DEPARTMENT OF MANAGEMENT

Oxytocin does not make a face appear more trustworthy but improves the accuracy of trustworthiness judgments

Bruno Lambert, Carolyn H. Declerck & Christophe Boone

UNIVERSITY OF ANTWERP
Faculty of Applied Economics

Stadscampus
Prinsstraat 13, B.226
BE-2000 Antwerpen
Tel. +32 (0)3 265 40 32
Fax +32 (0)3 265 47 99
www.ua.ac.be/tew

FACULTY OF APPLIED ECONOMICS

DEPARTMENT OF MANAGEMENT

Oxytocin does not make a face appear more trustworthy but improves the accuracy of trustworthiness judgments

Bruno Lambert, Carolyn H. Declerck & Christophe Boone

RESEARCH PAPER 2013-011
JUNE 2013

University of Antwerp, City Campus, Prinsstraat 13, B-2000 Antwerp, Belgium
Research Administration – room B.226
phone: (32) 3 265 40 32
fax: (32) 3 265 47 99
e-mail: joeri.nys@ua.ac.be

The papers can be also found at our website:
www.ua.ac.be/tew (research > working papers) &
www.repec.org/ (Research papers in economics - REPEC)

D/2013/1169/011

Oxytocin does not make a face appear more trustworthy but improves the accuracy of trustworthiness judgments

Bruno Lambert*;
Department of Management and Department of Medical Sciences, University of Antwerp, Antwerp, Belgium

Carolyn H. Declerck;
Department of Management, University of Antwerp, Antwerp, Belgium

Christophe Boone;
Department of Management, University of Antwerp, Antwerp, Belgium

This paper has been presented at the NeuroPsychoEconomics Conference of 2013.

* Bruno Lambert; S.Z.407, Prinsstraat 13, 2000 Antwerp, Belgium;
Bruno.Lambert@ua.ac.be; Tel: +32 (0) 3 265 50 95; Fax: +32 (0) 3 265 50 79

Abstract:
Previous research on the relation between oxytocin and trustworthiness evaluations has yielded inconsistent results. The current study reports an experiment using artificial faces which allows manipulating the dimension of trustworthiness without changing factors like emotions or face symmetry. We investigate whether (1) oxytocin increases the average trustworthiness evaluation of faces (level effect), and/or whether (2) oxytocin improves the discriminatory ability of trustworthiness perception so that people become more accurate in distinguishing faces that vary along a gradient of trustworthiness.

In a double blind oxytocin/placebo experiment (N = 106) participants conducted two judgment tasks. First they evaluated the trustworthiness of a series of pictures of artificially generated neutral faces. Next they compared neutral faces with artificially generated faces that were manipulated to vary in trustworthiness.

The results indicate that oxytocin (relative to a placebo) does not affect the evaluation of trustworthiness in the first task. However, in the second task, misclassification of untrustworthy faces as trustworthy occurred significantly less in the oxytocin group. Furthermore, oxytocin improved the discriminatory ability of untrustworthy, but not trustworthy faces. We conclude that oxytocin does not increase trustworthiness judgments on average, but that it helps people to more accurately recognize an untrustworthy face.

Keywords: Oxytocin; Perceived Trustworthiness; Social Perception; Face Evaluation

1. Introduction

Oxytocin (OT), the hormone well-known for its involvement in parturition and lactation, has recently seen a surge of interest with respect to its role in regulating social behaviour. Among humans, its prosocial effects have become well-documented. For example, after intranasal administration of OT, people tend to be more trusting (Kosfeld et al., 2005), more generous (Zak et al., 2007), and more cooperative towards in-group members (Kosfeld et al., 2005; De Dreu et al., 2010; De Dreu et al., 2011). However, the relation between OT and social behaviour does not appear to be straightforward, and when no social information is available, OT even makes people more cautious and uncooperative (Declerck et al., 2010).

