Psychoneuroendocrinology (2012) 37, 438—443

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/psyneuen

SHORT COMMUNICATION

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

Marinus H. Van IJzendoorn *, Marian J. Bakermans-Kranenburg

Centre for Child and Family Studies, Leiden University, The Netherlands

Received 11 April 2011; received in revised form 29 June 2011; accepted 5 July 2011

KEYWORDS
Oxytocin;
Intranasal administration;
Human behavior;
Trust;
Meta-analysis

Summary The neuropeptide oxytocin has a popular reputation of being the 'love' hormone. Here we test meta-analytically whether experiments with intranasal administration of oxytocin provide support for the proposed effects of oxytocin. Three psychological effects were subjected to meta-analysis: facial emotion recognition (13 effect sizes, $N = 408$), in-group trust (8 effect sizes, $N = 317$), and out-group trust (10 effect sizes; $N = 505$). We found that intranasal oxytocin administration enhances the recognition of facial expressions of emotions, and that it elevates the level of in-group trust. The hypothesis that out-group trust is significantly decreased in the oxytocin condition was not supported. It is concluded that a sniff of oxytocin can change emotion perception and behavior in trusting relationships.

The neuropeptide oxytocin has a strong popular and scientific reputation as the 'love' hormone that creates warm feelings for offspring (Carter, 1998; Feldman et al., 2007, 2010; Galbally et al., 2011; Insel, 1992, 2010) and supports empathic concern for conspecifics (MacDonald and MacDonald, 2010) through better recognition of emotional facial expression (Bartz et al., 2010; Hurlemann et al., 2010; Kirsch et al., 2005; Marsh et al., 2010). Moreover, it would elevate the level of trust in other human beings (De Dreu et al., 2010; Kosfeld et al., 2005). Experimental studies on oxytocin have contributed to our knowledge of its associations with human perception and behavior. Here we test meta-analytically whether experiments with intranasal administration of oxytocin indeed confirm the proposed effects of oxytocin.

Whereas early behavioral experiments with intravenous administration of oxytocin were short-lived due to disappointing results (e.g., Bruins et al., 1992), in recent years the number of experiments using intranasal administration of oxytocin to study human perception, emotion, and behavior has increased dramatically. The reason is that intranasal administration indeed seems to induce replicable changes in brain functioning (Perry et al., 2010; Riem et al., 2011), perception (Theodoridou et al., 2009), and behavior (Naber et al., 2010), in contrast to intravenously administered oxytocin for which the blood—brain barrier might have been difficult to pass. Nevertheless, salivary oxytocin might not be an adequate indicator of levels of oxytocin in the brain, and experimental effects of intranasally administered oxytocin

* Corresponding author at: Centre for Child and Family Studies, Leiden University, PO Box 9555, 2300 RB Leiden, The Netherlands.
E-mail address: vanijzen@fsw.leidenuniv.nl (M.H. Van IJzendoorn).

0306-4530/$ — see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.psyneuen.2011.07.008

may be due to the participants' awareness of the administration. Double blind experiments are crucial to counteract these potential biases (Eisenegger et al., 2010).

In oxytocin experiments two areas of human functioning have been investigated most intensively: recognition of facial expression of emotions such as fear, anger, happiness; and feelings of trust in other human beings. Recently trust of members of the in-group and out-group has been differentiated (De Dreu et al., 2010). From an evolutionary perspective it is suggested that oxytocin may enhance the inclination to protect offspring against predators (Carter, 1998), and thus increase (defensive) aggression against threats from out-group members.

Here we take stock of the first wave of experiments with intranasal oxytocin administration, and test whether intranasally administered oxytocin leads to better recognition of facial expressions and more trust in conspecifics, except when they are labeled as out-group members, in which case trust may even decrease after oxytocin administration (De Dreu et al., 2010; but see Chen et al., 2011). We will explore the moderating influence of the following design features on the outcomes of the oxytocin experiments: within-subject versus between-subject design; the use of a saline placebo or a placebo with all ingredients of the oxytocin spray except for the neuropeptide; time delay between oxytocin administration and test of effect; gender of the participants, and their awareness of the experimental manipulation.

