Supporting Information

Striepens et al. 10.1073/pnas.1208852109

SI Text

Subjects. ***Sample characteristics.*** Subjects were recruited by local advertisement at the University Bonn and provided written informed consent before study enrollment. Eighty healthy right-handed male volunteers [oxytocin (OXT) group: age range, 19–31, mean = 24 ± 3 y; placebo (PLC) group: age range 21–32, mean = 25 ± 3 y] participated in Exp. 1, and 73 healthy right-handed male volunteers (OXT group: age range 20–25, mean = 26 ± 4 y; PLC group: age range 20–25, mean = 26 ± 4 y) participated in Exp. 2. The subjects were free of current and past 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). All participants were naive to prescription-strength psychoactive medication and had not taken any over-the-counter psychoactive medication in the past 4 wk. Participants were asked to maintain their regular bed and wake times and to abstain from caffeine and alcohol intake on the day of the experiment. Tobacco smokers were excluded from participation. In Exp. 2, contraindications for MRI scanning were additional exclusion criteria. In both experiments OXT- and PLC-treated subjects showed no a priori differences regarding age, education, and pretreatment neuropsychological performance (all P values > 0.05) (for details see Tables S1 and S2). To control for potentially confounding effects of OXT on attention and anxiety, all subjects completed the d2 Test of Attention (Aufmerksamkeits- und Belastungstest d2) (1) and the State-Trait Anxiety Inventory (STAI) (2) immediately before the start of the experimental tasks. Analysis of these variables revealed no significant differences between the PLC- and OXT-treated subjects in both experiments (all P values > 0.05) (Tables S1 and S2). Thus, between-group differences cannot be attributed to potential confounding effects of OXT on attention or anxiety. After the experiments, participants were asked to guess whether they had received OXT or PLC; again, there was no significant group difference [Exp. 1: $\chi^2(1) = 0.88$, $P = 0.35$, Cramer-V = 0.12; Exp. 2: $\chi^2(1) = 0.61$, $P = 0.80$, Cramer-V = 0.06].

Neuropsychological screening. To control for pretreatment differences in cognitive performance, all participants completed a comprehensive neuropsychological test battery. Neuropsychological testing included the German version of the RAVLT (Rey Auditory Verbal Learning Test) (3, 4) to assess verbal learning skills, the DST (digit-span test) derived from the revised Wechsler adult intelligence scale (5) to assess working memory performance, the LPS-4 (Leistungspruefsystem Subtest 4) (6) to assess nonverbal reasoning IQ, the MWT-B (Mehrfach-Wortschatz-Intelligenztest Teil B) (7) to assess verbal IQ based on lexical decisions, and the trail-making test (TMT) part A and B (8) to assess visual attention and task-switching abilities. Because a previous study has reported that individual differences in the behavioral inhibition and approach system could account for some of the variance in the emotional modulation of the acoustic startle reflex (ASR) (9), we additionally administered a German version (10) of the questionnaire developed by Carver and White (11) to assess these personality dimensions.

Experiment 1. ***Set-up and stimuli.*** Subjects were seated ∼100 cm in front of a computer screen in a slightly reclined chair with a headrest and instructed to view the pictures presented on-screen and to disregard noises they might hear. The semantic contents of the pictures used in Exp. 1 comprised attractive women (in one picture together with a child) and erotic heterosexual couples in the pleasant condition, household (e.g., knobs, clothespins, a screwdriver) and kitchen objects (e.g., a spoon, a cup) in the neutral condition, and attacking humans (e.g., knife and gun assaults), injured humans (e.g., an accident victim, a starving child, a burning man), and mutilated bodies in the negative condition. Using an in-house programmed script the pictures were adjusted to closely resemble each other in luminance. An independent pilot study involving eight healthy men, none of whom participated in the main experiment, confirmed that the positive and negative pictures were equivalent in terms of arousal ratings. The arousal and valence scores were as follows: negative stimuli (mean = 5.8 ± 0.8 and mean = 2.7 ± 0.6); neutral stimuli (mean = 2.7 ± 1.0 and mean = 5.4 ± 0.4); positive stimuli (mean = 5.0 ± 1.1 and mean = 6.9 ± 0.4). A repeated-measures ANOVA yielded a significant main effect of arousal [$F_{(2, 14)} = 40.16$, $P < 0.01$, $\eta^2 = 0.85$], but Bonferroni-corrected post hoc tests revealed no difference between the negative and positive category ($P = 0.37$). Furthermore, there was a significant linear trend for valence [i.e., negative > neutral > positive; $F_{(1, 7)} = 292.30$, $P < 0.01$, $\eta^2 = 0.98$]. In the ASR paradigm pictures were presented in a pseudorandomized order. Each picture was presented for 5 s. Pictures were presented in two separate runs with 10 pictures of each valence category (30 pictures per run). Pictures were separated by a fixation cross, which was presented in the center of the screen for a randomly generated time interval, ranging from 7 to 17 s (mean, 12 s).

