Received Date : 28-Aug-2015

Revised Date : 09-Mar-2016

Accepted Date : 09-Mar-2016

Article type : Original Article

Is there a publication bias in behavioral intranasal oxytocin research on humans? Opening the file drawer of one lab

Running title: "Possible file drawer problem in behavioral OT research"

Anthony Lane*

Université catholique de Louvain – UCL

Psychological Sciences Research Institute

10, Place Cardinal Mercier

B-1348 Louvain-La-Neuve, Belgium

E-mail. Anthony.Lane@uclouvain.be

Fax. +32 10 47 48 34

Tel. +32 10 47 45 11

Olivier Luminet

Université catholique de Louvain – UCL

Psychological Sciences Research Institute

10, Place Cardinal Mercier

B-1348 Louvain-La-Neuve, Belgium

Gideon Nave

California Institute of Technology

Computation & Neural Systems

Caltech 228 – 77

Pasadena 91125, California, USA

Moïra Mikolajczak

Université catholique de Louvain – UCL

Psychological Sciences Research Institute

10, Place Cardinal Mercier

B-1348 Louvain-La-Neuve, Belgium

*Corresponding author

Abstract

The neurohormone oxytocin (OT) has been one the most studied peptides in behavioral sciences over the past two decades. Primarily known for its crucial role in labor and lactation, a rapidly growing literature suggests that intranasal OT (IN-OT) may also play a role in humans' emotional and social lives. However, the lack of a convincing theoretical framework explaining IN-OT's effects that would also allow to predict which moderators exert their effects and when, has raised healthy skepticism regarding the robustness of human behavioral IN-OT research. The poor knowledge of OT's exact pharmacokinetic properties, crucial statistical and methodological issues and the absence of direct replication efforts may have lead to a publication bias in IN-OT literature with many unpublished studies with null results lying in laboratories' drawers. Is there a file drawer problem in IN-OT research? If this is the case, it may also be the case in our laboratory. This paper aims to answer that question, document the extent of the problem and discuss its implications for OT research. Through eight studies (including 13 dependent variables overall, assessed through 25 different paradigms) performed in our lab between 2009 and 2014 on 453 subjects, results were too often not those expected. Only five publications emerged from our studies and only one of these reported a null-finding. After realizing that our publication portfolio has become less and less

representative of our actual findings and because the non-publication of our data might contribute to generating a publication bias in IN-OT research, we decided to get these studies out of our drawer and encourage other laboratories to do the same.

Keywords: Intranasal Oxytocin, file drawer, lab report

Introduction

Behavioral scientists have been investigating the psychosocial effects of the neuropeptide oxytocin (OT) in humans for over two decades, making it one of the most studied hormones in the social sciences. A rapidly growing literature suggests that OT - that has a well-established physiological role in labor and lactation - may also play a role in humans' emotional and social lives.

During the past two decades, preliminary findings have suggested that intranasal OT (IN-OT) administration increases trust toward strangers (1, 2), promotes self-confidence (3, 4), improves recognition of familiar faces (5), enhances emotional recognition (6) and facilitates mind reading (7). Other studies proposed that IN-OT also fosters sharing of emotions with others (8), makes people more sensitive to others' feelings (9), promotes altruism (10), enhances perceived trustworthiness and attractiveness and facilitates parent-infant (11) and romantic (12) attachments. These findings helped to build OT's reputation as the prosocial hormone *par excellence*, and the popular press has largely reinforced this reputation.

Nevertheless, several findings have tempered this idealistic view of IN-OT. For example, it has been proposed that IN-OT might also promote anti-social behavior such as aggression (13), ethnocentrism (14) and gloating (15). These findings questioned the mainstream theory of IN-OT as an affiliative/prosocial hormone (16), and motivated the proposal of several new hypotheses. Two of them in particular have been studied in depth: the first postulates that IN-OT increases the salience of social cues (16); the second conjectures that IN-OT increases social approach behaviors, whether good

or bad (17). Studies to date have not clearly favored one theory over the others. Some findings have been consistent with one (or more) of these theories, but others do not sit easily with either (18).

Another proposition that has emerged from the behavioral IN-OT literature is that IN-OT's influences are strongly moderated by environmental context and personal characteristics. A recent review (19) has concluded that the majority of IN-OT studies do not yield a main IN-OT treatment effects. To account for their findings, the authors proposed that IN-OT's effect might occur only under certain circumstances or only in as a function of specific personality traits - reflecting the plausible complexity of the interaction between IN-OT, environment and genotype. The lack of a convincing theoretical framework that allows to predict which moderators exert their effects and when, has raised healthy skepticism regarding the robustness of human behavioral IN-OT research (20, 21).

One source of skepticism is that the vast IN-OT research enterprise has relied on the pharmacokinetic properties of arginine vasopressin (AVP) administration - a peptide that is structurally similar, yet not identical to OT (22-24). IN-OT pharmacokinetics are not fully understood and the only study conducted to date (with a very small sample size) found that IN-OT does not yield elevated cerebrospinal fluid (CSF) OT levels 45 minutes after administration (the time window following administration at which most behavioral tasks took place) (25). Moreover, it is uncertain whether the standard doses used in OT research (between 24 and 40 IU) can deliver sufficient quantities of OT to the brain in order to produce significant changes in individuals, especially as OT is avidly degraded in brain tissue (24). Future studies investigating the penetration of IN-OT into brain and its pharmacokinetic properties in human are crucially needed.

A second source of skepticism concerns statistics. A recent meta-analysis of published studies involving IN-OT in humans (21) demonstrated that most studies are dramatically underpowered[1] and report overestimated effects. The meta-analysts estimated (using information on power, pre-study odds and the alpha level) that the false discovery rate in the IN-OT literature is over 80%.

