

# Effects of Oxytocin and Vasopressin on Preferential Brain Responses to Negative Social Feedback

Marta Gozzi<sup>1</sup>, Erica M Dashow<sup>1</sup>, Audrey Thurm<sup>1</sup>, Susan E Swedo<sup>1</sup> and Caroline F Zink<sup>\*,2,3</sup>

<sup>1</sup>Pediatrics & Developmental Neuroscience Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA; <sup>2</sup>Genes, Cognition, and Psychosis Program, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA; <sup>3</sup>Clinical Sciences, Lieber Institute for Brain Development, Johns Hopkins Medicine, Baltimore, MD, USA

Receiving negative social feedback can be detrimental to emotional, cognitive, and physical well-being, and fear of negative social feedback is a prominent feature of mental illnesses that involve social anxiety. A large body of evidence has implicated the neuropeptides oxytocin and vasopressin in the modulation of human neural activity underlying social cognition, including negative emotion processing; however, the influence of oxytocin and vasopressin on neural activity elicited during negative social evaluation remains unknown. Here 21 healthy men underwent functional magnetic resonance imaging in a double-blind, placebo-controlled, crossover design to determine how intranasally administered oxytocin and vasopressin modulated neural activity when receiving negative feedback on task performance from a study investigator. We found that under placebo, a preferential response to negative social feedback compared with positive social feedback was evoked in brain regions putatively involved in theory of mind (temporoparietal junction), pain processing (anterior insula and supplementary motor area), and identification of emotionally important visual cues in social perception (right fusiform). These activations weakened with oxytocin and vasopressin administration such that neural responses to receiving negative social feedback were not significantly greater than positive social feedback. Our results show effects of both oxytocin and vasopressin on the brain network involved in negative social feedback, informing the possible use of a pharmacological approach targeting these regions in multiple disorders with impairments in social information processing.

*Neuropsychopharmacology* (2017) **42**, 1409–1419; doi:10.1038/npp.2016.248; published online 30 November 2016

## INTRODUCTION

Receiving negative social evaluation such as negative performance feedback, particularly by a superior, can conjure negative emotions and elicit social anxiety. Indeed, negative social feedback can affect mood, self-esteem, behavior, and physiology (Akinola and Mendes, 2008; Dickerson and Kemeny, 2004b; Mendes *et al.*, 2008), and even lead to negative effects on health in case of repeated or prolonged experience (Dickerson *et al.*, 2004a). Furthermore, the impact and power of negative feedback are often more potent than positive feedback (Baumeister *et al.*, 2001). Clinically, fear of negative evaluation is a prominent feature of social anxiety displayed in social anxiety disorder (SAD; Rapee and Heimberg, 1997), which may be comorbid with other disorders such as autism spectrum disorder (ASD; Meyer *et al.*, 2006) and depression (O'Connor *et al.*, 2002; Wang *et al.*, 2012). It has also been associated with delusional ideation in schizophrenia (Kinoshita *et al.*, 2011).

Oxytocin and vasopressin—two structurally similar neuropeptides that are found in the human and mammalian brain—have both been considered potential treatment targets for social impairments and anxiety in clinical populations such as SAD and ASD due to their suggested role in social cognitive functions (Heinrichs *et al.*, 2009). Specifically, one proposed role of oxytocin and vasopressin is in the modulation of human behavior and neural activity related to negative socioemotional processing. Most consistently, an oxytocin-induced reduction of amygdala responses to negative-valenced social stimuli (eg, faces with negative affect) has been demonstrated in men (Domes *et al.*, 2007; Gamer *et al.*, 2010; Kirsch *et al.*, 2005; Petrovic *et al.*, 2008; Striepens *et al.*, 2012). Depending on the experimental task, however, oxytocin also influences neural activity in other brain regions, including temporoparietal junction (TPJ) (Domes *et al.*, 2007), ventral prefrontal cortex (PFC), fusiform (Petrovic *et al.*, 2008), and insula (Striepens *et al.*, 2012), suggesting that oxytocin's regional neural influence is context-specific. Such oxytocin-induced modulations of neural activity may underlie oxytocin's anxiolytic and prosocial effects in humans (Evans *et al.*, 2010; Heinrichs *et al.*, 2003; Kis *et al.*, 2013; Kosfeld *et al.*, 2005; Norman *et al.*, 2011; Petrovic *et al.*, 2008; Theodoridou *et al.*, 2009). Human vasopressin studies are less abundant; however, several investigations support opposing roles of oxytocin and

\*Correspondence: Dr C Zink, Clinical Sciences, Lieber Institute for Brain Development, Johns Hopkins Medicine, 855 N Wolfe Street, Suite 300, Baltimore, MD 21205, USA, Tel: 410 955 0815, Fax: 410 955 1044, E-mail: caroline.zink@libd.org

Received 10 May 2016; revised 28 September 2016; accepted 29 September 2016; accepted article preview online 31 October 2016

vasopressin, with the latter being associated with aggression and anxiogenic behaviors (Ebstein *et al*, 2009; Thompson *et al*, 2004). Accordingly, in men, vasopressin, as opposed to oxytocin, has been shown to increase amygdala (Brunnlieb *et al*, 2013) and ventral PFC (Zink *et al*, 2010) reactivity to negative socioemotional stimuli relative to placebo. On the other hand, studies have also revealed similar human neural influences of oxytocin and vasopressin. For example, both oxytocin (Domes *et al*, 2007) and vasopressin (Zink *et al*, 2011) decrease signal in the TPJ during implicit face processing, and both neuropeptides increase ventral PFC activity when cooperation is not reciprocated in an economic game (Rilling *et al*, 2012). Together, these previous investigations of the neural influence of oxytocin and vasopressin in humans suggest that the brain regions affected are at least in part context-dependent, and context and/or affected brain regions have a role in determining whether oxytocin and vasopressin have opposing or similar neural influences.

To date, most of these studies have been performed using non-evaluative and non-self-threatening contexts (Gamer *et al*, 2010; Kirsch *et al*, 2005; Zink *et al*, 2011; Zink *et al*, 2010). The influence of oxytocin and vasopressin on neural activity elicited during negative performance evaluation, a relatively more complex and self-relevant socioemotional context, remains unknown. Therefore, the goal of the current study was to determine whether oxytocin and vasopressin influence preferential neural responses to negative social feedback and if so, in opposing or similar directionality. Because of the potential clinical implications and the consequences of negative social feedback on well-being, it is important to determine whether the influence of oxytocin and vasopressin on neural activity during the processing of negative socioemotional stimuli demonstrated in previous studies extends to an influence on neural activity during negative performance evaluation. A better understanding of the role of these neuropeptides in modulating typical processing of negative social feedback may also provide valuable insights for treating social anxiety in clinical populations.

We investigated patterns of brain activity preferentially related to negative social feedback in our paradigm under placebo, oxytocin, and vasopressin. Under placebo, we expected that exposure to negative social feedback compared with positive social feedback would elicit greater activation of brain areas involved in emotional pain (Craggs *et al*, 2007), social rejection (Eisenberger *et al*, 2003), and theory of mind or judgment (Saxe and Kanwisher, 2003; Shamay-Tsoory, 2011). Based on the role of oxytocin and vasopressin in modulating social cognitive processes, including negative emotions, we expected both neuropeptides to influence neural activity associated with negative social feedback.

