence between bodily states and emotional experience and so OXT may directly affect emotional experience [Zaki et al., 2012]. Interestingly, the precuneus and the ACC exhibit significantly enhanced activation in response to self-intended violations of moral norms [Berthoz et al., 2002]. In our current study, self-benefit moral dilemmas were rarely endorsed under PLC, suggesting that their endorsement would reflect a norm violation. Thus, by inducing a bias in self-benefit moral decision making in men, OXT may be promoting norm violation behavior in the context of self-protection.

On the face of it, the finding that OXT enhances both self-processing and selfish decision making in the context of moral dilemmas in men might seem to suggest an egocentric antisocial role, which is very much at odds with its established importance for promoting maternal and partner bonds [Scheele et al., 2012, 2013; Striepens et al., 2014]. However, it has become clear from both behavioral and fMRI experiments that self-processing extends to incorporate significant others, and this obviously includes partners and children [Aron and Fraley, 1999; Krienen et al., 2010]. Thus, exhibiting potentially costly selfish behavior during moral decision making is not only benefiting self but also can be considered to benefit others with whom an individual has close social bonds. In this way, by increasing selfish behavior OXT is also acting to promote such bonds by increasing protective and supportive behaviors, which will aid survival of partners and close family even at a risk of evoking hostility from other group members, thereby effectively redefining in-group membership. Previous work has shown consistent OXT effects on facilitating in-group favoritism [Van Ijzendoorn and Bakermans-Kranenburg, 2012]. Interestingly a recent study on male rhesus monkeys also reported a facilitatory effect of OXT on self-reward behavior in the context of a choice between rewarding self or other. However, it also increased giving vicarious rewards to other monkeys when self-reward was not an option [Chang et al., 2012].

The opposite effect of OXT in women compared to men is surprising and may indicate sex-specific evolutionary mechanisms by which survival of in-group members is promoted. For women, a strategy appears to be adaptive in which OXT release triggers other-regarding behavior, even if this means forfeiting personal interests. Our finding is consistent with a recent study showing that OXT makes women, but not men, less forgiving of betrayal in the trust game [Yao et al., 2014]. Put differently, OXT appears to augment the punishment of selfish behavior in women. Sex-dimorphic effects of OXT have been reported in several previous studies [Domes et al., 2010; Preckel et al., 2014; Rilling et al., 2014] and might be related to sex differences in human OXT receptor distributions. Likewise, sex-specific effects of OXT may be at least partially mediated by dynamic interactions between OXT and steroid hormones such as testosterone [Weisman et al., 2014]. Distinct OXT effects in women and men may also reside in differential dose–response relationships, which may require a weight-dependent dose titration in future studies. The absence of an OXT effect on selfish decision making in non-moral contexts, in the current study, may simply reflect the fact that self-benefit decision making under these circumstances is associated with a low risk of blame and possible retribution from others.

Currently, there are two influential and only partially compatible models of moral reasoning. The dual-process model [Greene et al., 2004] posits that an emotional system favors a prepotent rejection of "utilitarian" decisions, for instance, pushing an innocent man onto the tracks in front of a runaway trolley to save the life of five other men – a prepotent response that can be overridden by cognitive control. By contrast, the corticolimbic integration model [Moll et al., 2005] argues that competition will occur among cognitive–emotional options. The preserved RT difference for self-benefit dilemmas after OXT administration in men suggests that OXT does not interfere with the generation of emotional responses, which infuse moral judgment, but may rather sharpen self-referential emotion differentiation, thus enabling subjects to behave more consistently with pre-existing ethical preferences. The endorsement rate of self-benefit dilemmas observed for women is further decreased by OXT and a proxy for an internal conflict, the RT difference between accepted and rejected dilemmas, is enhanced.