[1]Walum and colleagues' results indicate that the average study investigating intranasal OT in healthy subjects has a statistical power of 16%.

A third source of skepticism is a striking absence of efforts towards direct replication. As far as we know, almost none of the findings in the literature underwent direct replication attempts, despite the obvious importance of such efforts (26). Moreover, the seminal, highly cited study associating IN-OT with trust (1), that inspired much of the subsequent research, failed several times to replicate (20). Our lab has also failed to replicate a promising initial finding relating IN-OT with increased trust in a non monetary behavioral task (see (2) for the original study, see (27) for the failed replication). Furthermore, a recent study failed to replicate seminal findings associating IN-OT with mind-reading (see (7) for the original study, see (28) for the failed replication).

Finally, the methodological challenges accompanying behavioral OT research are not unique to the use of IN-OT administration: the literature using peripheral OT measurements also relies on OT assay methods that are considered by many researchers as bio-analytically invalid (29-31).

In the light of these concerns and after failing to replicate our own IN-OT trust-enhancing effect (2), we put forward four, non mutually exclusive, hypotheses regarding the true association between IN-OT and social behavior (27):

(A) The effects reported in the literature reflect the true state of the world, and failed replications are due to underpowered studies or methodological errors/differences.

(B) The effects found in the literature are indicative of an effect of IN-OT in humans, but the true effect of IN-OT on human behavior is much smaller than the impression given by published studies. Replications and highly powered studies would therefore allow to adjust the real effect size.

(C) The effects found in the literature are type I errors that reflect a publication bias of positive results (32), which is possible as we generally accept 5% rate of type I error.

(D) The effects of IN-OT do not truly exist but are artificially created (e.g., by extensive degree of researcher freedom (33), study misconduct).

If either of the two last hypotheses is true, there should exist many unpublished studies with null results lying in laboratories' drawers (32).

Is there a file drawer problem in IN-OT research? If this is the case, it may also be the case in our laboratory. This paper aims to answer that question, document the extent of the problem, and discuss its implications for IN-OT research. We present eight studies (including 13 dependent variables overall, assessed through 25 different paradigms) that were performed in our lab from 2009 until 2014 on a total of 453 subjects. All our studies relied on theoretical and experimental accounts of IN-OT's role in social behavior that had been published to date. As we will demonstrate below, the results were too often not those expected. Only four studies (most often a part of them) of the eight were submitted for publication, yielding five articles (2, 8, 27, 34, 35). Of these five article, only one (27) reports a null-finding. We submitted several studies yielding null-findings to different journals (from general interest in psychology to specialized in biological psychology and in psychoenodcrinolgy) but they were rejected time and time again[2]. After realizing that our publication portfolio has become less and less representative of our actual findings, and because the non-publication of our null-findings might contribute to generating a publication bias in IN-OT research, we decided to get these studies out of our drawer, hoping that other laboratories will do the same.

To avoid an overly pessimistic view by only presenting the null results obtained, we instead present a complete overview of the research performed in our lab since we started studying IN-OT in 2009. This will allow readers to form their own opinion about the findings and allow us to meta-analyze the cumulative effects.

[2] We submitted four articles that were rejected at least once (IN-OT and conformity to peer pressure, submitted once and rejected after review; IN-OT and mimetic desire, submitted twice and rejected twice after review; IN-OT and compassion, submitted twice and rejected twice after review; failed replication of IN-OT effect on trust, submitted twice, rejected once after review and then accepted in another journal).

Methods and results

Methods

We will present eight studies assessing 13 dependent variables (emotional, cognitive, behavioral or physiological) through 25 different paradigms, performed in our lab over the past seven years, in chronological order. The methodological details of our studies are summarized in Table 1, and a full description of the studies including each behavioral task appears in Appendix 1. In each study, the tasks were conducted in a fixed order determined by the importance we attributed to each paradigm: the most important target variable was tested in the first task in order to eliminate the potential of spillover effects from other tasks[3]. All studies met the guidelines for ethical conduct of research and were conducted in accordance with the Declaration of Helsinki. The Biomedical ethics committee of the Université catholique de Louvain approved the protocols. Exclusion criteria included medical or psychiatric condition, substance dependence and female gender (except for the Study 8 on jealousy which involved couples and focused on female reactions). The number of subjects varied between 12 and 95[4] (see Table 1, column 4). All studies followed a between-subject design (except for Study 3 on sleep) and were either single or double blind (see Table 1, column 7). The dose of IN-OT (Syntocinon spray, between 24 and 40 IU in order to get through the dosing spectrum found in IN-OT literature) and the provider varied across studies (see Table 1, column 6). The placebo was always a saline solution administered in a bottle similar to IN-OT one. Each spray bottle was numbered and covered with sticky paper that covered the product label. The timing of the tasks was set according to the current norms in behavioral IN-OT research. Thus, the first task took place at the earliest approximately 35 minutes after IN-OT administration (usually 45 minutes), and when there were several tasks in the same study, the last task ended no later than 85 minutes after IN-OT administration (see Table 1, columns 3 and 8). Generally, the subjects performed the experiment alone unless the

[3] The use of more than one task is common practice because of the imperative to maximize the knowledge gained from each subject undergoing pharmacological treatment

[4] Based on the standard found in IN-OT literature

presence of a confederate was required (Table 1, column 9). Across all studies, there were no differences between the treatments groups (OT vs. PL) with respect to all baseline measures (all *p*s > .05) that were focused on self-reported questionnaires regarding the dependent variables relevant to each study, (specified for each study in Appendix 1). All studies also involved a personality questionnaire and collected demographic information.

