Journal of Neuroendocrinology

Journal of Neuroendocrinology, 2016, **28**, 10.1111/jne.12384

ORIGINAL ARTICLE

Is there a Publication Bias in Behavioural Intranasal Oxytocin Research on Humans? Opening the File Drawer of One Laboratory

A. Lane*,†, O. Luminet*,†, G. Nave‡ and M. Mikolajczak*

**Psychological Sciences Research Institute, Université catholique de Louvain – UCL, Louvain-La-Neuve, Belgium.*

†National Fund for Scientific Research - FNRS, Brussels, Belgium.

‡California Institute of Technology, Computation & Neural Systems, Pasadena, CA, USA.

Journal of Neuroendocrinology

The neurohormone oxytocin (OT) has been one the most studied peptides in behavioural sciences over the past two decades. Primarily known for its crucial role in labour and lactation, a rapidly growing literature suggests that intranasal OT (IN-OT) may also play a role in the emotional and social lives of humans. However, the lack of a convincing theoretical framework explaining the effects of IN-OT that would also allow the prediction of which moderators exert their effects and when has raised healthy skepticism regarding the robustness of human behavioural IN-OT research. Poor knowledge of the exact pharmacokinetic properties of OT, as well as crucial statistical and methodological issues and the absence of direct replication efforts, may have lead to a publication bias in the IN-OT literature, with many unpublished studies with null results remaining buried in laboratory drawers. Is there a file drawer problem in IN-OT research? If this is the case, it may also be true in our own laboratory. The present study aims to answer this question, document the extent of the problem and discuss its implications for OT research. For eight studies (including 13 dependent variables overall, as assessed through 25 different paradigms) performed in our laboratory between 2009 and 2014 on 453 subjects, the results obtained were too often not those that were expected. Only five publications emerged from our studies and only one of these reported a null finding. After realising that our publication portfolio has become less and less representative of our actual findings and because the nonpublication of our data might contribute to generating a publication bias in IN-OT research, we decided to retrieve these studies from our drawer and encourage other laboratories to do the same.

Key words: intranasal oxytocin, file drawer, laboratory report

doi: 10.1111/jne.12384

Correspondence to: A. 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)

Behavioural 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, which has a well-established physiological role in labour and lactation, may also play a role in the emotional and social lives of humans.

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 the feelings of others (9), promotes altruism (10), enhances perceived trustworthiness and attractiveness and facilitates parent–infant (11) and romantic (12) attachments. These findings helped to build the reputation of OT 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 antisocial behaviour, such as aggression (13), ethnocentrism (14) and gloating (15). These findings have 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) and the second conjectures that IN-OT increases social approach behaviours, whether good or bad (17). Studies to date have not clearly favoured one theory over the others. Some findings have been consistent

with one (or more) of these theories, although others do not sit easily with either (18).

Another proposition that has emerged from the behavioural IN-OT literature is that the influences of IN-OT 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 such findings, it was proposed that the effect of IN-OT might occur only under certain circumstances or only 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 would exert their effects and when has raised healthy skepticism regarding the robustness of human behavioural 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 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 min after administration (the time window following administration at which most behavioural 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 to produce significant changes in individuals, especially because 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 [e.g. the results of Walum *et al.* (21) indicate that the average study investigating intranasal OT in healthy subjects has a statistical power of 16%] and report overestimated effects. The meta-analysts estimated (using information on power, pre-study odds and the α level) that the false discovery rate in the IN-OT literature is over 80%.

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), which inspired much of the subsequent research, failed to replicate several times (20). Our laboratory has also failed to replicate a promising initial finding relating IN-OT to increased trust in a non-monetary behavioural task [for the original study, see Mikolajczak *et al.* (2); for the failed replication, see Lane *et al.* (27)]. Furthermore, a recent study failed to replicate seminal findings associating IN-OT with mind-reading [for the original study, see Domes *et al.* (7); for the failed replication, see Radke and de Bruijn (28)].

Finally, the methodological challenges accompanying behavioural 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 proposed four, non-mutually exclusive, hypotheses regarding the true association between IN-OT and social behaviour (27):

- The effects reported in the literature reflect the true state of the world, and failed replications are a result of underpowered studies or methodological errors/differences.
- The effects found in the literature are indicative of an effect of IN-OT in humans, whereas the true effect of IN-OT on human behaviour is much smaller than the impression given by published studies. Replications and highly powered studies would therefore allow adjustment of the real effect size.
- The effects found in the literature are type I errors that reflect a publication bias of positive results (32), which is possible because a rate of type I error of 5% is generally accepted.
- The effects of IN-OT do not truly exist but are artificially created (e.g. by extensive degree of researcher freedom (33) and study misconduct, etc.).