Given the complex relationship between OT and social functions, some authors have suggested that the multitude of reported influences on social cognition and behaviour might be the ultimate result of a few more basal processes that are influenced by OT, such as perception, motivation, and anxiety (Churchland and Winkielman, 2012). One of the well-established facts regarding the neural functions of OT is that it reduces amygdala activation when it is exogenously administered (Kirsch et al., 2005). This in turn lowers social anxiety, which appears to facilitate trust (Baumgartner et al., 2008) and social approach behaviour (Kemp and Guastella, 2011). However, trusting behaviour is not only a function of reduced anxiety, but it is also moderated by the perception of trustworthiness (Adolphs, 2003; Frith and Frith, 2006; Krumhuber et al., 2007). In fact, perceptions of the trustworthiness of the partner have already been shown to matter greatly in the relation between OT and trust-related or cooperative behaviours (Mikolajczak et al., 2010). Thus, to fully understand how OT affects prosocial behaviour, given that it lowers social anxiety, one also needs to understand how OT affects the perception of trustworthiness.

This study addresses if and how OT influences the perception of faces that have been manipulated to vary on the dimension of trustworthiness. We investigate (1) if OT has a main level-effect on the perception of trustworthiness by which it would cause neutral faces to be perceived as more trustworthy, and (2) if OT affects the discriminatory ability of people who are asked to judge faces that vary only in their dimension of trustworthiness.

Previous research that has examined the relation between OT and the perception of trustworthiness has yielded inconsistent results. In the study by Theodoridou et al. (2009) participants who received OT judged faces as more trustworthy compared to those who received a placebo. In contrast, other studies found no significant effect of OT on trustworthiness evaluations (Guastella et al., 2008; Rimmele et al., 2009). These studies, however, used pictures of real faces, making it difficult to isolate the dimension of trustworthiness. Other factors such as facial symmetry are difficult to control for in natural facial expressions and may be accidentally introduced as confounds between comparison groups. Also emotional expressions (which may interact with the perception of trustworthiness) are difficult to exclude in real faces. To avoid some of these problems the current study makes use of artificially generated faces that have been validated in previous research (Oosterhof and Todorov, 2008; De Dreu et al., 2012). Therefore, the first objective of the current study is to replicate the above research and test if OT increases, on average, trustworthiness judgments of artificially generated, neutral faces that are devoid of recognizable emotional expressions.

An alternative means by which OT might influence the perception of trustworthiness of faces is by refining the perceptual accuracy of people's judgments, so that OT facilitates discriminating faces that vary in increments along the dimension of trustworthiness. Previous

studies have already revealed greater accuracy and increased processing speed in people who received OT compared to placebo, leading to better recognition of mental states and emotional expressions (Domes et al., 2007; Schulze et al., 2011; Lischke et al., 2012a; van Ijzendoorn and Bakermans-Kranenburg, 2012; Fischer-Shofty et al., 2013). However, as far as we know, none of these studies has experimentally manipulated the gradient along which perceptual judgments were tested. Therefore, the second objective of this study is to investigate if OT improves the recognition of trustworthiness of artificially generated faces that vary incrementally from less trustworthy to more trustworthy, but are otherwise emotionally neutral. Based on previous research that suggests that OT makes people more cautious (Striepens et al., 2012), we hypothesize that participants who received OT will make less mistakes in classifying faces, especially those that are perceived to be untrustworthy.

2. Methods

2.1 Participants

We recruited participants by e-mail and invited them to participate in a behavioural experiment that evaluated the effects of a hormone on on evaluative judgments. A total of 112 students of the University of Antwerp registered to participate in exchange for monetary remuneration. We used the results of 106 participants (61 females, 45 males; mean age = 22, S.D. = 2.5) in subsequent analyses: five individuals were deleted because they did not correctly complete the experimental task and one participant did not sufficiently commanded the Dutch language in which the study was conducted.

Inclusion criteria for participation included the abstinence of alcohol and nicotine 12h, and the use of medication other than anti-conception 24h prior to the study. Participants were free of neurological or psychological disorders, and had no nasal obstruction or colour vision deficiency. To exclude the administration of OT to pregnant women, we distributed a pregnancy test to all female participants, which they took anywhere between 1 and 48 hours prior to the experiment.

All participants gave written informed consent to the study procedures which were in accordance with the Declaration of Helsinki and were approved by the Ethical Commission of the University of Antwerp. Debriefing occurred by sending participants an e-mail referring them to a website where the methods, results and conclusions were explained.

Participants received a show-up fee of €10 which was increased with the earnings from an interactive game which was held at the end of the experiment[1]. Mean profit was €18.17.

2.2 Sessions

Seventeen sessions were held in computer rooms with no less than 4 and no more than 10 participants in each session. All sessions took place between 0945h and 1500h and took around 75 minutes to complete. Face to face contact between participants was kept to a minimum and no conversations were allowed during the experimental tasks.