## MATERIALS AND METHODS

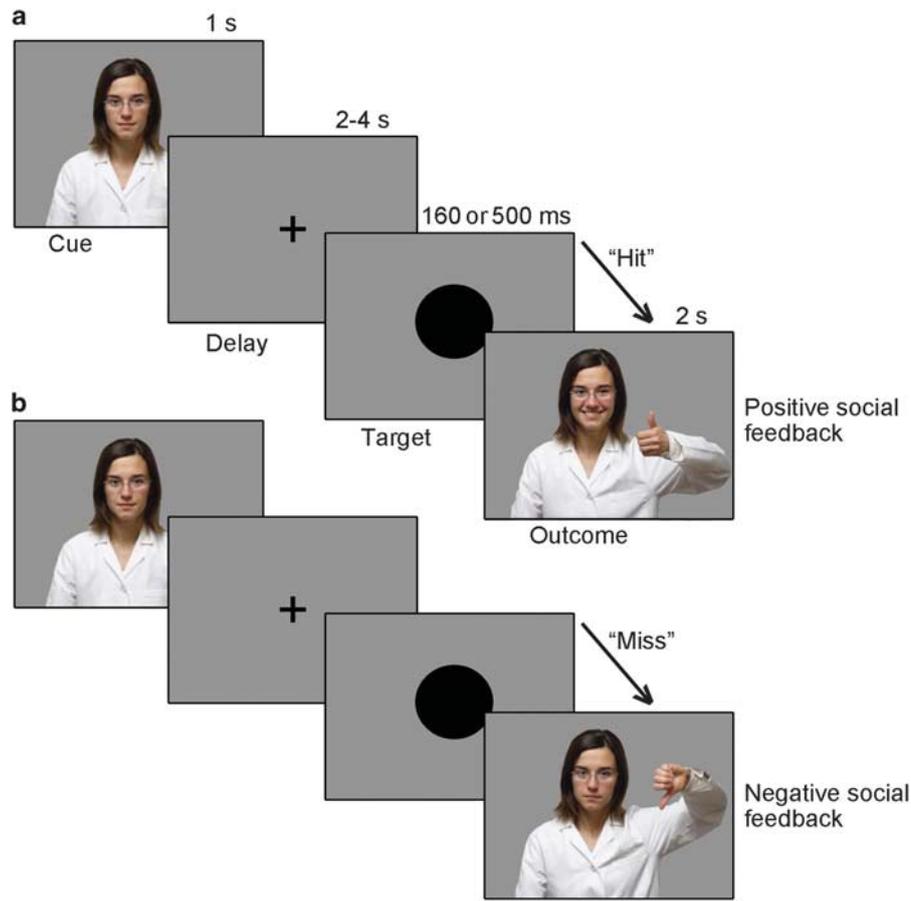
### Task, Participants, and Procedure

A total of 25 healthy men (mean age 26.04 years, range 20–39 years; all right-handed except for one) participated in this double-blind, placebo-controlled crossover study. Data from four participants were excluded because of missing visits ( $N=3$ ) or technical problems ( $N=1$ ), leaving a sample size

of  $N=21$  participants (all males; mean age 26.57 years, range 21–39 years). Screening procedures and inclusion/exclusion criteria are provided in the Supplementary Material and Methods. All participants gave their informed consent and were paid for their participation. The work was approved by the Institutional Review Board at the NIH.

All eligible individuals completed three scanning sessions, separated by at least 1 week. The procedure was the same for each session, whereas the drug administered differed. At each visit, participants self-administered a nasal spray containing oxytocin (24 IU total), vasopressin (40 IU total), or placebo under the supervision of a physician or nurse. These doses are similar to the doses that have been used in the previous studies (Evans *et al*, 2010; Zink *et al*, 2011). Intranasal medication was administered through two puffs in each nostril (four puffs total), alternating nostrils between each puff and waiting 40 s between each puff (ie, right nostril one puff, wait for 40 s, left nostril one puff, wait 40 s, right nostril one puff, wait 40 s, and left nostril one puff). Drug order was counterbalanced across participants and drug assignment was randomized. The NIH Pharmacy Department prepared the medications and maintained the blind. Participants were asked to refrain from smoking cigarettes or drinking alcohol and caffeine for 12 h before scanning. Blood pressure and heart rate were measured before drug administration and after scanning.

Scanning began 45 min after drug administration to capture peak levels of the neuropeptides in the cerebrospinal fluid (Born *et al*, 2002), and in accordance with previous functional magnetic resonance imaging (fMRI) studies using these neuropeptides (Labuschagne *et al*, 2010; Zink *et al*, 2011). Using the software Presentation (Neurobehavioral Systems, Inc., San Francisco, CA), the current task was administered as part of a neuroimaging task battery. After drug administration, participants received task instructions and practice trials before scanning. The task design (Figure 1) was as follows. Participants were instructed to look straight ahead at the fixation cross on the screen and to press a button, using their right thumb, as quickly as possible when a black circle (the target) appeared in the center of the screen. The amount of time that the black circle remained on the screen was either 160 or 500 ms, varied across trials. Unknown to the participants, the task was designed so that in 33% of the trials, the black circle disappeared very quickly (after 160 ms), making it highly unlikely for participants to respond in time, whereas in 67% of the trials (ie, 40 trials in each run), the black circle remained on the screen for a longer time (500 ms), making it likely for participants to respond in time, confirmed behaviorally. If participants responded while the black circle was still on the screen, the investigator gave them a thumbs up (positive social feedback); if participants did not respond quickly enough (ie, after the black circle disappeared), the investigator gave them a thumbs down (negative social feedback). Participants were informed that if they pressed too quickly (ie, before the appearance of the circle on the screen) or did not respond at all, they would also automatically receive negative feedback. Feedback was given in a dynamic manner, rather than being displayed as a static picture, by presenting a video of the investigator moving from a neutral position to either thumbs up or down to add ecological validity. Importantly, when interacting with participants before and after each scanning



**Figure 1** Task design for social trials in which (a) the target is 'hit' (button press before target disappears) and positive social feedback is provided (ie, video of investigator giving thumbs up), and (b) the target is 'missed' (button press after target disappears) and negative social feedback is provided (ie, video of investigator giving thumbs down). The task was designed so that participants should get positive feedback in  $\sim 2/3$  of the trials (target duration up to 500 ms) and negative feedback in  $\sim 1/3$  of the trials (target duration up to 160 ms). The cue was presented at the beginning of each trial for 1 s. The target was presented for either 160 or 500 ms. Feedback was presented for 2.0 s.

session, the investigator's look (hairstyle, glasses, and lab uniform) and verbal exchange was fixed across participants. Participants were specifically informed that the investigator would be watching, and keeping track of, their performance from a computer in the control room. They were also told that based on these observations, after completion of the task, participants would receive a certificate from the investigator (representing a gold, silver, or bronze medal) depending on how well they did. An actual certificate was given to each subject at the end of each scanning session based on their true performance on the task.

Of note, the experimental paradigm also included monetary trials (variation of the standard Monetary Incentive Delay Task (Knutson *et al*, 2001)) in which money, rather than the investigator, was used to provide feedback on task performance; the data from these monetary trials are not central to the research focus here regarding negative social feedback, however, findings from equivalent analyses performed on the monetary trials are presented in the Supplementary Information for completeness and comparison purposes. Events in the monetary trials are modeled in the current analysis as events of non-interest to statistically remove their influence on the current findings (see imaging analysis section below).

The timing was as follows: (a) the cue at the beginning of each trial to inform participants of the trial type (money or social) was presented for 1.0 s; (b) the fixation cross was displayed for a mean of 5.0 s, jittered between 4.0 and 6.0 s; (c) the black circle was presented for either 160 or 500 ms (see above) and participants were asked to respond as quickly as possible; (d) feedback was presented for 2.0 s. The duration of the task was  $\sim 20$  min for each participant, with 60 social trials divided into two runs.