Electrodermal activity. Electrodermal activity (EDA) was measured by two EDA electrodes attached to the thenar and hypothenar of the nondominant hand. The electrodes were filled with a nonhydrating NaCl paste and a constant current of 0.5 V was applied. A commercial system (Contact Precision Instruments) was used for stimulus delivery and psychophysiological recordings. EDA was recorded at a sampling rate of 1,000 Hz. EDA was analyzed for each subject individually for a 10-s period from picture onset. A skin conductance response (SCR) was defined as the first wave in a time window between 1 and 4 s after stimulus onset, with a phasic increase in conductance of more than 0.02 μS. The peak amplitude was baseline-corrected by subtracting the amplitude at stimulus onset. SCR magnitudes were only calculated for trials without startle probe. To normalize the data a log transformation was used. Data from a participant were included in the SCR analysis if at least one nonzero response was available. Eight participants had to be classified as nonresponders and had to be excluded from further analysis. A repeated-measures ANOVA for the SCR magnitude [log(μS + 1)] with "valence" (negative, neutral, or positive) as within-subjects factor and "treatment" (OXT or PLC) as between-subjects factor yielded a significant valence effect [$F_{(2, 118)} = 19.65$, $P < 0.01$, $\eta^2 = 0.25$], but no interaction between valence and treatment [$F_{(2, 118)} = 0.33$, $P = 0.72$, $\eta^2 = 0.01$]. Bonferroni-corrected post hoc tests revealed that both positive (mean = 0.04 ± 0.04, $P < 0.01$) and negative stimuli (mean = 0.03 ± 0.03, $P < 0.01$) elicited a stronger SCR than neutral scenes (mean = 0.01 ± 0.03), with positive pictures even leading to a greater electrodermal response than negative pictures ($P = 0.01$).

Emotion-modulated startle. The startle stimulus consisted of a single 50-ms burst of white noise (100 dB) with nearly instantaneous rise and was delivered binaurally via headphones during 60% of the pictures (i.e., 12 from each category) at 2–4 s after picture onset. A 70-dB white noise background was present throughout the experiment. Facial electromyographic (EMG) activity was recorded from two Ag/AgCl electrodes placed over the orbicularis oculi muscle below the left eye (12). A ground electrode was placed behind the subjects' left ear. A commercial system (Contact Precision Instruments) was used for stimulus delivery and psychophysiological recordings. In addition 18 of 59 interstimulus intervals (ISIs) were

accompanied by startle probes to reduce predictability. To account for early habituation, the experiment started with the presentation of five startle probes in 2-s intervals with no picture and five startle probes during the presentation of a neutral picture. The facial EMG signal was digitized at a rate of 1,000 Hz and amplified with a high-pass filter of 30 Hz and a low-pass filter of 500 Hz. EMG data were rectified and smoothed by a four-point moving average. Startle eyeblink reflex was calculated as the difference between the maximum increase of EMG activity in a time interval between 20 and 100 ms after startle-probe onset and the mean EMG of the 50-ms baseline directly preceding the onset. All EMG data were *z*-transformed within-subjects and then converted into *t*-scores to reduce between-subjects variability and skew. The EMG recordings were visually inspected, and trials with excessive noise were excluded from further analysis. Trials with no perceptible eye-blink reflex were assigned a magnitude of zero and included in the analysis. Subjects displaying fewer than 25% satisfactory blink responses in the paradigm (OXT group, $n = 2$; PLC group, $n = 4$) were excluded. Startle latencies were analyzed in a repeated-measures ANOVA with "group" as between-subjects factor (OXT vs. PLC) and "valence" (negative, neutral, or positive) as within-subjects factor. There was neither a main nor an interaction effect (all Ps > 0.34). Furthermore, we also analyzed the reaction times for the arousal and valence ratings of the stimuli used in the startle paradigm. The reaction times of all judgments were comparable between both treatment groups, except for the arousal ratings for neutral pictures. In this category, the OXT group was significantly slower than the PLC group [$t_{(59.55)} = 2.09$, $P = 0.04$, $\eta^2 = 0.06$] (Table S3).