2.3 Procedure

[1] As an additional manipulation check to determine if OT was effective in this experiment, we replicated Kosfeld et al. (2005) by testing if OT affected investment in a trust game. The results, indicating a significant effect of OT on trusting behaviour but only for cautious individuals, can be obtained from the corresponding author.

Participants were instructed to self-administer an intranasal dose of 24 IU OT (Syntocinon, Novartis; tree puffs per nostril with one minute in-between puffs) or placebo following a double-blind random design. The placebo contained the same active ingredients except for OT and was prepared by the pharmacy of the University Hospital of Antwerp.

After inhalation, participants waited 35 minutes before starting the actual experimental task. Meanwhile, the participants completed a trial version of the experimental task. The stimuli in this trial version also comprised artificially generated neutral faces, but with different facial identities from those used in the experimental task. The room was darkened to provide optimal viewing conditions and to reduce visual distraction and glare.

The task took around 20 minutes to complete. Afterwards, the participants played an interactive game and filled in a post-experimental questionnaire.

2.4 Experimental Paradigm

Participants were asked to evaluate the trustworthiness of two series of pictures showing artificially created faces displayed on the screen[2]. The software used to present the stimuli and to record the evaluation scores was Affect version 4.0 (Spruyt et al., 2010). The faces were selected out of a database of 175 faces that were created to vary in trustworthiness. The dataset was created by Oosterhof and Todorov (2008) using FaceGen Modeller version 3.1 (Singular Inversions, 2007). To create artificial faces that express different levels of trustworthiness, the authors relied on pictures of real faces against which the artificial faces were validated. To make sure that the colours of the faces did not interfere with judgement, the L*a*b* values of the faces were changed so that the mean values were equal between pictures[3]. The pictures were shown on a black background. To evaluate trustworthiness, participants used a left-mouse click to assign a score on a digital scale shown at the bottom of the computer screen. The scoring was self-paced. A fixation cross lasting 1.5s was shown between each trial and there was a one minute interval between the two series.

The first series (testing objective 1) consisted of five neutral faces each with a different identity and displayed one by one. Participants were asked to indicate on a scale ranging from 0 (not at all trustworthy) to 9 (very trustworthy) how trustworthy they thought the person was.

The second series (testing objective 2) comprised 35 trials (five face identities times seven variations of trustworthiness) in which two faces were displayed on the computer screen at the same time. The left face was always a neutral face, while the right face, depicting the same identity, was manipulated to vary in trustworthiness. To manipulate trustworthiness, we used seven variations of each face that varied from untrustworthy to trustworthy. These variations were included in the data set created by Oosterhof and Todorov (2008). The participants were asked to give a score ranging from -4 (right face more

[2] Two additional series of pictures were also included in this experiment which were manipulated by the author to vary in facial redness. These pictures were to be evaluated on perceived health. The order in which the trustworthiness and health series were shown was random. At the end of the experiment, the participants also evaluated pictures of scenes on the level of disgust. The results pertaining health and disgust evaluations will not be further elaborated on in this paper because these data were collected in order to test different hypotheses that are unrelated to the current one.

[3] L*, a* and b* are colour dimensions in the CIE 1976 colour space determining the lightness, the redness (in contrast to green) and the blueness (in contrast to yellow).

untrustworthy than left face) to +4 (right face more trustworthy than left face) with 0 indicating no difference between the left and the right face.

3. Results

To validate the trustworthiness scale created by Oosterhof and Todorov (2008) for the current study, we first averaged all of the participants' trustworthiness scores assigned to each of the neutral faces. Figure 1 shows that four out of the five faces were perceived to be neutral (score five on a scale from 0 to 9). Because face 3 deviated significantly from the four other scores we excluded this stimulus from further analysis. Next, we plotted the averaged perceived trustworthiness scores for each of the stimuli of the second series (comparison between a neutral and a manipulated face) relative to the trustworthiness scale dimension (see fig. 2). Visual inspection shows that, in accordance with the results of Oosterhof and Todorov (2008), the relation is linear. More importantly, we notice no apparent effect of OT on perceived trustworthiness above the effect of the scale dimension. To test this statistically, we pooled the data of the four face identities for each participant and conducted a regression analysis on the perceived scores. The independent variables in this regression are the trustworthiness dimension (ranging from -3 to 3), the treatment (1 = OT, 0 = placebo) and sex of the participant. The latter is included because recent publications indicate that sex can be a moderator on the effect of OT (Domes et al., 2010; Lischke et al., 2012b). The trustworthiness dimension appears to be the sole predictor for perceived trustworthiness ($B = 0.534$, Std. Error = 0.015, $p < 0.001$).

