Neuron
Article

Oxytocin Shapes the Neural Circuitry of Trust and Trust Adaptation in Humans

Thomas Baumgartner,[1,*] Markus Heinrichs,[2] Aline Vonlanthen,[1] Urs Fischbacher,[1] and Ernst Fehr[1,3,*]
[1]Center for the Study of Social and Neural Systems, Institute for Empirical Research in Economics, University of Zurich, Blumlisalpstrasse 10, CH-8006 Zurich, Switzerland
[2]Department of Psychology, Clinical Psychology and Psychobiology, University of Zurich, Binzmuhlestrasse 14/Box 8, CH-8050 Zurich, Switzerland
[3]Collegium Helveticum, Schmelzbergstrasse 25, CH-8092 Zurich, Switzerland
*Correspondence: efehr@iew.uzh.ch (E.F.), t.baumgartner@iew.uzh.ch (T.B.)
DOI 10.1016/j.neuron.2008.04.009

SUMMARY

Trust and betrayal of trust are ubiquitous in human societies. Recent behavioral evidence shows that the neuropeptide oxytocin increases trust among humans, thus offering a unique chance of gaining a deeper understanding of the neural mechanisms underlying trust and the adaptation to breach of trust. We examined the neural circuitry of trusting behavior by combining the intranasal, double-blind, administration of oxytocin with fMRI. We find that subjects in the oxytocin group show no change in their trusting behavior after they learned that their trust had been breached several times while subjects receiving placebo decrease their trust. This difference in trust adaptation is associated with a specific reduction in activation in the amygdala, the midbrain regions, and the dorsal striatum in subjects receiving oxytocin, suggesting that neural systems mediating fear processing (amygdala and midbrain regions) and behavioral adaptations to feedback information (dorsal striatum) modulate oxytocin's effect on trust. These findings may help to develop deeper insights into mental disorders such as social phobia and autism, which are characterized by persistent fear or avoidance of social interactions.

INTRODUCTION

In nonhuman mammals the neuropeptide oxytocin (OT) plays a central role in the ability to form social attachments and affiliations, including parental care, pair bonding, and social memory (Carter, 1998, 2003; Ferguson et al., 2002; Insel and Young, 2001; Lim and Young, 2006; Young and Wang, 2004). In addition, OT shows significant binding in the limbic system, including the amygdala (Huber et al., 2005; Landgraf and Neumann, 2004) and decreases stress responses and anxiety in social interactions (Bale et al., 2001; Neumann et al., 2000; Parker et al., 2005; Uvnas-Moberg, 1998a, 1998b). Initial behavioral experiments indicate that OT also seems to be a potent modulator of social interaction behavior and social cognition in humans (Bartz and Hollander, 2006; Heinrichs and Domes, 2008). OT has recently been shown to influence a person's ability to infer another's mental state, an ability that is referred to as "mind-reading" (Domes et al., 2007b). Moreover, a recent study has shown that OT increases people's willingness to trust others (Kosfeld et al., 2005). Interestingly, OT's effect on trust was not due to a general increase in the readiness to bear risks. Instead, OT specifically affected individuals' willingness to take *social* risks arising through interpersonal interactions. The behavioral impact of OT on trust offers a unique chance to gain a deeper understanding into the neural circuitry of trust and trust adaptation after betrayal of trust by combining behavioral experiments with pharmacological intervention and neuroimaging methods. To date, however, no study on the effects of this peptide on the neural circuitry associated with human trusting behavior is available. Moreover, it is not known how OT affects trust after subjects experienced that their trust had been betrayed, i.e., we do not know whether subjects receiving OT respond to this betrayal with a decrease in trust or whether they maintain their trusting behavior. We thus examined the effects of intranasally administered OT on both brain activity and individuals' decisions in a trust and a risk game with real monetary stakes after subjects received feedback indicating that their trust had been betrayed or that their risky investment resulted in no payback in about half the cases.

Our work is based on the combination of neuroscientific tools with economic experiments which recently gained momentum through the neuroeconomics research agenda (Camerer et al., 2005; Cohen and Blum, 2002; De Quervain et al., 2004; Delgado et al., 2004; Fehr and Camerer, 2007; Glimcher, 2002; Glimcher and Rustichini, 2004; Hsu et al., 2005; King-Casas et al., 2005; Knoch et al., 2006; Knutson et al., 2007; Kuhnen and Knutson, 2005; Montague and Berns, 2002; Sanfey et al., 2003; Spitzer et al., 2007). We apply, in particular, a suitably modified version of the trust game (Berg et al., 1995; Camerer and Weigelt, 1988; Fehr et al., 1993) to address our research questions. In a trust game (Figure 1), two subjects interacting anonymously are in the role of an investor (who is in the scanner) and a trustee. The investor first has the chance of choosing a costly trusting action by giving money to the trustee. If the investor transfers money, the total amount available for distribution between the two players increases because the experimenter triples the investor's transfer. Initially, however, the trustee reaps the whole

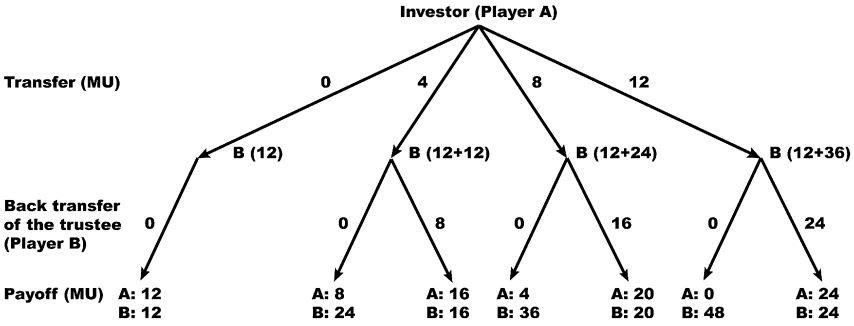

Figure 1. The Trust Game

At the beginning of each new trust period, investor and trustee receive an initial endowment of 12 money units (MUs). The investor then can decide to keep all MUs or to send 4, 8, or 12 MUs to the trustee. The experimenter triples the transferred money. The trustee then has the option of keeping the whole amount he received or sending back a payoff equalizing amount of money. For example, if the investor sends 8 MUs, the trustee receives 24 MUs, giving him in total 36 MUs (12 MUs own endowment + 24 MUs tripled transfer) while the investor has only 4 MUs at this stage of the game. Then the trustee can chose a back transfer of zero or a back transfer of 16 MUs. The experimenter does not triple the back transfer. Thus, if the trustee chooses a back transfer of zero MUs, he earns 36 MUs in the current period, while the investor only earns 4 MUs. If the trustee, however, chooses a back transfer of 16 MUs, both players end up with the same total amount of 20 MUs. In the risk game the investor faced the same investment opportunities as in the trust game, i.e., he could invest 0, 4, 8, or 12 MUs, and for every positive investment the computer chooses a zero investment return or a return equal to that which could be achieved in the trust game. The investment returns were drawn randomly from the probability distribution generated by the trustees' behavior in the trust game. Thus, investors in the trust and the risk game faced the same objective risk, but no social betrayal could occur in the risk game because no trustees were involved in the back transfers.

increase. Then the trustee is informed about the investor's transfer and can then honor the investor's trust by sending back money ("sharing"), so that both subjects earn the same amount of money. Thus, if the investor gives money to the trustee and the latter shares, both players end up with a higher and equal monetary payoff. However, the trustee also has the option of violating the investor's trust by not sharing the money. In this case, the investor loses all the money he sent to the trustee, an event that investors typically interpret as a breach or betrayal of trust (Bohnet and Zeckhauser, 2004). Since sharing the money is costly for the trustee, a selfish trustee will never honor the investor's trust because the investor and the trustee interact only once in the experiment.

In the risk (lottery) game, the investor faces the same choices and exactly the same probabilistic risk as in the trust game, but a random computer mechanism implements the payback and no interaction with a trustee takes place. Thus, the only difference between the two games is that the investor's risk in the trust game arises from the uncertainty regarding the trustee's behavior—that is, a social interaction with a specific trustee constitutes the risk—whereas a nonsocial random mechanism determines the investor's risk in the lottery game. The risk game constitutes an important control condition because economic theories (Falk and Fischbacher, 2006; Fehr and Schmidt, 1999; Rabin, 1993) and previous empirical research (Bohnet and Zeckhauser, 2004) has shown that many people have an aversion against being betrayed when they trust another person (Bohnet and Zeckhauser, 2004), but betrayal aversion cannot play a role in situations involving nonsocial risks. In addition, OT has been shown to increase trust but not nonsocial risk taking, suggesting the conjecture that OT reduces the special fears that are associated with social betrayal (Kosfeld et al., 2005). Therefore, OT may affect brain activity in these two games differently.