Before drug administration and after scanning, participants completed questionnaires investigating an acute effect of oxytocin and vasopressin on mood, anger, and anxiety, including the Self-Assessment Manikin (SAM; Bradley and Lang, 1994); the Positive and Negative Affect Schedule (PANAS; Watson *et al*, 1988); the state portion of the State-Trait Anxiety Inventory (Spielberger *et al*, 1983), and the state portion of the State-Trait Anger Expression Inventory (Spielberger, 1988).

### Imaging Acquisition and Analysis

The study was conducted on a 3T GE Signa scanner equipped with an eight-channel receiver head coil. Head motion during scanning was minimized by placing foam

pads around participants' heads. The stimuli were projected onto a screen that participants viewed through a mirror mounted on the head coil of the scanner.

In each run, 340 functional volumes were acquired for each participant to measure the T2\*-weighted blood oxygenation level-dependent effect (gradient-recall echo planar imaging; TR = 2000 ms; TE = 30 ms; flip angle = 90°; FOV = 240 mm; 32 slices; slice thickness 3.5 mm without gap; matrix size = 64 × 64; voxel size = 3.75 × 3.75 × 3.5 mm). Five additional scans were acquired at the beginning of each run to allow for steady-state magnetization and discarded from subsequent analyses. The fMRI data were preprocessed and analyzed using Statistical Parametric Mapping (SPM8; <http://www.fil.ion.ucl.ac.uk/spm>) implemented in MATLAB 7.10.0 (Mathworks Inc., Natick, MA). Images were slice-time corrected, realigned, and unwarped, normalized to the EPI template in MNI space, and spatially smoothed using a 6.0 mm full width at half maximum Gaussian kernel.

Statistical analyses were performed following a two-level procedure. At the first level, a separate general linear model was defined for each participant for each session, including the following regressors time-locked to event onset: social cue, social hit outcome (positive social feedback), and social miss outcome (negative social feedback). Regressors of non-interest were also included for the equivalent events from the monetary trials (irrelevant to the current investigation) to account for their effect. The influence of oxytocin and vasopressin on receipt of negative monetary feedback (ie, not earning money) compared with positive monetary feedback (ie, receiving money) is presented in the Supplementary Information; Supplementary Tables S2 and S3. We separately modeled errors in which participants responded too soon (ie, on the cue or fixation), more than once, or not at all in a given trial. The event-related design was modeled using a canonical hemodynamic response function. The data were high-pass filtered (128 s cut-off) and serial correlations were accounted for by an autoregressive model of the first order. Because we were specifically interested in preferential neural responses to negative social feedback, for each participant under each drug condition, we calculated contrast images for social miss outcome ('negative social feedback') and social hit outcome ('positive social feedback') events vs baseline. We then employed a full-factorial design to compare preferential neural responses to negative social feedback (negative social feedback > positive social feedback) within and between drug conditions. Specifically, the individual contrast images from the placebo, oxytocin, and vasopressin sessions were entered into a second-level whole-brain random effects analysis to determine the brain regions involved with the preferential processing of negative social feedback in each drug condition separately, as well as the

interactions to compare across drug conditions (placebo vs oxytocin, placebo vs vasopressin, and oxytocin vs vasopressin). Correction for multiple comparisons was applied to all resultant maps using a voxel-level threshold of  $p < 0.001$  and a cluster level threshold of  $p < 0.05$ , family-wise error corrected. Beta values for negative social feedback and positive social feedback were extracted from resultant peak voxels and subjected to paired *t*-tests using SPSS software to further assess the influence and directionality of oxytocin and vasopressin on preferential neural responses to negative social feedback determined under placebo. Significance thresholds for these statistical analyses were set to  $p < 0.05$ , corrected for multiple comparisons.

## RESULTS

### Behavioral Results

The data on participants' task performance during each drug condition are provided in Table 1. As expected by virtue of task design, the hit rate was ~60% in all sessions. A repeated-measures one-way ANOVA revealed no significant main effect of drug on either hit rate ( $F(2,40) = 0.766$ ,  $p = 0.472$ ) or reaction time ( $F(2,40) = 0.356$ ,  $p = 0.703$ ). The data on participants' emotional states rating scales collected at each visit pre- and post-scanning are reported in Supplementary Table S1. Repeated-measures ANOVAs with within-factors 'drug' (placebo, oxytocin, and vasopressin) and 'time point' (pre-scan and post-scan) revealed a significant main effect of time point on experienced pleasure (measured with SAM scale,  $F(1,19) = 8.21$ ,  $p = 0.010$ ), arousal (measured with SAM scale,  $F(1,19) = 27.11$ ,  $p < 0.001$ ), and positive affect (measured with PANAS scale  $F(1,18) = 26.19$ ,  $p < 0.001$ ). *Post hoc* paired *t*-tests showed that experienced pleasure, arousal, and positive affect were all significantly lower post-scan compared with pre-scan (pleasure:  $t(59) = 3.11$ ,  $p = 0.003$ ; arousal:  $t(59) = 5.07$ ,  $p < 0.001$ ; positive affect:  $t(56) = -5.86$ ,  $p < 0.001$ ). No significant main effect of drug, time point, or interaction was found for any other emotional measures (eg, dominance, negative affect, anger, and anxiety levels),  $p > 0.05$ .

### Neuroimaging Results

The fMRI results and corresponding statistics from the contrast negative social feedback > positive social feedback under placebo, oxytocin, and vasopressin are reported in Table 2. The whole-brain analysis under placebo revealed significant activations corresponding to a negative social feedback bias in the right anterior insula, right supplementary motor area (SMA)/pre-SMA, right fusiform/inferior temporal gyrus (ITG), and bilateral TPJ/posterior superior temporal sulcus (pSTS) (Figure 2). Preferential neural responses to negative social feedback were weakened by both oxytocin and vasopressin; under both oxytocin and vasopressin there were no significant activations for the contrast, negative social feedback > positive social feedback. An additional analysis of peak betas from placebo confirmed a highly significant ( $p < 0.001$ ) preferential response to negative compared with positive social feedback under placebo, whereas the difference between negative and positive feedback beta values was not significant under

**Table 1** Task-Performance Data for Social Trials

	Placebo	Oxytocin	Vasopressin
Hit rate (% hit responses)	58.02 (5.28)	56.59 (8.04)	58.33 (6.19)
Reaction time (ms)	276.04 (41.76)	276.55 (30.49)	280.91 (29.80)