To test the main-level effect of treatment (OT versus placebo) on the perception of neutral faces (objective 1), we conducted a 4 (face identities) * 2 (treatment) * 2 (sex) repeated ANOVA analysis on perceived trustworthiness with face identity as a within-subject factor. No significant effect of face identity, treatment or sex was found, neither did any of the interactions between these variables prove to be significant.

To test if OT affects the accuracy of trustworthiness perception (objective 2), we compared the number of misclassifications made between the OT and the placebo group. A misclassification is defined as giving a positive score to a face that is registered as untrustworthy or vice versa. Table 1 shows that the number of misclassifications versus correctly evaluated faces differed by treatment, but only in the case of untrustworthy faces: relatively less mistakes were made in the OT-group (Fisher's exact test; $p = 0.0062$). When trustworthy faces were evaluated, or when trustworthy and untrustworthy faces were pooled, OT had no significant influence (Fisher's exact test; $p = 0.65$).

As a robustness check, we conducted regression analyses on the number of misclassifications of each participant (see table 2). We fitted six different negative binomial models:

First, we conducted a regression analysis on the number of untrustworthy faces misclassified as trustworthy and again we took sex (female = 0, male = 1) into account as a possible moderator. Model 1 shows a significant effect of OT ($B = -0.46$, p-value = 0.04). Model 2 indicates that OT does not interact significantly with sex. Second, we did the same for the misclassifications of trustworthy faces (model 3 and 4) but this did not yield any significant results. Third, we pooled the trustworthy and untrustworthy faces and structured the data in panel form. We included a dummy regressor indicating if the face is untrustworthy (coded 0) or trustworthy (coded 1). Model 5 shows that the main effect of treatment was not significant. Finally, in model 6 we investigate the interaction effect of OT and trustworthiness. This interaction is found to be significant ($B = 0.57$, p-value = 0.03) which corroborates that the effect of OT depends on whether the face is trustworthy or untrustworthy. Because the

amount of misclassifications for untrustworthy and trustworthy faces were both overdispersed ($D_{untrustworthy}$ = 2.29; $D_{trustworthy}$ = 2.25), we fitted a negative binomial in each model using the statistical package Stata 9 (StataCorp, 2011).

4. Discussion

Two conclusions can be drawn from these data. First, the results indicate that OT, compared to placebo, does not, on average, improve the evaluation of the trustworthiness of faces. Participants who were given a single intranasal dose of 24 IU OT did not perceive an artificially generated, neutral face as being more trustworthy than participants who received a placebo. Second, the accuracy to discriminate between trustworthy and untrustworthy faces is significantly improved by OT: an untrustworthy face was misclassified as trustworthy less often in the OT group relative to the placebo group.

The absence of a main effect of OT on trustworthiness perception contradicts the results obtained by Theodoridou et al. (2009), who reported that participants in their experiment rated pictures of neutral faces as more trustworthy and attractive if the participants received OT rather than placebo. A first and straightforward explanation for the different results between the two studies may be the use of different facial stimuli. While Theodoridou et al. (2009) used pictures of real faces, we used pictures of computer generated faces. Although these faces captured variations in trustworthiness in a consistent way, they may have been perceived to be artificial and lacking a social component because they are not *'real'*. Previous research already points to the importance of subtle social information on the effect of OT. The study of Declerck et al. (2010) indicates that face to face contact with a partner is an important moderator in the relation between OT and trusting behaviour. This is also shown in the study of Mikolajczak et al. (2010), who found that people would not plainly trust more when they received OT, but that instead OT made them more considerate for the information on vignettes describing their partner. The absence of "real" information in an artificial face could have rendered the functions of OT to be obsolete.

Alternatively, the absence of a main effect of OT on trustworthiness perception could possibly be attributed to the moderating effect of individual differences. This would mean that the marginal effect of OT would be greater for people that are either low in endogenous OT, or for individuals lacking certain social skills. For example, Declerck et al. (2013) found that OT does not affect overall levels of cooperation in a prisoner's dilemma game, but that it is dependent on a three-way interaction: OT only significantly boosted cooperative behaviour of individuals who were a-priori classified to have a proself value orientation, and only when they had enjoyed prior contact with their partners. Similarly, the results of Fischer-Shofty et al. (2013) showed that the effect of OT on recognition of kinship and intimacy is only apparent in schizophrenia patients who are less socially competent and not in a control group of healthy people.