Overall our findings add support to the idea that archaic neurohumoral systems can influence the sophisticated sphere of human morality, but in men, the direction of this effect points to an increased self-referential focus instead of an enhanced response to moral disgust. These results argue against an OXT-induced non-specific generosity effect [Zak et al., 2007], which has been obtained for decisions in a neuroeconomic task and could not be replicated in an extended version of the game [Radke and de Bruijn, 2012]. Our findings resonate well with a clinical trial [Guastella et al., 2009] showing that OXT as an adjunct to exposure therapy can facilitate a more accurate appraisal of speech performance in patients with social anxiety disorder typically displaying exaggerated negative mental self-representation. Furthermore, intranasal OXT administration has been reported to sharpen self-other perceptual boundaries in a facial morphing procedure [Colonnello et al., 2013] and to intensify self-perception assessed by self-report personality measurements [Cardoso et al., 2012]. Thus, OXT may strengthen emotional granularity, that is, it may enable subjects to represent emotional experiences more precisely and more specifically. In other words, OXT may improve the ability to elicit and/or register somatic markers [Damasio and Carvalho, 2013]. Even the empathogenic actions of OXT [Hurlemann et al., 2010] may be partially mediated by such a neural mechanism, suggesting that increased empathy is a result, or even by-product, of sharpened interoceptive awareness [Ernst et al., 2014].

The specificity of our neural findings for disgusted expressions fits with the idea that the OXT system has evolved not only to maintain and promote attachment-

related functions but also to defend the organism against environmental perils and threats. Recently, faster responses to disgusted faces in an approach/avoidance task after OXT administration have been interpreted as a form of preventive and adaptive behavior [Theodoridou et al., 2013]. Here we show that this evolutionary function may exert its influence also in the culturally more sophisticated domain of moral decision making, even in the absence of an impending physical threat.

One limitation to our study is that all female subjects took oral contraceptives. By recruiting such a homogenous sample, we controlled for possible menstrual cycle-related changes in endogenous OXT levels [Salonia et al., 2005], but we cannot exclude that altered gonadal steroid concentrations may have contributed to our results [Montoya et al., 2013]. Furthermore, there is some evidence indicating that judgments of moral permissibility vary with age [Moran et al., 2012], and it is currently unclear whether sex-specific response patterns remain stable across the lifespan, precluding an extrapolation of our findings to aged populations.

Taken together, we show that in men intranasal administration of OXT may induce a self-referential processing bias, evident not only in the neural response to social signals of disgust but also in the sophisticated sphere of human morality. We argue that increased self-focus evoked by OXT is compatible both with enhanced processing of salient negative or ambiguous emotional cues and also with increased selfishness to promote survival both of self and of an individual's own family group. On the other hand, in women, OXT may act upon enhancing other-regarding behavior to facilitate offspring survival by promoting altruistic decisions and wider social group cooperation, perhaps reflecting the importance of communal care. This gender difference also resonates with genomic imprinting theories. Some imprinted genes have been shown to influence brain OXT expression, and there is some evidence to suggest that paternally imprinted genes promote altruistic behaviors in females, whereas maternally imprinted ones may promote selfishness in males [Tollkuhn et al., 2010; Ubeda and Gardner, 2010]. Our results also have an implication for the use of OXT in clinical trials investigating its potential as an augmenting adjunct to therapies for mental illness, since many psychiatric disorders are characterized by problems with self-processing [Zhao et al., 2013].

FINANCIAL DISCLOSURES

The authors report no competing biomedical financial interests or personal affiliations in connection with the content of this manuscript.

AUTHOR CONTRIBUTIONS

D.S., N.S., K.M.K., and R.H. designed the experiments; D.S., N.S., C.S., J.N., and A.W. conducted the experiments; D.S., N.S., K.M.K., C.S., A.W., and R.H. analyzed the data; D.S., N.S., K.M.K., C.S., T.E.S., W.M., and RH wrote the paper.

REFERENCES

Aron A, Fraley B (1999): Relationship closeness as including other in the self: Cognitive underpinnings and measures. Soc Cogn 17:140–160.