An investor either received OT or placebo. We had a total of 49 investors, each of whom played 12 risk periods (games) and 12 trust periods (games) that took place in a random order. The investor faced a new trustee in every trust period. Figure 2 depicts a timeline for one period of a trust and a lottery game for subjects in the scanner. After the first 12 periods, in which six risk games and six trust games took place, the investors received feedback that informed them how often their investment was successful in the risk game and how often the trustee paid back money in the trust game. Thus, the investors received feedback only once—at the end of the first 12 periods. The feedback information told the investors that their investment led to a return or their trust was repaid only in about 50% of the cases (see Figure 2 and Experimental Procedures section for detailed description of feedback administration). In the following, we refer to the first six risk and six trust games as the *prefeedback* phase, while the periods after the feedback are referred to as the *postfeedback* phase. Taken together our experimental design creates thus three main variables: group (OT, placebo), phase (prefeedback, postfeedback), and game (trust, risk).

Previous findings from neuroimaging and lesion studies led us to hypothesize that subcortical brain structures such as the amygdala and brainstem effector sites—that process fear, danger, and perhaps also risk of social betrayal—are involved in trusting behaviors. The amygdala has been shown to exhibit increased activation in social avoidance and phobia (Stein et al., 2002; Tillfors et al., 2001) and while viewing untrustworthy faces (Winston et al., 2002). Decreased amygdala activation has also been linked to genetic hypersociability (Meyer-Lindenberg et al., 2005), and lesion studies have indicated that patients with bilateral amygdala damage are impaired in judging the trustworthiness of other people's faces. These patients all judged other people to look more trustworthy and more approachable than did normal viewers (Adolphs et al., 1998). Finally, during the processing of fearful stimuli, subjects receiving OT have reduced amygdala activation and reduced connectivity of the amygdala with brainstem regions involved in automatic fear reactivity (Domes et al., 2007a; Kirsch et al., 2005). This finding is in agreement with a recent animal study that demonstrates in vitro that OT acts on the central amygdala by inhibiting excitatory information from the amygdala to brainstem sites mediating the autonomic fear response (Huber et al., 2005). Given that the amygdala is crucially involved in the processing of risks

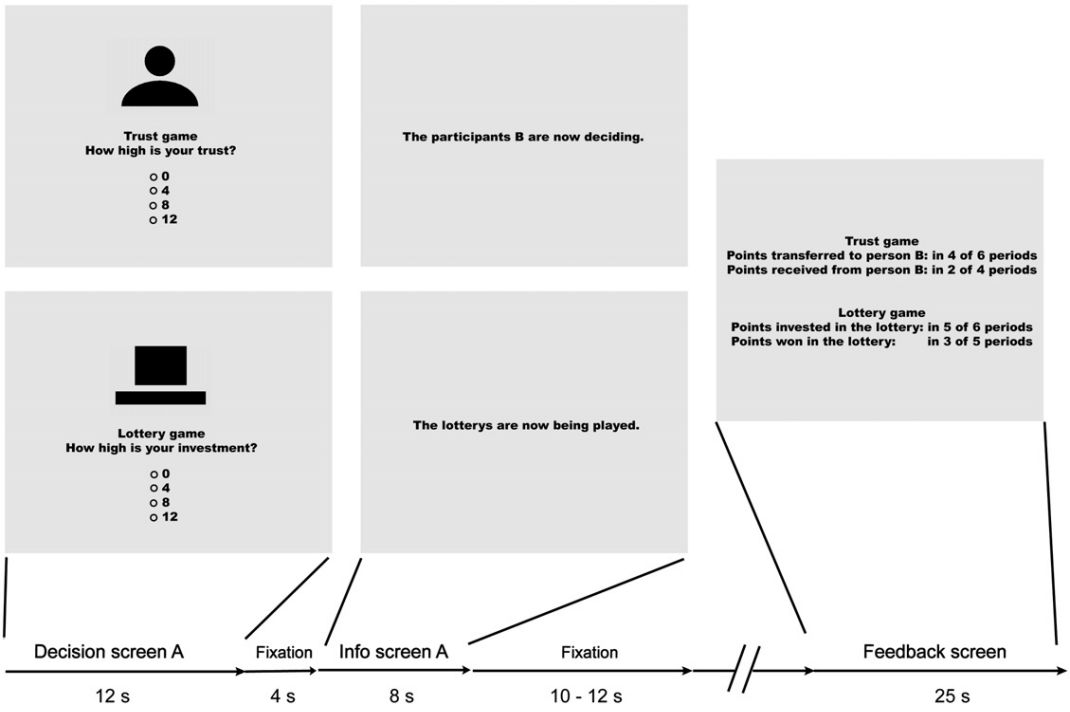

Figure 2. Timeline for Trust and Lottery Periods

Timeline for one period of the trust (top) and lottery game (bottom) for the investor (Player A) whose brain was observed in the scanner. A total of 12 trust and 12 lottery periods were played. At the beginning of each new period, a decision screen indicated whether a trust or lottery period will be played. The "trust screen" depicted a schematic picture of a human being while the "lottery screen" depicted a schematic picture of a computer. After 4 s, four buttons representing the four response options appeared on the screen, indicating that the subject now had 8 s to implement a decision. After the subjects made their choice, a fixation cross appeared for 4 s, after which a screen indicated a waiting epoch of 8 s. During that time, the subjects in the scanner were informed that the trustees (Player B) are now deciding or that the random mechanism determines the returns from the lottery. Finally, periods were separated by a screen depicting a fixation cross with a variable duration of 10–12 s. Importantly, after half of the played trust and lottery periods, a feedback screen appeared for 25 s, consisting of meager feedback indicating that only in about 50% of the cases the trustees shared the money or the lottery did yield a return, respectively.

arising in social situations, we hypothesized that oxytocin might affect the amygdala response to these social risks, thereby facilitating prosocial approach behavior—such as trust.

Other relevant evidence comes from neuroimaging studies using economic experiments involving social interaction paradigms (Delgado et al., 2005a; King-Casas et al., 2005; Rilling et al., 2002). These studies suggest the striatum could play a role in our experiment; it is thought to be part of a neural circuit that guides and adjusts future behavior on the basis of reward feedback (O'Doherty et al., 2004; Tricomi et al., 2004). Moreover, a study of reward-related (nonsocial) probability learning has shown that activation in this region in response to reward feedback diminishes, as cues learned through trial and error begin to predict how actions and outcomes are related, thus making feedback less informative (Delgado et al., 2005b). This finding has been extended to a repeated-interaction trust game in which participants faced the same opponent and gradually learned whether their partner is trustworthy through experience. Over time, as the partner's response became more predictable, the activity in the dorsal striatum decreased in this game (King-Casas et al., 2005). Trial and error is not the sole method for learning predictability, however. It has been recently shown that the perceptions of moral character alone suffice to modulate the dorsal striatum during both the decision and the outcome phases of a trust game (Delgado et al., 2005a). Participants made risky choices about whether to trust hypothetical trading partners after having read vivid descriptions of life events indicating praiseworthy, neutral, or suspicious moral character. Activations in the striatum differentiated between positive and negative feedback as well as between no-trust and trust decisions, but only for the "neutral" partner, whereas no differential activity was observed for the "good" partner despite the fact that (by experimental design) neutral and good partners responded in the same way to trusting decisions. This finding suggests that prior social and moral information can diminish reliance on brain structures such as the dorsal striatum that are important for behavioral adaptations to feedback information. These brain structures may also be recruited in our experiment as the subjects may feel the need to adjust their behaviors following meager information feedback. Therefore, if OT indeed diminishes the behavioral adaptation to this feedback, such an effect might be modulated by a diminished reliance on brain structures involved in behavioral adaptation.

Regarding the two different phases of the study (pre- and postfeedback phase), it is important to note that subjects in the prefeedback phase have an incentive to explore different

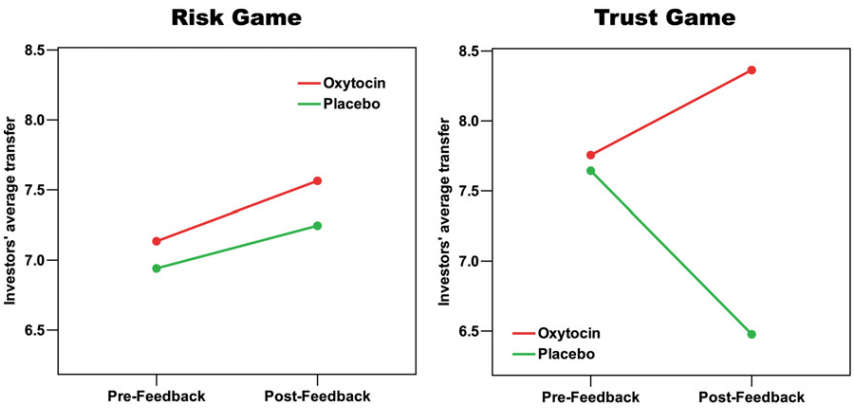

Figure 3. Investors' Average Transfer in the Risk and Trust Game across the Prefeedback and the Postfeedback Phases

Each data point represents the average over six decision periods (please see Table S1 and Figure S1 for SEM). The figure shows the interaction effect in the trust game (group × phase: $F_{(1,40)} = 4.06$, $p = 0.05$, $ETA^2 = 0.11$), indicating a differential adaptation in trusting behavior after the meager feedback in the OT (red color) compared to the placebo group (green color). While subjects receiving OT demonstrate no significant change in their trusting behavior, subjects receiving the placebo decrease theirs. In contrast, subjects receiving OT and the placebo respond in the same way to the feedback in the risk game by keeping their investments almost constant.

strategies in order to maximize the informativeness of the feedback. In order to maximize learning about the trustworthiness of the trustees' subject pool, for example, the investors may have an incentive to make more trusting decisions in the prefeedback phase because the gleaned knowledge about the trustees' trustworthiness can be valuable for behavioral adaptation after the feedback. An additional motive for trusting choices in the prefeedback phase thus exists that is absent in the postfeedback phase. Note too that in order to enable a clean comparison between the two games in the postfeedback phase, subjects received the same feedback in the risk and the trust game because the probability distribution of investment successes in the risk game replicated the probability distribution of the trustees' responses in the trust game (see Experimental Procedures section for details).