Experiment 2. ***Set-up and stimuli.*** We decided to implement a free-recall test for two reasons. First, it has been shown that subsequent memory effects are greater for a free-recall than a recognition test (13). Second, it has been suggested that OXT effects on empathy are more pronounced for difficult compared with easy task items (14), and free-recall tests are invariably more difficult than recognition tests (15). The behavioral data for the free-recall test are presented in Table S4. Each stimulus comprised a colored picture depicting a socially salient situation and its verbal descriptor. Using an in-house programmed script, the pictures were adjusted to closely resemble each other in size and luminance. The verbal descriptor (presented in arial font) was a noun with 4–14 letters that was semantically identical to the picture. Two categories of pictures (negative and neutral) were mainly selected from the International Affective Picture System (16) based on their standard normative scores for emotional arousal and valence. The semantic contents of the pictures used in Exp. 2 comprised daily living situations (e.g., cooking, buying groceries), waiting humans (e.g., on a market, on a street), and working humans (e.g., secretary, musician) in the neutral condition; and attacking or threatening humans (e.g., knife and gun assaults), injured humans or animals (e.g., an accident victim, a starving child, a burning man, extirpation of animals), mutilated bodies, and a snarling dog in the negative condition. In total, 21 of 24 pictures used portrayed human victims of violence/disease. In addition, stimuli were rated by an independent sample of 13 male volunteers on a nine-point self-assessment manikin scale for arousal (1, calm; 9, excited), valence (1, negative; 9, positive), and semantic congruency of the picture and its verbal descriptor (1, lowest; 9, highest). Only stimuli with the highest semantic congruency ($M = 7.1 \pm 1.0$) were included in the final version of the functional MRI (fMRI) paradigm. Arousal and valence scores for the stimuli confirmed the normative emotional arousal and valence ratings (negative stimuli: arousal mean = 6.7 ± 1.2; valence mean = 1.9 ± 0.3; neutral stimuli: arousal mean = 3.1 ± 1.4; valence mean = 6.2 ± 0.6.). Paired-sample *t* tests yielded significant differences between negative and neutral stimuli in arousal [$t_{(12)} = 12.51$, $P < 0.01$] and valence [$t_{(12)} = -23.73$, $P < 0.01$].