Bartz J, Simeon D, Hamilton H, Kim S, Crystal S, Braun A, Vicens V, Hollander E (2011): Oxytocin can hinder trust and cooperation in borderline personality disorder. Soc Cogn Affect Neurosci 6:556–563.

Berthoz S, Armony JL, Blair RJ, Dolan RJ (2002): An fMRI study of intentional and unintentional (embarrassing) violations of social norms. Brain 125:1696–1708.

Cabanis M, Pyka M, Mehl S, Muller BW, Loos-Jankowiak S, Winterer G, Wolwer W, Musso F, Klingberg S, Rapp AM, Langohr K, Wiedermann G, Herrlich J, Walter H, Wagner M, Schnell K, Vogeley K, Kockler H, Sha NJ, Stöcker T, Thienel R, Pauly K, Krug A, Kircher T (2013): The precuneus and the insula in self-attributional processes. Cogn Affect Behav Neurosci 13:330–345.

Cardoso C, Ellenbogen MA, Linnen AM. (2012): Acute intranasal oxytocin improves positive self-perceptions of personality. Psychopharmacology (Berl) 220:741–749.

Caseras X, Murphy K, Mataix-Cols D, Lopez-Sola M, Soriano-Mas C, Ortriz H, Pujol J, Torrubia R (2013): Anatomical and functional overlap within the insula and anterior cingulate cortex during interoception and phobic symptom provocation. Hum Brain Mapp 34:1220–1229.

Cauda F, D'Agata F, Sacco K, Duca S, Geminiani G, Vercelli A (2011): Functional connectivity of the insula in the resting brain. Neuroimage 55:8–23.

Chang SW, Barter JW, Ebitz RB, Watson KK, Platt ML (2012): Inhaled oxytocin amplifies both vicarious reinforcement and self reinforcement in rhesus macaques (*Macaca mulatta*). Proc Natl Acad Sci USA 109:959–964.

Chapman HA, Anderson AK (2012): Understanding disgust. Ann N Y Acad Sci 1251:62–76.

Chapman HA, Kim DA, Susskind JM, Anderson AK (2009): In bad taste: Evidence for the oral origins of moral disgust. Science 323:1222–1226.

Colonnello V, Chen FS, Panksepp J, Heinrichs M (2013): Oxytocin sharpens self-other perceptual boundary. Psychoneuroendocrinology 38:2996–3002.

Damasio A, Carvalho GB (2013): The nature of feelings: Evolutionary and neurobiological origins. Nat Rev Neurosci 14:143–152.

De Dreu CK, Greer LL, Handgraaf MJ, Shalvi S, Van Kleef GA, Baas M, Ten Velden FS, Van Dijk E, Feith SW (2010): The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. Science 328:1408–1411.

De Dreu CK, Greer LL, Van Kleef GA, Shalvi S, Handgraaf MJ (2011): Oxytocin promotes human ethnocentrism. Proc Natl Acad Sci USA 108:1262–1266.

Declerck CH, Boone C, Kiyonari T (2010): Oxytocin and cooperation under conditions of uncertainty: The modulating role of incentives and social information. Horm Behav 57:368–374.

Domes G, Lischke A, Berger C, Grossmann A, Hauenstein K, Heinrichs M, Herpertz SC (2010): Effects of intranasal oxytocin on emotional face processing in women. Psychoneuroendocrinology 35:83–93.

Eckstein M, Hurlemann R (2013): Oxytocin: Evidence for a therapeutic potential of the social neuromodulator. Nervenarzt 84: 1321–1328.

Eckstein M, Scheele D, Weber K, Stoffel-Wagner B, Maier W, Hurlemann R (2014): Oxytocin facilitates the sensation of social stress. Hum Brain Mapp 35:4741–4750.

Ernst J, Boker H, Hattenschwiler J, Schupbach D, Northoff G, Seifritz E, Grimm S (2014): The association of interoceptive awareness and alexithymia with neurotransmitter concentrations in insula and anterior cingulate. Soc Cogn Affect Neurosci 9:857–863.