RESULTS

Behavioral Results

Our main behavioral measure consists of each individual's average investment in the risk and the trust game during the prefeedback and the postfeedback phase, yielding four observations per individual. We performed a two-way repeated-measures ANOVA based on these observations for the risk and the trust game (controlling for potential personality differences in general trust [M. Siegrist, C. Keller, T.C. Earle, and M. Hanselmann, personal communication], general risk seeking propensity [Zuckerman and Link, 1968], and feedback information; see Experimental Procedures section for details). This analysis reveals a significant interaction effect in the trust game (group × phase, $F_{(1,40)} = 4.06$, $p = 0.05$, $ETA^2 = 0.11$), which is absent in the risk game (group × phase, $F_{(1,40)} = 0.04$, $p = 0.85$, $ETA^2 = 0.001$). Subjects receiving the placebo decreased their trusting behavior after they were informed that their interaction partner did not pay back in about 50% of the cases, whereas subjects receiving OT demonstrated no change in their trusting behavior in the postfeedback phase (see Figure 3 and see Table S1 available online) despite having received the same information. In the risk game, however, both groups showed no behavioral adaptation to the feedback information. Thus, it seems that OT only affects the behavioral adaptation to the feedback information if social risks are involved, but not if nonsocial risks are involved.

The specific impact of OT on behavioral adaptation in the trust but not in the risk game is complemented by similarly specific response time differences between the OT and placebo group. Individuals in the OT group exhibit considerably smaller response times than those in the placebo group ($t = -2.77$, $p < 0.01$) during the postfeedback phase, whereas no significant differences ($t = 0.51$, $p = 0.61$) are present during the prefeedback phase (see Table S2). This difference in the postfeedback phase is, in particular, due to the significant decline in responses times in the OT group in the postfeedback phase compared to the prefeedback phase ($t = 2.72$, $p < 0.05$). The response time effect of OT in the trust game contrasts with the risk game, in which we observe no significant differences between the two groups in the prefeedback and in the postfeedback phase.

In order to control for nonspecific effects that might be associated with OT administration, we explicitly measured mood, calmness, and wakefulness before substance administration and 10 min after the end of the scanning session (after subjects had played both the risk and trust game). We observed no significant group differences (independent t tests, all $p > 0.26$; see Table S3) neither before the scanning session nor afterwards. Finally, we asked subjects at the end of the experiment whether they believed they had received OT or the placebo. Thirty-four percent of subjects in the placebo group and 30% of subjects in the OT group reported the impression they had received OT. A correlation between this belief question and the effective administration of OT or the placebo showed no significant correlation (Spearman correlation, $r = -0.107$, $p = 0.470$), thus clearly indicating that neither subjects in the OT nor in the placebo group recognized whether they had received OT or a placebo.

Neuroimaging Results

We conducted analyses of functional magnetic resonance imaging (fMRI) data for the decision phase of the trust and the risk game. A random-effects general linear model (GLM) analysis was conducted using each condition (trust and risk game) and period (six prefeedback periods and six postfeedback periods) as predictors. We generated statistical maps contrasting the OT and the placebo groups using serial subtraction terms, separately for prefeedback and postfeedback periods. The serial subtraction term consisted either of *trust* > *risk* contrasts or *risk* > *trust* contrasts and was exclusively masked at $p < 0.05$

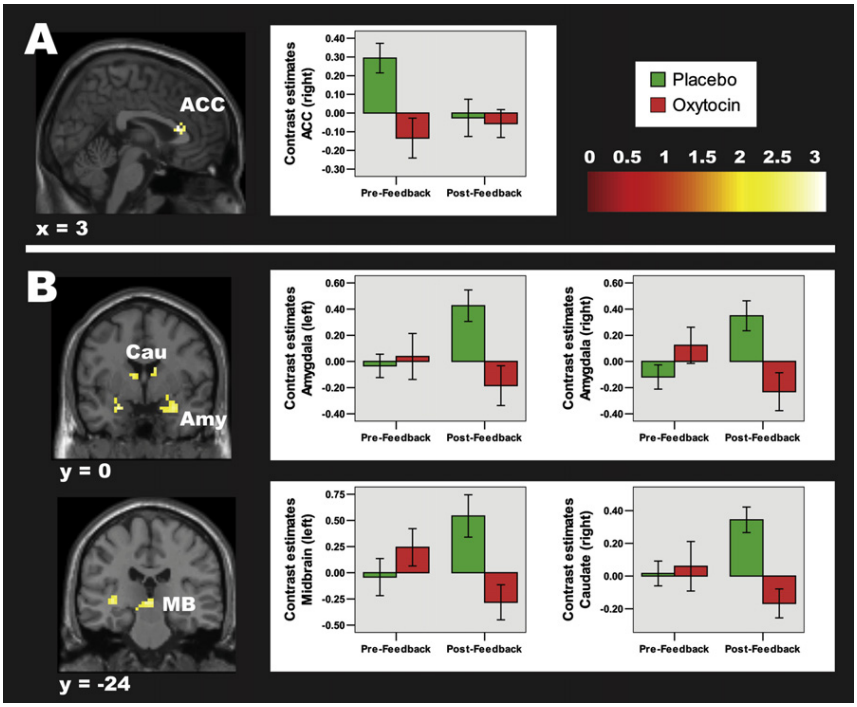

Figure 4. Brain Regions Showing Stronger Activation in the Placebo Compared to the Oxytocin Group

Depicted on sagittal or coronal slices is the increased activation in the placebo compared to the oxytocin group in trust periods played (A) prefeedback (including ACC) and (B) postfeedback (including bilateral amygdala, bilateral caudate nucleus and midbrain brain regions). All regions are significant at $p < 0.005$ with a cluster extent of ten voxels. However, for display purposes all regions are depicted at $p < 0.01$. Bar plots represent differences in contrast estimates (trust > risk) of functional ROI's (see Experimental Procedures section for details) for bilateral amygdala, right caudatus, midbrain regions and ACC, broken down for the oxytocin (in red color) and placebo group (in green color) as well as time phase (prefeedback/postfeedback). Univariate and repeated-measures ANOVAs calculated with and without control for potentially confounding variables (including general trust, sensation seeking, first feedback) confirmed for each depicted brain regions the interaction effect of *group* × *phase*, qualified by stronger activation in these brain regions in the placebo compared to the oxytocin group either *only* in prefeedback or *only* in postfeedback periods.

with the reversed second contrast of the serial subtraction term (see Experimental Procedures section for explanation). For example, one important statistical map in the postfeedback phase concerns $(\text{Trust} > \text{Risk})^{P} > (\text{Trust} > \text{Risk})^{OT}$, exclusively masked with the reverse second contrast $(\text{Risk} > \text{Trust})^{OT}$. Here, OT denotes oxytocin and P indicates placebo. The results of this map reveal the brain activations that are specific to trust taking (relative to risk taking) in the placebo (P) group, i.e., the extent to which OT reduces brain activations when individuals make choices in the trust game. Significant results are reported at $p < 0.005$ (uncorrected) with a cluster threshold of ten voxels. In case of significant unilateral activations in our main regions of interests, including the amygdala and striatum, we lowered the significance threshold to $p < 0.01$ (uncorrected) with the same voxel extent to verify whether a bilateral activation pattern could be found at this threshold.

Prefeedback Periods

A few differences were observed between the OT and placebo group during the decision phase of prefeedback periods, both in the trust and risk game. The OT group showed stronger activation of the thalamus (x = 15, y = −27, z = 6), whereas the placebo group demonstrated increased activation in the dorsal part of the ACC (x = 6, y = 24, z = 12; see Figure 4A) during prefeedback trust game periods (see Table 1). During the prefeedback risk game, increased activation in the OT group was found in the inferior temporal gyrus (x = 39, y = −54, z = −15), whereas the placebo group showed a relative increase of activation in parietal brain regions (precuneus/posterior cingulate, x = −15, y = −60, z = 21; superior parietal gyrus, x = 30, y = −60, z = 48). We will discuss these prefeedback activations in the Risk > Trust contrast in the Supplemental Data (Supplemental Discussion) because in the main text we are mainly interested in the Trust > Risk contrasts.