Results

The last two columns of the Table 1 summarize the main and interaction effects of IN-OT treatment on target behaviors. We found a statistically significant main IN-OT effect for only one of 25 tasks, and a significant interaction effect including the treatment condition (OT vs. PL) for only five out of 25 tasks across our 8 studies and 13 dependent variables (see full results and statistical details in Appendix 1). Table 1 (column 10) reports the effect sizes for each variable. Only 13 out of 25 task points estimating effect size reach the lower bound on a small affect size (Cohen's $d > 0.2$). Among those, one task reaches the lower bound of a moderate effect size (Cohen's $d > 0.5$); another reaches the lower bound of a large effect size (Cohen's $d > 0.8$) but this result has to be interpreted carefully as we have failed to replicate it twice (27). Furthermore, only one task rules out the zero effect size with a 95% confidence interval, but once again the results of this particular study did not replicate well (27).

In order to determine the extent of IN-OT's influence on human behavior in our studies, we meta-analyzed[5] the effects of IN-OT on cognitive, emotional or behavioral variables (excluding the studies of OT's effects on physiological processes, namely sleep and pain). The aggregated effect size was not reliably different from zero (Cohen's d = 0.003 [95% CI: -0.10;0.10]). We further aggregated IN-OT's effects on variables assessing behaviors, affect or cognition in isolation (see Table 2), and could not reliably reject the null hypothesis for either ($d_{behaviors}$= 0.09 [95% CI = -0.07;0.25]; $d_{affects}$ = -

[5] We computed the cumulative effect sizes using the "Comprehensive Meta-Analysis" software (36).

0.003 [95% CI = -0.20;0.24]; $d_{cognitions}$ = 0.1, [95% CI = -0.32;0.13]). Finally, aggregating our effect sizes in reference to the three major behavioral OT theories (i.e., OT as a hormone of affiliation (16); OT as a hormone of social salience (15) and OT as a hormone of social approach (17), see Table 2), did not yield any effects that were reliably different from zero ($d_{prosocial}$ = -0.04 [95% CI = -0.13;0.06]; $d_{social\ saliance}$ = -0.01 [95% CI = -0.11;0.10]; $d_{social\ approach}$ = -0.002 [95% CI = -0.11;0.11]).

Discussion

We reviewed eight studies testing the influence of IN-OT on human cognition and behavior, assessing 13 dependent variables through 25 different paradigms performed in our lab since 2009. We found a statistically significant main effect of IN-OT for only one out of 25 tasks and a significant interaction effect including the treatment condition (OT vs. PL) for only 5 out of 25 tasks. All of our hypotheses were derived from the three major behavioral IN-OT theories (i.e., OT as a hormone of affiliation (16); OT as a hormone of social salience (15) and OT as a hormone of social approach (17)).

This large proportion of "unexpected" null-findings (92% for IN-OT's main effect) raises concerns about the validity of what we know about the influence of IN-OT on human behaviors and cognition. As reported in the meta-analytic section, the aggregated effects are not reliably different from zero, regardless of how they have been pooled (by dependent variables, by theories or altogether). Our initial enthusiasm on IN-OT findings has slowly faded away over the years and the studies have turned us from "believers" into "skeptics". This led us to raise several questions.

If the published literature on IN-OT's behavioral effects does not reflect the true state of the world, how has the vast behavioral IN-OT literature accumulated? We reiterate here two possible accounts. First, the significant findings might be a consequence of a Type I error (the commonly accepted p-value to reach significance level allows a 5 % of false positive). If this is the case, much unpublished data must be lying in the drawers of laboratories studying IN-OT.

Second, the significant effect of IN-OT may be the result of methodological, measurement or statistical artifacts. As this has been demonstrated for peripheral OT measurements (29), it should not be excluded here, although the artifacts would be different. We see four potential sources of generating artifacts in IN-OT research: 1) small sample between subject-designs that might not be internally valid, 2) single blind designs 3) IN-OT pharmacokinetics and dosage and 4) statistical methods.

The massive use of between-subject designs of relatively small samples (about 30 participants per cell) carries the risk of attributing effects to IN-OT that are in fact generated by baseline group differences in various unobservable factors (e.g., personality)[6].

The use of single blind studies, where the subject is blind to the treatment condition but the experimenter is not, introduces the risk that the experimenter might unconsciously influence the subjects (37).

The dosage of IN-OT and typical timing of tasks following IN-OT administration is based on three assumptions that to our knowledge have not been directly or reliably (i.e. through several replications) tested: that IN-OT crosses the brain-blood barrier following administration, that 24-40 IU is a sufficient dose to produce behavioral changes, and that IN-OT pharmacokinetics mimics that of vasopressin (24).

Recent findings have demonstrated that IN-OT increases OT concentration in CSF in both human (25) and animal (38, 39). Furthermore, it has recently been demonstrated that IN-OT modulates amygdala responses in monkeys in a manner equivalent to humans (40). Taken together, those results suggest that IN-OT reaches, directly or indirectly (41), the central nervous system and would so produces observable affective, behavioral or cognitive modifications. However, if IN-OT produces a significant elevation of OT concentration in the CSF after 30 minutes in animals, this significant

[6] Note that within-subject designs also suffer from limitations such as reduced statistical power (e.g.: see Uri Simonsohn's post http://datacolada.org/2015/06/22/39-power-naps-when-do-within-subject-comparisons-help-vs-hurt-yes-hurt-power/)

elevation takes place 75 minutes after IN-OT in human, which is not consistent with the literature where most tasks start 40-45 minutes after IN-OT. Furthermore, in a recent research, Quintana and colleagues (42) suggest that the IN-OT doses commonly used (24 – 40 IU) may not be the most adequate as their results show that IN-OT effect on emotional recognition appears with an administration of 8 IU but not with 24 IU. Facing these challenges, further studies would be needed in order to strengthen our knowledge about IN-OT pharmacokinetic properties. Even if IN-OT reaches the brain, we cannot assure that the three assumptions on which IN-OT's literature is based are reliable.