fMRI data acquisition and analysis. Stimuli were presented for 5,000 ms and followed by a fixation cross, which served as a low-level baseline. The ISIs ranged between 2,500 ms and 4,500 ms to create jitter. Participants were instructed to fixate the stimuli and to index the valence category of the stimuli (aversive vs. neutral) by button-press. Stimuli were presented to the subjects by means of LCD video goggles (Nordic NeuroLab) connected to a PC running Presentation 14 (Neurobehavioral Systems). fMRI using blood oxygenation level-dependent (BOLD) contrast was carried out on a 1.5 Tesla Siemens Magnetom Espree MRI system (Siemens) using a T2*-weighted echo planar imaging sequence [imaging parameters: TR = 3,000 ms, TE = 50 ms, matrix size: 64×64, pixel size: $3 \times 3 \times 3$ mm, slice thickness = 3.0 mm, distance factor = 10%, field of view (FoV) = 210, flip angle = 90°, 35 axial slices]. In addition, high-resolution anatomical images were acquired on the same scanner using a T1-weighted 3D MPRAGE sequence (imaging parameters: TR = 1,660 ms, TE = 3.09, matrix size 256×256, pixel size 1×1 mm^2, slice thickness = 1.0 mm, FoV = 256, flip angle = 15°, 160 sagital slices). In total, 950 dynamic scans were recorded, and the task lasted ~28 min. fMRI data were preprocessed and analyzed using SPM8 software (Wellcome Trust Centre for Neuroimaging, London, United Kingdom; http://www.fil.ion.ucl.ac.uk/spm) implemented in Matlab 7 (MathWorks). The first five volumes of each functional time series were discarded to allow for T1 equilibration. Images were corrected for head movement between scans by an affine registration (17). For realignment, a two-pass procedure was used, by which images were initially realigned to the first image of the time-series and subsequently rerealigned to the mean of all images. For spatial normalization the mean EPI image of each subject was normalized to the current Montreal Neurological Institute (MNI) template (18, 19) using the unified segmentation function in SPM8. This algorithm combines image registration, tissue classification, and bias correction within the same generative model. All images were hereby transformed into standard stereotaxic space and resampled at $2 \times 2 \times 2$ mm^3 voxel size. The normalized images were spatially smoothed using an 8-mm FWHM Gaussian kernel. Raw time series were detrended by the application of a high-pass filter (cutoff period, 128 s). On the first level, the four conditions (neutral: later remembered; neutral: later forgotten; aversive: later remembered; aversive: later forgotten) were modeled by a stick function convolved with a hemodynamic response function (20). The movement parameters were included as confounds in the design matrix. Specific effects were assessed by applying appropriate linear contrasts to the parameter estimates of the experimental conditions resulting in *t*-statistics for each voxel. On the second level, effects of OXT were analyzed comparing the PLC and OXT treated subjects. To examine the modulatory effect of OXT on the difference due to memory (dm) effect for aversive and neutral items, we contrasted the OXT with the PLC group for the contrasts $\mathrm{OXT}^{\text{aversive: later remembered > aversive: later forgotten}} > \mathrm{PLC}^{\text{aversive: later remembered > aversive: later forgotten}}$ and $\mathrm{OXT}^{\text{neutral: later remembered > neutral: later forgotten}} > \mathrm{PLC}^{\text{neutral: later remembered > neutral: later forgotten}}$. In addition, groups were compared for the contrasts "aversive items > neutral items" (main effect of "valence"), "all later remembered items > all forgotten items" (main effect of "subsequent memory"), and "all items > low-level baseline" (main effect of "group"). Emotion-specific effects of OXT treatment were analyzed using the contrasts "aversive items > low-level baseline" and "neutral items > low-level baseline." Group means were tested using one sample *t* tests, and group differences were tested using two-sample *t* tests on the contrasts of interest (OXT > PLC, OXT < PLC) for the whole-brain with a significance threshold of $P < 0.05$ corrected for multiple comparisons based on family-wise error (FWE). The significant interaction effect was again computed and masked with the appropriate main effects thresholded at $P < 0.05$ (uncorrected) to confirm the independence of interaction effects from simple main effects. Given

the pivotal role of the amygdala in emotional memory-encoding and consolidation (21) and the modulatory effects of OXT treatment on amygdala responses to aversive stimuli (22–27), the comparisons were computed repeatedly, this time with an a priori regional focus on the amygdala. In this region-of-interest (ROI) approach the bilateral amygdala was anatomically defined using the Wake Forest University (WFU) Pickatlas (Version 3.0), which provides a method for generating ROI masks based on the Talairach Daemon database (28–30). The implemented atlases are available in MNI space with dimensions of $91 \times 109 \times 91$ sampled at 2-mm intervals, corresponding to the SPM MNI templates. ROI-based two-sample t tests were computed with a threshold of $P < 0.05$ and FWE-corrected for multiple comparisons based on the size of the ROI. The results of these fMRI analyses are presented in Table S5.