Postfeedback Periods

As hypothesized, the placebo group showed stronger activation in postfeedback trust game periods in the bilateral amygdala (x = 30, y = 3, z = −18; x = −24, y = 0, z = −21), bilateral caudatus (x = 12, y = 6, z = 9; x = −9, y = 0, z = 12), midbrain regions (x = −3, y = −24, z = −3), as well as arousal related structures such as the posterior insula (x = −33, y = −21, z = 0) and postcentral gyrus (x = −27, y = −54, z = 69; see Table 2 and Figure 4B). In contrast, not a single brain region showed group differences during the postfeedback risk game. Moreover, no brain region showed an activation increase in the OT compared to the placebo group during the postfeedback trust game.

ROI Analyses for Prefeedback and Postfeedback Periods

For all brain regions showing differing group activation either prefeedback or postfeedback (see Tables 1 and 2), we created combined functional and structural (based on the anatomic atlas by Tzourio-Mazoyer et al. [2002]) ROIs and calculated repeated-measures ANOVAs using participants' beta weights (for details see Experimental Procedures section). These analyses allow a deeper examination of (1) the lateralization pattern in the amygdala and the caudatus, (2) the effect of the prefeedback and the postfeedback phase on activation patterns in all brain regions, as well as (3) the influence on brain activation of the potentially confounding variables used in the behavioral analysis (including general trust [M. Siegrist, C. Keller, T.C. Earle, and M. Hanselmann, personal communication], sensation seeking [Zuckerman and Link, 1968], and feedback information; see Experimental

Table 1. Brain Activation Differences between Oxytocin and Placebo in Prefeedback Game Periods

Condition	Contrasts	Brain Regions	BA	Side	x	y	z	Max t Score		Voxels
Risk Game	Oxytocin > placebo $(R\text{-}T)^{OT} - (R\text{-}T)^{P}$	Temporal Lobe								
		Inferior temporal gyrus	37	R	39	−54	−15	3.16	*	14
	Placebo > oxytocin $(R\text{-}T)^{P} - (R\text{-}T)^{OT}$	Parietal Lobe								
		Precuneus/posterior cingulum	31	L	−15	−60	21	3.09	*	11
		Superior parietal gyrus	7	R	30	−60	48	3.08	*	21
Trust Game	Oxytocin > placebo $(T\text{-}R)^{OT} - (T\text{-}R)^{P}$	Subcortical Structures								
		Thalamus, pulvinar		R	15	−27	6	3.53	***	13
	Placebo > oxytocin $(T\text{-}R)^{P} - (T\text{-}R)^{OT}$	Frontal Lobe								
		ACC	24	R	6	24	12	3.18	**	20
		° *ACC*	24	R	15	33	9	2.83	*	

The coordinates of activated brain regions are given according to the MNI space together with the t scores and significance thresholds (*$p < 0.005$, **$p < 0.001$, ***$p < 0.0005$, uncorrected for multiple comparisons). R denotes risk game, T denotes trust game, OT denotes oxytocin and P denotes placebo. Minimum cluster size ten voxels. All observed maxima are reported. ° indicates a subpeak in the same cluster of voxels. All serial contrasts are masked exclusively at $p < 0.05$ using the reversed second contrast of the serial subtraction term. Regions of interest discussed in the paper are italic.

Procedures section). These ANOVAs confirm the previously reported results in Tables 1 and 2. In particular, we find a robust interaction effect between *Group* and *Time Phase* in bilateral amygdala and the right caudatus indicating that subjects in the placebo group only show higher activations in these brain regions in the postfeedback phase of the trust game, where we observe the differences in behavioral adaptation across the OT and the placebo group (see Tables S5 and S6 for detailed information). This interaction between *Group* and *Time Phase* is also present if we control for potentially confounding variables, such as personality differences in general trust, sensation seeking, or mood; if we do not control for these differences then the interaction effect is also obtained in the *left* Caudatus. Finally, we calculated univariate ANOVAs to control for response time differences between the OT and placebo group observed in postfeedback periods. There was no effect of response times on the brain activation patterns in these analyses (see Table S7).

DISCUSSION

This is the first study showing how OT affects humans' behavioral adaptation to meager feedback information about the success of previous trust and risk taking. Our results indicate that intranasally administered OT indeed affects these behavioral adaptations in a specific way. If subjects face the nonsocial risks in the risk game, OT does not affect their behavioral responses to the feedback. Both subjects in the OT group and the placebo group do not change their willingness to take risks after the feedback. In contrast, if subjects face social risks, such as in the trust game, those who received placebo respond to the feedback with a decrease in trusting behavior while subjects with OT demonstrate no change in their trusting behavior although they were informed that their interaction partners did not honor their trust in roughly 50% of the cases.

These behavioral findings are consistent with the hypothesis that betrayal aversion is operative in the trust game (Bohnet and Zeckhauser, 2004) and that OT contributes to a reduction in the fear of social betrayal (Kosfeld et al., 2005). Subjects in the risk game need not fear that another individual will breach their trust because they only face the probabilistic risk arising from a preprogrammed computer, i.e., betrayal aversion could play no role in the risk game. In contrast, subjects face a human partner in the trust game who has the option of abusing their trust. The response time differences between the OT and the placebo group during the postfeedback phase also support this interpretation. Subjects in the OT group need significantly less time to make a trusting decision, consistent with the view that it is easier for them to overcome the trust-inhibiting force of betrayal aversion.

Our findings also conform with both animal and human studies showing that OT ameliorates the symptoms associated with social anxiety and stress (Heinrichs et al., 2003; Heinrichs and Domes, 2008; Insel and Young, 2001). We suggest that future studies could systematically manipulate feedback information to examine whether OT reduces reliance on feedback mechanism and thus increases trusting behavior regardless of how negative the feedback information is or whether extremely negative feedback information (e.g., reinforcement rates below 20 percent) will diminish or even abolish OT's effect on behavioral adaptation in situations requiring trust.

The finding that OT had no behavioral impact on trusting behavior in prefeedback periods might seem surprising at first. However, we already pointed out that subjects faced additional incentives to transfer money to the trustees in the prefeedback periods because they knew they would receive feedback. The only way to learn about the degree of trustworthiness in the population of trustees is to transfer money to them. If OT indeed reduces betrayal aversion, the incentive to explore the trustees' trustworthiness must obviously be weaker in the OT group than in the placebo group because subjects with OT are less afraid of betrayal. Thus, the placebo group has a stronger reason for extracting information from the feedback, implying that they also transfer more money relative to their natural inclination to trust. If placebo subjects experience this conflict between their natural inclination to trust and the incentive to explore their partners' trustworthiness, the brain should then represent this conflict. In

Table 2. Brain Activation Differences between Oxytocin and Placebo in Postfeedback Game Periods

Condition	Contrast	Brain Regions	BA	Side	x	y	z	Max t Score		Voxels
Risk game	Oxytocin > placebo $(R\text{-}T)^{OT} - (R\text{-}T)^{P}$	*No suprathreshold clusters*								
	Placebo > oxytocin $(R\text{-}T)^{P} - (R\text{-}T)^{OT}$	*No suprathreshold clusters*								
Trust game	Oxytocin > placebo $(T\text{-}R)^{OT} - (T\text{-}R)^{P}$	*No suprathreshold clusters*								
	Placebo > oxytocin $(T\text{-}R)^{P} - (T\text{-}R)^{OT}$	Temporal Lobe								
		Amygdala		R	30	3	−18	3.06	*	12
		Amygdala		L	−24	0	−21	3.12	*	14
		° *Amygdala*		L	−27	0	−12	2.48	°°	
		Parietal Lobe								
		Postcentral Gyrus	5	L	−27	−54	69	2.96	*	10
		Subcortical Structures								
		Putamen/insula		L	−33	−21	0	3.15	*	11
		Caudate body		R	12	6	9	3.08	*	16
		° *Caudate body*		R	12	3	18	2.88	*	
		° *Caudate head*		R	9	9	0	2.87	*	
		Caudate body		L	−9	0	12	2.76	°°	10
		Brainstem, midbrain, red nucleus		L	−3	−24	−3	3.03	*	11

The coordinates are given according to the MNI space together with its t scores and significant thresholds (°°$p < 0.01$, *$p < 0.005$, **$p < 0.001$, ***$p < 0.0005$ [all uncorrected for multiple comparisons]). R denotes risk game, T denotes trust game, OT denotes oxytocin, and P denotes placebo. Minimum cluster size 10 voxels. ° indicates a subpeak in the same cluster of voxels. All observed maxima are reported. All serial contrasts are masked exclusively at $p < 0.05$ using the reversed second contrast of the serial subtraction term. Regions of interests discussed in the paper are italic.

this context, it is therefore interesting to observe that the placebo subjects in the trust game exhibit higher activation in the dorsal anterior cingulate cortex (ACC), a brain region frequently implicated in conflict monitoring and cognitive control in social (Delgado et al., 2005a; Sanfey et al., 2003) and nonsocial paradigms (Botvinick et al., 1999; Carter et al., 1998). In fact, the dorsal ACC is the only brain region showing increased activation in the placebo compared to the OT group during the prefeedback periods of the trust game which strengthens our interpretation of the behavioral finding.

