Faculteit Geneeskunde en Gezondheidswetenschappen

A neuroendocrine study of cooperative behavior: the moderating effects of oxytocin, decision context, and social values

Een neuro-endocriene studie over coöperatief gedrag: de invloed van oxytocine, beslissingscontext en sociale waarden

Proefschrift voorgelegd tot het behalen van de graad van
doctor in de medische wetenschappen
aan de Universiteit Antwerpen te verdedigen door

Bruno LAMBERT

Promotoren
Prof. Dr. Carolyn H. Declerck
Prof. Dr. Christophe Boone
Prof. Dr. Paul M. Parizel

Antwerpen, 2017

Doctoral jury

Prof. dr. Carolyn H. Declerck (Supervisor)
University of Antwerp, Belgium

Prof. dr. Christophe Boone (Supervisor)
University of Antwerp, Belgium

Prof. dr. Paul M. Parizel (Supervisor)
University of Antwerp, Antwerp University Hospital, Belgium

Prof. dr. Geert Dom (Chairman doctoral committee)
University of Antwerp, Belgium

Prof. dr. Andrew Maas (Member doctoral committee)
University of Antwerp, Antwerp University Hospital, Belgium

Prof. dr. Ale Smidts (External jury member)
Erasmus University, The Netherlands

Prof. dr. Madelon M.E. Hendricx-Riem (External jury member)
Tilburg University, The Netherlands

Contact information

Bruno Lambert

University of Antwerp
Faculty of Applied Economics, Management Department
Kipdorp 61, office S.Z.407
2000 Antwerp, Belgium

+32 (0) 496 78 78 19

Bruno.lambert@uantwerp.be
Bruno-lambert@skynet.be

Opgedragen aan

Emiel Lambert

Vera De Mey

Dankwoord

Onderzoek doe je nooit alleen. Dat ik mijn doctoraat heb kunnen afleggen, is te danken aan de professionele begeleiding en onuitputtelijke kennis van mijn promotoren.

Carolyn Declerck, u hebt voor mij altijd de nodige tijd voorzien. Geen moeite was te veel en uw deur stond altijd open, ongeacht de aard van de problemen. Ik heb veel van u geleerd, zowel op persoonlijk als op academisch vlak. Hartelijk bedankt om in mij te geloven!

Christophe Boone, uw begeleiding heeft me vaak geholpen om de zaken terug in het juiste licht te kunnen zien. Vergaderen met u was niet altijd gemakkelijk, maar wel motiverend. Dankzij u ben ik als kritisch wetenschapper gegroeid. Dank u voor uw steun.

Paul Parizel, u hebt mij steeds alle mogelijkheden aangereikt om mijn onderzoek in goede banen te leiden. Hierdoor heb ik mezelf kunnen ontplooien als onderzoeker. Bedankt hiervoor.

Ik dank ook de leden van de doctoraatscommissie, prof. Dom en prof. Maas, om mij op te volgen en de externe juryleden, prof. Smidts en prof. Hendricx-Riem, voor de interesse in mijn werk.

Ik wil ook mijn erkentelijkheid tonen aan al de mensen die mij geholpen hebben bij het uitvoeren van mijn experimenten. Specifiek bedank ik hierbij Floris Vanhevel en Griet Emonds.

Ik kan ook niet verder gaan zonder de vele collega's te benoemen die mij tijdens al deze jaren vele mooie en plezierige momenten hebben bezorgd. Anja, je was een super bureaugenoot! Loren, bedankt voor de vele leuke gesprekken en je hulp bij de experimenten! Eline, Johanna, Eline, Wim, Wouter, maar recentelijke ook Lode, Jolien, Vicky, Kim, Annelies, Nele, Konrad en Danica, en alle andere ACED collega's en ex-collega's, bedankt voor de gezellige wandelingen, toffe babbels en de vele aangename middagpauzes!

Mijn vrienden en uitgebreide familie hebben bijgedragen door de vele etentjes, hulp op de momenten dat het nodig was en hun warme vriendschap.

Ik bedankt hierbij ook An. Jou liefde is steeds een grote steun en hulp geweest. Het maakte de moeilijke momenten dragelijk en de mooie momenten des te beter.

Ik kan hier ook niet afsluiten zonder mijn ouders te bedanken. Het spijt me dat ze hier nu niet bij kunnen zijn, want het is via hun jarenlange liefde, hulp en begeleiding, dat ik op dit punt in mijn leven ben gekomen.

Bruno

Table of contents

Introduction

1. Social judgments and decisions: what drives us to cooperate?

A defining feature of human social groups is the high level of cooperation, which has made it possible for us to form societies and build economies. But achieving cooperation is not an obvious or a simple task because our social groups are highly complex and often interdependent. That is, the decisions we make in social interactions do not only affect ourselves, but also impact others. And vice versa, my own outcome is dependent on the actions of myself as well as those of others. This makes social decisions strategic: we decide in function of what we anticipate the others will do. To do so quickly and effectively, the decision to cooperate, or not, will depend on many different factors.

First, humans have a strong self-interest motive that drives them to select behavior that maximizes positive outcomes for themselves. Cooperative behavior is very likely when both parties know that cooperation leads to synergy and that they are better off by working together. But when resources are scarce, and the gains of one person equals the loss of another (zero-sum situation), the self-interest motive should drive people to compete. Similarly, in situations where it is possible to benefit from the cooperative efforts of others, the temptation to profit without contributing may undermine cooperation and lead to free-riding. In this vain, traditional economists view mankind as *Homo economicus*, making rational decisions based solely on the anticipated gain of a social interaction. Cooperation in this case will only occur if it pays off to do so. In experimental economics and social psychology, (non-)cooperative behavior has been extensively studied using economic games that manipulate pay-offs, creating decision contexts that can either incentivize cooperation, competition, or free-riding (Van Lange, Joireman, Parks, & Van Dijk, 2013).

Second, humans also have social motives that prompt behavior that benefits others and does not seem to maximize personal gains. Humans are social beings and, from birth on, depend on group living to survive. As a result group belonging is highly valued and encourages behavior that sustains the well-being of the group (Baumeister & Leary, 1995). These social motives can interfere with self-interest and drive some people to cooperate even at a cost to themselves. This behavior has been documented on all continents and in all cultures (Henrich et al., 2005). Psychologists have paid much attention to individual differences in so-called human social value orientation (SVO), a stable trait that predicts individuals' intrinsic tendencies to engage in prosocial behaviors and cooperate purely for the internal satisfaction they obtain from helping others (Bogaert, Boone, & Declerck, 2008). Social value orientation can be considered a compass in social decision making, introducing a bias favoring mutually beneficial outcomes. But cooperating for social rewards in the absence of

cooperative incentives makes individuals vulnerable to betrayal. Therefore an additional factor that affects cooperative decision making is the extent to which people expect that others can be trusted. Trust is a prerequisite to establish and maintain cooperation when pay-offs are such that free-riding and/or betrayal are possible.

Therefore, the third factor that affects cooperative decision making is the presence of social cues that relay information about the interaction partners. When pay-offs are uncertain, cues that help to assess the intentions of others are important. Especially individuals with a prosocial value orientation who are intrinsically motivated to cooperate are known to be sensitive to trustworthiness cues to reduce their fear of betrayal (Boone, Declerck, & Kiyonari, 2010). But also when pay-offs incentivize competition, social cues can be helpful to infer the strategy of one's adversary, impacting the decision to either engage in the interaction or withdraw from it altogether if winning is perceived to be unlikely.

The proposition that cooperative behavior is driven by a combination of selfish motives (triggered by the expected pay-offs), an individual's social value orientation, and contextual social cues (see figure 1), is consistent with recent literature in decision neuroscience highlighting the importance of heuristics in social decision making (Bechara & Damasio, 2005; Kahneman, 2003). Heuristics are mental shortcuts that reframe complex decision environments into ones that are "fast and frugal" and amenable to real-life human reasoning (Gigerenzer & Todd, 1999). In this sense, pay-offs, social value orientation, and social cues introduce a bias in decision making by framing how an individual will interpret the environment and/or his interaction partners, facilitating or impeding cooperative behavior according to readily available, relevant information and personal values.

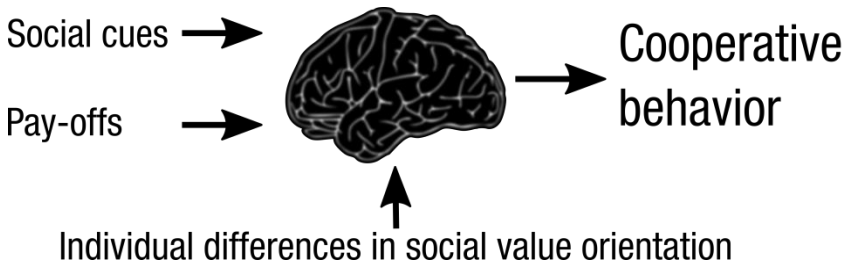

Fig. 1. Overview of different factors that affect deciding on cooperative behavior

While the role of these three factors with respect to human prosocial behavior has received much coverage in psychology, economics and decision neuroscience, they have been mostly investigated independently, with little attention to how they interact. In addition, the heuristics literature is silent about how the brain processes information to arrive at decisions. The aim of this dissertation is to fill

this gap and to study how combinations of the factors shown in figure 1 are integrated by the brain and contribute to heuristic processes and subsequent (non-)cooperative behavior.

In the first three chapters of this dissertation, we zoom in on the mechanisms behind heuristic processing by first studying how social cues are evaluated, and next how they automatically affect decision making when pay-offs incentivize either cooperation or competition. We focus in particular on the role of the neurohormone oxytocin (OT), and investigate how it affects 'fast and frugal' social judgments that are capable of significantly impacting behavior without conscious reasoning. In the fourth and final chapter, we investigate how cooperation can develop over time as individuals are accumulating more and more trustworthiness information of different partners. By means of neuroimaging, we identify which brain regions are recruited in the process of learning to cooperate.

2. Oxytocin: modulator of social behavior

An important contribution of this dissertation is that, in order to understand heuristic processing, we use an experimental approach in which we manipulate oxytocin (OT) neuromodulation.

While the nonapeptide OT (see figure 2 for the structure) is best known for its hormonal functions in reproduction, in the last decade its role as a neurotransmitter regulating social behavior has been increasingly highlighted. In animal studies (McCall & Singer, 2012), it is well known that OT neurotransmission in the central nervous system regulates not only maternal, but also other social behavior such as social recognition (in rodents (Ferguson, Young, & Insel, 2002) and sheep (Keverne & Kendrick, 1992)), pair bonding (prairie voles (Ross et al., 2009)), flocking (zebra finch (Goodson, Schrock, Klatt, Kabelik, & Kingsbury, 2009)), grooming (macaques (Maestripieri, Hoffman, Anderson, Carter, & Higley, 2009)), sharing and cooperating (vampire bat (Carter & Wilkinson, 2015) and chimpanzees (Wittig et al., 2014)) among others.

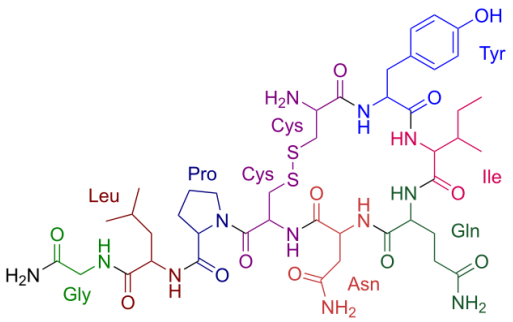

Fig. 2. The chemical structure of oxytocin, with 9 amino acids: cysteine (cys), tyrosine (tyr), isoleucine (ile), glutamine (gln), asparagine (asn), proline (pro), leucine (leu) and glycine (gly) (Oxytocin, 2016, juli 14).