Functional connectivity analysis of fMRI data. To address OXT effects on functional connectivity, a psychophysiological interactions (PPIs) analysis was performed. This analysis models condition-dependent changes in connectivity from a chosen seed region to each voxel in the whole brain. We decided to use a generalized form of context-dependent PPIs (gPPIs) (31). Compared with standard PPIs implementation in SPM, gPPIs allows modeling of more than two task conditions in the same PPIs model by spanning the entire experimental space and potentially improves model fit, specificity to true-negative findings, and sensitivity to true-positive findings (31). In a first analysis, we examined the modulation effects of OXT on functional connectivity with the amygdala. For this aim, we extracted the mean time series for each subject from the left and right amygdala defined using the WFU Pickatlas (28–30). Findings from the analysis of functional activation revealed significant effects of OXT treatment in the left anterior insula. To further examine the modulatory effects of OXT on the interplay between this region and other brain regions, mean time series were extracted from 6-mm radius spheres centered at the coordinates of the maximum t-values group × memory interaction effect for negative items in the left anterior insula (MNI: $x = -38, y = 22, z = -4$). Hemodynamic deconvolution was performed on the extracted time series to remove the effects of canonical hemodynamic response function (HRF). The resulting time-series were multiplied by the psychological variables and reconvolved with the HRF to obtain the PPIs interaction terms. gPPIs analysis for each subject was performed on the first level and included regressors for "neutral: later remembered," "neutral: later forgotten," "aversive: later remembered," and "aversive: later forgotten." Separate PPIs models for the three seed regions were computed. On the first level individual contrasts of interest (for the contrasts implemented see fMRI BOLD activation analysis) were computed and submitted to second level two-sample t tests to test for group differences (OXT > PLC; OXT < PLC). Sensitivity to detect OXT effects was increased by restricting between-group comparisons to regions with known functional connections to the amygdala, such as the superior and medial frontal gyri, insula, thalamus, globus pallidus, and anterior cingulate cortex (32). For the anterior insula between-group comparisons were restricted to regions with known functional (33, 34) and structural (35) connections to the anterior insula, namely the superior, middle, and inferior frontal gyri, anterior cingulate cortex, thalamus, and basolateral amygdala. ROIs were defined using the WFU Pickatlas (28–30) and in case of the basolateral amygdala using the Anatomy toolbox (36–38). Groups were compared using two-sample t tests with a significance threshold of $P < 0.05$ and FWE-corrected for multiple comparisons based on the size of the ROI.

Statistics. Demographical, neuropsychological, and psychophysiological data were analyzed using SPSS 19 (SPSS). Quantitative behavioral data were compared by repeated-measures ANOVAs. Partial η^2 was calculated as a measure of effect size. The assumption of sphericity was assessed with Mauchly's test, and for significant violations Greenhouse–Geisser's correction was applied. For qualitative variables, Pearson's χ^2 tests and Fisher's exact tests were used. All reported P values are two-tailed and P values of $P < 0.05$ were considered significant.