If either of the two last hypotheses is true, many unpublished studies with null results might have remained buried in laboratory's drawers (32).

Is there a file drawer problem in IN-OT research? If this is the case, it might also be the case in our own laboratory. This present study aimed 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, as assessed through 25 different paradigms) that were performed in our laboratory from 2009 until 2014 on a total of 453 subjects. All our studies relied on the theoretical and experimental accounts of the role of IN-OT in social behaviour that had been published to date. As we demonstrate below, the results obtained were too often not those that were 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) reported a null finding. We submitted several studies yielding null findings to different journals (from those with a general interest in psychology to those specialised in biological psychology and in psychoenodcrinolgy), although they were rejected time and time again [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)]. After realising that our publication portfolio has become less and less representative of our actual findings, and because the nonpublication of our null findings might contribute to generating a publication bias in IN-OT research, we decided to retrieve 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 laboratory 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-analyse the cumulative effects.

Methods and Results

Methods

We present eight studies assessing 13 dependent variables (emotional, cognitive, behavioural or physiological) through 25 different paradigms, performed in our laboratory over the past 7 years, in chronological order. The methodological details of our studies are summarised in Table 1, and a full description of the studies, including each behavioural task, is provided in the Supporting information (Appendix S1). 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 to eliminate the potential of spillover effects from other tasks (the use of more than one task is common practice because of the imperative to maximise the knowledge gained from each subject undergoing pharmacological treatment). 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 sex (except for the Study 8 on jealousy, which involved couples and focused on female reactions). The number of subjects varied between 12 and 95 (based on the standard found in IN-OT literature) (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 (Sandoz Pharmaceuticals Corporation East Hanover, NJ, USA); between 24 and 40 IU to achieve 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 behavioural IN-OT research. Thus, the first task took place at the earliest approximately 35 min after IN-OT administration (usually 45 min) and, when there were several tasks in the same study, the last task ended no later than 85 min after IN-OT administration (see Table 1, columns 3 and 8). Generally, the subjects performed the experiment alone unless the presence of a confederate was required (Table 1, column 9). Across all studies, there were no differences between the treatment groups [OT versus placebo (PL)] with respect to all baseline measures (all $P > 0.05$) that were focused on self-reported questionnaires regarding the dependent variables relevant to each study (as specified for each study in the Supporting information, Appendix S1). All studies also involved a personality questionnaire and collected demographic information.

Results

The last two columns of Table 1 summarise the main and interaction effects of IN-OT treatment on target behaviours. 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 versus PL) for only five out of 25 tasks across our eight studies and 13 dependent variables (full results and statistical details are provided in the Supporting information, Appendix S1). 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 effect 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 $d > 0.8$), although this result has to be interpreted carefully because we have failed to replicate it twice (27). Furthermore, only one task rules out the zero effect size with a 95% confidence interval (CI) but, once again, the results of this particular study did not replicate well (27).

To determine the extent of the influence of IN-OT on human behaviour in our studies, we meta-analysed [i.e. we computed the cumulative effect sizes using the 'Comprehensive Meta-Analysis' software (36)] the effects of IN-OT on cognitive, emotional or behavioural variables (excluding studies of the effects of OT 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 to 0.10). We further aggregated the effects of IN-OT on variables assessing behaviours, affect or cognition in isolation (Table 2) and could not reliably reject the null hypothesis for either ($d_{behaviours} = 0.09$; 95% CI = −0.07 to 0.25; $d_{affects} = -0.003$; 95% CI = −0.20 to 0.24; $d_{cognitions} = 0.1$; 95% CI = −0.32 to 0.13). Finally, aggregating our effect sizes with respect to the three major behavioural 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)] (Table 2) did not yield any effects that were reliably different from zero ($d_{prosocial} = -0.04$; 95% CI = −0.13 to 0.06; $d_{social\ saliance} = -0.01$; 95% CI = −0.11 to 0.10; $d_{social\ approach} = -0.002$; 95% CI = −0.11 to 0.11).