1. Method

For our meta-analysis we systematically searched the database Web of Science with the key words oxytocin, intranasal*, and administ* in the title or abstract (the asterisk indicating that the search contained the word or word fragment). We excluded intravenous administration studies, studies on the effects of oxytocin on parturition or breastfeeding (see for a meta-analysis Wei et al., 2009) non-experimental investigations of oxytocin, and studies on clinical samples (such as individuals with autism spectrum disorder, e.g., Hollander et al., 2007). We finished the search on January 1, 2011. We identified 23 original empirical papers with 31 pertinent effect sizes, providing data for three meta-analyses on effects of oxytocin on face recognition (13 effect sizes, $N = 408$), in-group trust (8 effect sizes, $N = 317$), and out-group trust (10 effect sizes; $N = 505$).

The Comprehensive Meta-Analysis (CMA; Borenstein et al., 2005) program was used to transform the results of the individual studies into the common metric of Cohen's d, or the standardized difference between the intervention and the control condition. Studies could contribute to all three meta-analyses but the same subject was never used twice in the same meta-analysis. The implication however was that some participants were included in two or more meta-analyses; which made it impossible to directly compare effect sizes across the three sets (i.e., effects on face recognition compared to in-group or out-group trust). Therefore the 85% confidence intervals for the point estimates of the combined effect sizes were computed: non-overlapping 85% CI's suggest a significant difference between combined effect sizes that are not independent (Goldstein and Healy, 1995; Van IJzendoorn et al., 2005).

Effect sizes in a set of studies may show smaller or larger variation, and the average or combined effect size across the studies might capture its central tendency more or less adequately. Heterogeneity across studies was assessed using the Q-statistic. Significance tests of combined effect sizes as well as categorical moderator effects were performed with the Q-statistic on the basis of a random effects model (Borenstein et al., 2005). Meta-regression was used to examine the effect of the delay in minutes between oxytocin administration and the test of a behavioral effect on the outcomes of the studies as this was a continuous moderator (Borenstein et al., 2005).

Studies with a small number of subjects and small effect sizes may have a lower chance to be published (publication bias), which might lead to an overestimation of the combined effect size. We used the "trim and fill" method that estimates the number and effect sizes of the potentially non-published studies (Duval and Tweedie, 2000a,b) to calculate the effect of potential data censoring or publication bias on the outcome of the meta-analyses (Sutton et al., 2000). We also computed the fail-safe number of studies needed to reduce a significant combined effect size to non-significance and compared it to Rosenthal's (1991) fail-safe number, $5k + 10$ (k = number of studies included). The fail-safe number is the lowest number of studies with null effects needed to reduce the combined effect size found in the current meta-analysis to non-significance. Moreover, we computed the combined effect size for awareness of condition, i.e., whether the subjects knew if they were administered oxytocin or placebo, as reported in part of the studies.

2. Results

The combined effect size for face recognition amounted to $d = 0.21$ ($p < .01$, 95% CI 0.07, 0.36), in a homogeneous set of studies (Q [$df = 12$] = 10.68). Trim-and-fill did not show a publication bias. Only 19 studies with null effect would be needed to bring the combined effect size down to a non-significant level, a considerably smaller number than Rosenthal's fail-safe criterion. Although the studies with a between-subjects design showed the significant and largest combined effect size ($d = 0.30$, $p = .01$), the difference with the combined outcome of the within-subjects experiments ($d = 0.16$, $p = .10$) was non-significant ($Q_{contrast} = 0.84$, $p = .36$). The other moderator contrasts could not be computed due to too small sets of studies (see Table 1). Time delay between oxytocin administration and behavioral test was not a significant moderator, $z = 0.43$, $p = .67$.

The combined effect size for the in-group trust experiments was $d = 0.48$ ($p < .01$, CI 0.19, 0.77) in a heterogeneous set of studies (Q [$df = 7$] = 15.09, $p < .05$). Trim-and-fill analysis showed a publication bias, and correcting for 1 missing study outcome the combined effect size amounted to $d = .40$ ($p < .05$, CI 0.10, 0.70). Forty-four studies with null effects would be needed to bring the combined effect size down to a non-significant level, still smaller than Rosenthal's fail-safe criterion. Again the six studies with a between-subjects design showed the largest combined effect size ($d = 0.63$, $p < .001$), but the significance of the difference with the outcome of the within-subjects experiments ($d = 0.12$, $p = .53$) could not be tested because only two