The brain activations we find in the postfeedback phase (Table 2) are also very informative with regard to the neural networks involved in the reduced behavioral adaptation to the meager feedback in the trust game and its absence in the risk game. There are no differences in brain activation between the OT and the placebo group in the risk game, where we observe no behavioral differences. In contrast, as hypothesized, we find differences between the placebo and the OT group in the amygdala, the midbrain, as well as the striatum in postfeedback trust periods, i.e., exactly in those periods in which we also observe differences in behavioral adaptation to the feedback information. More precisely, the bilateral amygdala and functionally connected brainstem effector sites showed significantly increased activation in the placebo compared to the OT group in postfeedback trust periods. A vast area of research in animal, human lesion, and neuroimaging studies points to the critical role of these brain areas in signaling and modulating fear responses (Adolphs et al., 2005; Amaral, 2003). Moreover, both human neuroimaging (Domes et al., 2007a; Kirsch et al., 2005) and animal studies (Huber et al., 2005) have shown that the neuropeptide OT decreases fear responses by modulating activation in the amygdala and brainstem effector sites. Finally, it has been reported that the amygdala shows increased activation during viewing faces of people that look untrustworthy (Winston et al., 2002) and that patients with bilateral amygdala damage judged other people to look more trustworthy and more approachable than did normal viewers or other patients with brain damage in other areas (Adolphs et al., 1998). Taken together, these findings are consistent with the view that OT reduces fear responses during the trust game by reducing activation in the amygdala and connected brainstem effector sites, which in turn enhances subjects' ability to trust in situations characterized by the risk of betrayal.

Animal studies indicate that increased availability of OT in the central nervous system facilitates approach behavior (affiliation and social attachment) by modulating brain circuits such as the nucleus accumbens (part of the striatum) and ventral pallidum (Insel and Young, 2001; Young et al., 2001) that are implicated in reward processing. In the light of these findings, it is interesting that in our experiment, OT reduces activations in a closely related striatal area—the caudate nucleus—in the trust game. Neuroimaging studies (O'Doherty et al., 2004; Tricomi et al., 2004) have shown that the caudate is critically involved in feedback processing and reward learning associated with behavioral adaptations to information about the action-outcome contingencies, and several studies document (Delgado et al., 2005b; King-Casas et al., 2005) that caudate activation is

reduced as learning progresses and rewards can be more reliably predicted. Thus, once subjects in these studies have learned the contingency between their actions and the associated outcomes, the caudate is less active. Moreover, a recent neuroimaging study (Delgado et al., 2005a) of investor behavior in the trust game has shown that the mere perception of a morally "good" character and to a lesser extent of a morally "bad" character (implemented by the attribution of vivid descriptions of life events—indicating morally good, neutral, or bad behaviors—to the faces of three different partners) already diminishes activations in caudate nucleus during the initial stages of the trust game. In contrast, the caudate is strongly activated during decision and outcome phases of the trust game when participants faced a morally "neutral" partner. Despite equivalent reinforcement rates (50%) for all three partners, subjects were more likely to trust the "good" partner, even during the later periods of the trust game. Thus, when playing with the "good" partner, subjects behave as if they "know" that the "good" partner is more trustworthy, i.e., they exhibit less behavioral adaptation to the same feedback information, and this lack of reliance on the feedback information is associated with less caudate activation.

It is interesting that OT administration and the explicit knowledge of facing a morally "good" partner both generate similar behavioral patterns and neural responses in the caudate nucleus. Recall, we also find less behavioral adaptation to feedback information in subjects with OT and—like in the other study (Delgado et al., 2005a)—this behavioral pattern is associated with diminished caudate activation. Thus, subjects with OT behave as if they implicitly "know" that they can trust their partners, and this may be the reason why the brain structure that is critical for learning the contingency between actions and outcomes—the caudate nucleus—shows diminished activation.

It is also important to note that the effect of OT on trust occurred without subjective awareness because subjects were completely unaware whether they received OT. There was also no difference in questionnaire measures of mood, calmness, and wakefulness between the placebo and the OT subjects. And finally, the differences in brain activation between placebo and OT subjects were only observed in subcortical structures as the amygdala, the midbrain, and the striatum. Those brain structures have each been associated with automatic and intuitive (Bechara et al., 1997; McClure et al., 2004) or even unconscious processes (Pasley et al., 2004). Various studies show that amygdala activation is also seen when fearful facial expressions of emotion are presented briefly and masked to prevent conscious perception (Whalen et al., 1998) or when presented in a cortically blind field (Morris et al., 2001; Pegna et al., 2005). Thus, taken together, these findings suggest that OT exerts its effect automatically or even unconsciously in subcortical brain structures which can be modulated without explicit awareness of the subjects.

Human societies are probably unique in the extent to which trust characterizes interpersonal interactions. Trust is indispensable in friendship, love, families, and organizations, and it is a lubricant of economic, political, and social exchange. However, whenever we trust there is also the possibility of trust betrayal. Despite the fact that most humans have experienced instances of breach of trust, they still remain capable of trusting others. In this study, we examined the neural circuitry of trust after breach of trust by studying how OT affects the behaviors and the brain networks of subjects whose trust has been broken in about 50% of the cases. OT significantly reduces subjects' behavioral adaptation to such meager feedback information. This behavioral effect is accompanied by a reduced activity in brain areas known to be involved in fear processing (amygdala, midbrain) and behavioral adaptation (caudate nucleus) in situations with unknown action-outcome contingencies. These effects of OT on brain activations are highly specific in the sense that they only occur when OT affects behavior. They are absent in the risk game, where OT does not affect the behavioral adaptation to the same feedback information.

Finally, our insights into the neural circuitry of trust adaptation, and oxytocin's role in trust adaptation, may also contribute to a deeper understanding of mental disorders such as social phobia or autism that are associated with social deficits. In particular, social phobia (which is the third most common mental health disorder) is characterized by persistent fear and avoidance of social interactions. We hope that our results will lead to further fertile research on such health disorders and the potential role of possible dysfunctions in neuroendocrine mechanisms such as the oxytocinergic system. Further progress in this area also requires more detailed knowledge about the mechanism of brain penetration of OT following different methods of administration and the relationship between plasma and central OT, including possible crosstalks of the neuropeptide at other central receptors (Heinrichs and Domes, 2008).

EXPERIMENTAL PROCEDURES

Subjects

A total of 49 healthy male students (mean age ± SD, 21.7 ± 2.5) from different universities in Zurich participated in the study. One reason for taking only one sex is that OT may strongly vary across male and female subjects, which introduces an additional source of noise. Subjects with chronic diseases, mental disorders, medication, or those who smoked or abused drugs or alcohol were excluded from the study. Participants abstained from food and drink (other than water) for 2 hr before the experiment, and from exercise, caffeine, and alcohol during the 24 hr before the session. In addition, we administered the brief symptom inventory scales (BSI), brief psychological self-reports that measure psychological symptoms; none of the subjects was in the clinical range, and there were no significant differences between the placebo and OT group in either scale (see Table S4). Participants were informed at the time of recruitment that the experiment evaluates the effects of a hormone on decision making. The study was carried out in accordance with the Declaration of Helsinki principles and approved by the institutional ethics committee. All subjects gave written, informed consent and were informed of their right to discontinue participation at any time.

Substance Administration

Subjects were randomly assigned to the OT or placebo group (double-blind, placebo-controlled study design). Recent research has shown that neuropeptides, such as vasopressin, gain access to the human brain after intranasal administration (Born et al., 2002), providing a useful method for studying the central nervous system effects of the neuropeptide oxytocin in humans (Heinrichs and Gaab, 2007). As oxytocin and vasopressin are closely related structurally, differing in only two amino acids, a pharmacokinetically similar mechanism regarding the pathway to the brain has been assumed for both peptides (Bartz and Hollander, 2006; Heinrichs and Domes, 2008). Subjects received a single dose of 24 IU OT (Syntocinon-Spray, Novartis; three puffs per nostril, each with 4 IU OT) intranasally or a placebo 50 min before the start of the trust

and risk experiment. In order to avoid any subjective substance effects other than those caused by OT (e.g., olfactory effects), the placebo contained all inactive ingredients except for the neuropeptide. Intranasal OT is widely prescribed for lactating women and has been used in several experimental studies in humans with no adverse side effects being reported (Heinrichs et al., 2003; Heinrichs et al., 2004). Because of potential diurnal variations in endogenous hormone secretion, we restricted the time of exogenous OT administration to 2 p.m. to 4 p.m.