Discussion

We have reviewed eight studies testing the influence of IN-OT on human cognition and behaviour, assessing 13 dependent variables through 25 different paradigms performed in our laboratory 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 versus PL) for only five out of 25 tasks. All of our hypotheses were derived from the three major behavioural 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 the main effect of IN-OT) raises concerns about the validity of what we know about the influence of IN-OT on human behaviours 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 for the 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 the behavioural IN-OT effects does not reflect the true state of the world, how has the vast behavioural 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 false positive rate of 5%). If this is the case, much unpublished data must remain buried in the drawers of laboratories studying IN-OT.

Second, the significant effect of IN-OT may be the result of methodological, measurement or statistical artefacts. Because this

Table 1. Presentation of the Studies, Including Methodology and Results.

Study	Dependant variable	Paradigm and time following product administration	Number of participants	Sex of the participants	Dose and product	Administration type and design	Time between administration and testing	Testing type	Oxytocin main effect	Interaction effect
Study 1: Oxytocin, trust and social sharing of the emotions (2009)	Trust (monetary)	Trust game 45 min after product administration	60 (30 OT and 30 PL)	Male	32 IU Syntocinon spray, Novartis, Basel Switzerland	Single-blind between-subject design	45 min	Participant alone	NS* Cohen's d = 0.13 OT possibly increases trust [95% CI: −0.38 to 0.64]†	Condition × Partner reliability: OT only increases trust for reliable partners
	Social sharing of the emotions	Self-reported willingness to share emotions 55 min after product administration							NS Cohen's d = 0.19 OT possibly increases the willingness to share emotions [95% CI: −0.33 to 0.70]†	Condition × Content of the sharing (facts versus emotions): OT only increases willingness to share emotions
	Trust (non-monetary)	Envelope task 65 min after product administration							Significant Cohen's d = 2.09 OT increases trust [95% CI: 0.80 to 3.38]	No

(continued)

Table 1. (continued)

Study	Dependant variable	Paradigm and time following product administration	Number of participants	Sex of the participants	Dose and product	Administration type and design	Time between administration and testing	Testing type	Oxytocin main effect	Interaction effect
Study 2: Oxytocin and empathy (2009)	Empathy	Reading the mind in the eyes test 45 min after product administration	60 (30 OT and 30 PL)	Male	32 IU Syntocinon spray, Novartis, Basel Switzerland	Single-blind between-subject design	45 min	Participant alone	NS Cohen's d = 0.26 OT possibly increases mind reading [95% CI: −0.26 to 0.78]†	Condition × 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 min after product administration							NS Cohen's d = −0.39 OT possibly decreases compassion [95% CI: −0.91 to 0.14]†	No
	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 min after product administration							NS Sympathy: Cohen's d = −0.42 OT possibly decreases sympathy [95% CI: −0.93 to 0.10]† Help: Cohen's d = −0.19 OT possibly decreases helping behaviours 95% CI: −0.70 to 0.32]†	No

(continued)

Table 1. (continued)

Study	Dependant variable	Paradigm and time following product administration	Number of participants	Sex of the participants	Dose and product	Administration type and design	Time between administration and testing	Testing type	Oxytocin main effect	Interaction effect
Study 3: Oxytocin and sleep (2011)	Sleep latency	Multiple sleep latency test 45 min after product administration	12	Male	32 IU Syntocinon spray, Novartis, Basel Switzerland	Single-blind within-subject design	45 min	Participant alone	NS Cohen's d = −0.14 OT possibly decreases sleep latency [95% CI: −0.94 to 0.66]†	No
	Sleep duration								NS Cohen's d = 0.27 OT possibly increases sleep duration [95% CI: −0.48 to 1]†	No
	Proportion of REM sleep								NS Cohen's d = 0.68 OT possibly increases REM sleep proportion [95% CI: −0.14 to 1.48]†	No
	Psychomotor vigilance	Psychomotor vigilance task							NS Cohen's d = −0.41 OT possibly decreases psychomotor vigilance [95% CI: −1.20 to 0.04]†	No

(continued)

Table 1. (continued)

Study	Dependant variable	Paradigm and time following product administration	Number of participants	Sex of the participants	Dose and product	Administration type and design	Time between administration and testing	Testing type	Oxytocin main effect	Interaction effect
Study 4: Oxytocin, pain and sensitivity to baby's cry (2011)	Pain threshold	Cold pressure test 45 min after product administration	60 (30 OT and 30 PL)	Male	32 IU Syntocinon spray, Novartis, Basel Switzerland	Double-blind between-subjects design	45 min	Participants alone	NS Cohen's d = −0.28 OT possibly decreases pain threshold [95% CI: −0.78 to 0.23]†	No
	Pain tolerance								NS Cohen's d = 0.16 OT possibly increases pain tolerance [95% CI: −0.35 to 0.66]†	No
	Willingness to endure pain								NS Cohen's d = 0.32 OT possibly increases willingness to endure pain [95% CI: −0.20 to 0.82]†	No
	Perceived pain intensity								N.S Cohen's d = 0.19 OT possibly increases perceived pain intensity [95% CI: −0.32 to 0.70]†	No
	Sensitivity to a baby's cry	Self-reported annoyance from baby's cry sound tracks 55 min after product administration							NS Cohen's d = 0.24 OT possibly increases sensitivity to baby's cry [95% CI: −0.27 to 0.75]†	No

