Neuropsychopharmacology (2010) 35, 2502–2509
© 2010 Nature Publishing Group All rights reserved 0893-133X/10 $32.00
www.neuropsychopharmacology.org

Oxytocin Decreases Aversion to Angry Faces in an Associative Learning Task

Simon Evans[1], Sukhwinder S Shergill[2,3] and Bruno B Averbeck*[,1,4]

[1]Sobell Department of Motor Neuroscience and Movement Disorders, Institute of Neurology, University College London, London, UK; [2]Department of Psychiatry, Kings College London, London, UK; [3]Institute of Psychiatry Cognition Schizophrenia and Imaging Laboratory, London, UK; [4]Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA

Social and financial considerations are often integrated when real life decisions are made, and recent studies have provided evidence that similar brain networks are engaged when either social or financial information is integrated. Other studies, however, have suggested that the neuropeptide oxytocin can specifically affect social behaviors, which would suggest separable mechanisms at the pharmacological level. Thus, we examined the hypothesis that oxytocin would specifically affect social and not financial information in a decision making task, in which participants learned which of the two faces, one smiling and the other angry or sad, was most often being rewarded. We found that oxytocin specifically decreased aversion to angry faces, without affecting integration of positive or negative financial feedback or choices related to happy *vs* sad faces.
Neuropsychopharmacology (2010) **35,** 2502–2509; doi:10.1038/npp.2010.110; published online 15 September 2010

Keywords: decision making; behavior; computational neuroscience; oxytocin; emotion

INTRODUCTION

Considerable work has defined an important role for the neuropeptide oxytocin in social behavior (Donaldson and Young, 2008, Insel and Young, 2001). In rodents, studies have shown that oxytocin is involved in social behaviors including approach, recognition, and bond formation (Lim and Young, 2006). Recently, work has begun to generalize these findings to human social behaviors (Bartz and Hollander, 2006). These studies have shown that intranasally delivered oxytocin can increase trust or generosity in economic games (Kosfeld *et al*, 2005; Zak *et al*, 2007), improve inference of emotions from images of the eyes (Domes *et al*, 2007b), and increase fixations to the eye region in human faces (Guastella *et al*, 2008a). Complimentary imaging work has shown that these behavioral effects may be mediated by a decreased response in the amygdala to aversive aspects of stimuli, including lack of reciprocity in trust games (Baumgartner *et al*, 2008), painful stimulation of one's own hand (Singer *et al*, 2008), images with either intrinsic or conditioned negative valence (Kirsch *et al*, 2005; Petrovic *et al*, 2008), or faces expressing emotions (Domes *et al*, 2007a).

*Correspondence: Dr BB Averbeck, Laboratory of Neuropsychology, National Institute of Mental Health, National Institute of Health, Bldg. 49, Room 1B-80, Bethesda, MD, 20892-4415, USA, Tel: +1 301 594 1126, Fax: +1 301 402 0046, E-mail: bruno.averbeck@nih.gov
Received 17 March 2010; revised 16 June 2010; accepted 22 June 2010

Interestingly, despite the fact that oxytocin receptors are found throughout brain areas involved in motivation and reward, including the amygdala and ventral striatum (Schorscher-Petcu *et al*, 2009; Skuse and Gallagher, 2009), the effects of oxytocin seem to be specific to social stimuli (Ferguson *et al*, 2001) and may not extend to financial or other rewards. This presents an interesting contrast, as several recent papers have suggested that social and financial reward stimuli are processed by similar brain networks (Izuma *et al*, 2008; Zink *et al*, 2008). This raises the question of whether social and non-social reward systems can be separated. To address this question, we examined the effects of oxytocin on a task in which participants integrate both social and financial reward feedback, when learning which of two faces is being most often rewarded within a block of trials (Averbeck and Duchaine, 2009). In each block of trials, both faces have the same identity, but one is smiling (positive social valence) and the other is angry or sad (negative social valence). After picking one of the faces within a trial, feedback is given in terms of winning (positive financial valence) or losing (negative financial valence) money. We have previously shown that participants have a preference for the happy face relative to angry or sad faces, such that they select it more often than they should, given the financial feedback (Averbeck and Duchaine, 2009) similar to preferences for attractive faces (Hayden *et al*, 2007). Here we examined whether oxytocin would specifically impact the social

preference, without affecting the integration of positive and negative financial feedback.

MATERIALS AND METHODS

Task, Participants, and Procedure

A total of 18 male participants (average age 26. 5) were recruited for this double-blind placebo-controlled crossover study of intranasal oxytocin. Standard consent and safety reporting procedures were followed, and ethical approval was obtained from the SLaM Institute of Psychiatry, King's College London REC. Participants were excluded from the study if they had clinical depression, anxiety or any other psychiatric history. In addition, participants were excluded if they had high blood pressure, heart problems or any major illness or hospitalization in the last year.

Each participant completed two sessions of testing separated by at least 1 week. The procedure was same for each session. On arrival, the participant self-administered a nasal spray consisting of either 24 IU oxytocin (Syntocinon, Novartis) or saline placebo. This is similar to the dose that has been used in the previous studies, which have shown effects (Guastella *et al*, 2008b; Kirsch *et al*, 2005; Rimmele *et al*, 2009; Savaskan *et al*, 2008). Each participant received both oxytocin and placebo and the order in which they received them was randomized. No serious adverse events occurred. One participant reported a mild headache after oxytocin administration.