Eskine KJ, Kacinik NA, Prinz JJ (2011): A bad taste in the mouth: Gustatory disgust influences moral judgment. Psychol Sci 22: 295–299.

Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD (2001): An fMRI investigation of emotional engagement in moral judgment. Science 293:2105–2108.

Greene JD, Nystrom LE, Engell AD, Darley JM, Cohen JD (2004): The neural bases of cognitive conflict and control in moral judgment. Neuron 44:389–400.

Guastella AJ, Howard AL, Dadds MR, Mitchell P, Carson DS (2009): A randomized controlled trial of intranasal oxytocin as an adjunct to exposure therapy for social anxiety disorder. Psychoneuroendocrinology 34:917–923.

Hurlemann R, Patin A, Onur OA, Cohen MX, Baumgartner T, Metzler S, Dziobek I, Gallinat J, Wagner M, Maier W, Kendrick KM (2010): Oxytocin enhances amygdala-dependent, socially reinforced learning and emotional empathy in humans. J Neurosci 30:4999–5007.

Insel TR (2010): The challenge of translation in social neuroscience: a review of oxytocin, vasopressin, and affiliative behavior. Neuron 65:768–779.

Koenigs M, Young L, Adolphs R, Tranel D, Cushman F, Hauser M, Damasio A (2007): Damage to the prefrontal cortex increases utilitarian moral judgements. Nature 446:908–911.

Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E (2005): Oxytocin increases trust in humans. Nature 435:673–676.

Krienen FM, Tu PC, Buckner RL (2010): Clan mentality: Evidence that the medial prefrontal cortex responds to close others. J Neurosci 30:13906–13915.

Lamm C, Singer T (2010): The role of anterior insular cortex in social emotions. Brain Struct Funct 214:579–591.

Lundqvist D, Flykt A, Ohman A (1998): Karolinska Directed Emotional Faces [Database of standardized facial images].

Moll J, Zahn R, de Oliveira-Souza R, Krueger F, Grafman J (2005): Opinion: the neural basis of human moral cognition. Nat Rev Neurosci 6:799–809.

Montoya ER, Terburg D, Bos PA, Will GJ, Buskens V, Raub W, van Honk J (2013): Testosterone administration modulates moral judgments depending on second-to-fourth digit ratio. Psychoneuroendocrinology 38:1362–1369.

Moran JM, Jolly E, Mitchell JP (2012): Social-cognitive deficits in normal aging. J Neurosci 32:5553–5561.

Olff M, Frijling JL, Kubzansky LD, Bradley B, Ellenbogen MA, Cardoso C, Bartz JA, Yee JR, van Zuiden M (2013): The role of oxytocin in social bonding, stress regulation and mental health: An update on the moderating effects of context and interindividual differences. Psychoneuroendocrinology 38: 1883–1894.

Preckel K, Scheele D, Kendrick KM, Maier W, Hurlemann R (2014): Oxytocin facilitates social approach behavior in women. Front Behav Neurosci 8:191.

Qin P, Northoff G (2011): How is our self related to midline regions and the default-mode network? Neuroimage 57:1221–1233.

Radke S, de Bruijn ER (2012): The other side of the coin: oxytocin decreases the adherence to fairness norms. Front Hum Neurosci 6:193.

Rilling JK, Demarco AC, Hackett PD, Chen X, Gautam P, Stair S, Haroon E, Thompson R, Ditzen B, Patel R and others (2014): Sex differences in the neural and behavioral response to intranasal oxytocin and vasopressin during human social interaction. Psychoneuroendocrinology 39:237–248.

Salonia A, Nappi RE, Pontillo M, Daverio R, Smeraldi A, Briganti A, Fabbri F, Zanni G, Rigatti P, Montorsi F (2005): Menstrual cycle-related changes in plasma oxytocin are relevant to normal sexual function in healthy women. Horm Behav 47:164–169.