(continued)

Table 1. (continued)

Study	Dependant variable	Paradigm and time following product administration	Number of participants	Sex of the participants	Dose and product	Administration type and design	Time between administration and testing	Testing type	Oxytocin main effect	Interaction effect
Study 5: The dark side of oxytocin: guilt, conformism and compliance to antisocial behaviours (2012)	Compliance to antisocial behaviours	Antisocial peer pressure 35 min after product administration	61 (31 OT and 30 PL)	Male	40 IU Syntocinon spray, Novartis, Basel Switzerland	Double-blind between-subject design	35 min	With 2 confederates	NS Cohen's d = 0.47 OT possibly increases compliance to peer's antisocial requests [95% CI: −0.05 to 0.98]†	No
	General conformism	Numeric estimation task 45 min after product administration						Alone	NS Cohen's d = −0.47 OT possibly decreases conformism [95% CI: −0.99 to 0.04]†	No
	Behavioural measure of guilt after guilt induction	Effective splitting of money with partner or charity to make amend 75 min after product administration						Alone	NS Cohen's d = 0.33 OT possibly increases guilt [95% CI: −0.18 to 0.83]†	No
	Guilt after guilt induction	self-reported questionnaire 85 min after product administration						With 1 confederate	NS Cohen's d = 0.41 OT possibly increases guilt [95% CI: −0.10 to 0.92]†	No

(continued)

Table 1. (continued)

Study	Dependant variable	Paradigm and time following product administration	Number of participants	Sex of the participants	Dose and product	Administration type and design	Time between administration and testing	Testing type	Oxytocin main effect	Interaction effect
Study 6: Oxytocin, mimetic desire, Visual perspective taking and Trust (2012)	Mimetic desire	Neutral painting evaluation task (looked at versus looked away) 45 min after product administration	95 (48 OT and 47 PL)	Male	32 IU Syntocinon spray, Fuerte Farma, Funchal, portugal	Double-blind between-subject design	45 min	Alone	NS Cohen's d = 0.19 OT possibly increases mimetic desire [95% CI: −0.22 to 0.60]†	No
	Self versus others' visual perspective taking	Visual perspective taking task (accuracy) 55 min after product administration							NS Cohen's d = −0.17 OT possibly decreases visual perspective accuracy [95% CI: −0.57 to 0.23]†	No
		Visual perspective taking task (reaction time) 55 min after product administration							NS Cohen's d = 0.01 OT does not influence visual perspective reaction time [95% CI: −0.39 to 0.41]†	No
	Trust (non-monetary)	Envelope task 65 min after product administration							NS Cohen's d = −0.10 OT possibly decreases trust [95% CI: −0.50 to 0.30]†	No

(continued)

Table 1. (continued)

Study	Dependant variable	Paradigm and time following product administration	Number of participants	Sex of the participants	Dose and product	Administration type and design	Time between administration and testing	Testing type	Oxytocin main effect	Interaction effect
Study 7: Oxytocin, compassion and trust (2013)	Vicarious experience of another's distress	Explicit measure of compassion: self-reported evaluation 45 min after product administration	61 (32 oxytocin and 29 PL)	Male	32 IU Syntocinon spray, Defiante Farmaceutica, Funchal, portugal	Double-blind between-subject design	45 min	Alone	NS Cohen's d = 0.10 OT possibly increases compassion [95% CI: −0.40 to 0.60]†	No
		Implicit measure of compassion: neutral painting evaluation 45 min after product administration							NS 0.031 [95% CI: −0.47 to 0.53]†	No
	Trust (non-monetary)	Envelope task 60 min after product administration							NS Cohen's d = −0.15 OT possibly decreases trust [95% CI: −0.65 to 0.36]†	No No No No
	Behavioural compassion	Number of gazes towards a suffering target 65 min after product administration							NS Cohen's d = −0.12 OT possibly decreases compassion [95% CI: −0.62 to 039]†	No
		Duration of gaze towards a suffering target 65 min after product administration							NS Cohen's d = 0.03 [95% CI: −0.48 to 0.53]†	
		Number of interaction with a suffering target 65 min after product administration							NS Cohen's d = 0.07 [95% CI: −0.44 to 0.57]†	
		Number of interaction with a suffering target 65 min after product administration							NS Cohen's d = −0.09 [95% CI: −0.59 to 0.41]†	