Finally, we note that the effect of OT on perception may be more subtle and difficult to detect when perception is decoupled from a behavioural goal. For example, De Dreu et al. (2012) showed that, while OT does not change the perception of threat (a combination of trustworthiness and dominance) displayed by faces, it may still influence behaviour. Males who received OT chose for faces expressing high threat to be their allies in a competitive setting while the placebo group chose faces with low threat. In other studies where perception does not hinge on an experimental tasks, null effects of OT have been reported several times. Guastella et al. (2008) asked participants to indicate their perceived trustworthiness of neutral, happy and angry faces. Rimmele et al. (2009) investigated how willingly participants approached faces with negative, neutral or positive emotional expression. In neither study was an effect of OT observed.

The finding that OT improves the detection of untrustworthy faces is compatible with a recent study (Striepens et al., 2012) that reported a facilitated acoustic startle reflex as well as improved memory for *negative stimuli* in response to exogenous OT administration. However, they found that the valence ratings of affective loaded pictures were not affected by OT. This is consistent with the proposition that OT is unlikely to influence perception unless the stimuli are environmentally salient. This is also substantiated by a recent study of Stallen et al. (2012): when participants were asked to evaluate the attractiveness of a series of symbols, OT only affected the evaluative scores when additional information was provided regarding the ratings of other people that were either on the same, or on another team. Given that people like to conform to others with whom they associate (the ingroup), OT influenced ratings only in those conditions when the in- and outgroup opinion differed, in which case it facilitated a conformist rating to the ingroup. OT did not influence the evaluative scores when no ratings of other people were available, or when the in- and outgroup provided similar ratings. In the latter case, the information loses saliency with respect to the desire to conform to the ingroup. Participants still conformed to the ratings, but no additional effect of OT (relative to placebo) was noted.

The bias to conform to one's ingroup (in the study by Stallen et al. (2012)), the increased sensitivity to negative stimuli (in the study by Striepens et al. (2012)) or to untrustworthy faces (in the current study) under influence of OT can be explained by Error Management Theory (Haselton and Nettle, 2006): throughout evolution an adaptive error bias has emerged which minimizes the number of false-positive or false-negative to assure the lowest cost to survival. In a social situation, there is less risk involved when conforming to the in-group or distrusting a trustworthy person than when conforming to the out-group or mistakenly trusting an untrustworthy face. Hence people should heuristically avoid misclassifying untrustworthy faces. Not surprisingly it is this type of error that appears to be reduced after OT administration: we found that the difference in accuracy of classification was significant for untrustworthy faces, but not for trustworthy.

A similar result was obtained by Di Simplicio et al. (2009). When they asked participants to label the emotion expressed by faces as fast and accurate as possible, they found that participants who received OT misclassified the emotion of "surprise" less often than those in the placebo group. Surprise is an emotion that is expressed when something in the environment does not fit the expectations. Detection of surprise in other people can possibly make people more aware of relevant changes or potential danger. Therefore, the improved detection of surprise in others may have adaptive value.

Together with our current findings that OT improves the detection of untrustworthiness, these findings fit the general conclusion of Striepens et al. (2012) that OT is promoting heightened caution, rather than trust.

In summary, the results of the current study do not support the notion that OT indiscriminately improves perception of trustworthiness. OT does not make people gullible. The effect of OT is more subtle: when perception is decoupled from an a actual task, OT may still facilitate awareness of social information, but only if it is relevant with respect to survival. Thus OT helps people to discriminate between trustworthy and untrustworthy faces while it does not need to change the evaluation of trustworthiness. Future research should continue to investigate the role of social information, environmental saliency, and individual differences when examining the relation between OT and perception.

Funding body agreements and policies
This research was funded by an ID-BOF grant (5340) from the University of Antwerp.
Conflict of Interest
All authors declare that they have no conflicts of interest.
Acknowledgements
We would like to thank Sofie Laureyssens, Eline Swolfs and Anja Waegeman for their assistance with supervising the experiments.
Contributors
All authors contributed equally to the design of the experiment, the analysis of the results and the writing of the manuscript. Bruno Lambert collected the data.