Design

In total, subjects played 12 rounds of a trust game against 12 different and anonymous human interaction partners and 12 rounds of a risk game in which a random mechanism determined the outcome of the game. The parameters of the experiment were determined with the help of behavioral pilot experiments. We wanted to ensure, in particular, that the trustees breach the investors' trust in the trust game in about 50% of the cases, which we did by using the trustees' choices from the pilot experiment as an input (i.e., as responses from the trustees) for the scanner experiment. Moreover, the computerized responses to the investment decisions in the risk game were drawn from a distribution that perfectly mimics the distribution of the trustees' choices in the pilot experiment. In this way, we ensured that the investors in the risk and the trust experiment received the same feedback. Trust and risk periods were presented counterbalanced and pseudorandomized. Rudimentary feedback information consisting of a reinforcement rate of 50% was revealed to the subjects in the scanner after half of the played trust and risk periods. In this feedback, subjects received the following information separately for the risk and trust game. First, they were informed in how many of six risk and six trust periods they invested money (regardless of the amount). In addition, they were told in how many of these periods they received a back transfer. For example, if the investor invested in four out of six periods in the trust game, he first was reminded of his transfer behavior and then he received the feedback information that a back transfer was executed in two out of the four periods in which he invested. If the investor invested money in an even number of periods, the reinforcement rate was exactly 50%. If money was invested in an odd number of periods, for example, in only three periods of the trust game, a random device determined with 50% probability whether trust was repaid in one or two out of the three trusting cases. This procedure ensures that the investor's trust was betrayed in about 50% of the cases (or that the investor's investment yielded a return in 50% of the cases in the risk game). After the feedback had been presented, investors played another six trust periods against six other human partners and six risk games without getting feedback information. Finally, they received detailed feedback information for each of the 12 trust and 12 risk periods at the end of the experiment.

Prior to scanning, subjects read written instructions describing the sequence of events, the payoff rules, details of the risk and the trust game, and were informed that they would receive feedback after the first 12 periods of a random sequence of trust and risk games. After the subjects had read the instructions, we checked whether they understood the payoff rules, the treatment conditions, and when they would receive feedback information by means of several hypothetical questions. All subjects answered the control questions correctly. Thus, all subjects knew that the whole experiment would last for 24 periods and that they would receive feedback information after 12 periods. Subjects received a lump sum payment of CHF 80 for participating in the experiment plus the additional money earned during the 24 risk/trust periods (exchange rate 5 money units = 1.- Swiss Franc; that is about $1.00). Subjects earned on average about 140.- Swiss Francs in the experiment.

Procedure

The computer screens that the subjects needed to see during the 24 decision trials were presented via a video projector onto a translucent screen that subjects viewed inside the scanner via a mirror. At the beginning of each period, the subjects were presented a fixation cross in the middle of the screen for 10 to 12 s (randomly jittered in the interval 10–12 s). The second screen in a period showed the treatment condition, indicating either the beginning of a trust period or a risk period for 4 s using a schematic picture of a human or computer, respectively (see Figure 2). After 4 s, four buttons representing the four response options were presented on the same screen, indicating that the subjects could now implement their decisions, with a time restriction of 8 s, by means of a four-button input device. On average, decisions were implemented 5.3 s (standard error, 0.103) after the treatment condition (either a trust or a risk period) appeared on the screen. After the subjects had made their decision, a fixation cross was presented for 4 s. After these 4 s, a fourth screen indicated a waiting epoch for 8 s, during which the subjects in the scanner received the information that the trustees are now deciding or that the lotteries are now being played. As mentioned above, feedback information was first shown after six risk and six trust periods had been played. After subjects had finished all 12 risk and 12 trust periods they received rudimentary feedback information for the postfeedback phase (i.e., for the 6 risk and 6 trust periods in that phase) and afterwards we gave them detailed feedback information for every of the 12 risk and 12 trust periods. The rudimentary feedback info in the middle and at the end of the experiment was depicted for 25 s each. The software package z-TREE (Fischbacher, 2007), a program for conducting behavioral experiments in combination with neuroimaging, was used for presenting screens and for collecting behavioral and timing data.

Questionnaire Measures

To measure alterations in the psychological state of the subjects throughout the course of the experiment, we assessed their mood, calmness, and wakefulness at the beginning of the experiment (before substance administration) and after the scanning session, by means of a suitable questionnaire (Steyer et al., 1997). Roughly 2 weeks after the experiment took place in the scanner, the subjects also completed personality questionnaires that assessed their general trust (M. Siegrist, C. Keller, T.C. Earle, and M. Hanselmann, personal communication) and sensation seeking (Zuckerman, 1996; Zuckerman and Link, 1968) behavior. The general trust scale is established using a ten-item questionnaire, where general trust is defined as the conviction that most people can be trusted most of the time. A person with a high level of general trust assumes, in the absence of other information, that the other person will be trustworthy. In other words, the general trust scale measures the general belief in human benevolence. Strong positive correlation between this general trust scale and an investor's trusting behavior in a one-shot trust game with anonymous partners is reported, whereas no such correlations were found in repeated trust games with the same partner or in one-shot games when participants received information about their partner's trustworthiness (M. Siegrist, C. Keller, T.C. Earle, and M. Hanselmann, personal communication). To measure sensation seeking, we used the SSS-V developed by Zuckerman (Zuckerman, 1996; Zuckerman and Link, 1968), which has proven validity and reliability. The SSS-V consists of 40 questions divided into four subscales and one complete scale. The four subscales measure subject's motivation in engaging in sports activities involving some physical danger or risk, the desire for uninhibited behavior in social situations, the desire to seek new experiences through unconventional friends and travel, and the aversion to repetition of any kind. The advantage of the delayed completion of the questionnaires is twofold. First, the administered hormone does not influence the completion of the questionnaires, and second, subjects were not aware that the completion of these questionnaires was associated with the experiment in the scanner and thus, carryover effects between the behavior in the experiment and the questionnaires are highly unlikely.

fMRI Acquisition and Analysis

The experiment was conducted on a 3 Tesla Philips Intera whole body MR Scanner (Philips Medical Systems) equipped with an eight-channel Philips SENSE head coil. Structural image acquisition consisted of 180 T_1-weighted transversal images (0.75 mm slice thickness). For functional imaging, a total of 310 volumes were obtained using a SENSitivity Encoded (SENSE [Pruessmann et al., 1999]) T_2*-weighted echo-planar imaging sequence with an acceleration factor of 2.0. 40 axial slices were acquired covering the whole brain with a slice thickness of 3 mm; no interslice gap; interleaved acquisition; TR = 3000 ms; TE = 35 ms; flip angle = 77°; field of view = 220 mm; matrix size = 80 × 80. In order to optimize functional sensitivity in orbitofrontal cortex and medial temporal lobes, we used a tilted acquisition in an oblique orientation at 30° to the AC-PC line.

For the preprocessing and statistical analyses, the statistical parametric mapping software package (SPM5, Wellcome Department of Cognitive Neurology, London, UK) implemented in Matlab (Version 7) were used. For analysis, all images were realigned to the first volume, corrected for motion artifacts and time of acquisition within a TR, normalized ($3 \times 3 \times 3$ mm^3) into standard stereotaxic space (template provided by the Montreal Neurological Institute), and smoothed using an 8 mm full-width-at-half-maximum Gaussian kernel. A band-pass filter, which was composed of a discrete cosine-basis function with a cutoff period of 128 s for the high-pass filter was applied. In order to increase signal-to noise-ratio, global intensity changes were minimized by scaling each image to the grand mean.

We performed random-effects analyses on the functional data for the decision phase. For that purpose, we defined a general linear model (GLM) that included four regressors of interests and nine other regressors. The four regressors of interests were modeled for the decision phase consisting of six decision periods with onsets at the time of treatment screen appearance (six trust periods prefeedback, six trust periods postfeedback, six risk periods prefeedback, and six risk periods postfeedback). Offsets of the decision phases (regressor's length) were individually modeled based on the subjects' button press. In addition, four regressors of noninterests were modeled for the waiting epoch (duration 8 s) and four for the fixation time between decision and waiting epoch (duration 4 s), again broken down for the two treatments (trust and risk game) and two phases (prefeedback and postfeedback). Finally, another regressor of noninterests modeled the three feedback periods (rudimentary feedback after 12 periods, rudimentary feedback after the second 12 periods, and full feedback after all 24 periods). All regressors were convolved with a canonical hemodynamic response function (HRF). The six scan-to-scan motion parameters produced during realignment were included as additional regressors in the SPM analysis to account for residual effects of scan to scan motion. The correction for multiple comparisons in whole-brain analyses was carried out using an uncorrected p value of 0.005 combined with a cluster-size threshold of 10 voxels (Forman et al., 1995). Furthermore, we focused in our analyses on specific a priori defined regions of interests, in particular the amygdala, midbrain regions and the caudate. Other brain regions, which are significant at the same threshold, are also reported (see Tables 1 and 2). However, we are reluctant to make strong interpretations based on these results because no a priori hypotheses have been made. In case of significant unilateral activations in our main regions of interests, including amygdala and striatum, we lowered the significance threshold to $p < 0.01$ (uncorrected) with the same voxel extent to verify whether a bilateral activation pattern could be found with this threshold.