Early experimental studies with humans showed that administering exogenous OT via a nasal spray increased trusting behavior (Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005). Subsequent studies further highlighted the role of OT in many aspects of social behavior such as judgment and perception (e.g. Domes, Steiner, Porges, & Heinrichs, 2013; Lambert, Declerck, & Boone, 2014; Leknes et al., 2012; Marsh, Yu, Pine, & Blair, 2010; Theodoridou, Rowe, Penton-Voak, & Rogers, 2009), approach or withdrawal (Harari-Dahan & Bernstein, 2014; Liu, Guastella, & Dadds, 2012; Preckel, Scheele, Kendrick, Maier, & Hurlemann, 2014; Radke, Roelofs, & de Bruijn, 2013; Scheele et al., 2012; Theodoridou, Penton-Voak, & Rowe, 2013), social anxiety (Kirsch et al., 2005; Labuschagne et al., 2010; Petrovic, Kalisch, Singer, & Dolan, 2008), social stress (Cardoso, Ellenbogen, Orlando, Bacon, & Joober, 2013; Kubzansky, Mendes, Appleton, Block, & Adler, 2012; Kumsta & Heinrichs, 2013) and group identity (De Dreu, 2012; De Dreu et al., 2010; Stallen, De Dreu, Shalvi, Smidts, & Sanfey, 2012). In general, these studies show that OT can have a facilitating effect on *prosocial* behaviors, but that in certain conditions it triggers competitive behavior (Declerck, Boone, & Kiyonari, 2010), ethnocentricity (De Dreu, Greer, Handgraaf, Shalvi, & Van Kleef, 2012), lying (Shalvi & De Dreu, 2014), violence (De Wall et al., 2014), and envy and schadenfreude (Shamay-Tsoory et al., 2009). Recently, theories have emerged that try to bind the different, sometimes seemingly contradictory, effects into a larger framework.

The current leading overarching theories suggest that OT administration facilitates the observed behavioral changes heuristically by its effect on underlying dopaminergic systems. OT targets brain regions that are evolutionary old and well conserved across species, such as the ventral tegmental area in the midbrain, the amygdala, and nucleus accumbens (Bartz, Zaki, Bolger, & Ochsner, 2011; Kemp & Guastella, 2011; Love, 2014; Shamay-Tsoory & Abu-Akel, 2015). By interacting with dopamine in these regions, OT has been proposed to enhance the encoding of social cues and improve social cognition, lower social anxiety, and link intrinsic motivation with social rewards (Love, 2014). While the number of studies that show these basal effects of OT is growing, the precise role of OT in integrating social cues and motivation in the decision making process is still largely unexplored.

3. Dissertation outline

The four chapters of this dissertation address social judgements and decision making in increasingly complex social settings. We start out in the first two chapters by investigating the role of OT on the *evaluation of social cues*, followed by a third chapter in which we study the role of OT when social information is processed to *arrive at a cooperative or competitive decision*. In the final chapter, we go beyond heuristic processing and investigate cooperative behavior in repeated social interactions (where individuals have the chance to meet and interact again). This setting allows us to better

understand *how trust and cooperation can be learned over time,* and how this interacts with the individual's social value orientation.

Chapter one, *"oxytocin does not make a face appear more trustworthy but improves the accuracy of trustworthiness judgments*", describes an experiment in which we test whether OT increases perceived trustworthiness of faces and/or whether it improves the discriminatory ability of trustworthiness perception (see figure 3). This chapter fits within a large body of literature that investigates the role of OT in establishing trust based on perceiving social cues such as faces (Theodoridou et al., 2009; van 't Wout & Sanfey, 2008).

In the second chapter, *"sexual dimorphism in oxytocin responses to health perception and disgust, with implications for theories on pathogen detection"*, we report on additional data that was collected during the experiment described in chapter one. Participants indicated (i) how (un)healthy they judged faces, and (ii) how disgusting they found certain situations (see figure 3). This allowed us to further assess whether OT increases cautious behavior as previously described in the literature (Lischke et al., 2012; Striepens et al., 2012). Improved health perception can facilitate the detection of infections and lead to avoiding contaminated others or unhealthy situations. Because OT is known to interact with estrogen in females, we also investigate if there are sex differences in the effect of OT administration (Bos, Panksepp, Bluthe, & van Honk, 2012; Choleris, Pfaff, & Kavaliers, 2013).

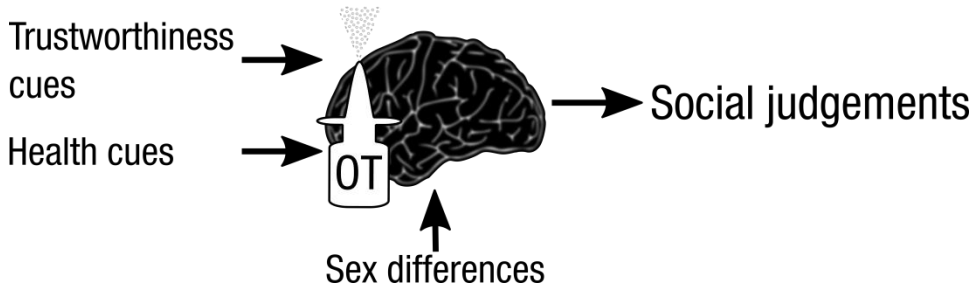

Fig. 3. Outline of the content discussed in the chapters one and two of this dissertation.

In the third chapter, *"a functional MRI study on how oxytocin affects decision making in social dilemmas: cooperate as long as it pays off, aggress only when you think you can win"*, we report the results of an experiment in which we combine functional magnetic resonance imaging (fMRI) with OT administration (see figure 4). We investigate if OT influences heuristic decision making in economic games that incentivize either cooperation or competition. We additionally include contextual social cues (expressions of happy or angry faces) that are informative with respect to assessing the trustworthiness/aggressiveness of interaction partners. This is the first study to combine different sources of social information in an fMRI experiment to investigate the context-dependent neural and behavioral effects of OT (Love, 2014; Shamay-Tsoory & Abu-Akel, 2015).

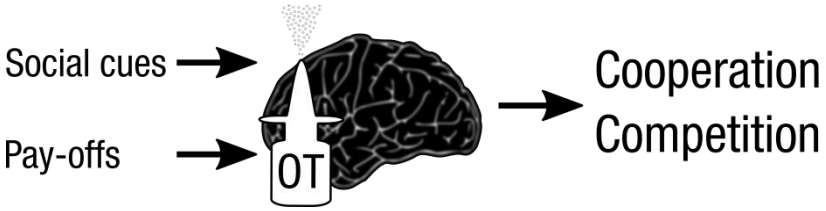

Fig. 4. Outline of the content discussed in chapter three of this dissertation.

In the fourth and last chapter (see figure 5), "*trust as commodity: social value orientation affects the neural substrates of learning to cooperate*", we rely on fMRI to investigate how trust becomes established during *repeated* interactions with anonymous members of a defined and transient group. In this experimental setting individuals can learn from previous encounters, something that was not possible in the one-shot interactions described in chapter three. We are particularly interested in how *individual differences in social value orientation* shape how trust is formed over time. Social value orientation describes how much individuals value fairness and consider the preferences of others in decision making (prosocials), versus how much they are interested in personal gains (proselfs).

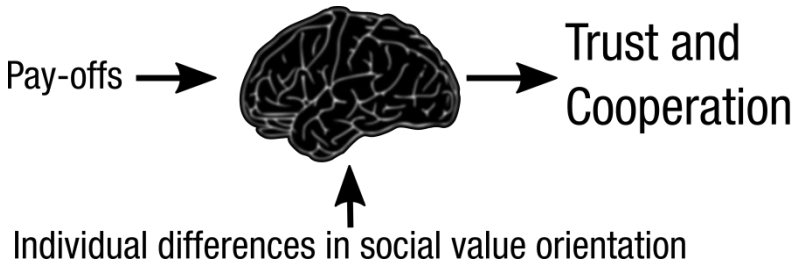

Fig. 5. Outline of the content discussed in chapter four of this dissertation.

Finally, in an epilogue we discuss how these four chapters contribute to both the study of social decision neuroscience and neuroeconomics. We have added to these fields by investigating how OT affects the judgment of social cues. We furthermore describe the effect of individual differences in social value orientation on the formation of trust. Finally, we correlate behavioral decisions in a social context with neuroimaging data and investigate how individual differences shape the learning process.

4. Research tools

Part of the innovative aspect of the dissertation lies in the fact that the research is multidisciplinary, relying on tools from disparate fields. We use functional MRI (fMRI) to correlate brain activity with behavior, administrate oxytocin to manipulate behavior, and make use of economic games derived from game theory to create decision contexts.

First, fMRI allows to monitor brain activation in real time in a non-invasive way. Individuals take place in a scanner, and, using magnetic coils, a signal is generated that is linked to the metabolic activity of neurons (neuro-vascular coupling): the BOLD signal. Using computational methods, it is possible to identify patterns in the collected BOLD signal data, and to correlate them to events of interest in an experimental design, e.g. to the exact moment when a person is making a decision or the interval that a significant cue is shown. This allows the identification of brain regions that are involved in different types of decisions and can thus be used to unravel different motivations for a decision.

Second, by administrating pharmaceuticals, one can directly manipulate behavior. Neurohormones such as oxytocin (OT) bind to specific receptor sites on neurons and have a direct effect on brain activation. In experimental studies, OT is administered in low doses through a nasal spray by the participants themselves. OT most likely reaches the brain by way of the olfactory sensory neurons and the olfactory bulb (Quintana, Alvares, Hickie, & Guastella, 2015).

Third, experimental paradigms derived from game theory have proven particularly useful to investigate different aspects of social decision making, such as trust, reciprocity, cooperation, fairness, and competition (Rilling & Sanfey, 2011). In these economic game, participants are shown a decision matrix and select their decisions on the basis of the potential outcome, which is depicted as numerical values. By changing these pay-offs, the incentives to cooperate or compete are manipulated.

For example, a prototypical economic game is the prisoner's dilemma (PD) which is used in chapter 4 (see figure 6). It is a dyadic game in which both participants decide simultaneously without knowing from each other which option they choose. It is also a mixed motive game in which the fear of betrayal and self-interested greed tend to pull people toward non-cooperative behavior. Each participant has two options: to cooperate (option C) or to defect (option D). The outcome is dependent on the actions of both participants and is written between brackets. If participant A cooperates, participant B can receive either 7 (when he also cooperates) or 9 monetary units (when he defects). When participant A defects, participant B can earn 3 (when he cooperates) or 5 units (when he also defects). For both participants, defect is the dominant strategy, either out of greed (to earn more) or out of fear (to minimize loss). If both participants follow this strategy, a Nash equilibrium is reached: defect-defect is an outcome in which neither of the two players can gain more by unilaterally changing their choice. But this Nash equilibrium is a suboptimal solution for both of the participants (they each earn 5 rather than 7 had they both cooperated), even though they both decided in their best self-interest.

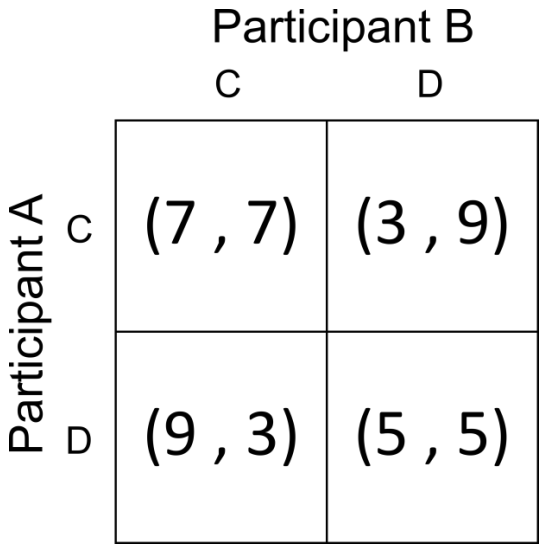

Fig. 6. A prisoner's dilemma, an example of an economic game

The outcome of these games is highly sensitive to a range of factors such as the pay-off matrix (if the pay-off for defect increases, people will cooperate less and vice versa), individual preferences (e.g. social value orientation), the sequence of gameplay (simultaneously or sequentially) and whether the game is repeated or not. In chapter three, we use economic games in which the pays-offs are changed so that participants are motivated to both cooperate (a coordination game) or are motivated to select the opposite behavior of their partner (an anti-coordination game). In chapter four, we investigate the effect of social value orientation on decision making in economic games where the player goes first, knowing that their choices are disclosed to their partners. Furthermore, the game is repeatedly played with returning partners.