5. References

Adolphs, R., 2003. Cognitive neuroscience of human social behaviour. Nature Reviews Neuroscience 4, 165-178.

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.

Churchland, P.S., Winkielman, P., 2012. Modulating social behavior with oxytocin: How does it work? What does it mean? Hormones and Behavior 61, 392-399.

De Dreu, C.K.W., Greer, L.L., Handgraaf, M.J.J., Shalvi, S., Van Kleef, G.A., 2012. Oxytocin modulates selection of allies in intergroup conflict. Proc. R. Soc. B-Biol. Sci. 279, 1150-1154.

De Dreu, C.K.W., Greer, L.L., Handgraaf, M.J.J., Shalvi, S., Van Kleef, G.A., Baas, M., Ten Velden, F.S., Van Dijk, E., Feith, S.W.W., 2010. The Neuropeptide Oxytocin Regulates Parochial Altruism in Intergroup Conflict Among Humans. Science 328, 1408-1411.

De Dreu, C.K.W., Greer, L.L., Van Kleef, G.A., Shalvi, S., Handgraaf, M.J.J., 2011. Oxytocin promotes human ethnocentrism. Proceedings of the National Academy of Sciences of the United States of America 108, 1262-1266.

Declerck, C.H., Boone, C., Kiyonari, T., 2010. Oxytocin and cooperation under conditions of uncertainty: The modulating role of incentives and social information. Hormones and Behavior 57, 368-374.

Declerck, C.H., Boone, C., Kiyonari, T., 2013. The effect of oxytocin on cooperation in a prisoner's dilemma depends on the social context and a person's social value orientation. Social Cognitive and Affective Neuroscience.

Di Simplicio, M., Massey-Chase, R., Cowen, P.J., Harmer, C.J., 2009. Oxytocin enhances processing of positive versus negative emotional information in healthy male volunteers. Journal of Psychopharmacology 23, 241-248.

Domes, G., Heinrichs, M., Michel, A., Berger, C., Herpertz, S.C., 2007. Oxytocin improves "mind-reading" in humans. Biological Psychiatry 61, 731-733.

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

Fischer-Shofty, M., Brüne, M., Ebert, A., Shefet, D., Levkovitz, Y., Shamay-Tsoory, S.G., 2013. Improving social perception in schizophrenia: The role of oxytocin. Schizophrenia Research 146, 357-362.

Frith, C.D., Frith, U., 2006. How we predict what other people are going to do. Brain Res. 1079, 36-46.

Guastella, A.J., Mitchell, P.B., Mathews, F., 2008. Oxytocin enhances the encoding of positive social memories in humans. Biological Psychiatry 64, 256-258.