To examine non-specific effects of drug on mood, participants completed a questionnaire on arrival (ie, before the spray) and immediately before undergoing testing, following delivery of oxytocin or placebo. The questionnaire used was the Brief Mood Inventory Scale, comprising 17 items (Mayer and Gaschke, 1988). Participants rated 16 dimensions of mood on a scale of 1–4, followed by overall mood on a scale of −10 to +10. Scores were compared between oxytocin and Placebo conditions, on overall and single dimensions of mood. There was no main effect of drug ($F_{1,17} = 0.764$, $p = 0.394$) and no interaction between drug and before/after ($F_{1,17} = 0.112$, $p = 0.742$), but there was a main effect of before/after ($F_{1,17} = 5.27$, $p = 0.035$), with participant's mood decreasing slightly after drug administration.

Following administration, behavioral testing commenced after a 50 min delay. It has been shown that vasopressin, which is closely related to oxytocin, reaches peak effects in 30–50 min when administered transnasally (Born *et al*, 2002), and a 50 min delay between drug administration and the start of testing has been used in previous studies with oxytocin (Domes *et al*, 2007b; Fischer-Shofty *et al*, 2010; Kirsch *et al*, 2005). Each participant did four blocks of 26 trials in both the happy *vs* angry and happy *vs* sad tasks. The order of the tasks was randomized across participants. In each block there were two faces with the same identity but different expressions. In experiment 1, one face had a happy expression and the other had an angry expression, and in experiment 2 the expressions were happy and sad. Two different male identities were used in each experiment, but they were the same two across experiments. Faces were taken from the Ekman series (Ekman and Friesen, 1971). The identities alternated across blocks, and we counter-balanced whether identity A or B was shown in the first block. In each block, one of the faces paid off 40% of the time when selected and the other paid off 60% of the time. Participants were instructed to make decisions to maximize their rewards. The order of high reward *vs* low reward associated with the happy and angry/sad face was randomized across participants, such that some participants started with the happy face being rewarded most often, followed by the angry/sad face being rewarded most often, whereas other participants had the opposite order. Participants were told when the block switched by the computer task and also that probabilities would be remapped to the expressions. In addition, the identity always changed across blocks.

We chose 26 trials, because an ideal observer is able to identify the most highly rewarded face correctly in 85% of blocks of this length with the probability values we used (Averbeck and Duchaine, 2009). Thus, there is a sufficient feedback evidence in most blocks to identify the correct face. This point is not directly relevant to this analysis, however, because we modeled the participants' belief trial-by-trial, and therefore, we do not rely on whether or not the participants know which face 'should' be correct in each block. Thus, if for some reason the sequence of rewards favors the face being rewarded stochastically less often, this analysis takes that into account.

Every participant was given the following instructions on the task: 'On each trial in this task you will be presented with two faces. You will have to select one of the faces. Press 'z' to select the left face, '/' to select the right face. Your task is to try to figure out which face in each block has the highest probability of winning and pick that face as many times as possible. You will be told when the block switches, and at each switch the faces will be associated with new probabilities of winning.'

Within an individual trial, the happy and angry faces were presented pseudo-randomly on either the left or the right side of the screen (Figure 1). Participants were given an unlimited duration to make their decision and the faces were present until the participants responded. After the participants made their decision, they pressed one of two buttons to indicate whether they had chosen the left or the right face. The chosen face was then presented at the center of the screen, and below it was text indicating whether they had 'won' or 'lost' in that trial, with a win worth 10 pence, and a loss worth nothing. A 5 KHz tone was played when they won, and a 2.5 KHz tone was played when they lost. Feedback was given for 3 s and there was a 1 s inter-trial interval during which the screen was blank. All participants were paid the same amount, which was greater than their actual winnings, but they were not informed about this until after the experiment had ended. Thus, during the experiment they believed their performance related to the amount they would be paid.

Data Analysis

All data analysis was carried out in Matlab. As the actual outcomes in the experiment were stochastic, it was possible for the face, which had a lower probability of being rewarded in an individual block to actually be rewarded more often, especially over a short run of trials. Therefore,

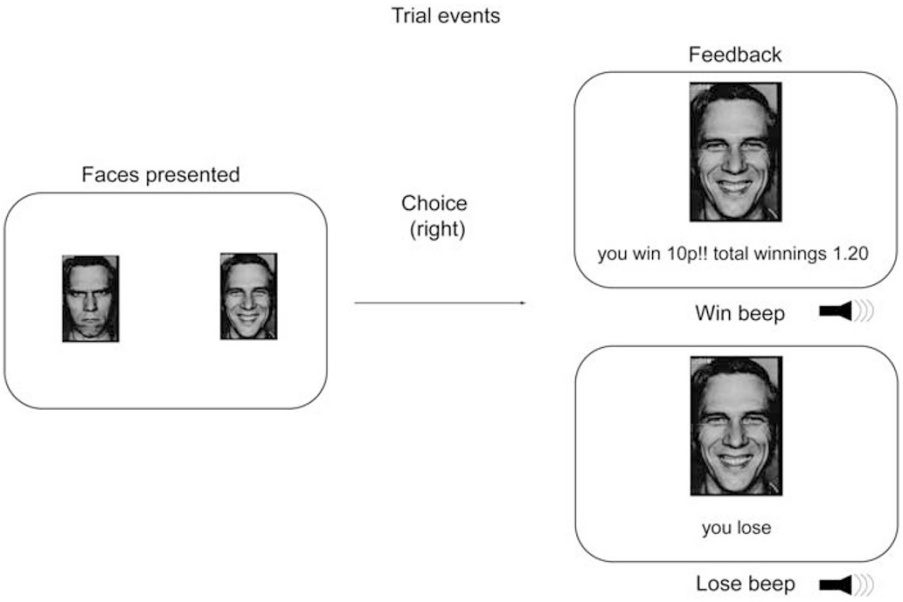

Figure 1 Task events. Faces were presented on the left and right of the screen. Participants then selected either the right or left face, after which they were given auditory and visual feedback about whether they had 'won' or 'lost'. If they won, their total winnings were incremented by 10 *p*.