1. Brickenkamp R (1995) Aufmerksamkeits-Belastungs-Test 'd2', erweiterte und neu gestaltete Auflage. *Diagnostica* 41:291–296.
2. Spielberger C, Gorsuch R, Lushene R (1970) *Manual for the State-Trait Anxiety Inventory* (Consulting Psychologist Press, Palo Alto, CA).
3. Helmstaedter C, Lendt M, Lux S (2001) *VLMT—Verbaler Lern- und Merkfähigkeitstest [VLMT—Verbal learning and memory test]* (Beltz Test GmbH, Göttingen, Germany).
4. Rey LB (1941) L'examen psychologique dans les cas d'encephalopathie traumatique. *Arch Psychol* 28:55.
5. Wechsler D (1997) *Wechsler Adult Intelligence Scale. Administration and Scoring Manual* (Psychological Corporation, San Antonio, TX), 3 ed.
6. Horn W (1983) *Leistungsprüfsystem L-P-S [Performance testing system L-P-S]* (Hogrefe, Göttingen, Germany).
7. Lehrl S (1978) *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B [Multiple-choice lexical intelligence test MWT-B]* (Dr. Med. Straube, Erlangen, Germany).
8. Raitan RM (1958) Validity of the trail making test as an indication of organic brain damage. *Percept Mot Skills* 8:271–276.
9. Caseras FX, et al. (2006) Influence of individual differences in the Behavioral Inhibition System and stimulus content (fear versus blood-disgust) on affective startle reflex modulation. *Biol Psychol* 72(3):251–256.
10. Hartig JM, Moosbrugger H (2003) The ARES-Scales as a measurement of individual BIS and BAS sensitivity: Development of a long and a short questionnaire version. *Z Differ Diagn Psychol* 24(4):293–310.
11. Carver CS, White TL (1994) Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *J Pers Soc Psychol* 67(2):319–333.
12. Fridlund AJ, Cacioppo JT (1986) Guidelines for human electromyographic research. *Psychophysiology* 23(5):567–589.
13. Paller KA, McCarthy G, Wood CC (1988) ERPs predictive of subsequent recall and recognition performance. *Biol Psychol* 26(1–3):269–276.
14. Domes G, Heinrichs M, Michel A, Berger C, Herpertz SC (2007) Oxytocin improves "mind-reading" in humans. *Biol Psychiatry* 61(6):731–733.
15. Haist F, Shimamura AP, Squire LR (1992) On the relationship between recall and recognition memory. *J Exp Psychol Learn Mem Cogn* 18(4):691–702.
16. Lang PJ, Bradley MM, Cuthbert BN (2005) *International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual. Technical Report A-6* (Univ of Florida, Gainesville, FL).
17. Ashburner J, Friston KJ (2003) *Human Brain Function*, ed Frackowiak RS (Academic, London, UK), 2nd Ed.
18. Evans AC, et al. (1992) Anatomical mapping of functional activation in stereotactic coordinate space. *Neuroimage* 1(1):43–53.
19. Holmes CJ, et al. (1998) Enhancement of MR images using registration for signal averaging. *J Comput Assist Tomogr* 22(2):324–333.
20. Friston KJ, et al. (1995) Statistical parametric maps in functional imaging: A general linear approach. *Hum Brain Mapp* 2(4):189–210.
21. Murty VP, Ritchey M, Adcock RA, LaBar KS (2010) fMRI studies of successful emotional memory encoding: A quantitative meta-analysis. *Neuropsychologia* 48 (12):3459–3469.
22. Kirsch P, et al. (2005) Oxytocin modulates neural circuitry for social cognition and fear in humans. *J Neurosci* 25(49):11489–11493.
23. Petrovic P, Kalisch R, Singer T, Dolan RJ (2008) Oxytocin attenuates affective evaluations of conditioned faces and amygdala activity. *J Neurosci* 28(26):6607–6615.
24. Domes G, et al. (2007) Oxytocin attenuates amygdala responses to emotional faces regardless of valence. *Biol Psychiatry* 62(10):1187–1190.
25. Singer T, et al. (2008) Effects of oxytocin and prosocial behavior on brain responses to direct and vicariously experienced pain. *Emotion* 8(6):781–791.
26. Domes G, et al. (2010) Effects of intranasal oxytocin on emotional face processing in women. *Psychoneuroendocrinology* 35(1):83–93.
27. Gamer M, Zurowski B, Büchel C (2010) Different amygdala subregions mediate valence-related and attentional effects of oxytocin in humans. *Proc Natl Acad Sci USA* 107(20):9400–9405.
28. Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* 19(3):1233–1239.
29. Maldjian JA, Laurienti PJ, Burdette JH (2004) Precentral gyrus discrepancy in electronic versions of the Talairach atlas. *Neuroimage* 21(1):450–455.
30. Tzourio-Mazoyer N, et al. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15(1):273–289.
31. McLaren DG, Ries ML, Xu G, Johnson SC (2012) A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. *Neuroimage* 61(4):1277–1286.
32. Roy AK, et al. (2009) Functional connectivity of the human amygdala using resting state fMRI. *Neuroimage* 45(2):614–626.
33. Cauda F, et al. (2012) Meta-analytic clustering of the insular cortex: Characterizing the meta-analytic connectivity of the insula when involved in active tasks. *Neuroimage* 62 (1):343–355.