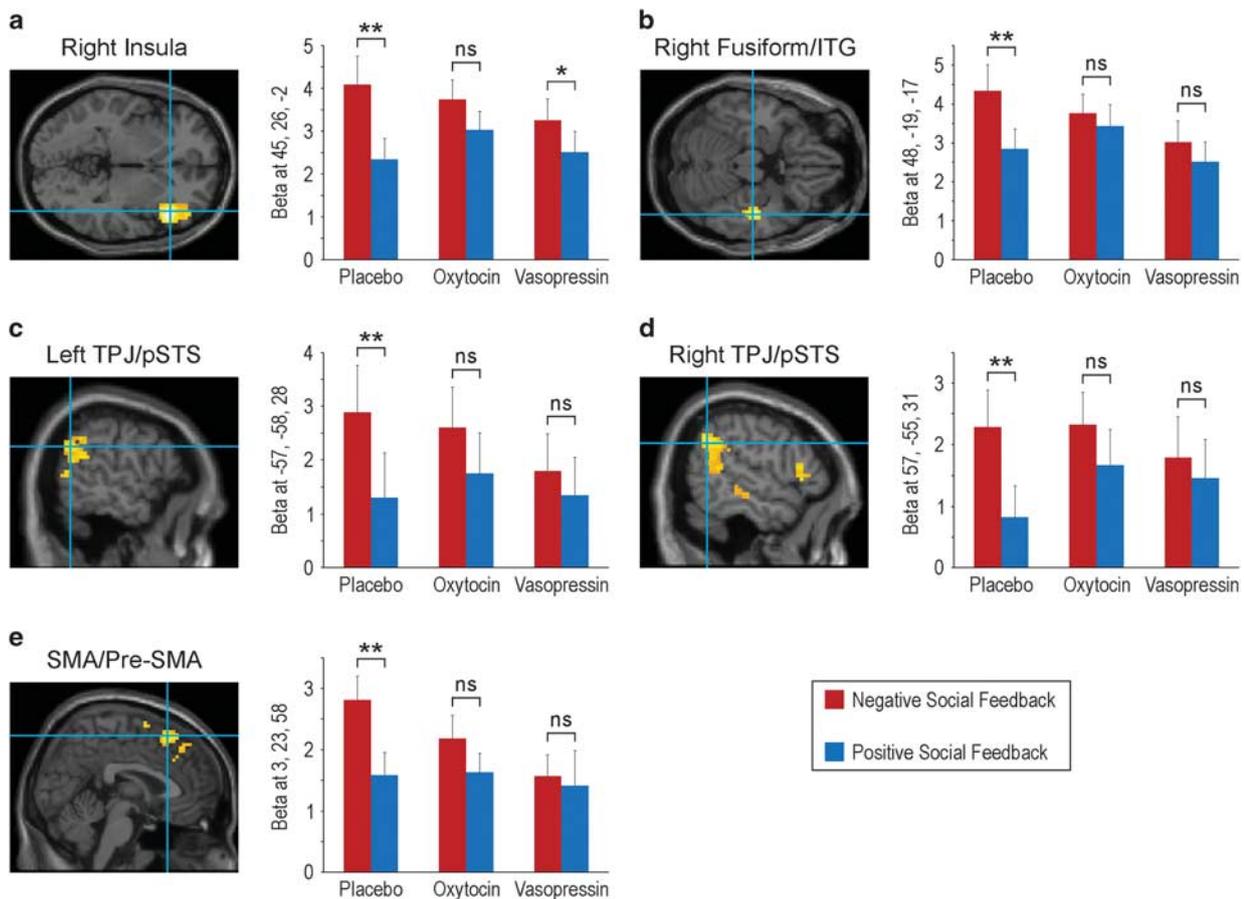
Mean data are given (SD in parenthesis).

$N = 21$ .

**Table 2** Significant ( $p < 0.05$ , corrected) Preferential Neural Responses to Negative Social Feedback (Negative Social Feedback > Positive Social Feedback) Under Each Drug Condition

Brain region	MNI coordinates X, Y, Z	Cluster size	Peak T-value	Cluster corrected p-value
<i>Placebo</i>				
Right anterior insula	45, 26, -2	303	6.11	<0.001
SMA incl.	9, 5, 70	211	5.21	0.001
Pre-SMA	3, 23, 58		4.60	
Right fusiform/ITG	48, -19, -17	102	5.86	0.028
Right TPJ/pSTS	57, -55, 31	186	5.02	0.002
Right TPJ/pSTS	-57, -58, 28	133	4.37	0.011
<i>Oxytocin</i>				
No significant activation				
<i>Vasopressin</i>				
No significant activation				

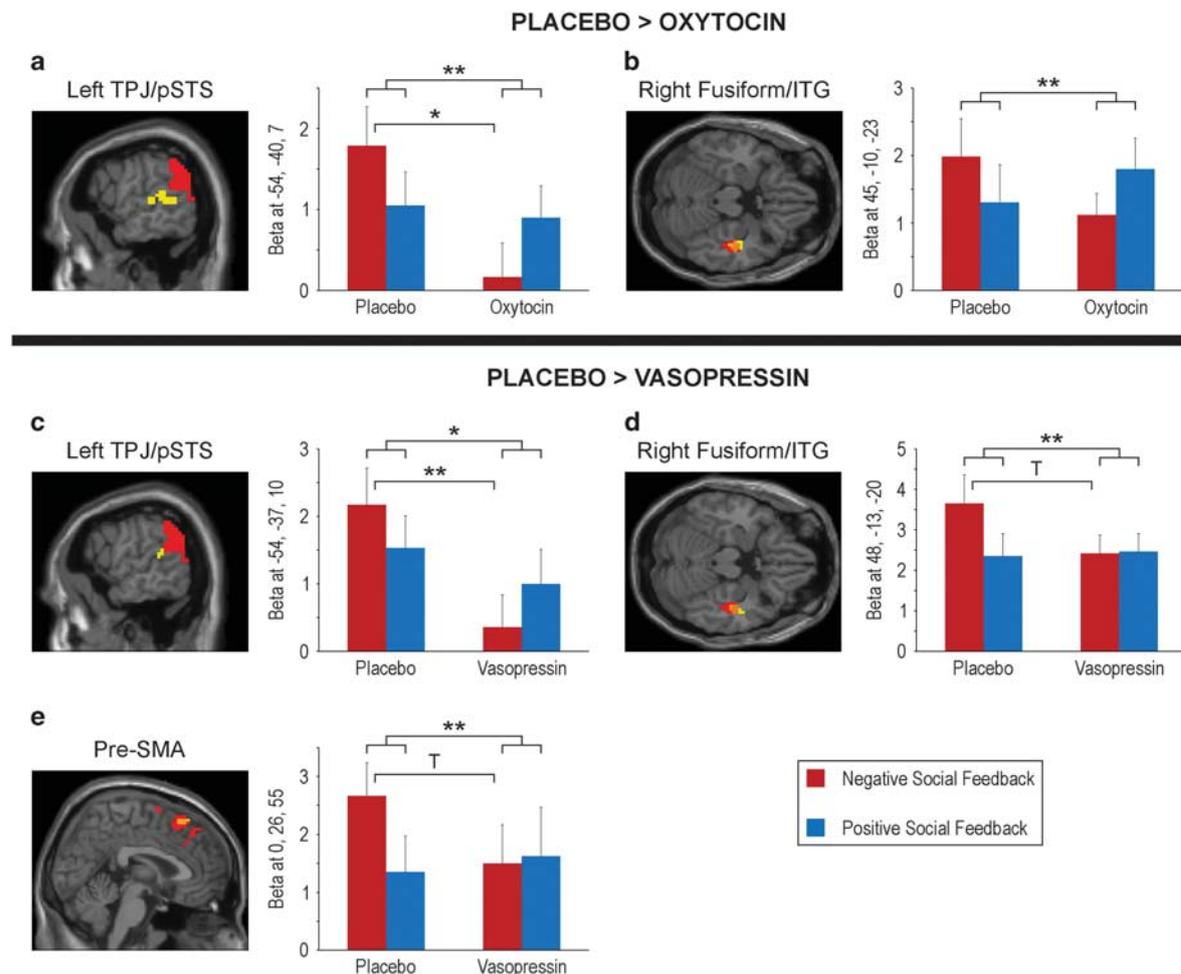
Abbreviations: ITG, inferior temporal gyrus; pSTS, posterior superior temporal sulcus; SMA, supplementary motor area; TPJ, temporoparietal junction.