we referenced all decisions of the participants to an ideal observer model. The ideal observer was modeled using a Bernoulli distribution for each of the two faces. Thus, the likelihood function was given by:

$$p\,(D|\theta_i) = \theta_i^{r_i}(1-\theta_i)^{N_i-r_i} \tag{1}$$

Where θ_i is the probability that face i (eg, angry or happy) is rewarded, r_i is the number of times face i was rewarded, and N_i is the number of times face i was selected in the current block. The vector D represents the data, which in this case are the values of r and N from the current block. The probability that face i was more often rewarded than face j is given by:

$$p\,(\theta_i > \theta_j) = \int_0^{\theta_1} p\,(\theta_i|D) \int_0^{\theta_i} p\,(\theta_j|D)\,d\theta_j\,d\theta_i \tag{2}$$

We have here used the posterior, $p(\theta_i|D)$, as we numerically normalized the distributions before carrying out the integral. The ideal choice or decision rule ($\hat{f}$) was then given by the face which is probably most highly rewarded, given by

$$\begin{cases} p\,(\theta_i > \theta_j) > 0.5 & \hat{f} = i \\ p\,(\theta_i > \theta_j) < 0.5 & \hat{f} = j \end{cases} \tag{3}$$

For the data analysis, in which we compared the model's choice to the participant's choice, when probabilities were tied we incremented both options for the model's choice by 0.5, thus spreading the model choice evenly. Although it could be argued that participants should explore in this task and therefore, that referencing choices to an ideal observer is not optimal, exploration should be minimized to perform optimally, as we used stationary distributions and therefore, this is a two-armed bandit task (Tsitsiklis, 1994).

To examine the differential effect of positive and negative feedback, we added parameters to the basic model, as we have carried out previously (Averbeck and Duchaine, 2009). This analysis applies only to Figures 2b and f. All other analyses were referenced to the ideal observer, without these parameters. In the extended model, for rewarded trials the reward value was calculated as:

$$r\,(t) = 0.5 + a \tag{4}$$

whereas for unrewarded trials it was calculated as

$$r\,(t) = 0.5 - b \tag{5}$$

The variables a and b were fit as free parameters in the model. The parameter a measured the amount that positive rewards were weighted and the parameter b measured the amount that no reward ('you lose') was weighted. For the ideal observer these terms are both 0.5. The total reward up to trial T substituted into equation (1) was then calculated as the sum across trials:

$$r_i = \sum_{t=1}^{T} r_i(t) \tag{6}$$

The parameters a and b were fit by maximizing the likelihood of the sequence of decisions of each participant:

$$p\,(D^*|a,\,b) = \prod_{k=1}^{N} (p_k(\theta_i > \theta_j)l_k + (1 - p_k(\theta_i > \theta_j))(1 - l_k)) \tag{7}$$

Where $l = 0$ if the participant selected stimulus j and $l = 1$ if the participant selected stimulus i. Here, D^* is the series of decisions of the participant, as opposed to the series of outcomes, which is collected in D in equation (1). We maximized the log of the likelihood using fminsearch in Matlab. We started from multiple initial values to minimize the chance of falling into a local minima.