Haselton, M.G., Nettle, D., 2006. The Paranoid Optimist: An Integrative Evolutionary Model of Cognitive Biases. Personality and Social Psychology Review 10, 47-66.
Kemp, A.H., Guastella, A.J., 2011. The Role of Oxytocin in Human Affect: A Novel Hypothesis. Current Directions in Psychological Science 20, 222-231.
Kirsch, P., Esslinger, C., Chen, Q., Mier, D., Lis, S., Siddhanti, S., Gruppe, H., Mattay, V.S., Gallhofer, B., Meyer-Lindenberg, A., 2005. Oxytocin Modulates Neural Circuitry for Social Cognition and Fear in Humans. The Journal of Neuroscience 25, 11489-11493.
Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U., Fehr, E., 2005. Oxytocin increases trust in humans. Nature 435, 673-676.
Krumhuber, E., Manstead, A.S.R., Cosker, D., Marshall, D., Rosin, P.L., Kappas, A., 2007. Facial dynamics as indicators of trustworthiness and cooperative Behavior. Emotion 7, 730-735.
Lischke, A., Berger, C., Prehn, K., Heinrichs, M., Herpertz, S.C., Domes, G., 2012a. lntranasal oxytocin enhances emotion recognition from dynamic facial expressions and leaves eye-gaze unaffected. Psychoneuroendocrinology 37, 475-481.
Lischke, A., Gamer, M., Berger, C., Grossmann, A., Hauenstein, K., Heinrichs, M., Herpertz, S.C., Domes, G., 2012b. Oxytocin increases amygdala reactivity to threatening scenes in females. Psychoneuroendocrinology 37, 1431-1438.
Mikolajczak, M., Gross, J.J., Lane, A., Corneille, O., de Timary, P., Luminet, O., 2010. Oxytocin Makes People Trusting, Not Gullible. Psychological Science 21, 1072-1074.
Oosterhof, N.N., Todorov, A., 2008. The functional basis of face evaluation. Proceedings of the National Academy of Sciences 105, 11087-11092.
Rimmele, U., Hediger, K., Heinrichs, M., Klaver, P., 2009. Oxytocin Makes a Face in Memory Familiar. Journal of Neuroscience 29, 38-42.
Schulze, L., Lischke, A., Greif, J., Herpertz, S.C., Heinrichs, M., Domes, G., 2011. Oxytocin increases recognition of masked emotional faces. Psychoneuroendocrinology 36, 1378-1382.
Singular Inversions. (2007) [Computer Program] Facegen Main Software Development Kit. (Version 3.1) Vancouver, BC, Canada.
Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F., Hermans, D., 2010. Affect 4.0: A Free Software Package for Implementing Psychological and Psychophysiological Experiments. Experimental psychology 57, 36-45.
Stallen, M., De Dreu, C.K.W., Shalvi, S., Smidts, A., Sanfey, A.G., 2012. The Herding Hormone: Oxytocin Stimulates In-Group Conformity. Psychological Science 23, 1288-1292.
StataCorp. (2011) [Computer Program] Stata Statistical Software. (Version 12) College Station, TX: StataCorp LP.
Striepens, N., Scheele, D., Kendrick, K.M., Becker, B., Schäfer, L., Schwalba, K., Reul, J., Maier, W., Hurlemann, R., 2012. Oxytocin facilitates protective responses to aversive social stimuli in males. Proceedings of the National Academy of Sciences 109, 18144-18149.
Theodoridou, A., Rowe, A., PentonVoak, I., Rogers, P., 2009. Oxytocin and social perception: Oxytocin increases perceived facial trustworthiness and attractiveness. Hormones and Behavior 56, 128-132.
van Ijzendoorn, M.H., Bakermans-Kranenburg, M.J., 2012. A sniff of trust: Meta-analysis of the effects of intranasal oxytocin administration on face recognition, trust to in-group, and trust to out-group. Psychoneuroendocrinology 37, 438-443.
Zak, P.J., Stanton, A.A., Ahmadi, S., 2007. Oxytocin Increases Generosity in Humans. PLoS ONE 2, e1128.

Figures and tables:

Table 1: *Fisher's exact tests on number of misclassifications*

	All faces		**Untrustworthy faces**		**Trustworthy faces**	
	OT	Placebo	OT	Placebo	OT	Placebo
Correct	1239	1043	631	515	608	528
Misclassified	129	133	53	73	76	60
odds ratio	1.22		1.69		0.91	
p-value	0.13		0.0062		0.65	

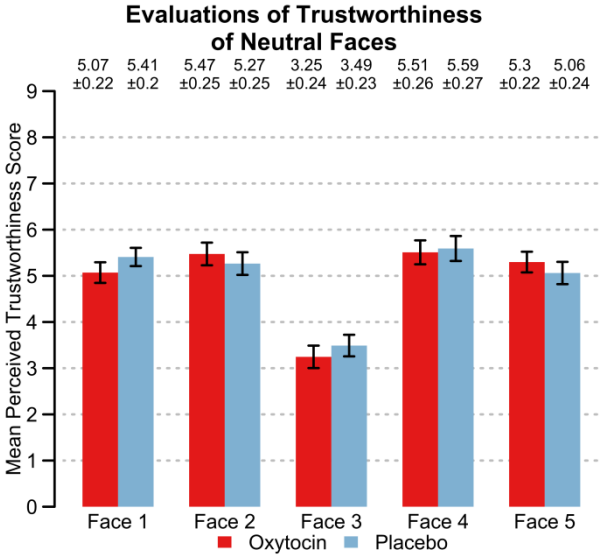
The number of misclassifications of untrustworthy faces, trustworthy faces and all faces and the results of the Fisher's exact tests.