The combination of these three research tools (exogenous OT administration, fMRI and economic games) is an innovative way to explore the roots of cooperative behavior and adds to both the novelty and sophistication of this dissertation.

References

Bartz, J. A., Zaki, J., Bolger, N., & Ochsner, K. N. (2011). Social effects of oxytocin in humans: Context and person matter. *Trends in Cognitive Sciences, 15*(7), 301-309. doi: 10.1016/j.tics.2011.05.002

Baumeister, R. F., & Leary, M. R. (1995). The need to belong - Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*(3), 497-529. doi: 10.1037/0033-2909.117.3.497

Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior, 52*(2), 336-372. doi: 10.1016/j.geb.2004.06.010

Bogaert, S., Boone, C., & Declerck, C. (2008). Social value orientation and cooperation in social dilemmas: A review and conceptual model. *British Journal of Social Psychology, 47*, 453-480. doi: 10.1348/014466607x244970

Boone, C., Declerck, C. H., & Kiyonari, T. (2010). Inducing cooperative behavior among proselfs versus prosocials: The moderating role of incentives and trust. *Journal of Conflict Resolution, 54*(5), 799-824. doi: 10.1177/0022002710372329

Bos, P. A., Panksepp, J., Bluthe, R. M., & van Honk, J. (2012). Acute effects of steroid hormones and neuropeptides on human social-emotional behavior: A review of single administration studies. *Frontiers in Neuroendocrinology, 33*(1), 17-35. doi: 10.1016/j.yfrne.2011.01.002

Cardoso, C., Ellenbogen, M. A., Orlando, M. A., Bacon, S. L., & Joober, R. (2013). Intranasal oxytocin attenuates the cortisol response to physical stress: A dose-response study. *Psychoneuroendocrinology, 38*(3), 399-407. doi: 10.1016/j.psyneuen.2012.07.013

Carter, G. G., & Wilkinson, G. S. (2015). Intranasal oxytocin increases social grooming and food sharing in the common vampire bat Desmodus rotundus. *Hormones and Behavior, 75*, 150-153. doi: 10.1016/j.yhbeh.2015.10.006

Choleris, E., Pfaff, D. W., & Kavaliers, M. (2013). *Oxytocin, vasopressin and related peptides in the regulation of behaviour*. Cambridge: Cambridge University Press.

De Dreu, C. K. W. (2012). Oxytocin modulates cooperation within and competition between groups: An integrative review and research agenda. *Hormones and Behavior, 61*(3), 419-428. doi: 10.1016/j.yhbeh.2011.12.009

De Dreu, C. K. W., Greer, L. L., Handgraaf, M. J. J., Shalvi, S., & Van Kleef, G. A. (2012). Oxytocin modulates selection of allies in intergroup conflict. *Proceedings of the Royal Society B-Biological Sciences, 279*(1731), 1150-1154. doi: 10.1098/rspb.2011.1444

De Dreu, C. K. W., Greer, L. L., Handgraaf, M. J. J., Shalvi, S., Van Kleef, G. A., Baas, M., . . . Feith, S. W. W. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science, 328*(5984), 1408-1411. doi: 10.1126/science.1189047

De Wall, C. N., Gillath, O., Pressman, S. D., Black, L. L., Bartz, J. A., Moskovitz, J., & Stetler, D. A. (2014). When the love hormone leads to violence: Oxytocin increases intimate partner violence inclinations among high trait aggressive people. *Social Psychological and Personality Science, 5*(6), 691-697. doi: 10.1177/1948550613516876

Declerck, C. H., Boone, C., & Kiyonari, T. (2010). Oxytocin and cooperation under conditions of uncertainty: The modulating role of incentives and social information. *Hormones and Behavior, 57*(3), 368-374. doi: 10.1016/j.yhbeh.2010.01.006

Domes, G., Steiner, A., Porges, S. W., & Heinrichs, M. (2013). Oxytocin differentially modulates eye gaze to naturalistic social signals of happiness and anger. *Psychoneuroendocrinology, 38*(7), 1198-1202. doi: 10.1016/j.psyneuen.2012.10.002

Ferguson, J. N., Young, L. J., & Insel, T. R. (2002). The neuroendocrine basis of social recognition. *Frontiers in Neuroendocrinology, 23*(2), 200-224. doi: 10.1006/frne.2002.0229

Gigerenzer, P., & Todd, P. M. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Goodson, J. L., Schrock, S. E., Klatt, J. D., Kabelik, D., & Kingsbury, M. A. (2009). Mesotocin and nonapeptide receptors promote estrildid flocking behavior. *Science, 325*(5942), 862-866. doi: 10.1126/science.1174929

Harari-Dahan, O., & Bernstein, A. (2014). A general approach-avoidance hypothesis of oxytocin: Accounting for social and non-social effects of oxytocin. *Neuroscience and Biobehavioral Reviews, 47*, 506-519. doi: 10.1016/j.neubiorev.2014.10.007

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., . . . Tracer, D. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences, 28*(6), 795-+.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review, 93*(5), 1449-1475. doi: 10.1257/000282803322655392

Kemp, A. H., & Guastella, A. J. (2011). The role of oxytocin in human affect: A novel hypothesis. *Current Directions in Psychological Science, 20*(4), 222-231. doi: 10.1177/0963721411417547

Keverne, E. B., & Kendrick, K. M. (1992). *Oxytocin facilitation of maternal behavior in sheep* (Vol. 652).

Kirsch, P., Esslinger, C., Chen, Q., Mier, D., Lis, S., Siddhanti, S., . . . Meyer-Lindenberg, A. (2005). Oxytocin modulates neural circuitry for social cognition and fear in humans. *The Journal of Neuroscience, 25*(49), 11489-11493. doi: 10.1523/jneurosci.3984-05.2005

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature, 435*(7042), 673-676. doi: 10.1038/nature03701

Kubzansky, L. D., Mendes, W. B., Appleton, A. A., Block, J., & Adler, G. K. (2012). A heartfelt response: Oxytocin effects on response to social stress in men and women. *Biological Psychology, 90*(1), 1-9. doi: 10.1016/j.biopsycho.2012.02.010

Kumsta, R., & Heinrichs, M. (2013). Oxytocin, stress and social behavior: Neurogenetics of the human oxytocin system. *Current Opinion in Neurobiology, 23*(1), 11-16. doi: 10.1016/j.conb.2012.09.004

Labuschagne, I., Phan, K. L., Wood, A., Angstadt, M., Chua, P., Heinrichs, M., . . . Nathan, P. J. (2010). Oxytocin attenuates amygdala reactivity to fear in generalized social anxiety disorder. *Neuropsychopharmacology, 35*(12), 2403-2413. doi: 10.1038/npp.2010.123

Lambert, B., Declerck, C. H., & Boone, C. (2014). Oxytocin does not make a face appear more trustworthy but improves the accuracy of trustworthiness judgments. *Psychoneuroendocrinology, 40*(0), 60-68. doi: 10.1016/j.psyneuen.2013.10.015

Leknes, S., Wessberg, J., Ellingsen, D.-M., Chelnokova, O., Olausson, H., & Laeng, B. (2012). Oxytocin enhances pupil dilation and sensitivity to “hidden” emotional expressions. *Social Cognitive and Affective Neuroscience*. doi: 10.1093/scan/nss062

Lischke, A., Gamer, M., Berger, C., Grossmann, A., Hauenstein, K., Heinrichs, M., . . . Domes, G. (2012). Oxytocin increases amygdala reactivity to threatening scenes in females. *Psychoneuroendocrinology, 37*(9), 1431-1438. doi: 10.1016/j.psyneuen.2012.01.011

Liu, J. C. J., Guastella, A. J., & Dadds, M. R. (2012). Effects of oxytocin on human social approach measured using intimacy equilibriums. *Hormones and Behavior, 62*(5), 585-591. doi: 10.1016/j.yhbeh.2012.09.002

Love, T. M. (2014). Oxytocin, motivation and the role of dopamine. *Pharmacology, Biochemistry, and Behavior, 119*, 49-60. doi: 10.1016/j.pbb.2013.06.011

Maestripieri, D., Hoffman, C. L., Anderson, G. M., Carter, C. S., & Higley, J. D. (2009). Mother–infant interactions in free-ranging rhesus macaques: Relationships between physiological and behavioral variables. *Physiology & Behavior, 96*(4–5), 613-619. doi: 10.1016/j.physbeh.2008.12.016

Marsh, A. A., Yu, H. H., Pine, D. S., & Blair, R. J. R. (2010). Oxytocin improves specific recognition of positive facial expressions. *Psychopharmacology, 209*(3), 225-232. doi: 10.1007/s00213-010-1780-4

McCall, C., & Singer, T. (2012). The animal and human neuroendocrinology of social cognition, motivation and behavior. *Nature Neuroscience, 15*(5), 681-688.

Oxytocin. (2016, juli 14). Wikipedia, the free encyclopedia.

Petrovic, P., Kalisch, R., Singer, T., & Dolan, R. J. (2008). Oxytocin attenuates affective evaluations of conditioned faces and amygdala activity. *The Journal of Neuroscience, 28*(26), 6607-6615. doi: 10.1523/jneurosci.4572-07.2008

Preckel, K., Scheele, D., Kendrick, K. M., Maier, W., & Hurlemann, R. (2014). Oxytocin facilitates social approach behavior in women. *Frontiers in Behavioral Neuroscience, 8*(191). doi: 10.3389/fnbeh.2014.00191

Quintana, D. S., Alvares, G. A., Hickie, I. B., & Guastella, A. J. (2015). Do delivery routes of intranasally administered oxytocin account for observed effects on social cognition and behavior? A two-level model. *Neuroscience and Biobehavioral Reviews, 49*, 182-192. doi: 10.1016/j.neubiorev.2014.12.011

Radke, S., Roelofs, K., & de Bruijn, E. R. A. (2013). Acting on anger: Social anxiety modulates approach-avoidance tendencies after oxytocin administration. *Psychological Science, 24*(8), 1573-1578. doi: 10.1177/0956797612472682

Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. In S. T. Fiske, D. L. Schacter & S. E. Taylor (Eds.), *Annual Review of Psychology* (Vol. 62, pp. 23-48). Palo Alto: Annual Reviews.

Ross, H. E., Freeman, S. M., Spiegel, L. L., Ren, X., Terwilliger, E. F., & Young, L. J. (2009). Variation in oxytocin receptor density in the nucleus accumbens has differential effects on affiliative behaviors in monogamous and polygamous voles. *Journal of Neuroscience, 29*(5), 1312-1318. doi: 10.1523/jneurosci.5039-08.2009

Scheele, D., Striepens, N., Gunturkun, O., Deutschlander, S., Maier, W., Kendrick, K. M., & Hurlemann, R. (2012). Oxytocin modulates social distance between males and females. *Journal of Neuroscience, 32*(46), 16074-16079. doi: 10.1523/jneurosci.2755-12.2012

Shalvi, S., & De Dreu, C. K. W. (2014). Oxytocin promotes group-serving dishonesty. *Proceedings of the National Academy of Sciences of the United States of America, 111*(15), 5503-5507. doi: 10.1073/pnas.1400724111

Shamay-Tsoory, S. G., & Abu-Akel, A. (2015). The social salience hypothesis of oxytocin. *Biological Psychiatry, 79*(3), 194-202. doi: 10.1016/j.biopsych.2015.07.020

Shamay-Tsoory, S. G., Fischer, M., Dvash, J., Harari, H., Perach-Bloom, N., & Levkovitz, Y. (2009). Intranasal administration of oxytocin increases envy and schadenfreude (gloating). *Biological Psychiatry, 66*(9), 864-870. doi: 10.1016/j.biopsych.2009.06.009

Stallen, M., De Dreu, C. K. W., Shalvi, S., Smidts, A., & Sanfey, A. G. (2012). The herding hormone: Oxytocin stimulates in-group conformity. *Psychological Science, 23*(11), 1288-1292. doi: 10.1177/0956797612446026

Striepens, N., Scheele, D., Kendrick, K. M., Becker, B., Schäfer, L., Schwalba, K., . . . Hurlemann, R. (2012). Oxytocin facilitates protective responses to aversive social stimuli in males. *Proceedings of the National Academy of Sciences, 109*(44), 18144-18149. doi: 10.1073/pnas.1208852109