34. Cauda F, et al. (2011) Functional connectivity of the insula in the resting brain. *Neuroimage* 55(1):8–23.
35. Flynn FG, Benson DF, Ardila A (1999) Anatomy of the insula—Functional and clinical correlates. *Aphasiology* 13(1):55–78.
36. Eickhoff SB, et al. (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25(4):1325–1335.
37. Eickhoff SB, et al. (2007) Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *Neuroimage* 36(3):511–521.
38. Amunts K, et al. (2005) Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: Intersubject variability and probability maps. *Anat Embryol (Berl)* 210(5-6):343–352.

Table S1. Exp. 1: Demographics and neuropsychological performance

	OXT group	PLC group			
Demographic	Mean (± SD)	Mean (± SD)	*t*	*P*	η^2
Age (y)	24.28 (2.86)	25.03 (2.56)	−1.15	0.26	<0.02
Years of education	16.49 (2.66)	16.85 (1.58)	−0.68	0.49	<0.01
RAVLT					
Trial 1–5	61.72 (5.69)	61.24 (6.45)	0.33	0.74	<0.01
Trial 6 retention	14.19 (1.14)	13.67 (2.30)	1.22	0.23	<0.02
Trial 7 delayed recall	14.50 (0.94)	13.84 (2.29)	1.58	0.12	<0.04
LPS-4	31.28 (3.58)	31.21 (4.55)	0.07	0.95	<0.01
MWT-A	30.08 (3.15)	31.15 (2.95)	−1.45	0.15	<0.03
d2	191.40 (37.11)	189.96 (34.60)	0.01	0.99	<0.01
TMT-A	26.92 (8.24)	24.65 (6.95)	0.95	0.35	<0.01
TMT-B	66.64 (19.32)	60.38 (20.69)	0.91	0.37	<0.02
Digit-span, forward	8.97 (2.00)	8.81 (2.05)	0.36	0.72	<0.01
Digit-span, backward	8.77 (2.27)	9.06 (2.46)	−0.53	0.60	<0.01
BDI	4.81 (4.70)	3.00 (3.48)	1.80	0.08	<0.05
STAI-X1	42.58 (4.09)	41.29 (5.56)	1.09	0.28	<0.02
STAI-X2	43.72 (3.70)	43.91 (4.56)	−0.19	0.85	<0.01

Verbal declarative memory performance was assessed using a German adaption of the RAVLT and included: learning performance across five trials ("Trial 1–5" maximum possible score 75); susceptibility to interference ("Trial 6 retention" maximum possible score 15); and delayed recall ("Trial 7," maximum possible score 15). Nonverbal reasoning IQ was assessed by the LPS-4 (maximum possible score 40). Verbal IQ based on lexical decisions was assessed by the MWT-A (maximum possible score 37); visual attention and concentration was assesses using the d2; visual attention and task-switching was assessed using the TMT-A and TMT-B (results displayed in seconds); working memory performance was assessed using the digit-span forward and backward tests (maximum possible score 14). Depressive symptoms were assessed by the self-report BDI (Beck's Depression Scale, Version II), and anxiety symptoms by the STAI.

Table S2. Exp. 2: Demographics and neuropsychological performance

	OXT group	PLC group			
Demographics	Mean (± SD)	Mean (± SD)	*t*	*P*	η^2
Age (y)	25.53 (4.16)	26.06 (3.54)	−0.57	0.57	<0.01
Years of education	17.59 (3.00)	18.71 (3.38)	−1.47	0.15	<0.03
RAVLT					<0.01
Trial 1–5	59.68 (10.86)	59.29 (7.42)	0.17	0.86	<0.01
Trial 6 retention	12.76 (2.41)	12.51 (2.14)	0.46	0.65	<0.01
Trial 7 delayed Recall	12.79 (2.79)	12.74 (2.14)	0.09	0.93	<0.01
LPS-4	32.56 (3.91)	31.97 (3.83)	0.64	0.53	<0.01
MWT-A	30.56 (2.31)	30.20 (2.44)	0.63	0.53	<0.01
d2	230.09 (37.78)	221.71 (36.11)	0.94	0.35	<0.01
TMT-A	22.97 (6.35)	24.17 (8.89)	−0.65	0.52	<0.01
TMT-B	57.07 (14.68)	61.11 (13.68)	−1.07	0.29	<0.01
Digit-span, forward	8.91 (1.76)	8.63 (1.63)	0.69	0.49	<0.01
Digit-span, backward	8.59 (2.05)	8.20 (1.71)	0.85	0.40	<0.01
BDI	2.68 (3.87)	3.26 (3.34)	−0.67	0.51	<0.01
STAI-X1	41.29 (3.89)	42.97 (4.76)	−1.06	0.29	<0.01
STAI-X2	40.43 (4.49)	42.20 (4.54)	−1.11	0.27	<0.01