Finally, the use of too small samples (21) and the vast amount of candidate factors that could potentially moderate IN-OT's behavioral effects (19, 20) might inflate the false discovery rate unless direct replication efforts and correction for multiple hypotheses are applied.

Two alternative hypotheses can also explain the seemingly puzzling results described in this paper.

First, our studies, like most published studies on IN-OT, might be underpowered (21). Thus, the fact that effects of IN-OT observed in our studies are non-significant does not mean that they are point estimates of a zero effect. For example, some of our studies do not rule out a small effect size (Cohen's d = 0.2)[7]. In order to detect such effects, or even a moderate effect, a sample size between 120 (Study 9, jealousy assessment through the word completion task, Cohen's d = 0.518) and 468 (Study 2, empathy assessment through the RMEt, Cohen's d = 0.260) participants would be required to reliably detect an IN-OT effect with 80% of power. Such sample sizes are much greater than the norm in both the IN-OT field and our lab. Therefore, several of our findings could potentially have turned significant in well-powered experiments. Yet, as shown in Table 1, their significance would not always have been in the expected direction.

[7] We have excluded the highest effect size found, in Study 1 - non monetary trust assessment, as it has been questioned by Lane et al.(27)..

A second proposition is that IN-OT effects do exist, but that they are strongly moderated by various factors, making them appear large in some circumstances but not others. Through the literature, more and more findings suggest that IN-OT influences behaviors by interacting with several moderators (for a review see (19)). Arguably, our findings do not rule out the possibility that the effects of IN-OT are moderated by various factors – a proposition that will be difficult to rule out, given the infinitely large set of factors that could potentially moderate IN-OT's behavioral influences (genes, personality or environmental factors). Unfortunately, as far as we know, candidate moderators do not seem to replicate from one study to another[8] and appear most often to represent post-hoc data fits rather than a-priori hypotheses[9]. Indeed, one can be sure to find a "significant" interaction in any data set, simply by conducting many statistical tests, even in the absence of a true signal in the data, unless the test level alpha is corrected for multiple hypothesis testing (43, 44).

We can either believe that these interactions are statistical artifacts (see above) or believe that they are real. If we believe that they are real, it means that there is no such "general effect of IN-OT on behavior" but that IN-OT effects are always context dependent (for a review see (19)). In the studies reported in this article, the relevant potential moderators have been taken into account and only provided five interaction effects. Yet, it is possible that less obvious moderators, or moderators that we did not measure, would have provided more significant effects.

As we write these lines, we do not know which of the four hypotheses is true; IN-OT might not influence human behaviors at all or may influence it only under specific circumstances. In any case, falsifiable theories must emerge in order to progress in our understanding of IN-OT's behavioral influences, as no current theory seems to yield robust behavioral predictions - and almost every

---

[8] For example, in their failed replication of IN-OT's influence on the RMEt, Radke and de Bruijn (28) did not find any moderating effect of items' difficulty as demonstrated by Domes and colleagues (7).

[9] And we do not make exception to the rule: it is because we could not replicate IN-OT effect on the RMET that we looked for personality moderators and found a significant interaction with alexithymia (34)

behavioral effect can be explained by one of the theories ex-post. Along this line, although the value of replications cannot be over-estimated for increasing the reliability of scientific findings (26, 45), replication attempts are almost absent in IN-OT research, and the only attempts made to replicate high profile publications did not yield the expected effects (e.g.: trust game investment (46); non-monetary trust (27); empathy through the RMEt (28)).

To our view, nothing can be taken for granted with IN-OT and some non-replicable findings might have biased the development of existing theories. Hopefully, incorporating null findings and failed replications into the theoretical process would allow to draw lines between robust, replicable IN-OT effects and facilitate the development of falsifiable theories. It is therefore crucial that non-significant findings and failed replications are published[10]. Every piece of evidence, even experiments that did not yield "significant" effects, should be taken into account and weighted according to its evidential value.

In the present case, only 5 articles (2, 8, 27, 34, 35) have been published across the 13 dependent variables we have assessed, producing a publication rate of 38.5%. If our lab is a representative sample of IN-OT research, then for 626 search results found in Scopus by entering "oxytocin" and "human" as research keys (and limiting the outputs to "Psychology"), approximately 1000 potential studies have remained in labs' drawers. Unraveling these 1000 data sets is extremely important for understanding whether IN-OT exerts reliable effects on humans and under which circumstances.

We believe that a systematic shift in the IN-OT publication process is essential in order to reveal the true state of the world. Pre-registration of ex-ante hypotheses, replication attempts of the findings before their submission and submission of null results and failed replication for publication, especially when the studies are well-powered to detect the original findings, should be encouraged. Review processes should insist on fully reporting all of the of the candidate moderators that were measured and tested and encourage publication of well-conducted studies, whatever their results (47).

[10] http://psychfiledrawer.org

Many labs do report their work transparently. But as far as the editorial process does not sufficiently promote non-significant results and failed replications, it is difficult to obtain a complete overview of IN-OT research field. One way to improve the standards is by institutionalization[11]: as suggested by Leng and Ludwig (24): journals could oblige researchers to preregister trials, declare hypotheses and primary outcomes in advance, specify statistical methods to be applied and fully disclose the data, including tasks that did not yield results and assessed moderators that did not moderate the findings. This would help to drastically decrease reporting bias (i.e., picking significant results from a battery of tests and only reporting these). Moreover, authors could easily test the robustness of their findings by adjusting the alpha level to the number of tests that were performed (e.g. if the subjects were asked to perform three tasks, the level of significance would be 0.05/3 = 0.016, instead of 0.05).