Theodoridou, A., Penton-Voak, I. S., & Rowe, A. C. (2013). A direct examination of the effect of intranasal administration of oxytocin on approach-avoidance motor responses to emotional stimuli. *PLoS ONE, 8*(2). doi: 10.1371/journal.pone.0058113

Theodoridou, A., Rowe, A. C., Penton-Voak, I., & Rogers, P. (2009). Oxytocin and social perception: Oxytocin increases perceived facial trustworthiness and attractiveness. *Hormones and Behavior, 56*(1), 128-132. doi: citeulike-article-id:6088836

van ’t Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition, 108*(3), 796-803. doi: 10.1016/j.cognition.2008.07.002

Van Lange, P. A. M., Joireman, J., Parks, C. D., & Van Dijk, E. (2013). The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes, 120*(2), 125-141. doi: 10.1016/j.obhdp.2012.11.003

Wittig, R. M., Crockford, C., Deschner, T., Langergraber, K. E., Ziegler, T. E., & Zuberbuhler, K. (2014). Food sharing is linked to urinary oxytocin levels and bonding in related and unrelated wild chimpanzees. *Proceedings of the Royal Society B-Biological Sciences, 281*(1778). doi: 10.1098/rspb.2013.3096

Chapter 1: Oxytocin does not make a face appear more trustworthy but improves the accuracy of trustworthiness judgments

This chapter has been published as:

Lambert, B., Declerck, C. H., & Boone, C. (2014). Oxytocin does not make a face appear more trustworthy but improves the accuracy of trustworthiness judgments. *Psychoneuroendocrinology, 40*(0), 60-68. doi: 10.1016/j.psyneuen.2013.10.015

Abstract

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

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

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

1. Introduction

Oxytocin (OT), the hormone well-known for its involvement in parturition and lactation, has recently seen a surge of interest with respect to its role in regulating social behaviour. Among humans, its prosocial effects have become well-documented. For example, after intranasal administration of OT, people tend to be more trusting (Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005), more generous

(Zak, Stanton, & Ahmadi, 2007), and more cooperative towards in-group members (De Dreu et al., 2010; De Dreu, Greer, Van Kleef, Shalvi, & Handgraaf, 2011). However, the relation between OT and social behaviour does not appear to be straightforward. For example, in social interactions that involve monetary losses and gains, OT-exposed men confronted with a threatening out-group become more competitive so as to protect the in-group (De Dreu, Shalvi, Greer, Van Kleef, & Handgraaf, 2012). Similarly, when no social information is available, OT tends to make people more cautious and uncooperative in monetary games (Declerck, Boone, & Kiyonari, 2010).

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

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

Previous research that has examined the relation between OT and the perception of trustworthiness has yielded inconsistent results. In the study by Theodoridou, Rowe, Penton-Voak, and Rogers (2009) participants who received OT judged faces as more trustworthy compared to those who received a placebo. In contrast, other studies found no significant effect of OT on trustworthiness evaluations (Guastella, Mitchell, & Mathews, 2008; Rimmele, Hediger, Heinrichs, & Klaver, 2009). All these studies, however, used pictures of real faces, making it difficult to isolate the dimension of

trustworthiness. Other factors such as facial symmetry – shown to influence judgements of health and personality (Fink, Neave, Manning, & Grammer, 2006) – are difficult to control for in natural facial expressions and may be accidentally introduced as confounds between comparison groups. Also emotional expressions (which may interact with the perception of trustworthiness) are difficult to exclude in real faces. Possibly, such hard to control differences in natural faces could even explain the discrepant findings between these studies. To avoid some of these problems the current study makes use of artificially generated faces that have been validated in previous research (De Dreu, Greer, Handgraaf, Shalvi, & Van Kleef, 2012; Oosterhof & Todorov, 2008). Therefore, the first objective of the current study is to replicate the above research and test if OT increases, on average, trustworthiness judgments of artificially generated, neutral faces that are devoid of recognizable emotional expressions.

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

2. Methods

2.1 Participants

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

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

At the time of registering, participants filled in an on-line questionnaire consisting of five parts: (i) personal (demographic) information, (ii) a generalised trust and caution questionnaire (Yamagishi & Yamagishi, 1994), (iii) an assessment of Social Value Orientation (Van Lange, Otten, De Bruin, & Joireman, 1997), (iv) a trait anxiety questionnaire (GAD7 by Spitzer, Kroenke, Williams, and Löwe (2006)) and (v) questions regarding how manipulative they are (a shortened version of the Mach-IV questionnaire by Christie and Geis (1970)).

All participants gave written informed consent to the study procedures which were in accordance with the Declaration of Helsinki and were approved by the Ethical Commission of the University of Antwerp. Debriefing occurred by sending participants an e-mail referring them to a website where the methods, results and conclusions were explained. Participants received a show-up fee of €10 which was increased with the earnings from an interactive game which was held at the end of the experiment and served as a manipulation check (further details to follow). Mean total profit of the entire experiment was €18.17.

2.2 Sessions

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

2.3 Procedure

Participants were instructed to self-administer an intranasal dose of 24 IU OT (Syntocinon, Novartis; tree puffs per nostril with one minute in-between puffs) or placebo following a double-blind random design. The placebo contained the same active ingredients except for OT and was prepared by the pharmacy of the University Hospital of Antwerp. The OT group consisted of 57 participants (21 males, 36 females) and the placebo group consisted of 49 participants (24 males, 25 females).

After inhalation, participants waited 35 minutes before starting the actual experimental task. Meanwhile, the participants completed a trial version of the experimental task. The stimuli in this trial version also comprised artificially generated neutral faces, but with different facial identities

from those used in the experimental task. The room was darkened to provide optimal viewing conditions and to reduce visual distraction and glare. The task took around 20 minutes to complete. Afterwards, the participants played an interactive game and filled in a post-experimental questionnaire.

2.4 Experimental Paradigm

Participants were asked to evaluate the trustworthiness of two series of pictures showing artificially created faces displayed on the screen. The software used to present the stimuli and to record the evaluation scores was Affect version 4.0 (Spruyt, Clarysse, Vansteenwegen, Baeyens, & Hermans, 2010). The faces were selected out of a database of 25 male (all bald) facial identities, each varying along 7 dimensions of trustworthiness (yielding a total of 175 faces that varied in trustworthiness). The dataset was created by Oosterhof and Todorov (2008) using Facegen Modeller version 3.1 (Singular Inversions, 2007). As is described by these authors, the face model that is used in Facegen is based on a database of 3D laser-scans of faces. An average face is represented by a collection of vertex positions of a polygonal model of fixed mesh topology in Facegen. Fifty principal components (PC) were constructed to extract the components that account for most of the variance in face shape. A face is expressed as the average face plus a linear combination of these PCs. The coefficients of this linear combination form the face vector α. Oosterhof and Todorov (2008) generated 300 artificial faces and collected judgements of trustworthiness, dominance and threat. Based on mean judgement scores the authors constructed a vector t in the 50-dimensional face space (composed of the weights for each PC) that is optimal for changing trustworthiness score. The features specific for trustworthiness can then be changed by a factor X times the standard deviations (SD) in a face by adding X times the normalized version of the t vector to the face vector α for that face.

We selected ad-hoc five facial identities from those created by Oosterhof and Todorov (2008) that we subsequently used in both the first and the second series of pictures. Each facial identity had seven variations in which the features specific for trustworthiness were changed a fixed amount of standard deviations. This yielded a total of 35 facial stimuli varying along a trustworthiness dimension ranging from -3 SD change (least trustworthy), -2 SD, -1 SD, 0 SD (neutral), +1 SD, +2 SD and +3 SD (most trustworthy). To make sure that the colours of the faces did not interfere with judgment, the L*a*b* values of the faces were changed so that the mean values were equal between pictures. These values were based on the CIE 1976 colour space and respectively determine the lightness, the redness (in contrast to green) and the blueness (in contrast to yellow). The pictures were shown on a black background. To evaluate trustworthiness, participants used a left-mouse click to assign a score on a digital scale shown at the bottom of the computer screen. The scoring was self-

paced. A fixation cross lasting 1.5s was shown between each trial and there was a one minute interval between the two series.

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

The first series of pictures (testing objective 1) consisted of five faces, neutral in trustworthiness (0 SD in the database), each with a different identity and displayed one by one. Participants were asked to indicate on a scale ranging from 0 (not at all trustworthy) to 9 (very trustworthy) how trustworthy they thought the person was. The second series of pictures (testing objective 2) comprised 35 trials (five face identities times seven variations of trustworthiness) in which two faces of the same identity were simultaneously displayed on the computer screen. The left face was always a neutral face (one of the faces used in series 1 with 0 SD in the database), while the right face, depicting the same identity, was manipulated to vary in trustworthiness. The trustworthiness manipulation (the independent variable) comprised the seven variations of each face that were included in the data set created by Oosterhof and Todorov (2008). The participants were asked to give a score ranging from -4 (right face more untrustworthy than left face) to +4 (right face more trustworthy than left face) with 0 indicating no difference between the left and the right face. We note that the scale of the scoring (ranging from -4 to +4) is not the same as the scale of the manipulated trustworthiness dimension (ranging from -3 to +3, reflecting a fixed amount of standard deviation changes). We opted to broaden the scoring scale to allow for more variation in the dependent variable.

2.5 Manipulation Check

When all participants had completed the main task described above, they received instructions for an interactive game which was to proceed with paper and pencil rather than computers. The game was a replication of the experiment by Kosfeld et al. (2005), investigating how OT affects investments in a trust game. We included this replication as an independent manipulation check to determine if OT was effective in this experiment.

The instructions stated that each participant was matched with another participant of the same session, but that neither of them would know who the partner was. Both parties were given an endowment of 5 euro's, which could be altered depending on their decision as well as their partner's

decision throughout this interactive game they would play (See Berg, Dickhaut, and McCabe (1995) for detail on the trust game). Each participant would play the role of trustor (deciding how much of the allotment to invest with the other party) and as trustee (deciding how much of the received money they would return), but with different partners. Both decisions were recorded on a scoring sheet (using the strategy method for the trustee decision). However, payments were computed based only on one of the two decisions (randomly determined).

To assess the influence of OT, we conducted a logistic regression on a dummy indicating if the participant invested the entire endowment (coded 1, investing less than the maximum possible was coded 0). A positive and significant effect would indicate that OT enhances the tendency to trust in line with the findings of Kosfeld et al. (2005). The independent variables were OT (coded 1, placebo = 0), Sex (female = 0, male = 1) and three personality variables obtained from the questionnaires which were filled in at the time of registering for the experiment (generalized trust, generalized caution and Machiavellianism).

3. Results

The results of the manipulation check indicate a significant effect of OT on trusting behaviour but only for cautious individuals (interaction of OT*generalized caution in a logistic regression: B = 1.11, Std. Error = 0.47, z = 2.38, p = 0.017). There was neither a main effect of OT, nor any interaction with other personality variables that proved to be significant. The finding that OT influences trust mainly among cautious people is completely in line with other recent studies that investigated how the behavioural effects of OT in social interactions might depend on individual characteristics. It appears that OT increases the level of cooperation especially among those who are by nature less inclined to cooperate because they are either more anxious (De Dreu, 2012), or they hold individualistic and competitive social values (Declerck, Boone, & Kiyonari, 2013).

To validate the trustworthiness scale created by Oosterhof and Todorov (2008) for the current study, we first averaged all of the participants' trustworthiness scores assigned to each of the neutral faces presented in the first series of pictures. Figure 1 shows that four out of the five faces were rated as having an average trustworthiness score (score five on a scale from 0 to 9). Face 3 appears to deviate significantly from the four other scores: we verified this by means of a repeated measures ANOVA model (main effect of faces: $F_{(3.77, 396.28)} = 41.58$, $p < 0.001$, partial $\eta^2 = 0.28$; Note that the degrees of freedom in this analysis have been adjusted to correct for violation of the sphericity assumption) and pairwise comparisons among estimated marginal means of the within factor, i.e. face identities. Therefore, we excluded this stimulus from further analysis.