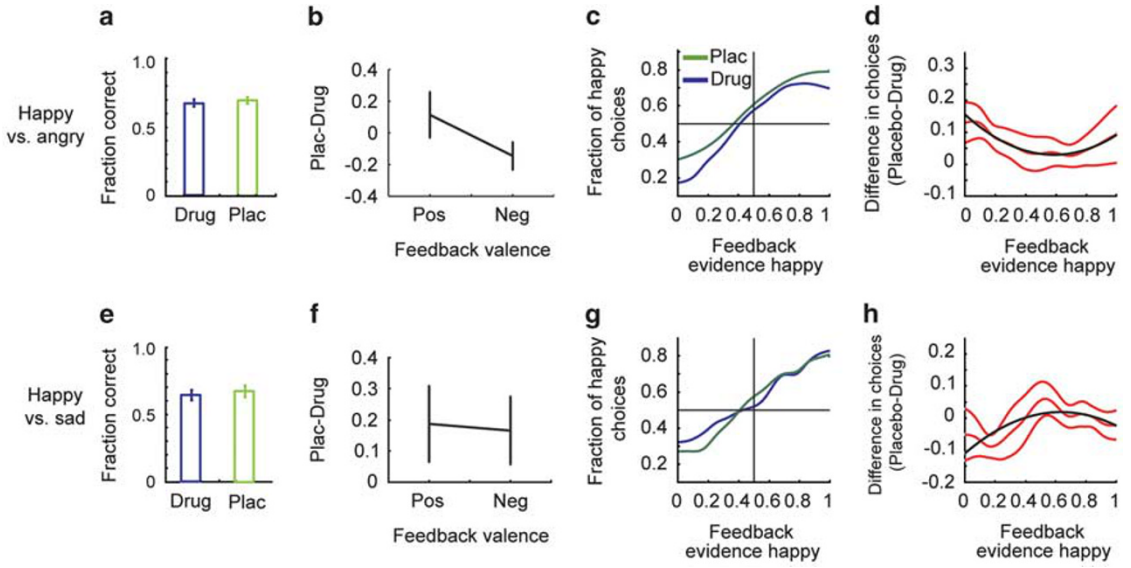

Figure 2 Effects of oxytocin on learning and emotion preference. (a) Fraction correct, referenced to the ideal observer when choosing between happy and angry faces in drug and placebo conditions. Error bars are +/− 1 SEM. (b). Difference in learning from positive and negative feedback off and on oxytocin. Error bars are +/− 1 SEM (c) Evidence *vs* choice probability curves. Probability that the participants picked the happy face (*y*-axis) *vs* the evidence for the happy face (*x*-axis), under drug and placebo conditions. The evidence comes from the ideal observer model, and is given by $p(\theta_{happy} > \theta_{angry})$ from equation (2). (d) Participant-by-participant difference between drug and placebo conditions, as a function of the evidence. Red lines are mean +/−1 SEM (SEM calculated participant difference curves, $n = 18$) and black line is best-fit regression line (see results). e–h same as a–d except in the happy *vs* sad condition.

It would be possible to model learning in this task using a reinforcement learning model, and these model are related to ours in many respects. The main disadvantage of a reinforcement learning model is that it has free parameters, usually a learning rate parameter and a temperature parameter, that have to be fit.

The evidence *vs* choice probability curves were generated by calculating a moving average. Differences were then taken bin-by-bin for statistical analysis.

RESULTS

A total of 18 participants were run in a double-blind study, in which they were given either intranasal oxytocin or placebo in separate sessions and then engaged in two associative learning tasks, in which they had to select between a happy and an angry face, or a happy and a sad face (see Methods and Figure 1). Within each block of 26 trials, one of the faces was rewarded 60% of the time and the other 40% of the time. Participants were asked to learn which face was most often being rewarded and pick it as many times as possible. We compared the fraction of correct decisions (ie, percent correct divided by 100) participants made in drug and placebo conditions. As the feedback was probabilistic, correct was defined as making the same decision as an ideal observer, which always made the optimal choice given the previous financial feedback. Referencing to an ideal observer, which based its decisions on exactly the same information that the participant received on a trial-by-trial basis, controlled for the fact that the feedback could be more positive for the face, which had a lower probability of being rewarded in the block. Participants were above chance for the happy *vs* angry task in the placebo (mean 0.69,$p < 0.001$, $t_{17} = 7.1$) and drug (mean 0.68, $p < 0.001$, $t_{17} = 8.2$) conditions. Similarly, they were above chance for the happy *vs* sad task in the placebo (0.67, $p < 0.001$, $t_{17} = 6.5$) and drug (mean 0.64, $p < 0.001$, $t_{17} = 6.3$) conditions. The difference between placebo and drug conditions, however, did not reach significance in either the happy *vs* angry ($t_{17} = 0.7$, $p = 0.473$) or the happy *vs* sad ($t_{17} = 1.5$, $p > 0.149$) tasks (Figures 2a and d). Thus, there was no overall effect of drug on learning to respond to the most rewarded target.

Next we examined the face preference (ie, an effect of emotion valence) by comparing how often the participants picked the happy face when they should have picked (given the ideal observer prediction) the angry face, to the number of times the participants picked the angry face when they should have picked the happy face. A mixed-effects ANOVA with repeated measures on valence (ie, picking angry when they should pick happy, *vs* picking happy when they should pick angry) and drug (oxytocin *vs* placebo) in the happy *vs* angry task, had a main effect of valence ($F_{1,17} = 5.45$, $p = 0.032$), but no main effect of drug ($F_{1,17} = 0.34$, $p = 0.573$) and no valence by drug interaction ($F_{1,17} = 1.84$, $p = 0.192$). Thus, participants picked the happy face when they should have picked the angry face, more often than they picked the angry face when they should have picked the happy face. We also carried out planned comparisons on the valence effect separately in the drug and placebo conditions and found that the valence effect was significant under placebo (mean difference = 0.159, $t_{17} = 2.9$, $p = 0.009$) but not under the drug (mean difference = 0.069, $t_{17} = 1.2$, $p = 0.261$). We carried out the same analysis in the happy *vs* sad condition and found that there was a significant main effect of valence ($F_{1,17} = 8.5$, $p = 0.010$), but no main effect of drug ($F_{1,17} = 1.11$, $p = 0.307$) and no valence by drug

interaction ($F_{1,17} = 0.32$, $p = 0.581$). Planned comparisons showed that there was a significant preference in both the placebo (mean difference = 0.089, $t_{17} = 2.27$, $p = 0.037$) and drug conditions (mean difference = 0.126, $t_{17} = 2.20$, $p = 0.042$).

We also examined whether oxytocin affected the amount that participants relied on positive or negative feedback during learning. We fit a parameterized model (see methods), which was a direct extension of the ideal observer, to individual participants which assessed the amount that positive and negative feedback drove future decisions, and compared parameters within participants between drug and placebo conditions for both tasks (Figures 2b and f). In the happy *vs* angry task we found a main effect of valence ($F_{1,17} = 31.51$, $p < 0.001$), but no main effects of drug ($F_{1,17} = 0.03$, $p = 0.854$) or drug by valence interaction ($F_{1,17} = 1.08$, $p = 0.306$). In the happy *vs* sad task, we found a main effect of valence ($F_{1,17} = 56.47$, $p < 0.001$), but again, no main effects of drug ($F_{1,17} = 2.89$, $p = 0.108$) or drug by valence interactions ($F_{1,17} = 0.02$, $p = 0.890$). Thus, participants learned more from positive feedback than negative feedback across tasks, but oxytocin did not modify this learning.