Linear contrasts of regression coefficients were computed at the individual subject level and then taken to a group level random effects analysis of variance. The following four different t contrast images were calculated for the different analyses of the decision phase at the individual level using the four regressors of interests: trust periods prefeedback > risk periods prefeedback, trust periods postfeedback > risk periods postfeedback, risk periods prefeedback > trust periods prefeedback, and risk periods postfeedback > trust periods postfeedback. For second-level random effects analysis, the single-subject contrasts were entered into two two-way ANOVAs with the following factors: "time phase" (prefeedback, postfeedback as a within subject factor) and "group" (placebo, oxytocin as a between subject factor), and separately for the single subjects contrasts trust > risk (prefeedback and postfeedback) and risk > trust (prefeedback and postfeedback). Based on these ANOVAs, we calculated the following eight serial subtraction contrasts, focusing on the first four contrasts, based on our hypothesis of differential brain activation patterns between OT and placebo in the trust and *not* risk game.

Trust Game

- OT group (trust prefeedback > risk prefeedback) > placebo group (trust prefeedback > risk prefeedback)
- OT group (trust postfeedback > risk postfeedback) > placebo group (trust postfeedback > risk postfeedback)
- Placebo group (trust prefeedback > risk prefeedback) > OT group (trust prefeedback > risk prefeedback)
- Placebo group (trust postfeedback > risk postfeedback) > OT group (trust postfeedback > risk postfeedback)

Risk Game

- OT group (risk prefeedback > trust prefeedback) > placebo group (risk prefeedback > trust prefeedback)
- OT group (risk postfeedback > trust postfeedback) > placebo group (risk postfeedback > trust postfeedback)
- Placebo group (risk prefeedback > trust prefeedback) > OT group (risk prefeedback > trust prefeedback)
- Placebo group (risk postfeedback > trust postfeedback) > OT group (risk postfeedback > trust postfeedback)

All described serial subtraction terms were exclusively masked at $p < 0.05$ with the reversed second contrast of the serial subtraction term in order to make sure that the observed differences between the two groups are not due to differences in the reversed second contrast (Bermpohl et al., 2006).

We created ROIs using the MarsBaR software for all regions showing significant activations based on the described serial subtraction terms (depicted in Tables 1 and 2), including our main regions of interests (amygdala, striatum, midbrain regions and anterior cingulate cortex). ROIs for amygdala and striatum were defined by a combined functional and anatomical criterion by selecting all voxels in the anatomical brain regions (according to the anatomical atlas of Tzourio-Mazoyer et al. [2002]) that were significantly ($p < 0.01$) activated in the corresponding serial subtraction term (please see Table 2). In addition, we created functional ROIs for all other brain regions by selecting all voxels that were significantly activated at $p < 0.005$ in the corresponding serial subtraction terms (please see Tables 1 and 2). Using these ROIs and the software package SPSS (version 13), we conducted repeated-measures ANOVAs using participants' mean beta weights to further investigate lateralization patterns and time effects. In addition, we calculated repeated-measures and univariate ANOVAs to control for potentially confounding variables that we used in the analysis of investors' choices and response times.

Behavioral and Psychometrical Analysis

For the behavioral data of the risk and trust game (transfer decisions), we first created two trust and two risk indexes, consisting of the average transfer during six trust or risk periods either played in the prefeedback or the postfeedback phase. Using these four behavioral indexes, two two-way repeated-measures ANOVAs were performed (separately for the trust and risk indexes) with the following factors: "time phase" (prefeedback, postfeedback as a within subject factor) and "group" (placebo, OT as a between subject factor). We controlled for general trust (M. Siegrist, C. Keller, T.C. Earle, and M. Hanselmann, personal communication) and sensation seeking (Zuckerman and Link, 1968) scores (using questionnaire measures described above) in these ANOVAs and for the feedback information after the first six trust and six risk periods. We controlled for these variables due to the following reasons. First, general trust and sensation seeking have been shown to correlate positively with the trusting behavior in trust games (Schechter, 2007; M. Siegrist, C. Keller, T.C. Earle, and M. Hanselmann, personal communication). Second, because of an integer problem, the reinforcement rate of 50% could imply slightly different feedback information if the number of positive investment decisions (i.e., investments greater than 0) were odd. In these cases, a random mechanism determined whether to reinforce more or less than 50%. For example, if a subject invested three times in the first six periods of the trust game, the subject received information that the trustees repaid either in one of the three cases or in two of the three trusting cases. On average, these differences in feedback information cancel out (i.e., all treatments are affected in the same way) but we nevertheless control for these differences in our ANOVA's to rule out any influence of this feature of our experiments on our results.

For the psychometrical mood questionnaire (Steyer et al., 1997), a two-way repeated-measures ANOVA was performed with the following factors: "time" (PreScan, before substance administration; PostScan, 10 min after the end of the experiment; as a within subject factor) and "group" (placebo, OT, as a between subject factor). Finally, responses time differences (prefeedback and postfeedback for the risk and trust game) as well as subjective rating scales of the BIS were analyzed using independent t tests with group (OT, placebo) as a between subject factor. Results were considered significant at the level of $p < 0.05$ (two-tailed). In case of a significant multivariate effect, post hoc

paired t tests were computed using the Bonferroni correction according to Holm (1979). As effect size measure ETA^2 is reported. Psychometrical and behavioral analyses were performed using the statistical software package SPSS 13 for PC (SPSS Inc.).

SUPPLEMENTAL DATA

The Supplemental Data for this article can be found online at http://www.neuron.org/cgi/content/full/58/4/639/DC1/.

ACKNOWLEDGMENTS

This work is part of Project 9 of the National Competence Center for Research (NCCR) in Affective Sciences. The NCCR is financed by the Swiss National Science Foundation. E.F. and M.H. also gratefully acknowledge support from the research priority program at the University of Zurich on the "Foundations of Human Social Behavior." M.H. also received support from the Swiss National Science Foundation (grant No. PP001-114788).

Received: September 30, 2007
Revised: February 10, 2008
Accepted: April 7, 2008
Published: May 21, 2008

REFERENCES

Adolphs, R., Tranel, D., and Damasio, A.R. (1998). The human amygdala in social judgment. Nature *393*, 470–474.

Adolphs, R., Gosselin, F., Buchanan, T.W., Tranel, D., Schyns, P., and Damasio, A.R. (2005). A mechanism for impaired fear recognition after amygdala damage. Nature *433*, 68–72.

Amaral, D.G. (2003). The amygdala, social behavior, and danger detection. Ann. N. Y. Acad. Sci. *1000*, 337–347.

Bale, T.L., Davis, A.M., Auger, A.P., Dorsa, D.M., and McCarthy, M.M. (2001). CNS region-specific oxytocin receptor expression: importance in regulation of anxiety and sex behavior. J. Neurosci. *21*, 2546–2552.

Bartz, J.A., and Hollander, E. (2006). The neuroscience of affiliation: forging links between basic and clinical research on neuropeptides and social behavior. Horm. Behav. *50*, 518–528.

Bechara, A., Damasio, H., Tranel, D., and Damasio, A.R. (1997). Deciding advantageously before knowing the advantageous strategy. Science *275*, 1293–1295.

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social-history. Games Econ. Behav. *10*, 122–142.

Bermpohl, F., Pascual-Leone, A., Amedi, A., Merabet, L.B., Fregni, F., Gaab, N., Alsop, D., Schlaug, G., and Northoff, G. (2006). Dissociable networks for the expectancy and perception of emotional stimuli in the human brain. Neuroimage *30*, 588–600.

Bohnet, I., and Zeckhauser, R. (2004). Trust, risk and betrayal. J. Econ. Behav. Organ. *55*, 467–484.

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

Botvinick, M., Nystrom, L.E., Fissell, K., Carter, C.S., and Cohen, J.D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. Nature *402*, 179–181.

Camerer, C., and Weigelt, K. (1988). Experimental tests of a sequential equilibrium reputation model. Econometrica *56*, 1–36.

Camerer, C., Loewenstein, G., and Prelec, D. (2005). Neuroeconomics: how neuroscience can inform economics. J. Econ. Lit. *43*, 9–64.

Carter, C.S. (1998). Neuroendocrine perspectives on social attachment and love. Psychoneuroendocrinology *23*, 779–818.

Carter, C.S. (2003). Developmental consequences of oxytocin. Physiol. Behav. *79*, 383–397.

Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, M.M., Noll, D., and Cohen, J.D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. Science *280*, 747–749.

Cohen, J.D., and Blum, K.I. (2002). Reward and decision. Neuron *36*, 193–198.

De Quervain, D.J.F., Fischbacher, U., Treyer, V., Schelthammer, M., Schnyder, U., Buck, A., and Fehr, E. (2004). The neural basis of altruistic punishment. Science *305*, 1254–1258.

Delgado, M.R., Stenger, V.A., and Fiez, J.A. (2004). Motivation-dependent responses in the human caudate nucleus. Cereb. Cortex *14*, 1022–1030.