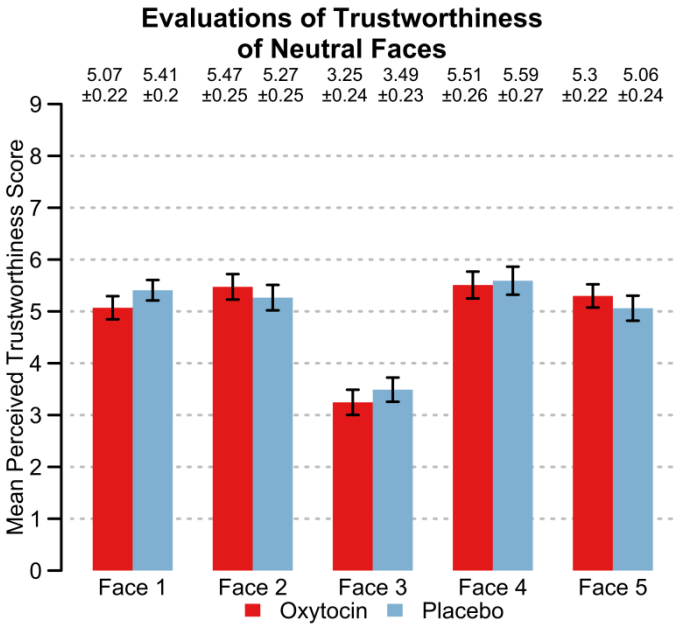

Fig. 1. Evaluations of trustworthiness of neutral faces. All 106 participants evaluated the trustworthiness of five neutral faces of different identity (Face 1 to 5; obtained from the dataset created by Oosterhof and Todorov (2008)). Error bars represent the standard error of the mean.

Second, we plotted the averaged perceived trustworthiness scores for each of the stimuli of the second series (comparison between a neutral and a manipulated face) relative to the trustworthiness scale dimension (see fig. 2). Visual inspection shows that, in accordance with the results of Oosterhof and Todorov (2008), the relation is linear. We furthermore notice no apparent effect of OT on perceived trustworthiness above the effect of the scale dimension. To corroborate these visual observations statistically, we pooled the data of the four face identities for each participant and conducted a regression analysis on the perceived scores (ranging from -4 to 4). The independent variables in this regression are the trustworthiness dimension (ranging from -3 SD to 3 SD), the treatment (1 = OT, 0 = placebo) and sex of the participant (0 = female, 1 = male). The latter is included because recent publications indicate that sex can be a moderator of the effect of OT (Domes et al., 2010; Lischke, Gamer, et al., 2012). The trustworthiness dimension appears to be the sole predictor for perceived trustworthiness ($B = 0.534$, Std. Error = 0.015, $z = 35.18$, $p < 0.001$, $R^2 = 0.611$), as we could have inferred from figure 2. Neither treatment ($B = -0.077$, Std. Error = 0.076, $z = -1.02$, $p = 0.307$) nor sex ($B = 0.11$, Std. Error = 0.076, $z = 1.47$, $p = 0.142$) proved to have a significant effect.

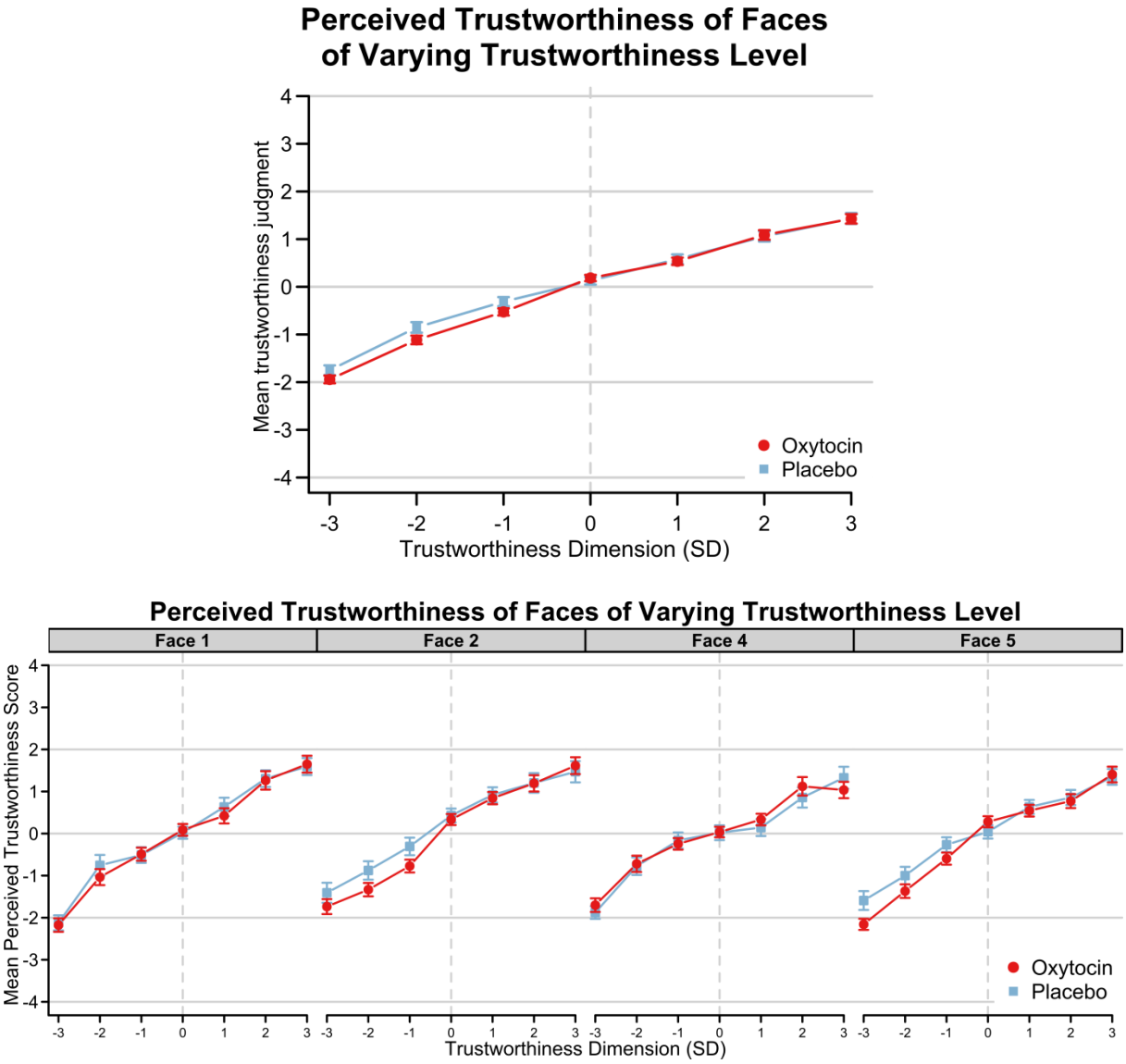

Fig. 2. Perceived trustworthiness of faces of varying trustworthiness level. All 106 participants evaluated 28 faces (face identity 3 is dropped; see text) in comparison to a neutral face of the same face identity and gave them a score between -4 and +4 on perceived trustworthiness. (a) Mean scores with standard error are depicted for each value of the trustworthiness dimension (ranging from -3 SD to 3 SD, see Oosterhof and Todorov (2008)) when all faces of each identity are pooled. (b) Mean scores are depicted for each face identity separately.

To test the main-level effect of treatment (OT versus placebo) on the perception of neutral faces (objective 1), we conducted a 4 (face identities) * 2 (treatment) * 2 (sex) repeated ANOVA analysis on perceived trustworthiness with face identity as a within-subject factor. No significant effect of face identity ($F_{(2.95, 300.95)} = 1.55$, $p = 0.201$, partial $\eta^2 = 0.015$), treatment ($F_{(1, 102)} = 0.044$, $p = 0.835$, partial $\eta^2 < 0.001$) or sex ($F_{(1, 102)} = 1.14$, $p = 0.288$, partial $\eta^2 = 0.011$) was found, neither did any of the interactions between these variables prove to be significant (face*treatment: $F_{(2.95, 300.95)} = 0.92$, $p =$

0.433, partial $\eta^2 = 0.009$; face*sex: $F_{(2.95, 300.95)} = 0.37$, $p = 0.773$, partial $\eta^2 = 0.004$; treatment*sex: $F_{(1, 102)} = 0.15$, $p = 0.704$, partial $\eta^2 = 0.001$). Furthermore, using a non-parametric test as a robustness check, the overall mean of the trustworthiness ratings of the four neutral faces presented in series 1 (excluding Face 3) taken together, does not differ between treatments (Mann-Whitney, U = 34831.5, z = -0.047, p = 0.963; OT mean = 4.92, SD = 1.99; placebo mean = 4.96, SD = 1.82).

To test if OT affects the accuracy of trustworthiness perception (objective 2), we compared the number of misclassifications made between the OT and the placebo group. A misclassification is defined as giving a positive score (deeming it more trustworthy than the neutral faces) to a face that is registered as untrustworthy (with a -3 SD, -2 SD or -1 SD change from the neutral face), or vice versa. Table 1 shows that the number of misclassifications versus correctly evaluated faces differed by treatment, but only so in the case of untrustworthy faces: relatively less mistakes were made in the OT-group (Fisher's exact test; p = 0.0062). OT had no significant effect on misclassifications when trustworthy faces were evaluated (Fisher's exact test; p = 0.65), or when trustworthy and untrustworthy faces were pooled (Fisher's exact test; p = 0.13). When untrustworthy faces were evaluated, the OT group made on average 0.93 (SD = 1.38) mistakes and the placebo group made 1.49 mistakes (SD = 1.89), a significant difference (Wilcoxon rank sum test: p = 0.043, one-tailed). When the participants evaluated trustworthy faces, the mean number of mistakes was similar across the two groups (OT group, mean = 1.33, SD = 1.74; placebo group, mean = 1.22, SD = 1.67; Wilcoxon rank sum test: p = 0.63; one-tailed).

Table 1. Fisher's exact tests on number of misclassifications

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

Results of the Fisher's exact tests show that the number of misclassifications versus correctly evaluated faces differed by treatment only in the case of untrustworthy faces: relatively less mistakes were made in the OT-group than in the placebo-group.

As a robustness check, we conducted regression analyses on the number of misclassifications of each participant (see table 2). We fitted six different negative binomial models: First, we conducted a regression analysis on the number of untrustworthy faces misclassified as trustworthy and again we took sex (female = 0, male = 1) into account as a possible moderator. Model 1 shows a significant effect of OT (B = -0.46, Std. Error = 0.26, z = -1.75, p = 0.04). Model 2 indicates that OT does not interact significantly with sex. Second, we did the same for the misclassifications of trustworthy faces (model 3 and 4) but this did not yield any significant results. Third, we pooled the trustworthy and untrustworthy faces and structured the data in panel form. We included a dummy regressor indicating if the face is untrustworthy (coded 0) or trustworthy (coded 1). Model 5 shows that the main effect of treatment was not significant. Finally, in model 6 we investigate the interaction effect of OT and trustworthiness. This interaction is significant (B = 0.58, Std. Error = 0.30, z = -1.91, p = 0.03) which corroborates that the effect of OT depends on whether the face is trustworthy or untrustworthy. Because the amount of misclassifications for untrustworthy and trustworthy faces were both overdispersed ($D_{untrustworthy}$ = 2.29; $D_{trustworthy}$ = 2.25), we fitted a negative binomial in each model using the statistical package Stata 9 (StataCorp, 2011).

The methodological decision to use only facial stimuli depicting males is a limitation of the study, as we cannot be sure that women rate men (opposite sex) the same as men rate men (same sex). However, as we reported earlier in this section when we tested the main level-effect of OT, the interaction effect of treatment*sex on trustworthiness scores proves to be not significant (repeated ANOVA; $F_{(1, 102)}$ = 0.15, p = 0.704, partial η^2 = 0.001). Hence we suspect that there is no differential influence of OT on same-sex or opposite-sex evaluation. But it is of course still possible that women rate other women differently than men rate men.

Table 2. Negative binomial regression models on the number of misclassifications.