(continued)

Table 1. (continued)

Study	Dependant variable	Paradigm and time following product administration	Number of participants	Sex of the participants	Dose and product	Administration type and design	Time between administration and testing	Testing type	Oxytocin main effect	Interaction effect
Study 8: Oxytocin and jealousy in woman (2014)	Jealousy	Self-reported mood (PANAS 75 min after product administration	44 (22OT and 22 PL)	Female	24 IU Syntocinon spray, Defiante Farmaceutica, Funchal, portugal	Double-blind between-subject design	45 min	With life partner and 1 female confederate	NS Positive affects: Cohen's d = 0.13 OT possibly increases positive affects [95% CI: −0.58 to 0.83]† Negative affects: Cohen's d = −0.07 [95% CI: −0.66 to 0.52]†	No
		Behavioural jealousy: the mime game 80 min after product administration							NS Cohen's d = −0.35 OT possibly decreases jealousy [95% CI: −0.94 to 0.25]†	No
		Implicit cognitive measure: word completion 85 min after product administration							NS Cohen's d = −0.52 OT possibly decreases jealousy [95% CI: −1.12 to 0.08]†	No
		Implicit cognitive measure: positive versus negative valence words recall 90 min after product administration							NS Cohen's d = −0.03 [95% CI: −0.62 to 0.56]†	No

*Even if our original findings reported in *Psychological Science* were significant, we have been informed subsequently 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 versus reliable human partner versus unreliable human partner) as within-subjects variables and with condition [OT versus placebo (PL)] as between-subjects factor, we do not find a significant effect of OT ($F_{2,57} = 1.24$, $P = 0.294$). Therefore, our published results appear 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 (CI) includes zero. PANAS, Positive and Negative Affect Schedule.

Table 2. Computed Effect Sizes for Main Variables and Theories.

	Cohen's d	95% Confidence interval	Size of the effect according to Cohen's norms
Variable			
Trust (in Studies 1, 6 and 7)	0.04	−0.22 to 0.30	Null effect size
Compassion (in Studies 2 and 7)	−0.05	−0.21 to 0.14	Null effect size
Empathy (in Study 2)	−0.12	−0.42 to 0.18	Null to small negative effect size
Conformism (in Study 5)	−0.003	−0.36 to 0.36	Null effect size
Jealousy (in Study 8)	−0.12	−0.39 to 0.14	Null to small negative effect size
Affects: feeling of sympathy (Study 2), feeling of compassion (Studies 2 and 7), feeling of guilt (Study 5) and mimetic desire (Study 6)	With jealousy = −0.02* Without jealousy = −0.003	−0.19 to 0.14 −0.20 to 0.24	Null effect size
Behaviours: trust (Studies 1, 6 and 7), compassion (Study 7), guilt (Study 5) and antisocial conformism (Study 5)	With jealousy = 0.06 Without jealousy = 0.09	−0.10 to 0.22 −0.07 to 0.25	Null effect size
Cognition: RMEt (Study 2), conformism (Study 5) and visual perspective taking (Study 6)	−0.10	−0.32 to 0.13	Null to small negative effect size
Theory			
Prosocial theory [all variables excepted antisocial conformism (Study 5)[	−0.04	−0.13 to 0.06	Null effect size
Social salience theory [all variables excepted social sharing of the emotions (Study 1)]	−0.01	−0.11 to 0.10	Null effect size
Social approach theory [all variables excepted RMEt (Study 2) and visual perspective taking (Study 6)]	−0.002	−0.11 to 0.11	Null effect size

*Because OT could either promotes or decreases jealousy regarding the adopted theoretical approach, we considered it important to report both results.

has been demonstrated for peripheral OT measurements (29), it should not be excluded here, although the artifacts would be different. We recognise four potential sources of generating artefacts in IN-OT research: (i) small sample between subject-designs that might not be internally valid; (ii) single-blind studies; (iii) IN-OT pharmacokinetics and dosage; and (iv) statistical methods.