These considerations must be taken into account if we want to dispose of a solid theoretical background for interpreting and understanding the complex effects of IN-OT and to warrant all the efforts and resources invested in IN-OT research.

Acknowledgement

The authors of the manuscript have no conflicts of interest to declare. We want to thank the Cliniques Universitaires Saint-Luc (university hospital), which hosted our experiments. We also want to thank Gillian Rosner for her proofreading.

References

1. Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E. Oxytocin increases trust in humans. Nature. 2005;435(7042):673-6.

[11] As it is encourage by, notably, the American Psychological Association (http://www.apa.org/research/responsible/publication/index.aspx), the Association for Psychological Science (http://www.psychologicalscience.org/index.php/news/releases/psychological-science-sets-new-standards-for-research-reporting.html) and the NHI (http://grants.nih.gov/grants/policy/data_sharing/data_sharing_faqs.htm#900)

2. Mikolajczak M, Pinon N, Lane A, de Timary P, Luminet O. Oxytocin not only increases trust when money is at stake, but also when confidential information is in the balance. Biological Psychology. 2010;85(1):182-4.

3. Cardoso C, Linnen A, Joober R, Ellenbogen MA. Coping style moderates the effect of intranasal oxytocin on the mood response to interpersonal stress. Experimental and Clinical Psychopharmacology. 2012;20(2):84-91.

4. Heinrichs M, Baumgartner T, Kirschbaum C, Ehlert U. Social support and oxytocin interact to suppress cortisol and subjective responses to psychosocial stress. Biological Psychiatry. 2003;54(12):1389-98.

5. Rimmele U, Hediger K, Heinrichs M, Klaver P. Oxytocin makes a face in memory familiar. Journal of Neuroscience. 2009;29(1):38-42.

6. Domes G, Heinrichs M, Gläscher J, Büchel C, Braus DF, Herpertz SC. Oxytocin Attenuates Amygdala Responses to Emotional Faces Regardless of Valence. Biological Psychiatry. 2007;62(10):1187-90.

7. Domes G, Heinrichs M, Michel A, Berger C, Herpertz SC. Oxytocin Improves "Mind-Reading" in Humans. Biological Psychiatry. 2007;61(6):731-3.

8. Lane A, Luminet O, Rimé B, Gross JJ, de Timary P, Mikolajczak M. Oxytocin increases willingness to socially share one's emotions. International Journal of Psychology. 2013;48(4):676-81.

9. Hurlemann R, Patin A, Onur OA, Cohen MX, Baumgartner T, Metzler S, et al. Oxytocin enhances amygdala-dependent, socially reinforced learning and emotional empathy in humans. Journal of Neuroscience. 2010;30(14):4999-5007.

10. Zak PJ, Stanton AA, Ahmadi S. Oxytocin increases generosity in humans. PLoS ONE. 2007;2(11).

11. Naber F, van Ijzendoorn MH, Deschamps P, van Engeland H, Bakermans-Kranenburg MJ. Intranasal oxytocin increases fathers' observed responsiveness during play with their children: A double-blind within-subject experiment. Psychoneuroendocrinology. 2010;35(10):1583-6.

12. Ditzen B, Schaer M, Gabriel B, Bodenmann G, Ehlert U, Heinrichs M. Intranasal Oxytocin Increases Positive Communication and Reduces Cortisol Levels During Couple Conflict. Biological Psychiatry. 2009;65(9):728-31.

13. De Dreu CKW, Greer LL, Handgraaf MJJ, Shalvi S, Van Kleef GA, Baas M, et al. The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. Science. 2010;328(5984):1408-11.

14. De Dreu CKW, Greer LL, Van Kleef GA, Shalvi S, Handgraaf MJJ. Oxytocin promotes human ethnocentrism. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(4):1262-6.

15. Shamay-Tsoory SG, Fischer M, Dvash J, Harari H, Perach-Bloom N, Levkovitz Y. Intranasal Administration of Oxytocin Increases Envy and Schadenfreude (Gloating). Biological Psychiatry. 2009;66(9):864-70.

16. Insel TR. Oxytocin — A neuropeptide for affiliation: Evidence from behavioral, receptor autoradiographic, and comparative studies. Psychoneuroendocrinology. 1992;17(1):3-35.

17. Kemp AH, Guastella AJ. The role of oxytocin in human affect: A novel hypothesis. Current Directions in Psychological Science. 2011;20(4):222-31.

18. Bethlehem RA, Baron-Cohen S, van Honk J, Auyeung B, Bos PA. The oxytocin paradox. Frontiers in behavioral neuroscience. 2014;8.

19. Bartz JA, Zaki J, Bolger N, Ochsner KN. Social effects of oxytocin in humans: Context and person matter. Trends in Cognitive Sciences. 2011;15(7):301-9.

20. Nave G, Camerer C, McCullough M. Does Oxytocin increase trust in humans? Critical review of research. Perspectives on Psychological Science. 2015;10(6):772-89.

21. Walum H, Waldman ID, Young LJ. Statistical and methodological considerations for the interpretation of intranasal oxytocin studies. Biological Psychiatry. 2015.

22. Born J, Lange T, Kern W, McGregor GP, Bickel U, Fehm HL. Sniffing neuropeptides: a transnasal approach to the human brain. Nature neuroscience. 2002;5(6):514-6.