Next, we examined these effects in more detail by estimating the evidence *vs* choice probability curves, which show the fraction of times that the participants chose each face as a function of the strength of the (financial) evidence supporting each face. The evidence is given by the ideal observer: $p(\theta_{happy} > \theta_{angry})$ for happy *vs* angry and correspondingly for happy *vs* sad. For example, when the evidence was 0.5 for the happy face, the evidence equally supported both faces, and participants should have picked the happy and angry faces about 50% of the time if they were behaving as the ideal observer. What we found instead, was that participants picked the happy face about 60% of the time, and correspondingly the angry face about 40% of the time (Figure 2c), consistent with the significant preference effects reported above. This was true under both placebo and drug conditions. However, when the evidence strongly supported the angry face, participants in the placebo condition picked it just over 60% of the time, whereas participants in the drug condition picked it over 80% of the time. We carried out the same analysis for the happy *vs* sad condition and found that the preference near 50% was smaller than the preference in the happy *vs* angry condition (Figure 2g). However, when the evidence strongly supported the happy face, participants picked it about 80% of the time, but when the evidence strongly supported the sad face participants picked it about 70% of the time. Thus, participants were more likely to pick the happy face when the evidence strongly supported it than the sad face. Qualitatively, however, there were few differences between drug and placebo conditions for the happy *vs* sad task.

To quantify the difference in the evidence *vs* choice probability curves (Figures 2c and g) between drug conditions, we computed this curve for each individual participant, separately for the drug and placebo conditions for each task. Then, we took the difference between these curves for individual participants (placebo–drug), and examined this difference curve across the participants. We could carry out *t*-tests on individual bins, but this would lead to a multiple comparisons problem. Therefore, we quantified the effect by fitting the following regression equation to the difference curve of individual participants:

$$dchoices_{placebo-drug} = a_0 + a_1 evidence_{happy} + a_2 evidence^2_{happy}$$

Thus, this analysis gave us 3 regression coefficients for each participant, corresponding to the intercept (a_0), slope (a_1), and quadratic terms (a_2). Across participants we had three distributions, one for each parameter. Subsequently, we carried out *t*-tests on each of these distributions to assess significance of the corresponding parameter at a population level (Holmes and Friston, 1998). Thus, we were testing the hypothesis that the intercept term, across participants, was significantly different than zero, and correspondingly for the linear and quadratic terms. For the happy *vs* angry task we found that the intercept was significant (mean $a_0 = 0.17$, $t_{17} = 2.80$, $p = 0.012$) but the linear and quadratic terms just missed significance (mean $a_1 = -0.49$, $t_{17} = 2.01$, $p = 0.060$; mean $a_2 = 0.45$, $t_{17} = 2.00$, $p = 0.061$). Thus, there was a significant difference in how participants chose happy and angry faces between placebo *vs* drug conditions (Figure 2d). Most specifically, when the evidence strongly supported the angry face, participants on drug more often chose the angry face, as shown by the significant intercept term (a_0) across participants. For the happy *vs* sad face, however, none of the three parameters were significant (mean $a_0 = -0.11$, $t_{17} = 1.7$, $p = 0.10$; mean $a_1 = 0.41$, $t_{17} = 1.38$, $p = 0.18$; mean $a_2 = -0.32$, $t_{17} = 1.09$, $p = 0.29$). Therefore, there was no systematic difference between placebo and drug, in how the participants chose the happy *vs* the sad face, as a function of the evidence (Figure 2h).

DISCUSSION

Although oxytocin is often considered prosocial, whether it achieves these effects by increasing the appetitive effects of positively valenced stimuli, or decreasing the aversive effects of negatively valenced stimuli, or both has not been clear. In addition, whether it affects only social stimuli or a broader class of stimuli with positive and negative emotional valence has not been clear. Consistent with most previous studies, we found no effects of oxytocin on mood (Di Simplicio *et al*, 2009; Fischer-Shofty *et al*, 2010; Kirsch *et al*, 2005; Marsh *et al*, 2010). When participants given placebo were required to learn which of the two faces in a block of trials was being rewarded most often, they preferred the happy face relative to the angry or sad faces, replicating our previous results (Averbeck and Duchaine, 2009). We further found that administration of oxytocin reduced aversion to the angry face, such that when the participants had been given strong financial rewards, which suggested they should select the angry face, they did so significantly more often under drug than placebo. In fact, on drug, participants chose the angry face when the evidence strongly supported it about as often ($\sim$80%) as they chose the happy face when the evidence strongly supported it ($\sim$80%), on placebo. Interestingly, there was no effect of oxytocin in the happy *vs* sad task, although participants did show choice preference in this task as well. Additional analyses of the integration of positive and negative financial feedback, as well as assessment of overall task performance, showed that oxytocin did not significantly affect financial reward processing. Thus, this data

supports the hypothesis that oxytocin specifically decreases the aversive aspects of angry faces, while having no effect on sad faces, which also have negative emotional valence, or financial feedback, whether it be positive or negatively emotionally valenced. There are of course other classes of stimuli, which elicit emotions, beyond financial and social considerations. Therefore, more of these will have to be examined before the specificity of oxytocin to social emotional stimuli, and perhaps even more specifically to angry stimuli, can be established. It should be noted that these effects cannot be because of the simple changes in visual processing, as the visual discrimination is the same in all trials. The participants must simply determine which emotion is on each side of the screen. The effects we see are specifically related to how choice behavior is affected by both financial and social information, as the effects of decreased aversion are specific to cases in which the angry face should be chosen.