The massive use of between-subject designs of relatively small samples (approximately 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) (note that within-subject designs also suffer from limitations such as reduced statistical power;[1].

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

Recent findings have demonstrated that IN-OT increases OT concentration in CSF in both humans (25) and animals (38,39). Furthermore, it has recently been demonstrated that IN-OT modulates amygdala responses in primates in a manner equivalent to humans (40). Taken together, such results suggest that IN-OT reaches, directly or indirectly (41), the central nervous system and this would produce observable affective, behavioural or cognitive modifications. However, if IN-OT produces a significant elevation of OT concentration in the CSF after 30 min in animals, this significant elevation takes place 75 min after IN-OT in humans, which is not consistent with the literature, where most tasks start 40–45 min after IN-OT. Furthermore, in a recent study, Quintana *et al.* (42) suggest that the IN-OT doses commonly used (24–40 IU) may not be the most adequate. Their results show that the effect of IN-OT on emotional recognition appears with an administration of 8 IU but not with 24 IU. Facing these challenges, further studies are needed to strengthen our knowledge about IN-OT pharmacokinetic properties. Even if IN-OT reaches the brain, we cannot be sure whether the three assumptions on which the IN-OT literature is based are valid.

Finally, the use of too small samples (21) and the vast amount of candidate factors that could potentially moderate the behavioural effects of IN-OT (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 apparently puzzling results described in the present study.

[1]for example, 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)

First, our studies, similar to most published studies on IN-OT, might be underpowered (21). Thus, the fact that the effects of IN-OT observed in our studies are nonsignificant 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) [we have excluded the highest effect size found in Study 1 (non-monetary trust assessment) because it has been questioned by Lane *et al.* (27)]. To detect such effects, or even a moderate effect, a sample size of 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 (Read the Mind in th Eyes test), Cohen's d = 0.260) participants would be required to detect an IN-OT effect with 80% power. Such sample sizes are much greater than the norm in the IN-OT field. Therefore, several of our findings could potentially have turned significant in well-powered experiments. Yet, as shown in Table 1, their significance would might not always be in the expected direction.

A second proposition is that IN-OT effects do exist, although they are strongly moderated by various factors, making them appear large under some circumstances but not in others. Through the literature, more and more findings suggest that IN-OT influences behaviours by interacting with several moderators (19). Arguably, our findings do not rule out the possibility that the effects of IN-OT are moderated by various factors, which is a proposition that is difficult to rule out, given the infinitely large set of factors that could potentially moderate the behavioural influences of IN-OT (e.g. genes, personality or environmental factors). Unfortunately, as far as we know, candidate moderators do not appear to replicate from one study to another [e.g. in their failed replication of the influence of IN-OT on the RMEt, Radke and de Bruijn (28) did not find any moderating effect of the difficulty of an item as demonstrated by Domes *et al.* (7)] and appear most often to represent post-hoc data fits rather than *a priori* hypotheses [we do not make an exception to the rule: (34)]. Indeed, a 'significant' interaction can be found 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 α is corrected for multiple hypothesis testing (43,44).

We can either believe that these interactions are statistical artefacts (see above) or consider them to be real. If we believe that they are real, it means that there is no such 'general effect of IN-OT on behaviour' but that the effects of IN-OT are always context-dependent (19). In the studies reported within the present paper, relevant potential moderators have been taken into account and only provided five interaction effects. It is possible that less obvious moderators, or moderators that we did not measure, would have provided more significant effects.

At present, we still do not know which of the four hypotheses is true; IN-OT might not influence human behaviours at all or may influence it only under specific circumstances. In any case, falsifiable theories must emerge to enable progress in our understanding of the behavioural influences of IN-OT because no current theory appears to yield robust behavioural predictions, and almost every behavioural effect can be explained by one of the theories after the fact. 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)].

In our view, nothing can be taken for granted with IN-OT and some nonreplicable findings might have biased the development of existing theories. Hopefully, incorporating null findings and failed replications into the theoretical process will allow lines to be drawn between robust, replicable IN-OT effects and facilitate the development of falsifiable theories. It is therefore crucial that nonsignificant findings and failed replications are published (http://psychfiledrawer.org). Every piece of evidence, even for those experiments that did not yield 'significant' effects, should be taken into account and weighted according to its evidential value.