**Figure 2** Significant ( $p < 0.05$ , corrected) activation foci for negative social feedback > positive social feedback during placebo: (a) right anterior insula, (b) right fusiform/inferior temporal gyrus (ITG), (c) left temporoparietal junction (TPJ), (d) right TPJ, and (e) supplementary motor area (SMA)/pre-SMA. For each of these activations (peak voxel), mean beta estimates (and SEM) for negative social feedback and positive social feedback are plotted separately under placebo, oxytocin and vasopressin administration. \*\* $p < 0.001$ ; \* $p = 0.006$ ; ns, not significant.

oxytocin or vasopressin for any placebo peak with the exception of insula under vasopressin (albeit weaker than under placebo; Figure 2). Although the preferential neural responses to negative social feedback under placebo were not

significant under oxytocin and vasopressin, no interaction contrasts statistically comparing between drug conditions reached significance with corrected  $p$ -values. At a more lenient threshold ( $p \leq 0.001$  uncorrected,  $t > 3.0$ ), however,



**Figure 3** Activations corresponding to preferential neural responses to negative social feedback (ie, negative > positive social feedback) under placebo in regions that also have an outcome (negative–positive) by drug (placebo–oxytocin or placebo–vasopressin) interaction ( $p \leq 0.001$ , uncorrected,  $t > 3$ ). The preferential neural response to negative social feedback is greater under placebo than oxytocin in (a) left temporoparietal junction (TPJ)/posterior superior temporal sulcus (pSTS), and (b) right fusiform/inferior temporal gyrus (ITG). The preferential neural response to negative social feedback is greater under placebo than vasopressin in (c) left TPJ/pSTS, (d) right fusiform/ITG, and (e) pre-supplementary motor area (pre-SMA). Significant activations under placebo alone are displayed in red, activations for placebo–drug comparisons are displayed in yellow ( $p < 0.005$ , cluster size  $\geq 25$  for illustrative purposes), and where they overlap is displayed in orange. For each peak from the placebo–drug comparison, mean beta estimates (and SEM) for negative social feedback and positive social feedback are plotted separately under placebo and drug. The legend corresponds to the beta graphs. \*\* $p \leq 0.005$ ; \* $p \leq 0.01$ ; T indicates a trend for significance that does not survive correction for multiple comparisons ( $p = 0.016$ – $0.05$ ).

most regions (with the exception of insula) that were significantly activated by negative compared with positive social feedback under placebo were statistically more activated under placebo than oxytocin (ie, placebo > oxytocin) or vasopressin (ie, placebo > vasopressin), suggesting a trend for significance for those drug  $\times$  outcome interactions (Figure 3; Table 3). Specifically, preferential responses to negative compared with positive social feedback were greater under placebo than oxytocin in left TPJ/pSTS and right fusiform/ITG, and greater to placebo than vasopressin in left TPJ/pSTS, right fusiform/ITG, and pre-SMA (a full list of brain regions is provided in Table 3). Activation patterns to negative compared with positive social feedback in these regions did not significantly differ between the oxytocin and vasopressin conditions, even at the uncorrected threshold. An analysis of peak beta values revealed that the differences between placebo and oxytocin/vasopressin were mostly due to a drug-induced significant decrease in activation to

negative social feedback (as opposed to increase to positive social feedback; Figure 3). Beta values for positive social feedback did not significantly differ between placebo and oxytocin or vasopressin ( $p > 0.20$ ).

## DISCUSSION

In this study, we investigated the effects of oxytocin and vasopressin on brain activity during social feedback using a newly developed paradigm. In line with evidence that negative social feedback is more salient and potent than positive social feedback in terms of cognitive and neural processing (Baumeister *et al*, 2001), several brain regions were identified in which activity was preferentially associated with negative social feedback. These activations weakened with oxytocin and vasopressin administration such that

**Table 3** Drug Condition Comparisons ( $p \leq 0.001$ ,  $t > 3.0$ ) for Preferential Neural Responses to Negative Social Feedback (Negative Social Feedback > Positive Social Feedback)

Brain region	MNI coordinates X, Y, Z	Cluster size	Peak T-value	Peak uncorrected p-value
<i>Placebo &gt; Oxytocin</i>				
Left superior frontal cortex	-21, 32, 55	55	5.01	<0.001
Right superior temporal gyrus	60, -4, -14	36	4.55	<0.001
Left TPJ/pSTS <sup>a</sup>	-54, -40, 7	101	4.12	<0.001
Cingulate	-12, -22, 49	202	4.11	<0.001
Left fusiform	-24, -37, -20	89	4.05	<0.001
Left OFC	-39, 29, -17	60	3.93	<0.001
Right fusiform	24, -34, -17	39	3.85	<0.001
Right fusiform/ITG <sup>a</sup>	45, -10, -23	53	3.50	<0.001
Left middle temporal gyrus	-57, -13, -17	56	3.47	<0.001
Posterior cingulate	-9, -55, 22	30	3.30	0.001
<i>Oxytocin &gt; Placebo</i>				
No significant activation				
<i>Placebo &gt; Vasopressin</i>				
Left superior frontal incl.	-18, 35, 55	100	4.12	<0.001
Pre-SMA <sup>a</sup>	0, 26, 55		3.07	0.001
Anterior cingulate/MPFC	-9, 44, 28	52	3.76	<0.001
Right fusiform/ITG <sup>a</sup>	48, -13, -20	28	3.54	<0.001
Left OFC	-30, 26, -20	34	3.42	<0.001
Right OFC	45, 35, -11	50	3.33	0.001
MPFC	-3, 38, 40	26	3.31	0.001
Left TPJ/pSTS <sup>a</sup>	-54, -37, 10	46	3.29	0.001
<i>Vasopressin &gt; Placebo</i>				
No significant activations				
<i>Oxytocin &gt; Vasopressin</i>				
Anterior cingulate	9, 23, 25	25	3.50	<0.001
<i>Vasopressin &gt; Oxytocin</i>				
Right superior temporal gyrus	39, 14, -26	32	3.75	<0.001
Posterior cingulate	-3, -31, 10	34	3.60	<0.001

Abbreviations: ITG, inferior temporal gyrus; MPFC, medial prefrontal cortex; OFC, orbitofrontal cortex; pSTS, posterior superior temporal sulcus; TPJ, temporoparietal junction. Cluster size determined by  $p < 0.005$ , cluster size  $\geq 25$  for illustrative purposes in Figure 3.

<sup>a</sup>Brain region significantly activated for negative social feedback > positive social feedback under placebo (and not oxytocin or vasopressin).

neural responses to receiving negative social feedback were not significantly greater than positive social feedback.