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Oxytocin	-0.46 (0.26)	-0.51 (0.33)	0.07 (0.25)	0.02 (0.35)	-0.18 (0.22)	0.10 (0.27)
	p = 0.04	p = 0.06	p = 0.39	p = 0.48	p = 0.21	p = 0.36
Trustworthiness	-	-	-	-	0.09 (0.15)	-0.20 (0.21)
					p = 0.29	p = 0.18
Sex	0.02 (0.26)	-0.03 (0.35)	-0.03 (0.25)	-0.09 (0.35)	0.01 (0.22)	0.03 (0.23)
	p = 0.47	p = 0.46	p = 0.45	p = 0.40	p = 0.48	p = 0.45
Oxytocin*Sex	-	0.13 (0.54)	-	0.11 (0.49)	-	-
		p = 0.40		p = 0.41		
Oxytocin*	-	-	-	-	-	0.58 (0.30)
Trustworthiness						p = 0.03
Constant	0.38 (0.22)	0.41 (0.22)	0.23 (0.24)	0.25 (0.28)	1.20 (0.62)	1.21 (0.67)
	p = 0.04	p = 0.03	p = 0.16	p = 0.18	p = 0.03	p = 0.04
N	106	106	106	106	212	212

Unstandardized regression coefficients of the negative binomial regression analyses on the number of misclassifications. Dependent variable of Model 1 and 2: number of misclassifications of untrustworthy faces; dependent variable of Model 3 and 4: number of misclassifications of trustworthy faces. In Model 5 and 6, the number of misclassifications are pooled with trustworthiness added as a dummy variable (untrustworthy = 0; trustworthy = 1). Standard errors are given in parentheses. All p-values are one-tailed. Oxytocin (OT; coded 1; placebo = 0) and trustworthiness of the faces are the predictor variables of interest. Sex (female = 0; male = 1) is added to the models as control variable and is tested as a possible moderator on the effect of oxytocin. Participants (106) evaluated three trustworthy and three untrustworthy variations of four different face identities. Model 1 shows a significant effect of OT on the number of misclassifications of untrustworthy faces. Model 6 shows a significant interaction effect of OT and trustworthiness.

4. Discussion

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

will discuss each one in the light of other recent findings that are creating a complex picture of OT's role in modulating social behaviour.

4.1 No level effect of OT on trustworthiness judgments

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

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

Finally, we note that the effect of OT on perception may be more subtle and difficult to detect when perception is decoupled from a behavioural goal. For example, De Dreu, Greer, et al. (2012) showed that, while OT does not change the perception of threat (a combination of trustworthiness and dominance) displayed by faces, it may still influence behaviour. Males who received OT chose for faces expressing high threat to be their allies in a competitive setting while the placebo group chose faces with low threat. In other studies where perception does not hinge on an experimental tasks,

null effects of OT have been reported several times. Guastella et al. (2008) asked participants to indicate their perceived trustworthiness of neutral, happy and angry faces. Rimmele et al. (2009) investigated how willingly participants approached faces with negative, neutral or positive emotional expression. In neither study was an effect of OT observed.

4.2 OT improves the detection of untrustworthy faces

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

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

A similar result was obtained by Di Simplicio, Massey-Chase, Cowen, and Harmer (2009). When they asked participants to label the emotion expressed by faces as fast and accurate as possible, they found that participants who received OT misclassified the emotion of "surprise" less often than those

in the placebo group. Surprise is an emotion that is expressed when something in the environment does not fit the expectations. Detection of surprise in other people can possibly make people more aware of relevant changes or potential danger. Therefore, the improved detection of surprise in others may have adaptive value. Recent findings by (De Dreu, Greer, et al., 2012; Kret & De Dreu, 2013) similarly fit the theory of Error Management: both articles describe that OT leads people to select faces low in trust (and high in threat) more than faces high in trust (and low in threat) in the context of intergroup competition. Selecting high threat allies can be seen an evolutionary selected bias that improves the chances of ingroup success when hostile clans that do not know each other are competing for resources.

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

4.3 Conclusion

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

References

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience, 4*(3), 165-178. doi: 10.1038/nrn1056

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron, 58*(4), 639-650. doi: 10.1016/j.neuron.2008.04.009

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social-history. *Games and Economic Behavior, 10*(1), 122-142. doi: 10.1006/game.1995.1027

Christie, R., & Geis, F. (1970). *Studies in Machiavellianism*. New York: Academic Press.

Churchland, P. S., & Winkielman, P. (2012). Modulating social behavior with oxytocin: How does it work? What does it mean? *Hormones and Behavior, 61*(3), 392-399. doi: 10.1016/j.yhbeh.2011.12.003

De Dreu, C. K. W. (2012). Oxytocin modulates the link between adult attachment and cooperation through reduced betrayal aversion. *Psychoneuroendocrinology, 37*(7), 871-880. doi: 10.1016/j.psyneuen.2011.10.003

De Dreu, C. K. W., Greer, L. L., Handgraaf, M. J. J., Shalvi, S., & Van Kleef, G. A. (2012). Oxytocin modulates selection of allies in intergroup conflict. *Proceedings of the Royal Society B-Biological Sciences, 279*(1731), 1150-1154. doi: 10.1098/rspb.2011.1444

De Dreu, C. K. W., Greer, L. L., Handgraaf, M. J. J., Shalvi, S., Van Kleef, G. A., Baas, M., . . . Feith, S. W. W. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science, 328*(5984), 1408-1411. doi: 10.1126/science.1189047

De Dreu, C. K. W., Greer, L. L., Van Kleef, G. A., Shalvi, S., & Handgraaf, M. J. J. (2011). Oxytocin promotes human ethnocentrism. *Proceedings of the National Academy of Sciences of the United States of America, 108*(4), 1262-1266. doi: 10.1073/pnas.1015316108

De Dreu, C. K. W., Shalvi, S., Greer, L. L., Van Kleef, G. A., & Handgraaf, M. J. J. (2012). Oxytocin motivates non-cooperation in intergroup conflict to protect vulnerable in-group members. *PLoS ONE, 7*(11), e46751. doi: 10.1371/journal.pone.0046751

Declerck, C. H., Boone, C., & Kiyonari, T. (2010). Oxytocin and cooperation under conditions of uncertainty: The modulating role of incentives and social information. *Hormones and Behavior, 57*(3), 368-374. doi: 10.1016/j.yhbeh.2010.01.006

Declerck, C. H., Boone, C., & Kiyonari, T. (2013). The effect of oxytocin on cooperation in a prisoner's dilemma depends on the social context and a person's social value orientation. *Social Cognitive and Affective Neuroscience*, DOI: 10.1093/scan/nst1040.

Di Simplicio, M., Massey-Chase, R., Cowen, P. J., & Harmer, C. J. (2009). Oxytocin enhances processing of positive versus negative emotional information in healthy male volunteers. *Journal of Psychopharmacology, 23*(3), 241-248. doi: 10.1177/0269881108095705

Domes, G., Heinrichs, M., Michel, A., Berger, C., & Herpertz, S. C. (2007). Oxytocin improves "mind-reading" in humans. *Biological Psychiatry, 61*(6), 731-733. doi: 10.1016/j.biopsych.2006.07.015

Domes, G., Lischke, A., Berger, C., Grossmann, A., Hauenstein, K., Heinrichs, M., & Herpertz, S. C. (2010). Effects of intranasal oxytocin on emotional face processing in women. *Psychoneuroendocrinology, 35*(1), 83-93. doi: 10.1016/j.psyneuen.2009.06.016

Fink, B., Neave, N., Manning, J. T., & Grammer, K. (2006). Facial symmetry and judgements of attractiveness, health and personality. *Personality and Individual Differences, 41*(3), 491-499. doi: 10.1016/j.paid.2006.01.017

Fischer-Shofty, M., Brüne, M., Ebert, A., Shefet, D., Levkovitz, Y., & Shamay-Tsoory, S. G. (2013). Improving social perception in schizophrenia: The role of oxytocin. *Schizophrenia Research, 146*(1–3), 357-362. doi: 10.1016/j.schres.2013.01.006

Frith, C. D., & Frith, U. (2006). How we predict what other people are going to do. *Brain Research, 1079*, 36-46. doi: 10.1016/j.brainres.2005.12.126

Guastella, A. J., Mitchell, P. B., & Mathews, F. (2008). Oxytocin enhances the encoding of positive social memories in humans. *Biological Psychiatry, 64*(3), 256-258. doi: 10.1016/j.biopsych.2008.02.008

Haselton, M. G., & Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review, 10*(1), 47-66. doi: 10.1207/s15327957pspr1001_3

Kemp, A. H., & Guastella, A. J. (2011). The role of oxytocin in human affect: A novel hypothesis. *Current Directions in Psychological Science, 20*(4), 222-231. doi: 10.1177/0963721411417547

Kirsch, P., Esslinger, C., Chen, Q., Mier, D., Lis, S., Siddhanti, S., . . . Meyer-Lindenberg, A. (2005). Oxytocin modulates neural circuitry for social cognition and fear in humans. *The Journal of Neuroscience, 25*(49), 11489-11493. doi: 10.1523/jneurosci.3984-05.2005

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature, 435*(7042), 673-676. doi: 10.1038/nature03701

Kret, M. E., & De Dreu, C. (2013). Oxytocin-motivated ally selection is moderated by fetal testosterone exposure and empathic concern. *Frontiers in Neuroscience, 7*, 1. doi: 10.3389/fnins.2013.00001

Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion, 7*(4), 730-735. doi: 10.1037/1528-3542.7.4.730

Lischke, A., Berger, C., Prehn, K., Heinrichs, M., Herpertz, S. C., & Domes, G. (2012). Intranasal oxytocin enhances emotion recognition from dynamic facial expressions and leaves eye-gaze unaffected. *Psychoneuroendocrinology, 37*(4), 475-481. doi: 10.1016/j.psyneuen.2011.07.015

Lischke, A., Gamer, M., Berger, C., Grossmann, A., Hauenstein, K., Heinrichs, M., . . . Domes, G. (2012). Oxytocin increases amygdala reactivity to threatening scenes in females. *Psychoneuroendocrinology, 37*(9), 1431-1438. doi: 10.1016/j.psyneuen.2012.01.011

Mikolajczak, M., Gross, J. J., Lane, A., Corneille, O., de Timary, P., & Luminet, O. (2010). Oxytocin makes people trusting, not gullible. *Psychological Science, 21*(8), 1072-1074. doi: 10.1177/0956797610377343

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087-11092. doi: 10.1073/pnas.0805664105

Rimmele, U., Hediger, K., Heinrichs, M., & Klaver, P. (2009). Oxytocin makes a face in memory familiar. *Journal of Neuroscience, 29*(1), 38-42. doi: 10.1523/jneurosci.4260-08.2009

Schulze, L., Lischke, A., Greif, J., Herpertz, S. C., Heinrichs, M., & Domes, G. (2011). Oxytocin increases recognition of masked emotional faces. *Psychoneuroendocrinology, 36*(9), 1378-1382. doi: 10.1016/j.psyneuen.2011.03.011

Singular Inversions. (2007). Facegen Main Software Development Kit (Version 3.1) [Computer Program]. Vancouver, BC, Canada.

Spitzer, R. L., Kroenke, K., Williams, J. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166*(10), 1092-1097. doi: 10.1001/archinte.166.10.1092

Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2010). Affect 4.0: A free software package for implementing psychological and psychophysiological experiments. *Experimental Psychology, 57*(1), 36-45. doi: 10.1027/1618-3169/a000005

Stallen, M., De Dreu, C. K. W., Shalvi, S., Smidts, A., & Sanfey, A. G. (2012). The herding hormone: Oxytocin stimulates in-group conformity. *Psychological Science, 23*(11), 1288-1292. doi: 10.1177/0956797612446026

StataCorp. (2011). Stata Statistical Software (Version 12) [Computer Program]. College Station, TX: StataCorp LP.