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

We believe that a systematic shift in the IN-OT publication process is essential for revealing the true state of the world. Pre-registration of ex-ante hypotheses, replication attempts of the findings before their submission, and the submission of null results and failed replications 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, regardless of their results (47). Many laboratories do report their work transparently. However, because the editorial and review process does not sufficiently promote nonsignificant results and failed replications, it is difficult to obtain a complete overview of the IN-OT research field. One way of improving the standards is by institutionalisation [as is encouraged 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 NIH (http://grants.nih.gov/grants/policy/data_sharing/data_sharing_faqs.htm#900)], 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 those 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, researchers could easily test the robustness of their findings by adjusting the α 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 encourage a solid theoretical background for interpreting and understanding the complex effects of IN-OT and justify all of the efforts and resources invested in IN-OT research.

Acknowledgements

The authors of the manuscript have no conflicts of interest to declare. We thank the Cliniques Universitaires Saint-Luc (University Hospital), which hosted our experiments. We also thank Gillian Rosner for proofreading the version of the manuscript submitted for publication.

Received 28 August 2015,
revised 9 March 2016,
accepted 9 March 2016

References

1 Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E. Oxytocin increases trust in humans. *Nature* 2005; **435**: 673–676.
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. *Biol Psychol* 2010; **85**: 182–184.
3 Cardoso C, Linnen A, Joober R, Ellenbogen MA. Coping style moderates the effect of intranasal oxytocin on the mood response to interpersonal stress. *Exp Clin Psychopharmacol* 2012; **20**: 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. *Biol Psychiatry* 2003; **54**: 1389–1398.
5 Rimmele U, Hediger K, Heinrichs M, Klaver P. Oxytocin makes a face in memory familiar. *J Neurosci* 2009; **29**: 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. *Biol Psychiatry* 2007; **62**: 1187–1190.
7 Domes G, Heinrichs M, Michel A, Berger C, Herpertz SC. Oxytocin improves 'mind-reading' in humans. *Biol Psychiatry* 2007; **61**: 731–733.
8 Lane A, Luminet O, Rimé B, Gross JJ, de Timary P, Mikolajczak M. Oxytocin increases willingness to socially share one's emotions. *Int J Psychol* 2013; **48**: 676–681.
9 Hurlemann R, Patin A, Onur OA, Cohen MX, Baumgartner T, Metzler S, Dziobek I, Gallinat J, Wagner M, Maier W, Kendrick K. Oxytocin enhances amygdala-dependent, socially reinforced learning and emotional empathy in humans. *J Neurosci* 2010; **30**: 4999–5007.
10 Zak PJ, Stanton AA, Ahmadi S. Oxytocin increases generosity in humans. *PLoS ONE* 2007; **2**: e1128.
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**: 1583–1586.
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. *Biol Psychiatry* 2009; **65**: 728–731.
13 De Dreu CKW, Greer LL, Handgraaf MJJ, Shalvi S, Van Kleef GA, Baas M, Ten Velden F, Van Dijk E, Feith S. The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science* 2010; **328**: 1408–1411.
14 De Dreu CKW, Greer LL, Van Kleef GA, Shalvi S, Handgraaf MJJ. Oxytocin promotes human ethnocentrism. *Proc Natl Acad Sci USA* 2011; **108**: 1262–1266.
15 Shamay-Tsoory SG, Fischer M, Dvash J, Harari H, Perach-Bloom N, Levkovitz Y. Intranasal administration of oxytocin increases envy and schadenfreude (gloating). *Biol Psychiatry* 2009; **66**: 864–870.
16 Insel TR. Oxytocin – a neuropeptide for affiliation: evidence from behavioral, receptor autoradiographic, and comparative studies. *Psychoneuroendocrinology* 1992; **17**: 3–35.
17 Kemp AH, Guastella AJ. The role of oxytocin in human affect: a novel hypothesis. *Curr Dir Psychol Sci* 2011; **20**: 222–231.
18 Bethlehem RA, Baron-Cohen S, van HJ, Auyeung B, Bos PA. The oxytocin paradox. *Front Behav Neurosci* 2014; **8**: 48.