Table 2: *Negative Binomial Regression Models on the Number of Misclassifications.*

	Model 1: untrustworthy faces only	**Model 2: untrustworthy faces and sex*OT interaction**	**Model 3: trustworthy faces only**	**Model 4: trustworthy faces and sex*OT interaction**	**Model 5: panel data main effect**	**Model 6: panel data interaction**
Oxytocin	-0.46 (0.26) p = 0.04	-0.51 (0.33) p = 0.06	0.07 (0.24) p = 0.39	0.02 (0.35) p = 0.48	-0.18 (0.22) p = 0.21	0.10 (0.27) p = 0.36
Trustworthiness	-	-	-	-	0.09 (0.15) p = 0.29	-0.20 (0.21) p = 0.18
Sex	0.02 (0.26) p = 0.47	-0.03 (0.35) p = 0.46	-0.03 (0.24) p = 0.45	-0.09 (0.35) p = 0.40	0.01 (0.22) p = 0.48	0.03 (0.23) p = 0.45
Oxytocin*Sex	-	0.13 (0.54) p = 0.40	-	0.11 (0.49) p = 0.41	-	-
Oxytocin* Trustworthiness	-	-	-	-	-	0.57 (0.30) p = 0.03
Constant	0.38 (0.22) p = 0.04	0.41 (0.22) p = 0.03	0.23 (.22) p = 0.16	0.25 (0.28) p = 0.18	1.20 (0.62) p = 0.03	1.21 (0.67) p = 0.04
N	106	106	106	106	212	212

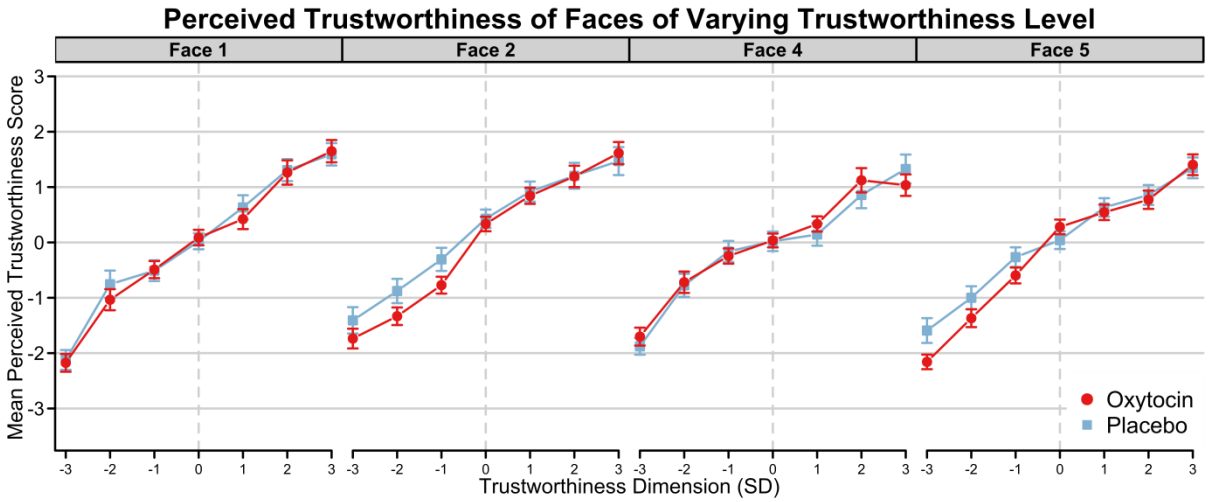
Unstandardized regression coefficients of the negative binomial regression analyses on the number of misclassifications. Standard errors are given in parentheses. All p-values are one-tailed . Oxytocin (OT; coded 1; placebo = 0) and trustworthiness of the faces (untrustworthy = 0; trustworthy = 1) are the predictor variables of interest. Sex (female = 0; male = 1) is added to the models as control variable and is tested as a possible moderator on the effect of oxytocin. Participants (106) evaluated three trustworthy and three untrustworthy variations of four different face identities.

Figure 1: *Evaluations of Trustworthiness of Neutral Faces*



All 106 participants evaluated the trustworthiness of five neutral faces of different identity (Face 1 to 5; obtained from the dataset created by Oosterhof and Todorov (2008)). Error bars represent the standard error of the mean.

Figure 2*: Perceived Trustworthiness of Faces of Varying Trustworthiness Level*



All 106 participants evaluated 35 faces in comparison to a neutral face of the same face identity and gave them a score between -4 and +4 on perceived trustworthiness (not whole range is depicted). Mean scores with standard error are depicted for each value of the trustworthiness dimension (ranging from -3 to 3, see Oosterhof and Todorov (2008)).