There are a number of limitations to this study that are important to consider. First, we did not include an explicit loss condition in which participants actually lost money. We simply showed the text, 'you lose' when the participants did not win. Thus, it is possible that oxytocin would have had effects in an explicit loss. Second, we used highly polar emotions, including happy *vs* angry and happy *vs* sad. We used these emotions, as we believed they would generate the largest behavioral bias. However, all of these effects are relative effects between these emotions. The inclusion of a neutral emotion condition may have clarified the specific effects of oxytocin on happy *vs* angry faces. Third, this study used only male participants, and there is evidence that male and female participants respond differently to oxytocin (Domes *et al*, 2010). Finally, we have conceptualized the effect in this task as a reduction in aversion to angry faces. However, it is also possible that participants find the happy faces less appetitive, as the effects we find are changes in relative choice preferences. In this case, however, we might have expected results in the happy *vs* sad task, which we did not see. It is possible that the lack of results in the happy *vs* sad task are because of the fact that oxytocin specifically affects approach/withdrawal behaviors (Kemp and Guastella, 2010), and sad faces do not elicit withdrawal responses as do angry faces. It could also be the case that participants are better at assigning reward to angry faces on oxytocin. Overall, however, we did not find that participants learned better on oxytocin, so they do not seem to be better at assigning reward across emotions. In fact, the improvement in performance for angry faces is offset by a decrease in performance for happy faces.

Although this task structure differs from previous studies, which used explicit evaluation of stimuli after pavlovian association with shock (Petrovic *et al*, 2008) or levels of investment in an economic exchange (Baumgartner *et al*, 2008; Kosfeld *et al*, 2005), aspects of these studies are consistent with the results of this study. Specifically, Petrovic *et al* (2008) found that faces which had been associated with shock were not judged more aversive than faces which had not been associated with shock following administration of oxytocin. Furthermore, decreased activation on drug in the amygdala was specific to faces with direct and not averted gaze. This experiment, however, did not include a positively valenced condition, so it is not clear whether oxytocin would have increased positive associations. In the experiment by Baumgartner *et al* (2008), oxytocin specifically increased investment following feedback that 50% of trustees had not returned an investment in the first round of play. Participants on placebo actually decreased investment following feedback, which suggests that these participants differentially considered the 50% trustee reciprocity to be unfair and aversive. Again, this effect seemed to be mediated by the amygdala. Thus, across studies requiring a behavioral response, oxytocin seems to mitigate aversive aspects of social stimuli on behavior. This task shows this effect more specifically, and also shows within the same paradigm that learning from financial feedback is not affected by oxytocin.

A recent note has conceptualized the effects of oxytocin as both increasing approach related behavior and decreasing withdrawal related behavior (Kemp and Guastella, 2010). The hypothesis as developed by this group, however, discusses anger from the point of view of the angry individual, not the individual toward whom the anger is directed. Thus, their conclusion that anger is an approach behavior does not map directly onto the effects of anger we see in this task. If being the object of anger generates a withdrawal response, however, these results would be consistent with this aspect of their hypothesis. The results of this study are further specific in that negative feedback, which also generates a withdrawal response in the participants of this study, is not affected by oxytocin.

Other tasks have reported results which address related, but different hypotheses about the effects of oxytocin and show that oxytocin can affect a broader class of behaviors. For example, it has been shown that oxytocin attenuates the amygdala response to emotional expressions, regardless of valence (Domes *et al*, 2007a), improves one's ability to infer emotions from images of the eyes (Domes *et al*, 2007b) and increases gaze to the eye region of human faces (Guastella *et al*, 2008a). In addition, some studies have seen improvements in memory for faces on oxytocin, although there has been some inconsistency in whether these effects are specific to positive or negative emotions potentially attributable to differences in task design (Guastella *et al*, 2008b; Rimmele *et al*, 2009; Savaskan *et al*, 2008). Finally, it has been shown that oxytocin increases feelings of both envy and schadenfreude in a task in which participants win more or less than other participants doing the same task (Shamay-Tsoory *et al*, 2009) and that oxytocin can increase sensitivity to social feedback (Hurlemann *et al*, 2010). Both of these studies suggest that oxytocin actually increases sensitivity to social stimuli, whereas this study seems to indicate a decreased sensitivity. The paradigms differ in important ways from ours, however. First, in the study by Shamay-Tsoory *et al*, (2009), participants are not required to make a decision based on their outcome. They are only required to indicate how they feel about the outcome. In the second paradigm, Hurlemann *et al* (2010), show that participants on oxytocin do not learn more from a cognitive form of feedback, which is consistent with the results of this study. However, they also show that on oxytocin participants are more sensitive to the social feedback. In their study, participants use the social feedback to learn the category membership of an abstract string of numbers, whereas in this study participants learn to associate rewards

to a social stimulus. Thus, the studies are examining fundamentally different questions, which may explain the difference in results.