Delgado, M.R., Frank, R.H., and Phelps, E.A. (2005a). Perceptions of moral character modulate the neural systems of reward during the trust game. Nat. Neurosci. *8*, 1611–1618.

Delgado, M.R., Miller, M.M., Inati, S., and Phelps, E.A. (2005b). An fMRI study of reward-related probability learning. Neuroimage *24*, 862–873.

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

Domes, G., Heinrichs, M., Michel, A., Berger, C., and Herpertz, S.C. (2007b). Oxytocin improves "mind-reading" in humans. Biol. Psychiatry *61*, 731–733.

Falk, A., and Fischbacher, U. (2006). A theory of reciprocity. Games Econ. Behav. *54*, 293–315.

Fehr, E., and Schmidt, K.M. (1999). A theory of fairness, competition, and cooperation. Q. J. Econ. *114*, 817–868.

Fehr, E., and Camerer, C.F. (2007). Social neuroeconomics: the neural circuitry of social preferences. Trends Cogn. Sci. *11*, 419–427.

Fehr, E., Kirchsteiger, G., and Riedl, A. (1993). Does fairness prevent market clearing? An experimental investigation. Q. J. Econ. *108*, 437–459.

Ferguson, J.N., Young, L.J., and Insel, T.R. (2002). The neuroendocrine basis of social recognition. Front. Neuroendocrinol. *23*, 200–224.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. *10*, 171–178.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., and Noll, D.C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn. Reson. Med. *33*, 636–647.

Glimcher, P. (2002). Decisions, decisions, decisions: Choosing a biological science of choice. Neuron *36*, 323–332.

Glimcher, P.W., and Rustichini, A. (2004). Neuroeconomics: the consilience of brain and decision. Science *306*, 447–452.

Heinrichs, M., and Gaab, J. (2007). Neuroendocrine mechanisms of stress and social interaction: implications for mental disorders. Curr. Opin. Psychiatry *20*, 158–162.

Heinrichs, M., and Domes, G. (2008). Neuropeptides and social behavior: effects of oxytocin and vasopressin in humans. Prog. Brain Res. *170*, 337–350.

Heinrichs, M., Baumgartner, T., Kirschbaum, C., and Ehlert, U. (2003). Social support and oxytocin interact to suppress cortisol and subjective responses to psychosocial stress. Biol. Psychiatry *54*, 1389–1398.

Heinrichs, M., Meinlschmidt, G., Wippich, W., Ehlert, U., and Hellhammer, D.H. (2004). Selective amnesic effects of oxytocin on human memory. Physiol. Behav. *83*, 31–38.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scand. J. Stat. *6*, 65–70.

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., and Camerer, C.F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. Science *310*, 1680–1683.

Huber, D., Veinante, P., and Stoop, R. (2005). Vasopressin and oxytocin excite distinct neuronal populations in the central amygdala. Science *308*, 245–248.

Insel, T.R., and Young, L.J. (2001). The neurobiology of attachment. Nat. Rev. Neurosci. *2*, 129–136.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C.F., Quartz, S.R., and Montague, P.R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. Science *308*, 78–83.

Kirsch, P., Esslinger, C., Chen, Q., Mier, D., Lis, S., Siddhanti, S., Gruppe, H., Mattay, V.S., Gallhofer, B., and Meyer-Lindenberg, A. (2005). Oxytocin modulates neural circuitry for social cognition and fear in humans. J. Neurosci. *25*, 11489–11493.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., and Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. Science *314*, 829–832.

Knutson, B., Rick, S., Wimmer, G.E., Prelec, D., and Loewenstein, G. (2007). Neural predictors of purchases. Neuron *53*, 147–156.

Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U., and Fehr, E. (2005). Oxytocin increases trust in humans. Nature *435*, 673–676.

Kuhnen, C.M., and Knutson, B. (2005). The neural basis of financial risk taking. Neuron *47*, 763–770.

Landgraf, R., and Neumann, I.D. (2004). Vasopressin and oxytocin release within the brain: a dynamic concept of multiple and variable modes of neuropeptide communication. Front. Neuroendocrinol. *25*, 150–176.

Lim, M.M., and Young, L.J. (2006). Neuropeptidergic regulation of affiliative behavior and social bonding in animals. Horm. Behav. *50*, 506–517.

McClure, S.M., Laibson, D.I., Loewenstein, G., and Cohen, J.D. (2004). Separate neural systems value immediate and delayed monetary rewards. Science *306*, 503–507.

Meyer-Lindenberg, A., Hariri, A.R., Munoz, K.E., Mervis, C.B., Mattay, V.S., Morris, C.A., and Berman, K.F. (2005). Neural correlates of genetically abnormal social cognition in Williams syndrome. Nat. Neurosci. *8*, 991–993.

Montague, P.R., and Berns, G.S. (2002). Neural economics and the biological substrates of valuation. Neuron *36*, 265–284.

Morris, J.S., DeGelder, B., Weiskrantz, L., and Dolan, R.J. (2001). Differential extrageniculostriate and amygdala responses to presentation of emotional faces in a cortically blind field. Brain *124*, 1241–1252.

Neumann, I.D., Torner, L., and Wigger, A. (2000). Brain oxytocin: differential inhibition of neuroendocrine stress responses and anxiety-related behaviour in virgin, pregnant and lactating rats. Neuroscience *95*, 567–575.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R.J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science *304*, 452–454.

Parker, K.J., Buckmaster, C.L., Schatzberg, A.F., and Lyons, D.M. (2005). Intranasal oxytocin administration attenuates the ACTH stress response in monkeys. Psychoneuroendocrinology *30*, 924–929.

Pasley, B.N., Mayes, L.C., and Schultz, R.T. (2004). Subcortical discrimination of unperceived objects during binocular rivalry. Neuron *42*, 163–172.

Pegna, A.J., Khateb, A., Lazeyras, F., and Seghier, M.L. (2005). Discriminating emotional faces without primary visual cortices involves the right amygdala. Nat. Neurosci. *8*, 24–25.

Pruessmann, K.P., Weiger, M., Scheidegger, M.B., and Boesiger, P. (1999). SENSE: sensitivity encoding for fast MRI. Magn. Reson. Med. *42*, 952–962.

Rabin, M. (1993). Incorporating fairness into game theory and economics. Am. Econ. Rev. *83*, 1281–1302.

Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., and Kilts, C. (2002). A neural basis for social cooperation. Neuron *35*, 395–405.

Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., and Cohen, J.D. (2003). The neural basis of economic decision-making in the Ultimatum Game. Science *300*, 1755–1758.

Schechter, L. (2007). Traditional trust measurement and the risk confound: an experiment in rural Paraguay. J. Econ. Behav. Organ. *62*, 272–292.

Spitzer, M., Fischbacher, U., Herrnberger, B., Gron, G., and Fehr, E. (2007). The neural signature of social norm compliance. Neuron *56*, 185–196.

Stein, M.B., Goldin, P.R., Sareen, J., Zorrilla, L.T., and Brown, G.G. (2002). Increased amygdala activation to angry and contemptuous faces in generalized social phobia. Arch. Gen. Psychiatry *59*, 1027–1034.

Steyer, R., Schenkmezger, P., Notz, P., and Eid, M. (1997). Der Mehrdimensionale Befindlichkeitsfragebogen (MDBF) (Göttingen: Hogrefe).

Tillfors, M., Furmark, T., Marteinsdottir, I., Fischer, H., Pissiota, A., Langstrom, B., and Fredrikson, M. (2001). Cerebral blood flow in subjects with social phobia during stressful speaking tasks: a PET study. Am. J. Psychiatry *158*, 1220–1226.

Tricomi, E.M., Delgado, M.R., and Fiez, J.A. (2004). Modulation of caudate activity by action contingency. Neuron *41*, 281–292.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage *15*, 273–289.

Uvnas-Moberg, K. (1998a). Antistress pattern induced by oxytocin. News Physiol. Sci. *13*, 22–25.

Uvnas-Moberg, K. (1998b). Oxytocin may mediate the benefits of positive social interaction and emotions. Psychoneuroendocrinology *23*, 819–835.

Whalen, P.J., Rauch, S.L., Etcoff, N.L., McInerney, S.C., Lee, M.B., and Jenike, M.A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. J. Neurosci. *18*, 411–418.

Winston, J.S., Strange, B.A., O'Doherty, J., and Dolan, R.J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. Nat. Neurosci. *5*, 277–283.

Young, L.J., and Wang, Z. (2004). The neurobiology of pair bonding. Nat. Neurosci. *7*, 1048–1054.

Young, L.J., Lim, M.M., Gingrich, B., and Insel, T.R. (2001). Cellular mechanisms of social attachment. Horm. Behav. *40*, 133–138.

Zuckerman, M., and Link, K. (1968). Construct validity for sensation-seeking scale. J. Consult. Clin. Psychol. *32*, 420–426.

Zuckerman, M. (1996). The psychobiological model for impulsive unsocialized sensation seeking: a comparative approach. Neuropsychobiology *34*, 125–129.