23. Meyer-Lindenberg A, Domes G, Kirsch P, Heinrichs M. Oxytocin and vasopressin in the human brain: social neuropeptides for translational medicine. Nature Reviews Neuroscience. 2011;12(9):524-38.

24. Leng G, Ludwig M. Intranasal oxytocin: myths and delusions. Biological Psychiatry. 2015.

25. Striepens N, Kendrick KM, Hanking V, Landgraf R, Wullner U, Maier W, et al. Elevated cerebrospinal fluid and blood concentrations of oxytocin following its intranasal administration in humans. Sci Rep. 2013;3.

26. Simons DJ. The value of direct replication. Perspectives on Psychological Science. 2014;9(1):76-80.

27. Lane A, Mikolajczak M, Treinen E, Samson D, Corneille O, de Timary P, et al. Failed Replication of Oxytocin Effects on Trust: The Envelope Task Case. PLoS One. 2015;10(9):e0137000.

28. Radke S, de Bruijn ER. Does oxytocin affect mind-reading? A replication study. Psychoneuroendocrinology. 2015.

29. McCullough ME, Churchland PS, Mendez AJ. Problems with measuring peripheral oxytocin: can the data on oxytocin and human behavior be trusted? Neuroscience & Biobehavioral Reviews. 2013;37(8):1485-92.

30. Szeto A, McCabe PM, Nation DA, Tabak BA, Rossetti MA, McCullough ME, et al. Evaluation of enzyme immunoassay and radioimmunoassay methods for the measurement of plasma oxytocin. Psychosomatic Medicine. 2011;73(5):393.

31. Christensen JC, Shiyanov PA, Estepp JR, Schlager JJ. Lack of association between human plasma oxytocin and interpersonal trust in a prisoner's dilemma paradigm. PLoS one. 2014;9(12):e116172.

32. Rosenthal R. The file drawer problem and tolerance for null results. Psychological bulletin. 1979;86(3):638.

33. Simonsohn U, Nelson LD, Simmons JP. P-curve: A key to the file-drawer. Journal of Experimental Psychology: General. 2014;143(2):534.

34. Luminet O, Grynberg D, Ruzette N, Mikolajczak M. Personality-dependent effects of oxytocin: Greater social benefits for high alexithymia scorers. Biological Psychology. 2011;87(3):401-6.

35. Mikolajczak M, Gross JJ, Lane A, Corneille O, de Timary P, Luminet O. Oxytocin Makes People Trusting, Not Gullible. Psychological Science. 2010;21(8):1072-4.

36. Borenstein M, Hedges L, Higgins J, Rothstein H. Comprehensive meta-analysis version 2. Englewood, NJ: Biostat. 2005;104.

37. Doyen S, Klein O, Pichon C-L, Cleeremans A. Behavioral priming: It's all in the mind, but whose mind. PloS one. 2012;7(1):e29081.

38. Neumann ID, Maloumby R, Beiderbeck DI, Lukas M, Landgraf R. Increased brain and plasma oxytocin after nasal and peripheral administration in rats and mice. Psychoneuroendocrinology. 2013;38(10):1985-93.

39. Freeman SM, Samineni S, Allen PC, Stockinger D, Bales KL, Hwa GG, et al. Plasma and CSF oxytocin levels after intranasal and intravenous oxytocin in awake macaques. Psychoneuroendocrinology. 2016.

40. Liu N, Hadj-Bouziane F, Jones KB, Turchi JN, Averbeck BB, Ungerleider LG. Oxytocin modulates fMRI responses to facial expression in macaques. Proceedings of the National Academy of Sciences. 2015:201508097.

41. Quintana DS, Alvares GA, Hickie IB, Guastella AJ. Do delivery routes of intranasally administered oxytocin account for observed effects on social cognition and behavior? A two-level model. Neuroscience & Biobehavioral Reviews. 2015;49:182-92.

42. Quintana D, Westlye L, Rustan Ø, Tesli N, Poppy C, Smevik H, et al. Low-dose oxytocin delivered intranasally with Breath Powered device affects social-cognitive behavior: a randomized four-way crossover trial with nasal cavity dimension assessment. Translational psychiatry. 2014;5:e602-e.

43. Shaffer JP. Multiple hypothesis testing. Annual review of psychology. 1995;46(1):561-84.

44. McKay B, Bar-Natan D, Bar-Hillel M, Kalai G. Solving the Bible Code Puzzle. Statistical Science. 1999;14(2):150-73.

45. Koole SL, Lakens D. Rewarding replications a sure and simple way to improve psychological science. Perspectives on Psychological Science. 2012;7(6):608-14.

46. Klackl J, Pfundmair M, Agroskin D, Jonas E. Who is to blame? Oxytocin promotes nonpersonalistic attributions in response to a trust betrayal. Biological Psychology. 2013;92(2):387-94.

47. John LK, Loewenstein G, Prelec D. Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. Psychological Science. 2012;23(5):524-32.