Verbal declarative memory performance was assessed using a German adaption of the RAVLT and included: learning performance across five trials ("Trials 1–5," maximum possible score 75), susceptibility to interference ("Trial 6," maximum possible score 15), and delayed recall ("Trial 7," maximum possible score 15). Nonverbal reasoning IQ was assessed by the LPS-4 (maximum possible score 40). Verbal IQ based on lexical decisions was assessed by the MWT-A (maximum possible score 37); visual attention and concentration was assesses using the d2; visual attention and task-switching was assessed using the TMT-A and TMT-B (results displayed in seconds); working memory performance was assessed using the digit-span forward and backward test (maximum possible score 14). Depressive symptoms were assessed by the self-report BDI and anxiety symptoms by the STAI.

Table S3. Exp. 1: Valence and arousal ratings

Valance and arousal	OXT group Mean (± SD)	PLC group Mean (± SD)	*t*	*P*	η^2
Valence					
Negative	2.65 (1.02)	2.66 (0.91)	−0.07	0.95	<0.01
Neutral	4.96 (0.60)	4.85 (0.80)	−0.70	0.49	<0.01
Positive	6.94 (1.37)	7.06 (1.19)	−0.41	0.69	<0.01
Arousal					
Negative	6.19 (1.33)	6.06 (1.39)	−0.41	0.68	<0.01
Neutral	2.04 (0.84)	1.98 (1.08)	−0.26	0.80	<0.01
Positive	5.74 (1.60)	5.82 (1.75)	−0.18	0.86	<0.01
Valence reaction time(s)					
Negative	3.22 (1.09)	3.46 (1.34)	−0.81	0.42	<0.01
Neutral	3.02 (1.05)	2.74 (1.13)	−1.09	0.28	<0.02
Positive	3.16 (1.26)	3.00 (1.07)	−0.57	0.57	<0.01
Arousal reaction time(s)					
Negative	2.90 (0.84)	2.87 (0.85)	−0.15	0.89	<0.01
Neutral	2.46 (0.88)	2.09 (0.55)	−2.09	0.04	<0.06
Positive	2.76 (1.05)	2.80 (1.04)	−0.16	0.88	<0.01

Table S4. Exp. 2: Performance in the surprise free-recall test 24 h postscan

Items	OXT group Mean (± SD)	PLC group Mean (± SD)	*t*	*P*	η^2
Total items remembered	16.46 (7.37)	16.58 (5.96)	−0.07	0.943	<0.01
Negative items remembered	10.66 (3.61)	6.89 (4.01)	−1.29	0.200	0.02
Neutral items remembered	5.8 (2.93)	9.69 (3.94)	−1.08	0.286	0.02
Edm	4.86 (2.78)	−2.80 (2.99)	−2.98	0.004	0.12

Edm, subsequently remembered negative items minus neutral items.

Table S5. Exp. 2: Brain activation differences between oxytocin and placebo groups

Contrast	Hemisphere	MNI coordinates *x*	*y*	*z*	Size*	*P*
Whole-brain analysis						
OXT > PLC						
Later remembered negative > later forgotten negative						
Anterior insular cortex	Left	−38	22	4	14	0.004
Amygdala ROI analysis						
OXT < PLC						
Negative and neutral > baseline						
Superficial subregion	Right	22	−6	−14	8	0.007
Negative > baseline						
Superficial subregion	Right	22	−6	−12	6	0.018
Neutral > baseline						
Superficial subregion	Right	22	−6	−14	7	0.009

*Cluster extent threshold of K (number of contiguous voxels) ≥ 5.