Results from non-human animal studies also support the hypothesis that oxytocin exerts its effects by decreasing the aversive aspect of social stimuli. For example, virgin female rats find pups aversive (Fleming and Anderson, 1987). However, after parturition, rats find pups appetitive (Lee *et al*, 1999), an effect blocked by injection of oxytocin antagonists into the ventricles (Fahrbach *et al*, 1985) among other manipulations. It is possible, however, that this effect is specific to maternal behaviors, as pair bonding after mating is also dependent on oxytocin (Williams *et al*, 1994), but it is not clear why pair bonding would be enhanced by reduced aversion to negative social stimuli.

Oxytocin interacts with other neurotransmitter systems, including the opioid (Vuong *et al*, 2010) and dopamine systems (Aragona *et al*, 2003; Aragona *et al*, 2006; Gingrich *et al*, 2000). With respect to dopamine, oxytocin had no significant effect on learning from positive or negative financial feedback, both of which have been shown to be dopamine dependent (Djamshidian *et al*, 2010; Frank *et al*, 2007; Frank *et al*, 2004; Pessiglione *et al*, 2006). Studies in prairie voles have shown that pair bonding, which is oxytocin dependent, is also dependent on dopamine transmission in the nucleus accumbens (Aragona *et al*, 2003; Aragona *et al*, 2006; Gingrich *et al*, 2000), a structure which has been shown to have oxytocin receptors (Schorscher-Petcu *et al*, 2009). Thus, the effects of oxytocin that are mediated by the dopamine system appear to be specific to social behaviors. Whether the reduction in the aversive aspect of the angry face in this task is also mediated by interactions with either the dopamine or opioid systems remains an open question.

Results from a range of studies are beginning to show an important role for oxytocin in social interactions and studies have already begun to show that oxytocin can improve deficits in social interactions in patient groups (Averbeck, 2010). For example, in participants with autism spectrum disorders (ASD), oxytocin can improve the recognition of emotions in the reading the mind from the eyes task (RMET; (Guastella *et al*, 2010), and it can increase saccades to the eye region when ASD patients view images of faces (Andari *et al*, 2010). Functional imaging results suggest that oxytocin increases eye movements toward the eye region in healthy participants by making them more salient or appetitive (Gamer *et al*, 2010). Thus, the improvement in performance in the RMET task may be because of increased attention to the informative eye region, and patients may avoid this region because they find the eyes aversive. We have also found, using the task reported in this study, that patients with schizophrenia are more averse to angry faces than healthy controls (unpublished data). Thus, it is possible that oxytocin might also be able to reverse this deficit, as we have seen that oxytocin increases the probability that healthy participants will pick an angry face, when it is associated with reward.

ACKNOWLEDGEMENTS

This work was supported in part by the Intramural Program of the NIH, National Institute of Mental Health and a Welcome Trust Project Grant to BBA. SE was supported by an MRC 3 Year PhD fellowship in Neuroscience at UCL.

DISCLOSURE

Over the last three years SSS has received compensation from Glaxo Smith Kline. The remaining authors declare no conflict of interest.

Author contributions: SE and BBA designed the experiment. SE and SSS carried out the behavioral testing and administered drugs. SE, SSS, and BBA carried out data analysis and wrote the article.

REFERENCES

Andari E, Duhamel JR, Zalla T, Herbrecht E, Leboyer M, Sirigu A (2010). Promoting social behavior with oxytocin in high-functioning autism spectrum disorders. *Proc Natl Acad Sci USA* **107**: 4389–4394.

Aragona BJ, Liu Y, Curtis JT, Stephan FK, Wang Z (2003). A critical role for nucleus accumbens dopamine in partner-preference formation in male prairie voles. *J Neurosci* **23**: 3483–3490.

Aragona BJ, Liu Y, Yu YJ, Curtis JT, Detwiler JM, Insel TR *et al* (2006). Nucleus accumbens dopamine differentially mediates the formation and maintenance of monogamous pair bonds. *Nat Neurosci* **9**: 133–139.

Averbeck BB (2010). Oxytocin and the salience of social cues. *Proc Natl Acad Sci USA* **107**: 9033–9034.

Averbeck BB, Duchaine B (2009). Integration of social and utilitarian factors in decision making. *Emotion* **9**: 599–608.

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

Baumgartner T, Heinrichs M, Vonlanthen A, Fischbacher U, Fehr E (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron* **58**: 639–650.

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

Di Simplicio M, Massey-Chase R, Cowen PJ, Harmer CJ (2009). Oxytocin enhances processing of positive versus negative emotional information in healthy male volunteers. *J Psychopharmacol* **23**: 241–248.

Djamshidian A, Jha A, O'Sullivan SS, Silveira-Moriyama L, Jacobson C, Brown P *et al* (2010). Risk and learning in impulsive and nonimpulsive patients with Parkinson's disease. *Movement Disorders*; e-pub ahead of print.

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

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

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

Donaldson ZR, Young LJ (2008). Oxytocin, vasopressin, and the neurogenetics of sociality. *Science* **322**: 900–904.

Ekman P, Friesen WV (1971). Constants across cultures in the face and emotion. *J Pers Soc Psychol* **17**: 124–129.

Fahrbach SE, Morrell JI, Pfaff DW (1985). Possible role for endogenous oxytocin in estrogen-facilitated maternal behavior in rats. *Neuroendocrinology* **40**: 526–532.

Ferguson JN, Aldag JM, Insel TR, Young LJ (2001). Oxytocin in the medial amygdala is essential for social recognition in the mouse. *J Neurosci* **21**: 8278–8285.