Table 1: Presentation of the studies, including methodology and results

Study	Dependant Variable	Paradigm & time following product administration	Number of participants	Sex of the participants	Dose & Product	Administration type and design	Time between administration and testing	Testing type	OT Main effect	Interaction effect
Study 1: Oxytocin, trust and social sharing of the emotions (2009)	Trust (monetary)	Trust game 45 minutes after product administration	60 (30 OT & 30 PL)	Male	32 IU Syntocinon Spray, Novartis, Basel Switzerland	Single Blind Between-subject design	45 minutes	Participant alone	N.S.[1] Cohen's *d* =0.13 OT possibly increases trust [95% CI: -0.38;0.64]*	Condition x Partner reliability: OT only increases trust for reliable partners
	Social Sharing of the Emotions	Self reported willingness to share emotions 55 minutes after product administration							N.S. Cohen's *d* = 0.19 OT possibly increases the willingness to share emotions [95% CI: -0.33;0.70]*	Condition x Content of the sharing (Facts vs. Emotions): OT only increases willingness to share emotions

	Trust (non monetary)	Envelope Task 65 minutes after product administration							Significant Cohen's *d* = 2.09 OT increases trust [95% CI: 0.80;3.38]	No
Study 2: Oxytocin and empathy (2009)	Empathy	Reading the Mind in the Eyes test 45 minutes after product administration	60 (30 OT & 30 PL)	Male	32 IU Syntocinon Spray, Novartis, Basel Switzerland	Single Blind Between-subject design	45 minutes	Participant alone	N.S. Cohen's *d* = 0.26 OT possibly increases mind reading [95% CI: -0.26;0.78]*	Condition x Level of Alexithymia: OT only increases empathy for participants with a high level of alexithymia
	Compassion	Explicit measurement of Compassion after something bad happens to someone in a story . 55 minutes after product							N.S. Cohen's *d* = -0.39 OT possibly decreases compassion [95% CI: -0.91;0.14]*	No

		administration								
	Empathy	Self reported empathic feeling and tendency to help someone who is first presented as a victim and then as a culprit in scenarios 65 minutes after product administration							N.S. Sympathy: Cohen's *d* = -0.42 OT possibly decreases sympathy [95% CI: -0.93;0.10]* Help: Cohen's *d* = -0.19 OT possibly decreases helping behaviors [95% CI: -0.70;0.32]*	No
Study 3: Oxytocin and sleep (2011)	Sleep latency	Multiple Sleep Latency test 45 minutes	12	Male	32 IU Syntocinon Spray,	Single Blind Within-subject	45 minutes	Participant alone	N.S. Cohen's *d* = -0.14 OT possibly	No

		after product administration			Novartis, Basel Switzerland	design			decreases sleep latency [95% CI: -0.94;0.66]*	
	Sleep duration								N.S.Marginally Significant (p = .097) Cohen's d = 0.27 OT possibly increases sleep duration [95% CI: -0.48;1]*	No
	Proportion of REM sleep								N.S.Tendency to Significant (p = .115) Cohen's d = 0.68 OT possibly increases REM sleep proportion [95% CI: -0.14;1.48]*	No

	Psychomotor vigilance	Psychomotor Vigilance Task							N.S.Marginally Significant (p = .083) Cohen's d = -0.41 OT possibly decreases psychomotor vigilance [95% CI: -1.20;0.04]*	No
Study 4: Oxytocin, pain and sensitivity to baby's cry (2011)	Pain threshold	Cold Pressure test 45 minutes after product administration	60 (30 OT & 30 PL)	Male	32IU Syntocinon Spray, Novartis, Basel Switzerland	Double Blind Between-subjects design	45 minutes	Participants alone	N.S. Cohen's d = -0.28 OT possibly decreases pain threshold [95% CI: -0.78;0.23]*	No
	Pain tolerance								N.S. Cohen's d = 0.16 OT possibly increases pain tolerance	No

									[95% CI: -0.35;0.66]*	
	Willingness to endure Pain								N.S. Cohen's *d* = 0,32 OT possibly increases willingness to endure pain [95% CI: -0.20;0.82]*	No
	Perceived pain intensity								N.S. Cohen's *d* = 0.19 OT possibly increases perceived pain intensity [95% CI: -0.32;0.70]*	No
	Sensitivity to a baby's cry	Self reported annoyance from baby's cry sound tracks							N.S. Cohen's *d* = 0.24 OT possibly increases	No

		55 minutes after product administration							sensitivity to baby's cry [95% CI: -0.27;0.75]*	
Study 5: The dark side of Oxytocin: guilt, conformism and compliance to antisocial behaviors (2012)	Compliance to anti-social behaviors	Anti-social peer pressure 35 minutes after product administration	61 (31 OT & 30 PL)	Male	40IU Syntocinon Spray, Novartis, Basel Switzerland	Double Blind Between-subject design	35 minutes	With 2 confederates	Marginally Significant (p = .078)N.S. Cohen's d = 0.47 OT possibly increases compliance to peer's anti-social requests [95% CI: -0.05;0.98]*	No
	General conformism	Numeric estimation task 45 minutes after product administration						Alone	N.S.Marginally Significant (p = .074) Cohen's d = -0.47 OT possibly decreases conformism	No

									[95% CI: -0.99;0.04]*	
	Behavioral measure of guilt after guilt induction	Effective splitting of money with partner or charity to make amend 75 minutes after product administration						Alone	N.S. Cohen's *d* = 0.33 OT possibly increases guilt [95% CI: -0.18;0.83]*	No
	Guilt after guilt induction	Self-reported questionnaire 85 minutes after product administration						With 1 confederate	N.S. Cohen's *d* = 0.41 OT possibly increases guilt [95% CI: -0.10;0.92]*	No
Study 6: Oxytocin, Mimetic Desire, Visual	Mimetic Desire	Neutral painting evaluation task (looked at vs looked away)	95 (48 OT & 47 PL)	Male	32 IU Syntocinon Spray, Fuerte Farma,	Double Blind Between-subject design	45 minutes	Alone	N.S. Cohen's *d* = 0.19 OT possibly increases	No

perspective taking and Trust (2012)		45 minutes after product administration			Funchal, portugal				mimetic desire [95% CI: -0.22;0.60]*	
	Self vs. Others' Visual perspective taking	Visual Perspective Taking task (accuracy) 55 minutes after product administration							N.S. Cohen's *d* = -0.17 OT possibly decreases visual perspective accuracy [95% CI: -0.57;0.23]*	No
		Visual Perspective Taking task (reaction time) 55 minutes after product administration							N.S. Cohen's *d* = 0.01 OT does not influence visual perspective reaction time [95% CI: -0.39;0,41]*	No
	Trust (non-	Envelope Task							N.S.	No