Scheele D, Striepens N, Güntürkün O, Deutschlander S, Maier W, Kendrick KM, Hurlemann R (2012): Oxytocin modulates social distance between males and females. J Neurosci 32:16074–16079.

Scheele D, Wille A, Kendrick KM, Stoffel-Wagner B, Becker B, Güntürkün O, Maier W, Hurlemann R (2013): Oxytocin enhances brain reward system responses in men viewing the face of their female partner. Proc Natl Acad Sci USA 110:20308–20313.

Shamay-Tsoory SG, Fischer M, Dvash J, Harari H, Perach-Bloom N, Levkovitz Y (2009): Intranasal administration of oxytocin increases envy and schadenfreude (gloating). Biol Psychiatry 66:864–870.

Striepens N, Kendrick KM, Hanking V, Landgraf R, Wullner U, Maier W, Hurlemann R (2013): Elevated cerebrospinal fluid and blood concentrations of oxytocin following its intranasal administration in humans. Sci Rep 3:3440.

Striepens N, Kendrick KM, Maier W, Hurlemann R (2011): Prosocial effects of oxytocin and clinical evidence for its therapeutic potential. Front Neuroendocrinol 32:426–450.

Striepens N, Scheele D, Kendrick KM, Becker B, Schafer L, Schwalba K, Reul J, Maier W, Hurlemann R (2012): Oxytocin facilitates protective responses to aversive social stimuli in males. Proc Natl Acad Sci USA 109:18144–18149.

Theodoridou A, Penton-Voak IS, Rowe AC (2013): A direct examination of the effect of intranasal administration of oxytocin on approach-avoidance motor responses to emotional stimuli. PLoS One 8:e58113.

Striepens N, Matusch A, Kendrick KM, Mihov Y, Elmenhorst D, Becker B, Lang M, Coenen HH, Maier W, Hurlemann R, Bauer A (2014): Oxytocin enhances attractiveness of unfamiliar female faces independent of the dopamine reward system. Psychoneuroendocrinology 39:74–87.

Tollkuhn J, Xu X, Shah NM (2010): A custody battle for the mind: Evidence for extensive imprinting in the brain. Neuron 67:359–362.

Ubeda F, Gardner A (2010): A model for genomic imprinting in the social brain: juveniles. Evolution 64:2587–2600.

Van Ijzendoorn MH, Bakermans-Kranenburg MJ (2012): A sniff of trust: Meta-analysis of the effects of intranasal oxytocin administration on face recognition, trust to in-group, and trust to out-group. Psychoneuroendocrinology 37:438–443.

Walter NT, Montag C, Markett S, Felten A, Voigt G, Reuter M (2012): Ignorance is no excuse: Moral judgments are influenced by a genetic variation on the oxytocin receptor gene. Brain Cogn 78:268–273.

Weisman O, Zagoory-Sharon O, Feldman R (2014): Oxytocin administration, salivary testosterone, and father-infant social behavior. Prog Neuropsychopharmacol Biol Psychiatry 49:47–52.

Wicker B, Keysers C, Plailly J, Royet JP, Gallese V, Rizzolatti G (2003): Both of us disgusted in My insula: The common neural basis of seeing and feeling disgust. Neuron 40: 655–664.

Yao S, Zhao W, Cheng R, Geng Y, Luo L, Kendrick KM (2014): Oxytocin makes females, but not males, less forgiving following betrayal of trust. Int J Neuropsychopharmacol doi:10.1017/S146114571400090X.

Zak PJ, Stanton AA, Ahmadi S (2007): Oxytocin increases generosity in humans. PLoS One 2:e1128.

Zaki J, Davis JI, Ochsner KN (2012): Overlapping activity in anterior insula during interoception and emotional experience. Neuroimage 62:493–499.

Zhao W, Luo L, Li Q, Kendrick KM (2013): What can psychiatric disorders tell us about neural processing of the self? Front Hum Neurosci 7:485.