19 Bartz JA, Zaki J, Bolger N, Ochsner KN. Social effects of oxytocin in humans: context and person matter. *Trends Cogn Sci* 2011; **15**: 301–309.
20 Nave G, Camerer C, McCullough M. Does Oxytocin increase trust in humans? Critical review of research. *Pers Psychol Sci* 2015; **10**: 772–789.
21 Walum H, Waldman ID, Young LJ. Statistical and methodological considerations for the interpretation of intranasal oxytocin studies. *Biol Psychiatry* 2016; **79**: 251–257.
22 Born J, Lange T, Kern W, McGregor GP, Bickel U, Fehm HL. Sniffing neuropeptides: a transnasal approach to the human brain. *Nat Neurosci* 2002; **5**: 514–516.
23 Meyer-Lindenberg A, Domes G, Kirsch P, Heinrichs M. Oxytocin and vasopressin in the human brain: social neuropeptides for translational medicine. *Nat Rev Neurosci* 2011; **12**: 524–538.
24 Leng G, Ludwig M. Intranasal oxytocin: myths and delusions. *Biol Psychiatry* 2016; **79**: 243–250.
25 Striepens N, Kendrick KM, Hanking V, Landgraf R, Wullner U, Maier W, Hurlemann R. Elevated cerebrospinal fluid and blood concentrations of oxytocin following its intranasal administration in humans. *Sci Rep* 2013; **3**: 3440.
26 Simons DJ. The value of direct replication. *Pers Psychol Sci* 2014; **9**: 76–80.
27 Lane A, Mikolajczak M, Treinen E, Samson D, Corneille O, de Timary P, Luminet O. Failed replication of oxytocin effects on trust: the envelope task case. *PLoS ONE* 2015; **10**: e0137000.
28 Radke S, de Bruijn ER. Does oxytocin affect mind-reading? A replication study *Psychoneuroendocrinology* 2015; **60**: 75–81.
29 McCullough ME, Churchland PS, Mendez AJ. Problems with measuring peripheral oxytocin: can the data on oxytocin and human behavior be trusted? *Neurosci Biobehav Rev* 2013; **37**: 1485–1492.
30 Szeto A, McCabe PM, Nation DA, Tabak BA, Rossetti MA, McCullough ME, Schneiderman N, Mendez AJ. Evaluation of enzyme immunoassay and radioimmunoassay methods for the measurement of plasma oxytocin. *Psychosom Med* 2011; **73**: 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**: e116172.
32 Rosenthal R. The file drawer problem and tolerance for null results. *Psychol Bull* 1979; **86**: 638.
33 Simonsohn U, Nelson LD, Simmons JP. P-curve: a key to the file-drawer. *J Exp Psychol Gen* 2014; **143**: 534.
34 Luminet O, Grynberg D, Ruzette N, Mikolajczak M. Personality-dependent effects of oxytocin: greater social benefits for high alexithymia scorers. *Biol Psychol* 2011; **87**: 401–406.
35 Mikolajczak M, Gross JJ, Lane A, Corneille O, de Timary P, Luminet O. Oxytocin makes people trusting not gullible. *Psychol Sci* 2010; **21**: 1072–1074.
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**: 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**: 1985–1993.
39 Freeman SM, Samineni S, Allen PC, Stockinger D, Bales KL, Hwa GG, Roberts JA. Plasma and CSF oxytocin levels after intranasal and intravenous oxytocin in awake macaques. *Psychoneuroendocrinology* 2016; **66**: 185–194.
40 Liu N, Hadj-Bouziane F, Jones KB, Turchi JN, Averbeck BB, Ungerleider LG. Oxytocin modulates fMRI responses to facial expression in macaques. *Proc Natl Acad Sci USA* 2015; **112**: E3123–E3130.
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. *Neurosci Biobehav Rev* 2015; **49**: 182–192.
42 Quintana D, Westlye L, Rustan Ø, Tesli N, Poppy C, Smevik H, Tesli M, Roine M, Mahmoud RA, Smerud KT, Djupesland PG, Andreassen OA. Low-dose oxytocin delivered intranasally with Breath Powered device affects social-cognitive behavior: a randomized four-way crossover trial with nasal cavity dimension assessment. *Transl Psychiat* 2014; **5**: e602.
43 Shaffer JP. Multiple hypothesis testing. *Annu Rev Psychol* 1995; **46**: 561–584.
44 McKay B, Bar-Natan D, Bar-Hillel M, Kalai G. Solving the Bible Code Puzzle. *Stat Sci* 1999; **14**: 150–173.
45 Koole SL, Lakens D. Rewarding replications a sure and simple way to improve psychological science. *Pers Psychol Sci* 2012; **7**: 608–614.
46 Klackl J, Pfundmair M, Agroskin D, Jonas E. Who is to blame? Oxytocin promotes nonpersonalistic attributions in response to a trust betrayal. *Biol Psychol* 2013; **92**: 387–394.
47 John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol Sci* 2012; **23**: 524–532.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1. Detailed presentation of the 8 studies.