	monetary)	65 minutes after product administration							Cohen's *d* = -0.10 OT possibly decreases trust [95% CI: -0.50;0.30]*	
Study 7: Oxytocin, Compassion and Trust (2013)	Vicarious experience of another's distress	Explicit measure of compassion: self reported evaluation 45 minutes after product administration	61 (32OT & 29 PL)	Male	32IU Syntocinon Spray, Defiante Farmaceutica, Funchal, portugal	Double Blind Between-subject design	45 minutes	Alone	N.S. Cohen's *d* = 0.10 OT possibly increases compassion [95% CI: -0.40;0.60]*	No
		Implicit measure of compassion: neutral painting evaluation 45 minutes after product administration							N.S. 0.031 [95% CI: -0.47;0.53]*	No
	Trust (non-	Envelope Task							N.S.	No

	monetary)	60 minutes after product administration							Cohen's *d* = -0.15 OT possibly decreases trust [95% CI: -0.65;0.36]*	No No No
	Behavioral compassion	Number of gazes towards a suffering target 65 minutes after product administration							N.S. Cohen's *d* = -0.12 OT possibly decreases compassion [95% CI: -0.62;039]*	No
		Duration of gaze towards a suffering target 65 minutes after product administration							N.S. Cohen's *d* = 0.03 [95% CI: -0.48;0.53]*	
		Number of interaction with a suffering target							N.S. Cohen's *d* = 0.07 [95% CI:	

		65 minutes after product administration							-0.44;0.57]*	
		Number of interaction with a suffering target 65 minutes after product administration							N.S. Cohen's *d* = -0.09 [95% CI: -0.59;0.41]*	
Study 8: Oxytocin and jealousy in woman (2014)	Jealousy	Self-reported mood (PANAS) 75 minutes after product administration	44 (22OT & 22 PL)	Female	24IU Syntocinon Spray, Defiante Farmaceutica, Funchal, portugal	Double Blind Between-subject design	45 minutes	With life partner & 1 female confederate	N.S. Positive affects: Cohen's *d* = 0.13 OT possibly increases positive affects [95% CI: -0.58;0.83]* Negative affects: Cohen's *d* = -	No

									0.07 [95% CI: -0.66;0.52]*	
		Behavioral jealousy: the mime game 80 minutes after product administration							N.S. Cohen's *d* = -0.35 OT possibly decreases jealousy [95% CI: -0.94;0.25]*	No
		Implicit cognitive measure: word completion 85 minutes after product administration							N.S.Marginally Significant (*p* = .098) Cohen's *d* = -0.52 OT possibly decreases jealousy [95% CI: -1.12;0.08]*	No

		Implicit cognitive measure: positive vs. negative valence words recall 90 minutes after product administration							N.S. Cohen's *d* = -0.03 [95% CI: -0.62;0.56]*	No

[1] Even if our original findings reported in Psychological Science were significant, we have been told afterward that the analysis recommended by our statistician was not controlling for the fact that observations coming from the same subject are dependent. When we perform a repeated measures ANOVA with the partner (computer vs. reliable human partner vs. unreliable human partner) as within-subjects variables and with condition (OT vs. PL) as between-subjects factor, we do not find a significant effect of OT ($F(2,57) = 1.24$, $p = .294$). Therefore our published results seem to be erroneous. The only significant effect of OT we have found was with the computer as partner ($F(1,58) = 4.61$, $p = 0.04$)

* Confidence interval includes zero

Table 2 : Computed effect sizes for main variables and theories

Variable	Cohen's *d*	95% Confidence interval	Size of the effect according to Cohen's norms
Trust (in Studies 1, 6 & 7)	0.04	-0.22 ; 0.30	Null effect size
Compassion (in Studies 2 & 7)	-0.05	-0.21 ; 0.14	Null effect size
Empathy (in Study 2)	- 0.12	-0.42 ; 0.18	Null to small negative effect size
Conformism (in Study 5)	-0.003	-0.36 ; 0.36	Null effect size
Jealousy (in Study 8)	-0.12	-0.39 ; 0.14	Null to small negative effect size

Affects: feeling of sympathy (Study 2), feeling of compassion (Studies 2 & 7), feeling of guilt (Study 5) & mimetic desire (Study 6)	With jealousy[12] = -0.02	-0.19 ; 0.14	
	Without jealousy = -0.003	-0.20 ; 0.24	Null effect size
Behaviors: trust (Studies 1, 6 & 7), compassion (Study 7), guilt (Study 5) & antisocial conformism (Study 5)	With jealousy = 0.06	-0.10 ; 0.22	
	Without jealousy = 0.09	-0.07 ; 0.25	Null effect size
Cognition: RMEt (Study 2), conformism (Study 5) & visual perspective taking (Study 6)	-0.10	-0.32 ; 0.13	Null to small negative effect size

Theory	Cohen's *d*	95% Confidence interval	Size of the effect according to Cohen's norms
Prosocial theory (all variables excepted antisocial conformism (Study 5))	-0.04	-0.13 ; 0.06	Null effect size

[12] As OT could either promotes or decreases jealousy regarding the adopted theoretical approach, we thought important to present both results

Social salience theory (all variables excepted social sharing of the emotions (Study 1))	-0.01	-0.11 ; 0.10	Null effect size
Social Approach theory (all variables excepted RMEt (Study 2) and visual perspective taking (Study 6))	-0.002	-0.11 ; 0.11	Null effect size