Under placebo, exposure to negative social feedback compared with positive social feedback enhanced activity in a variety of brain regions known to be involved in social cognitive functioning. One such region, the TPJ/pSTS, has been implicated in a variety of social-specific cognitive processes, including theory of mind processes (Bahnmann *et al*, 2010; Saxe and Kanwisher, 2003) and cognitive empathy (Shamay-Tsoory, 2011). The TPJ/pSTS is strongly engaged when inferring temporary states such as goals, intentions, and desires of other people (Van Overwalle, 2009) and when processing the communicative significance of other's behavior (Bahnmann *et al*, 2010), particularly when determined from biological motion (as the case with the feedback videos in the current study; Carrington and Bailey, 2009). Furthermore, TPJ/pSTS has been implicated in encoding and updating representations of other people's beliefs and feelings (Behrens *et al*, 2008; Cooper *et al*, 2014; Young and Saxe, 2008). A study using transcranial magnetic stimulation to transiently disrupt neural activity in the TPJ found that participants judged attempted harms as less morally forbidden and more morally permissible, suggesting that the TPJ may be especially involved in the attribution of negative beliefs (Young *et al*, 2010). Accordingly, we suggest that the preferential activation of the TPJ/pSTS to negative social feedback may reflect the attribution of negative beliefs and negative judgment from the investigator whom subjects were told was observing and keeping track of their performance in real time. Notably, both oxytocin and vasopressin weaken the preferential response of the TPJ/pSTS to negative social feedback, perhaps thus weakening the updated representation of negative judgment from the investigator. Interestingly, congruent with our interpretation of the TPJ/pSTS activation, a recent behavioral study found that intranasally administered oxytocin impaired updates of undesirable feedback but facilitated updates of desirable feedback (Ma *et al*, 2016). The drug-induced dampening of TPJ/pSTS activation in the current study is in accordance with the previous studies in men demonstrating that both oxytocin (Domes *et al*, 2007) and vasopressin (Zink *et al*, 2011) decrease TPJ reactivity during implicit processing of unfamiliar negative emotional faces. Here we extend previous findings by targeting a different negative social cognitive process, ie, negative social feedback, and demonstrate effects of both oxytocin and vasopressin.

The fusiform/ITG is another brain region more strongly activated by negative relative to positive social feedback under placebo, a response weakened by both oxytocin and vasopressin. As part of the ventral visual 'what' stream, the fusiform and ITG are specialized in the identification and recognition of visual stimuli, including faces (Haxby *et al*, 2001). Consistent with the current study in which feedback was provided by dynamic facial expressions and hand gestures in a video, activation of the fusiform/ITG has been implicated in the recognition of dynamic, affective facial expressions and gestures, with a right hemisphere dominance for processing affective body expressions (Prochnow *et al*, 2013). As further support for the current findings of preferential fusiform/ITG responses to negative social feedback, evidence suggests that the fusiform and ITG are more activated by dynamic faces and bodies portraying negative

emotion (Kret *et al*, 2011) and the fusiform/ITG has been associated with processing facial evocations of grief (Gundel *et al*, 2003). In addition, in line with our results, a recent study using a social exclusion paradigm (ie, a virtual ball-tossing game called 'cyberball') found that the fusiform area was more active in the exclusion than in the inclusion trials (Cristofori *et al*, 2013). The preferential response of fusiform/ITG for negative affective stimuli may represent an attentional/processing bias based on the increased emotional saliency of unpleasant compared with pleasant situations (Baumeister *et al*, 2001) that is reduced by oxytocin and vasopressin.

Exposure to negative compared with positive social feedback also augmented activation in anterior insula and SMA/pre-SMA under placebo. These brain structures have been shown to be involved in physical pain processing (Craggs *et al*, 2007; Lamm *et al*, 2011; Peyron *et al*, 2000; Treede *et al*, 1999), and our finding suggests that they may be involved in the emotional 'pain' associated with negative social evaluation as well, providing additional evidence in support of the idea that physical and emotional pain may share some of the same neurobiological substrates (Cristofori *et al*, 2013; Eisenberger, 2012; Eisenberger *et al*, 2003; Lamm *et al*, 2011). The response bias towards negative social feedback in the insula, which has also been associated with other socially relevant functions, including emotion processing (Pera-Guardiola *et al*, 2016) and affective empathy (Tusche *et al*, 2016), was not significantly reduced by oxytocin or vasopressin. Vasopressin did, however, weaken the preferential response of pre-SMA to negative social feedback. In addition to pain processing, SMA/pre-SMA has been implicated in learned fear, and activity in the SMA/pre-SMA correlates with sympathetic nervous system activity (ie, fear expression; Etkin *et al*, 2011). The effect of vasopressin on SMA/pre-SMA, ie, dampening response bias to negative social feedback, suggests that vasopressin modulates the emotional pain and/or fear associated with negative social evaluations by dampening activity; future behavior and imaging studies are necessary, however, to confirm this postulate.

Interestingly, activity in some of the brain regions modulated by both oxytocin and vasopressin in the current study (ie, TPJ and fusiform) has been shown to be affected by one or both of these neuropeptides in previous human studies during implicit or explicit processing of negative socioemotional stimuli (eg, emotional face processing and negative social scenes) (Domes *et al*, 2007; Petrovic *et al*, 2008; Striepens *et al*, 2012; Zink *et al*, 2011). The similarities between studies suggest not only that these regions have a role in negative socioemotional processing across contexts but that they also appear to be influenced by oxytocin and vasopressin. The directionality of influence (augment or dampen neural responses) is also consistent with many previous studies showing dampened neural responses (Domes *et al*, 2007; Petrovic *et al*, 2008; Zink *et al*, 2011). It should be noted, however, that some previously reported data reveal a neuropeptide-driven augmentation of neural activity (Rilling *et al*, 2012; Striepens *et al*, 2012). Regardless, it is clear that oxytocin and vasopressin do not generally change neural activity in one particular direction, but rather have specific context-dependent effects on activation in regions of interest. In support of context-dependent effects,

the neural activity profile for monetary feedback was more extensive, particularly in frontal cortical areas, than social feedback under placebo, and whereas oxytocin diminished the preferential response to negative monetary feedback in these regions, vasopressin did not (Supplementary Tables S2 and S3).

In this study, we used a within-subjects design, thus reducing error variance associated with individual differences. One potential disadvantage of the within-subjects design could be the presence of carry-over effects (eg, participants may improve task performance from the first to the last visit because of practice, regardless of drug condition); however, we can rule out this possibility because (a) drug order was counterbalanced; (b) we scheduled the visits at least 1 week apart to minimize carry-over effects; and (c) we examined potential carry-over effects and did not find significant differences in task performance between visits.

There are several limitations of the current study that should be addressed in future studies. The exact mechanism by which oxytocin and vasopressin alter brain activity during negative social feedback remains unknown. It is unclear from the current study whether these neuropeptides are acting directly on receptors in the influenced brain regions or via indirect pathways. Furthermore, although the event-related design of our paradigm is advantageous to isolate neural activity to unexpected negative social feedback, it limits our ability to identify behavioral or emotional reactions to negative social feedback that mirror our imaging findings because these reactions were not measured on a trial-by-trial basis during the task. A lack of correlation between oxytocin/vasopressin-related imaging and behavioral findings is not uncommon (either because not significant or behavioral correlate was not measured; Chen *et al*, 2016; Feng *et al*, 2015b; Kanat *et al*, 2015; Shin *et al*, 2015; Wittfoth-Schardt *et al*, 2012); however, it is important for future work to determine the influence of oxytocin and vasopressin on behavioral and emotional reactions to negative social feedback using paradigms specifically designed to do so. It should be noted that the lack of a measurable drug influence on emotion inventories measured pre- and post-session in the current study is consistent with many previous studies of oxytocin and vasopressin administration in men using similar doses despite drug-related changes in neural activity (Kirsch *et al*, 2005; Labuschagne *et al*, 2010; Zink *et al*, 2011). Also, oxytocin and vasopressin did not affect reaction time or accuracy in the current paradigm, which is important to ensure that the consequences of the drugs on fMRI signals were not secondary to task-performance differences. An additional important caveat is that the participants were all males, and therefore we cannot determine whether our findings may be generalized to women in the current study. Because of known behavioral and neural gender effects (Feng *et al*, 2015a; Thompson *et al*, 2006), the majority of previous studies have examined the effects of oxytocin and vasopressin in males only (Domes *et al*, 2007; Kirsch *et al*, 2005; Rilling *et al*, 2012; Zink *et al*, 2011; Zink *et al*, 2010), and a few studies have included females only (Domes *et al*, 2010; Groppe *et al*, 2013). It will be important to study the effects in females before assuming the current findings are universal across gender. Moreover, to increase the salience/validity of the social evaluation, the feedback was provided by an actual study investigator with whom each participant interacted in

a manner consistent across participants. In the current study, we do not have evidence to support a gender-bias effect, but because this investigator was female and participants were males, the possibility remained that culturally induced gender biases influenced the current findings. As such, the potential importance of investigator gender could be evaluated in a subsequent study. Also, as noted in the method section, the paradigm also included monetary (as opposed to social) feedback trials of non-interest for the current investigation. Although modeling monetary trials events in the analysis should have accounted for their potential influence analysis-wise, the inclusion of monetary trials may have evoked an unintentional context-related cognitive influence on the social events of interest, which could be teased out in future studies. Finally, although the interaction contrasts statistically comparing between drug conditions revealed interesting and potentially important trends (uncorrected for multiple comparisons within imaging space), they did not reach significance with corrected thresholds in imaging space; therefore, replication, ideally in a larger sample, would be necessary before inferring clinical applications.