Striepens, N., Scheele, D., Kendrick, K. M., Becker, B., Schäfer, L., Schwalba, K., . . . Hurlemann, R. (2012). Oxytocin facilitates protective responses to aversive social stimuli in males. *Proceedings of the National Academy of Sciences, 109*(44), 18144-18149. doi: 10.1073/pnas.1208852109

Theodoridou, A., Rowe, A. C., Penton-Voak, I., & Rogers, P. (2009). Oxytocin and social perception: Oxytocin increases perceived facial trustworthiness and attractiveness. *Hormones and Behavior, 56*(1), 128-132. doi: citeulike-article-id:6088836

van 't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition, 108*(3), 796-803. doi: 10.1016/j.cognition.2008.07.002

van Ijzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2012). A sniff of trust: Meta-analysis of the effects of intranasal oxytocin administration on face recognition, trust to in-group, and trust to out-group. *Psychoneuroendocrinology, 37*(3), 438-443. doi: 10.1016/j.psyneuen.2011.07.008

Van Lange, P. A. M., Otten, W., De Bruin, E. M. N., & Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology, 73*(4), 733-746. doi: 10.1037/0022-3514.73.4.733

Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United-States and Japan. *Motivation and Emotion, 18*(2), 129-166. doi: 10.1007/bf02249397

Zak, P. J., Stanton, A. A., & Ahmadi, S. (2007). Oxytocin increases generosity in humans. *PLoS ONE, 2*(11), e1128. doi: 10.1371/journal.pone.0001128

Chapter 2: Sexual dimorphism in oxytocin responses to health perception and disgust, with implications for theories on pathogen detection

This chapter has been published as:

Declerck, C. H., Lambert, B., & Boone, C. (2014). Sexual dimorphism in oxytocin responses to health perception and disgust, with implications for theories on pathogen detection. *Hormones and Behavior, 65*(5), 521-526. doi: 10.1016/j.yhbeh.2014.04.010

Abstract

In response to a recent hypothesis that the neuropeptide oxytocin might be involved in human pathogen avoidance mechanisms, we report the results of a study in which we investigate the effect of intranasal oxytocin on two behaviors serving as proxies for pathogen detection. Participants received either oxytocin or a placebo and were asked to evaluate (1) the health of Caucasian male computer-generated pictures that varied in facial redness (an indicator of hemoglobin perfusion) and (2) a series of pictures depicting disgusting scenarios. Men, but not women, evaluated all faces, regardless of color, as less healthy when given oxytocin compared to a placebo. Women, on the other hand, expressed decreased disgust when given oxytocin compared to a placebo. These results suggest that intranasal oxytocin administration does not facilitate pathogen detection based on visual cues, but instead reveal clear sex differences in the perception of health and sickness cues.

1. Introduction

Oxytocin (OT), the hormone involved in mammalian parturition and lactation, has recently attracted much attention with regards to its role in regulating social behavior. As a neurotransmitter in the central nervous system, OT has been shown to mediate a host of socially relevant behaviors, including social perception, social memory, trust, generosity, and cooperation (see reviews in Bos, Panksepp, Bluthe, and van Honk (2012); Campbell (2008); Meyer-Lindenberg, Domes, Kirsch, and Heinrichs (2011)). However, the effect of OT is not straightforward, and may include prosocial as well as anti-social effects depending on context (Bartz, Zaki, Bolger, & Ochsner, 2011; Declerck, Boone, & Kiyonari, 2010; Guastella & MacLeod, 2012). Such findings have prompted researchers to identify the conditions that determine when and how OT moderates behavior and thereby impacts the outcome of social interactions.

Because social interactions provide key opportunities for parasitic transmission, humans have evolved a rich repertoire of behavioral responses by which they minimize pathogen exposure (Oaten,

Stevenson, & Case, 2009). In a recent review article Kavaliers and Choleris (2011) put forth the hypothesis that OT may play a role in safeguarding people against pathogen infection by improving the ability to recognize and avoid contaminated others. Their arguments are mostly based on evidence from animal research. In rodents, OT has been specifically associated with detecting and responding to disease-infected conspecifics (Kavaliers, Choleris, Agmo, & Pfaff, 2004). For example, in mate-choice experiments, female mice are able to discriminate between the odors of healthy and infected males. But OT-gene knockout mice, or mice treated with an OT antagonist, lack this ability and show reduced aversion towards infected males. OT-gene knockout mice are furthermore unable to discriminate between the urinary odors of male mice that have been treated with a chemical substance containing elements of bacterial cell-walls, a procedure that is often used to simulate bacterial infections. However, while rodents rely primarily on olfaction, social cognition in humans is dominated by interpreting visual cues (Broad, Curley, & Keverne, 2006), making it difficult to extrapolate the findings of animal research regarding OT to humans. While rodents have evolved olfactory mechanisms to avoid stimuli that smell like disease, humans show behaviorally more elaborate responses to threatening social stimuli, including in-group favoritism, out-group exclusion, and specific emotions such as trust and schadenfreude, all of which have previously been associated with OT (Kavaliers & Choleris, 2011). For example, exogenous OT tends to promote in-group cooperation and ethnocentrism (reviewed in De Dreu (2012)), increase the perception of fear expression (Fischer-Shofty, Shamay-Tsoory, Harari, & Levkovitz, 2010), decrease trust when people are perceived to be untrustworthy (Mikolajczak et al., 2010) and decrease cooperation with anonymous others (Declerck et al., 2010). Such behaviors are useful to prevent contact with out-groups which may not only be hostile but are also potentially carriers of strange parasites and bacteria.

So far, we are not aware of studies that have empirically addressed if OT directly or indirectly facilitates the *detection* of pathogens in humans. This report is a first step towards filling this gap by presenting the results of an experiment whereby we test if intranasal OT affects sensory perceptions that may have to do with detecting pathogens. First, we tested if OT enhances the perception of health cues by discriminating between pictures of faces that vary in redness. There is evidence that facial redness is perceived to reflect physiological health of Caucasians, while pallor, indicating poor blood perfusion, has been associated with infections like malaria (Stephen, Coetzee, Law Smith, & Perrett, 2009). Hence if OT plays a role in pathogen detection, we expect that compared to a placebo, individuals given OT would assign lower health scores to pale individuals. Second, we tested if OT enhances feelings of disgust in response to pictures of repulsive situations that reveal a high pathogen load. In the field of evolutionary psychology, there is abundant literature to substantiate

that disgust is an adaptive response to avoid infestation by pathogens (Oaten et al., 2009). If OT facilitates avoiding infection, then we expect that OT will also increase feelings of disgust when exposed to pictures representing foul, bacteria-loaded conditions. Finally, because OT interacts with estrogen in females, making sex differences in OT regulation very probable (Bos et al., 2012; Choleris, Pfaff, & Kavaliers, 2013), we also investigate if the effects of OT on health perception and feelings of disgust are moderated by gender.

2. Methods

A student population (N = 106, 61 females, mean age = 22 ± 2.5) registered by e-mail to participate in exchange for a €10 remuneration fee. The study was described as an experiment that tested the effect of a hormone on evaluative judgments. Exclusion criteria included the use of alcohol and nicotine for 12 hrs, and any other medication other than anti-conception for 24 hrs prior to the study, any diagnosed neurological or psychological disorder, nasal obstruction or colorblindness. To make sure none of the female participants were pregnant, they took a pregnancy test between 1 and 48 hrs prior to the experiment. Participants consented to the procedures which were in accordance with the Declaration of Helsinki and approved by the Ethical Commission of the University. Debriefing occurred by sending each participant an e-mail referring to a website where the intent and results of the study were explained. This experiment was part of a larger study in which we investigated the effect of exogenous OT on two dimensions: trustworthiness and health evaluations.[1] Here we report only on the latter.

The experiment was conducted in computer rooms during 17 sessions with no less than 4 and no more than 10 participants in each session. They were instructed to self-administer an intranasal dose of 24 IU OT (Syntocinon, Novartis; three puffs per nostril) or placebo (with only the carrier) following a double-blind random design. Participants then waited 35 minutes before starting with the experiment. They were instructed to remain quiet and not leave the room. A room supervisor stayed with them the entire time.

The task consisted of evaluating two series of pictures. For the first series (testing the first objective) five artificially generated, male[2] compound faces were selected out of a large database created by

[1] Trustworthiness evaluations were assessed with different experimental stimuli than the ones presented here and are highly unlikely to have affected the current results. We do not report these data here because they were collected with the purpose of testing a different hypothesis that was unrelated to the current experiment. The order of presenting the two sets of stimuli was determined randomly. For more information regarding these additional data, please contact the corresponding author.

[2] Using only male faces simplifies the interpretation of the results by eliminating the possibility that perceptions of health depends on the sex of the target. A drawback, however, is that we introduce a bias (same sex

Oosterhof and Todorov (2008). The advantage of using artificial faces is that they are perfectly symmetrical and emotionally neutral. In a preliminary task (conducted with the same population) we asked the participants to rate the health of these five neutral faces on a scale from 0 to 9. From this we could deduce that the health of these "neutral standards" were rated higher by males (mean = 5.8, S.E. = .15) compared to females (mean = 5.3, S.E. = .15). OT, however, had no effect on the evaluation of neutral faces, neither for men, nor for women. Furthermore, there is no significant interactive effect between sex and OT on the evaluation of neutral faces (see Appendix figure 1 and table 1).

For the actual task, each face was manipulated by the authors to vary along a gradient of 7 shades of redness, going from an extremely pale face (given the value of -3) to an exaggerated red face (given the value of +3). The neutral variant was given value 0 and the original pictures were manipulated to ensure equal mean CIELab[3] values for each face before manipulating redness. A total of 35 pictures (5 facial identities in 7 color gradations) were shown one by one on a computer screen, always in combination with a neutral face (the "standard") of the same identity. The neutral face was always shown on the left side of the screen, and the face that varied in redness on the right side of the screen. The participants were asked to give a score ranging from -4 (right face less healthy than the left face) to +4 (right face more healthy than left face),[4] with 0 indicating no difference between the left and the right face. Scoring was self-paced and occurred by left-mouse clicking on a scale displayed at the bottom of the screen.

The second series (testing the second objective) consisted of five pictures depicting disgusting scenes. These included: (1) a diphtheria skin lesion on the leg, (2) a filthy bathroom, (3) a pot of worm-infested soup, (4) a festering arm-wound and (5) rotten teeth. These pictures, obtained from the internet, were chosen because they contain cues that are indicative of an adaptive, pathogen-related disgust response (Tybur, Lieberman, Kurzban, & DeScioli, 2013). Participants were asked to indicate on a scale of 0 to 9 how disgusted they felt when looking at each picture.

The dependent variables are the health scores (-4 to +4) assigned to the stimuli in series 1, and the disgust scores (0 – 9) in series 2. The independent variables of interest are OT (coded 1, versus placebo, coded 0) and sex (males coded 1).

---

evaluations versus opposite sex evaluations for males and females respectively), making an absolute comparison between the evaluations of men and women problematic. This is an additional reason why we analyze the results for men and women separately.

[3] The CIELab color system (L = light-dark; a = red-green, b = yellow-blue) is commonly used in research on visual perception and designed to be perceptually uniform (see Stephen, Smith, Stirrat, and Perrett (2009))

[4] Note that the range of possible answers (-4 to +4) was greater than the range of possible color values assigned to each picture (-3 to +3) in order to broaden the response variance.

3. Results

To test if OT affects health perception on the basis of facial redness (perceived hemoglobin perfusion), we plotted the health score of each face versus the redness gradient (see figure 1), separately for men and women. The linear relation between perceived health and redness is consistent with the results of Stephen, Coetzee, et al. (2009) and Stephen, Smith, et al. (2009) indicating that people optimize health appearance by increasing facial redness above basal levels. Visual inspection reveals that (1) for both men and women the relation between perceived health and redness is similar across the 5 facial identities, and (2) that for men, but not for women, OT appears to lower health scores for nearly all shades of red. To test the latter finding statistically we averaged the health scores corresponding to each redness grade over the five facial identities and conducted regression analyses on panel data (7 averaged scores per participant).[5] To control for potential differences in the a-priori evaluations of the five neutral faces, the average *neutral score* obtained from each individual in the preliminary task is added as a factor in the regression models. To control for unobserved heterogeneity between subjects, we use a random effects model that accounts for clustering of observations per individual. All reported *p*-values are two-tailed. The analyses of the female participants also include a dummy variable "contraceptive," indicating whether or not the participant was on hormonal contraceptives. From questionnaire data collected on the day of the experiment we knew that 66% of the females were taking hormonal contraceptives, and that one of the naturally cycling women reported an unusually long cycle of 221 days. This person was not included in further analyses, leaving N=60 for the female participants and N=105 for the total population.