Fischer-Shofty M, Shamay-Tsoory SG, Harari H, Levkovitz Y (2010). The effect of intranasal administration of oxytocin on fear recognition. *Neuropsychologia* **48**: 179–184.

Fleming AS, Anderson V (1987). Affect and nurturance: mechanisms mediating maternal behavior in two female mammals. *Prog Neuropsychopharmacol Biol Psychiatry* **11**: 121–127.

Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc Natl Acad Sci USA* **104**: 16311–16316.

Frank MJ, Seeberger LC, O'Reilly R C (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science* **306**: 1940–1943.

Gamer M, Zurowski B, Buchel C (2010). From the cover: different amygdala subregions mediate valence-related and attentional effects of oxytocin in humans. *Proc Natl Acad Sci USA* **107**: 9400–9405.

Gingrich B, Liu Y, Cascio C, Wang Z, Insel TR (2000). Dopamine D2 receptors in the nucleus accumbens are important for social attachment in female prairie voles (Microtus ochrogaster). *Behav Neurosci* **114**: 173–183.

Guastella AJ, Einfeld SL, Gray KM, Rinehart NJ, Tonge BJ, Lambert TJ *et al* (2010). Intranasal oxytocin improves emotion recognition for youth with autism spectrum disorders. *Biol Psychiatry* **67**: 692–694.

Guastella AJ, Mitchell PB, Dadds MR (2008a). Oxytocin increases gaze to the eye region of human faces. *Biol Psychiatry* **63**: 3–5.

Guastella AJ, Mitchell PB, Mathews F (2008b). Oxytocin enhances the encoding of positive social memories in humans. *Biol Psychiatry* **64**: 256–258.

Hayden BY, Parikh PC, Deaner RO, Platt ML (2007). Economic principles motivating social attention in humans. *Proc Biol Sci* **274**: 1751–1756.

Holmes AP, Friston KJ (1998). Generalisability, random effects and population inference. *Neuroimage* **7**: s754.

Hurlemann R, Patin A, Onur OA, Cohen MX, Baumgartner T, Metzler S *et al* (2010). Oxytocin enhances amygdala-dependent, socially reinforced learning and emotional empathy in humans. *J Neurosci* **30**: 4999–5007.

Insel TR, Young LJ (2001). The neurobiology of attachment. *Nat Rev Neurosci* **2**: 129–136.

Izuma K, Saito DN, Sadato N (2008). Processing of social and monetary rewards in the human striatum. *Neuron* **58**: 284–294.

Kemp AH, Guastella AJ (2010). Oxytocin: prosocial behavior, social salience, or approach-related behavior? *Biol Psychiatry* **67**: e33–e34.

Kirsch P, Esslinger C, Chen Q, Mier D, Lis S, Siddhanti S *et al* (2005). Oxytocin modulates neural circuitry for social cognition and fear in humans. *J Neurosci* **25**: 11489–11493.

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

Lee A, Clancy S, Fleming AS (1999). Mother rats bar-press for pups: effects of lesions of the mpoa and limbic sites on maternal behavior and operant responding for pup-reinforcement. *Behav Brain Res* **100**: 15–31.

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

Marsh AA, Yu HH, Pine DS, Blair RJ (2010). Oxytocin improves specific recognition of positive facial expressions. *Psychopharmacology (Berl)* **209**: 225–232.

Mayer JD, Gaschke YN (1988). The experience and meta-experience of mood. *J Pers Soc Psychol* **55**: 102–111.

Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* **442**: 1042–1045.

Petrovic P, Kalisch R, Singer T, Dolan RJ (2008). Oxytocin attenuates affective evaluations of conditioned faces and amygdala activity. *J Neurosci* **28**: 6607–6615.

Rimmele U, Hediger K, Heinrichs M, Klaver P (2009). Oxytocin makes a face in memory familiar. *J Neurosci* **29**: 38–42.

Savaskan E, Ehrhardt R, Schulz A, Walter M, Schachinger H (2008). Post-learning intranasal oxytocin modulates human memory for facial identity. *Psychoneuroendocrinology* **33**: 368–374.

Schorscher-Petcu A, Dupre A, Tribollet E (2009). Distribution of vasopressin and oxytocin binding sites in the brain and upper spinal cord of the common marmoset. *Neurosci Lett* **461**: 217–222.

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

Singer T, Snozzi R, Bird G, Petrovic P, Silani G, Heinrichs M *et al* (2008). Effects of oxytocin and prosocial behavior on brain responses to direct and vicariously experienced pain. *Emotion* **8**: 781–791.

Skuse DH, Gallagher L (2009). Dopaminergic-neuropeptide interactions in the social brain. *Trends Cogn Sci* **13**: 27–35.

Tsitsiklis JN (1994). A short proof of the gittins index theorem. *The Annals of Applied Probability* **4**: 194–199.

Vuong C, Van Uum SH, O'Dell LE, Lutfy K, Friedman TC (2010). The effects of opioids and opioid analogs on animal and human endocrine systems. *Endocr Rev* **31**: 98–132.

Williams JR, Insel TR, Harbaugh CR, Carter CS (1994). Oxytocin administered centrally facilitates formation of a partner preference in female prairie voles (Microtus ochrogaster). *J Neuroendocrinol* **6**: 247–250.

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

Zink CF, Tong Y, Chen Q, Bassett DS, Stein JL, Meyer-Lindenberg A (2008). Know your place: neural processing of social hierarchy in humans. *Neuron* **58**: 273–283.