In conclusion, to our knowledge, this is the first neuroimaging study examining the effects of oxytocin and vasopressin on preferential negative social evaluation-related brain activity. We found that both neuropeptides affected activation patterns by weakening response biases towards negative social feedback. Our results provide a potential neural basis for the influence of oxytocin and vasopressin on negative social feedback and suggest that both neuropeptides may be relevant for the development and treatment of disorders that involve social dysfunction such as SAD and ASD.

## FUNDING AND DISCLOSURE

This research was supported by the Intramural Research Program of the National Institute of Mental Health, NIH, ZIA MH002868, under NCT01093768, protocol 10-M-0068, and also by the Lieber Institute for Brain Development. The views expressed in this article do not necessarily represent the views of the NIMH, NIH, HHS, or the United States Government. The authors declare no conflict of interest.

## ACKNOWLEDGMENTS

We thank all physicians and nurses who helped us conducting this research, Gerald Overman and Judith Starling for pharmaceutical management, Lainey Bukowiec for manuscript and figure assistance, and Elizabeth Smith for thoughtful comments on the manuscript. We also thank all individuals who participated in our study.

## REFERENCES

Akinola M, Mendes WB (2008). The dark side of creativity: biological vulnerability and negative emotions lead to greater artistic creativity. *Pers Soc Psychol Bull* **34**: 1677–1686.

Bahnemann M, Dziobek I, Prehn K, Wolf I, Heekeren HR (2010). Sociotopy in the temporoparietal cortex: common versus distinct processes. *Soc Cogn Affect Neurosci* **5**: 48–58.

Baumeister RF, Bratslavsky E, Finkenauer C, Vohs KD (2001). Bad is stronger than good. *Rev Gen Psychol* **5**: 323–370.

Behrens TE, Hunt LT, Woolrich MW, Rushworth MF (2008). Associative learning of social value. *Nature* **456**: 245–249.

Born J, Lange T, Kern W, McGregor GP, Bickel U, Fehm HL (2002). Sniffing neuropeptides: a transnasal approach to the human brain. *Nat Neurosci* **5**: 514–516.

Bradley MM, Lang PJ (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *J Behav Ther Exp Psychiatry* **25**: 49–59.

Brunnlieb C, Munte TF, Tempelmann C, Heldmann M (2013). Vasopressin modulates neural responses related to emotional stimuli in the right amygdala. *Brain Res* **1499**: 29–42.

Carrington SJ, Bailey AJ (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Hum Brain Mapp* **30**: 2313–2335.

Chen X, Hackett PD, DeMarco AC, Feng C, Stair S, Haroon E et al (2016). Effects of oxytocin and vasopressin on the neural response to unreciprocated cooperation within brain regions involved in stress and anxiety in men and women. *Brain Imaging Behav* **10**: 581–593.

Cooper JC, Dunne S, Furey T, O'Doherty JP (2014). The role of the posterior temporal and medial prefrontal cortices in mediating learning from romantic interest and rejection. *Cereb Cortex* **24**: 2502–2511.

Craggs JG, Price DD, Verne GN, Perlstein WM, Robinson MM (2007). Functional brain interactions that serve cognitive-affective processing during pain and placebo analgesia. *Neuroimage* **38**: 720–729.

Cristofori I, Moretti L, Harquel S, Posada A, Deiana G, Isnard J et al (2013). Theta signal as the neural signature of social exclusion. *Cereb Cortex* **23**: 2437–2447.

Dickerson SS, Gruenewald TL, Kemeny ME (2004a). When the social self is threatened: shame, physiology, and health. *J Pers* **72**: 1191–1216.

Dickerson SS, Kemeny ME (2004b). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychol Bull* **130**: 355–391.

Domes G, Heinrichs M, Glascher J, Buchel C, Braus DF, Herpertz SC (2007). Oxytocin attenuates amygdala responses to emotional faces regardless of valence. *Biol Psychiatry* **62**: 1187–1190.

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

Ebstein RP, Israel S, Lerer E, Uzefovsky F, Shalev I, Gritsenko I et al (2009). Arginine vasopressin and oxytocin modulate human social behavior. *Ann NY Acad Sci* **1167**: 87–102.

Eisenberger NI (2012). The pain of social disconnection: examining the shared neural underpinnings of physical and social pain. *Nat Rev Neurosci* **13**: 421–434.

Eisenberger NI, Lieberman MD, Williams KD (2003). Does rejection hurt? An fMRI study of social exclusion. *Science* **302**: 290–292.

Etkin A, Egner T, Kalisch R (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cognit Sci* **15**: 85–93.

Evans S, Shergill SS, Averbach BB (2010). Oxytocin decreases aversion to angry faces in an associative learning task. *Neuropsychopharmacology* **35**: 2502–2509.

Feng C, Hackett PD, DeMarco AC, Chen X, Stair S, Haroon E et al (2015a). Oxytocin and vasopressin effects on the neural response to social cooperation are modulated by sex in humans. *Brain Imaging Behav* **9**: 754–764.

Feng C, Lori A, Waldman ID, Binder EB, Haroon E, Rilling JK (2015b). A common oxytocin receptor gene (OXTR) polymorphism modulates intranasal oxytocin effects on the neural response to social cooperation in humans. *Genes Brain Behav* **14**: 516–525.

- Gamer M, Zurowski B, Buchel C (2010). Different amygdala subregions mediate valence-related and attentional effects of oxytocin in humans. *Proc Natl Acad Sci USA* **107**: 9400–9405.
- Grope SE, Gossen A, Rademacher L, Hahn A, Westphal L, Grunder G et al (2013). Oxytocin influences processing of socially relevant cues in the ventral tegmental area of the human brain. *Biol Psychiatry* **74**: 172–179.
- Gundel H, O'Connor MF, Littrell L, Fort C, Lane RD (2003). Functional neuroanatomy of grief: an fMRI study. *Am J Psychiatry* **160**: 1946–1953.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**: 2425–2430.
- Heinrichs M, Baumgartner T, Kirschbaum C, Ehlert U (2003). Social support and oxytocin interact to suppress cortisol and subjective responses to psychosocial stress. *Biol Psychiatry* **54**: 1389–1398.
- Heinrichs M, von Dawans B, Domes G (2009). Oxytocin, vasopressin, and human social behavior. *Front Neuroendocrinol* **30**: 548–557.
- Kanat M, Heinrichs M, Schwarzwald R, Domes G (2015). Oxytocin attenuates neural reactivity to masked threat cues from the eyes. *Neuropsychopharmacology* **40**: 287–295.
- Kinoshita Y, Kingdon D, Kinoshita K, Kinoshita Y, Saka K, Arisue Y et al (2011). Fear of negative evaluation is associated with delusional ideation in non-clinical population and patients with schizophrenia. *Soc Psychiatry Psychiatric Epidemiol* **46**: 703–710.
- Kirsch P, Esslinger C, Chen Q, Mier D, Lis S, Siddhanti S et al (2005). Oxytocin modulates neural circuitry for social cognition and fear in humans. *J Neurosci* **25**: 11489–11493.
- Kis A, Kemerle K, Hernadi A, Topal J (2013). Oxytocin and social pretreatment have similar effects on processing of negative emotional faces in healthy adult males. *Front Psychol* **4**: 532.
- Knutson B, Fong GW, Adams CM, Varner JL, Hommer D (2001). Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport* **12**: 3683–3687.
- Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E (2005). Oxytocin increases trust in humans. *Nature* **435**: 673–676.
- Kret ME, Pichon S, Grezes J, de Gelder B (2011). Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. *Neuroimage* **54**: 1755–1762.
- Labuschagne I, Phan KL, Wood A, Angstadt M, Chua P, Heinrichs M et al (2010). Oxytocin attenuates amygdala reactivity to fear in generalized social anxiety disorder. *Neuropsychopharmacology* **35**: 2403–2413.
- Lamm C, Decety J, Singer T (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* **54**: 2492–2502.
- Ma Y, Li S, Wang C, Liu Y, Li W, Yan X et al (2016). Distinct oxytocin effects on belief updating in response to desirable and undesirable feedback. *Proc Natl Acad Sci USA* **113**: 9256–9261.
- Mendes WB, Major B, McCoy S, Blascovich J (2008). How attributional ambiguity shapes physiological and emotional responses to social rejection and acceptance. *J Pers Soc Psychol* **94**: 278–291.
- Meyer JA, Mundy PC, Van Hecke AV, Durocher JS (2006). Social attribution processes and comorbid psychiatric symptoms in children with Asperger syndrome. *Autism* **10**: 383–402.
- Norman GJ, Cacioppo JT, Morris JS, Karelina K, Malarkey WB, Devries AC et al (2011). Selective influences of oxytocin on the evaluative processing of social stimuli. *J Psychopharmacol* **25**: 1313–1319.
- O'Connor LE, Berry JW, Weiss J, Gilbert P (2002). Guilt, fear, submission, and empathy in depression. *J Affective Disord* **71**: 19–27.
- Pera-Guardiola V, Contreras-Rodriguez O, Batalla I, Kosson D, Menchon JM, Pifarre J et al (2016). Brain structural correlates of emotion recognition in psychopaths. *PLoS ONE* **11**: e0149807.
- Petrovic P, Kalisch R, Singer T, Dolan RJ (2008). Oxytocin attenuates affective evaluations of conditioned faces and amygdala activity. *J Neurosci* **28**: 6607–6615.
- Peyron R, Laurent B, Garcia-Larrea L (2000). Functional imaging of brain responses to pain. A review and meta-analysis (2000). *Neurophysiol Clin* **30**: 263–288.
- Prochnow D, Hoing B, Kleiser R, Lindenberg R, Wittsack HJ, Schafer R et al (2013). The neural correlates of affect reading: an fMRI study on faces and gestures. *Behav Brain Res* **237**: 270–277.
- Rapee RM, Heimberg RG (1997). A cognitive-behavioral model of anxiety in social phobia. *Behav Res Ther* **35**: 741–756.
- Rilling JK, DeMarco AC, Hackett PD, Thompson R, Ditzen B, Patel R et al (2012). Effects of intranasal oxytocin and vasopressin on cooperative behavior and associated brain activity in men. *Psychoneuroendocrinology* **37**: 447–461.
- Saxe R, Kanwisher N (2003). People thinking about thinking people. The role of the temporo-parietal junction in 'theory of mind'. *Neuroimage* **19**: 1835–1842.
- Shamay-Tsoory SG (2011). The neural bases for empathy. *Neuroscientist* **17**: 18–24.
- Shin NY, Park HY, Jung WH, Park JW, Yun JY, Jang JH et al (2015). Effects of oxytocin on neural response to facial expressions in patients with schizophrenia. *Neuropsychopharmacology* **40**: 1919–1927.
- Spielberger CD (1988). *Manual for the State-Trait Anger Expression Inventory (STAXI)*. Psychological Assessment Resources: Odessa, FL, USA.
- Spielberger CD, Gorsuch RL, Lushene PR, Vagg PR, Jacobs AG (1983). *Manual for the State-Trait Anxiety Inventory (Form Y)*. Consulting Psychologists Press: Palo Alto, CA.
- Striepens N, Scheele D, Kendrick KM, Becker B, Schafer L, Schwalba K et al (2012). Oxytocin facilitates protective responses to aversive social stimuli in males. *Proc Natl Acad Sci USA* **109**: 18144–18149.
- Theodoridou A, Rowe AC, Penton-Voak IS, Rogers PJ (2009). Oxytocin and social perception: oxytocin increases perceived facial trustworthiness and attractiveness. *Horm Behav* **56**: 128–132.
- Thompson R, Gupta S, Miller K, Mills S, Orr S (2004). The effects of vasopressin on human facial responses related to social communication. *Psychoneuroendocrinology* **29**: 35–48.
- Thompson RR, George K, Walton JC, Orr SP, Benson J (2006). Sex-specific influences of vasopressin on human social communication. *Proc Natl Acad Sci USA* **103**: 7889–7894.
- Treede RD, Kenshalo DR, Gracely RH, Jones AK (1999). The cortical representation of pain. *Pain* **79**: 105–111.
- Tusche A, Bockler A, Kanske P, Trautwein FM, Singer T (2016). Decoding the charitable brain: empathy, perspective taking, and attention shifts differentially predict altruistic giving. *The J Neurosci* **36**: 4719–4732.
- Van Overwalle F (2009). Social cognition and the brain: a meta-analysis. *Hum Brain Mapp* **30**: 829–858.
- Wang WT, Hsu WY, Chiu YC, Liang CW (2012). The hierarchical model of social interaction anxiety and depression: the critical roles of fears of evaluation. *J Anxiety Disord* **26**: 215–224.
- Watson D, Clark LA, Tellegen A (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers Soc Psychol* **54**: 1063–1070.
- Wittfoth-Schardt D, Grunding J, Wittfoth M, Lanfermann H, Heinrichs M, Domes G et al (2012). Oxytocin modulates neural reactivity to children's faces as a function of social salience. *Neuropsychopharmacology* **37**: 1799–1807.
- Young L, Camprodon JA, Hauser M, Pascual-Leone A, Saxe R (2010). Disruption of the right temporoparietal junction with

transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proc Natl Acad Sci USA* **107**: 6753–6758.

Young L, Saxe R (2008). The neural basis of belief encoding and integration in moral judgment. *Neuroimage* **40**: 1912–1920.

Zink CF, Kempf L, Hakimi S, Rainey CA, Stein JL, Meyer-Lindenberg A (2011). Vasopressin modulates social

recognition-related activity in the left temporoparietal junction in humans. *Transl Psychiatry* **1**: e3.

Zink CF, Stein JL, Kempf L, Hakimi S, Meyer-Lindenberg A (2010). Vasopressin modulates medial prefrontal cortex-amygdala circuitry during emotion processing in humans. *J Neurosci* **30**: 7017–7022.

Supplementary Information accompanies the paper on the Neuropsychopharmacology website (<http://www.nature.com/npp>)