The results of the analyses for men indicate a significant negative effect of *oxytocin* (versus placebo) on perceived health above the effect of *redness gradient* ($B = -.42$, S. E. = .17, $p < .007$). For women, this is not the case ($B = .15$, S. E. = .11, $p = .17$). Table 1 shows the full regression models, as well as further analyses indicating (1) no main effect of sex on health perception across the entire redness gradient, (2) a significant interactive effect of OT and sex ($B = -.56$, S. E. = .18, $p < .002$), as suggested by the separate male/female analyses presented above, and (3) neither for males, nor for females, does the effect of OT depend on redness. To test if OT differentially affects the health perception of pale versus red faces in males, we also tested models in which we replaced the variable "redness gradient" by a dummy variable: "pale" (face paler than neutral, coded 1) versus "red" (face more red than neutral, coded 0). The interactive effects of OT*pale is not significant (see Appendix table 2).

[5] We first tested the effect of OT on perceived health for each face separately and confirmed that the effect of OT did not differ between faces.

Hence males given OT rated pale as well as red faces less healthy then females. The latter is not consistent with the hypothesized pathogen detection effect of OT. In addition, we corroborated that OT did not affect the ratings of the neutral faces in males and performed a regression only on the neutral faces, with the mean rating (averaged over 5 facial identities) as the dependent variable, OT as the independent variable, and the average neutral score each individual gave to the neutral faces in the preliminary task as control variable. As expected, OT had no effect on the ratings of the neutral faces (B = -.22, S. E. = .22, p < .31).[6]

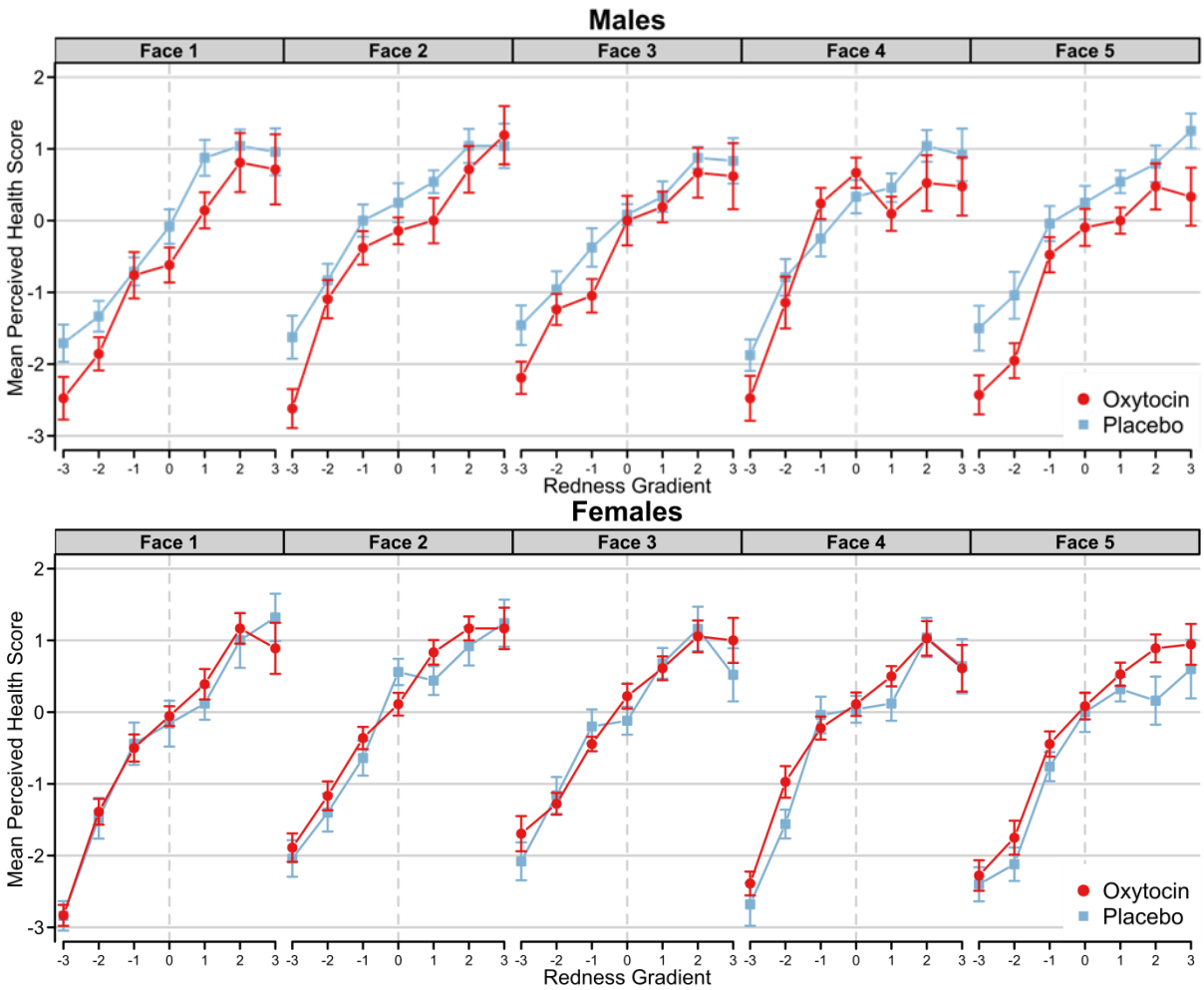

Fig. 1. Perceived health of Caucasian faces that have been manipulated to vary in redness (as an indicator of hemoglobin perfusion). Error flags represent standard errors.

[6] We also tested this for each facial identity separately by conducting 5 separate t- tests comparing the average rating of the oxytocin group with the average rating of the placebo group. For none of the five facial identities do the groups differ in their ratings of neutral faces.

Table 1. Unstandardized coefficients of random effects GLS regressions (clustered per individual) showing the effect of OT and other predictors on perceived health.

	Model I Males only	**Model II Males only**	**Model III Females only**	**Model IV Females only**	**Model V Males & females pooled**	**Model VI Males & females pooled**
Oxytocin (OT)	-.42 (.16)	-.42 (.16)	.15 (.11)	.15 (.11)	-.094 (.094)	.15 (.12)
	p=.007	p= .007	p=.17	p=.17	p=.32	p=.21
Sex					-.007 (.097)	.28 (.13)
					p=.94	p=.035
Contraceptive			-.039 (.12)	-.039 (.12)		
			p=.74	p=.74		
Redness	.48 (.024)	.45 (.033)	.55 (.023)	.55 (.035)	.52 (.017)	.52 (.017)
	p<.001	p<.001	p<.001	p<001	p<.001	p<.001
Neutral score	.068 (.081)	.068 (.081)	.024 (.048)	.024 (.048)	.028 (.044)	.041 (.042)
	p=.40	p=.40	p=.62	p=.62	p=.53	p=.33
OT*sex						-.56 (.18)
						p=.002
OT*redness		.052 (.049)		-.0044 (.046)		
		p=.29		p=.93		
Constant	-.42 (.47)	-.42 (.47)	-.42 (.30)	-.42 (.30)	-.33 (.25)	-.54 (.25)
	p=.38	p=.38	p =.15	p=.15	p=.18	p=.027
Wald Chi ^{2}	388.64	389.96	576.26	574.69	951.36	960.96
N	315	315	420	420	735	735

The data is in panel form with each individual assigning 7 health scores corresponding to the different grades of redness. These scores represent the average of five different facial identities. *Contraceptive* indicates whether the participant was currently taking hormonal contraceptives (coded 1). *Redness* represents a continuous variable ranging from -3 (extreme pale) to +3 (extreme red); *Neutral score* is the average health score each individual assigned to the neutral faces in the preliminary task; *oxytocin* is coded 1, placebo coded 0; *sex* is coded 1 for males, 0 for females. Standard errors are given in parentheses. All p-values are two-tailed.

To test if OT influences feelings of disgust when confronted with pictures of pathogen laden-conditions, we first plotted the disgust scores for each picture (see figure 2), separately for men and women. Visual inspection shows that (1) women tend to assign overall higher disgust scores compared to men, (2) OT tends to consistently lower disgust scores for women, but not for men. We

perform regression analyses with random effects on panel data (5 evaluations per participants), controlling for the different picture types. For women, the effect of OT on disgust is significant and negative (B = -1.35, S. E. = .43, p <. 002), while OT has no effect on men (B = .031, S. E. = .74, p = .97), see Table 2 which furthermore shows a large main effect of sex (B = -1.02, S. E. = .41, p < .013). The interactive effect of OT and sex is not significant, but the effect size is in the expected direction (B = 1.37, S. E.= .81, p = .088).

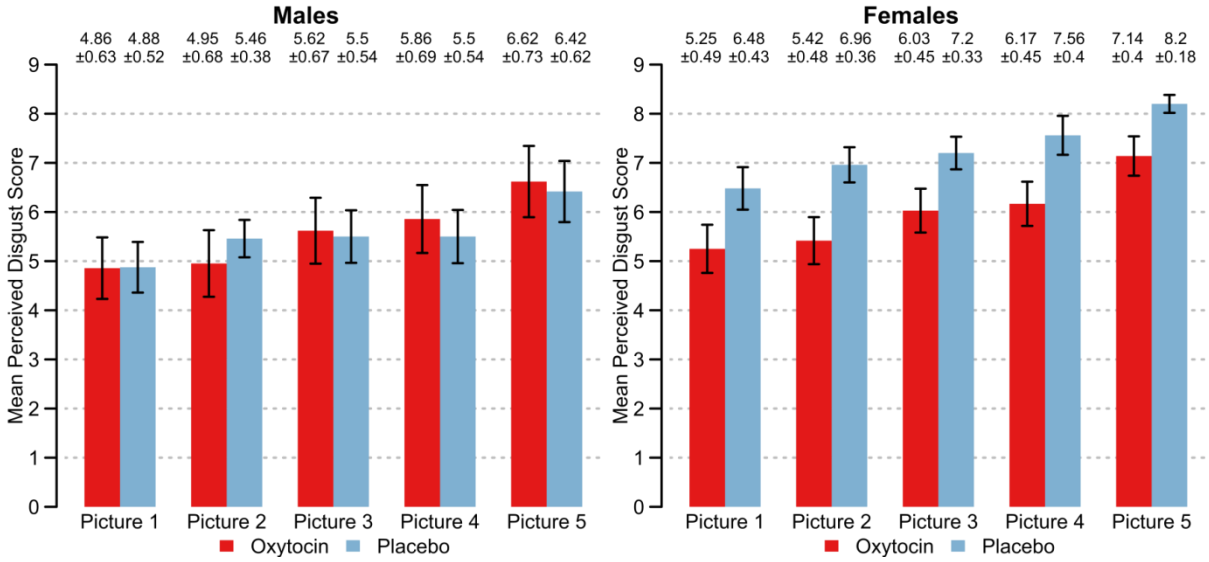

Fig. 2. Evaluations of feelings of disgust upon viewing 5 different pictures. Error flags represent standard error. Picture 1: a diphtheria skin lesion on the leg; Picture 2: a filthy bathroom; Picture 3: a pot of spoiled, worm-infested soup; Picture 4: a festering wound; Picture 5: rotten teeth.

Table 2. Unstandardized coefficients of GLS regressions with random effects on panel data (clustered on each individuals) showing the effect of oxytocin on disgust scores.

	Model I **Males only**	**Model II** **Females only**	**Model III** **Males & females pooled**	**Model IV** **Males & females pooled**
Picture 1	-1.64 (.32)	-1.85 (.35)	-1.76 (.24)	-1.76 (.24)
	p<.001	p<.001	p<.001	p<.001
Picture 2	-1.29 (.32)	-1.52 (.35)	-1.42 (.24)	-1.42 (.24)
	p<.001	p<.001	p<.001	p<.001
Picture 3	-.96 (.32)	-1.07 (.35)	-1.02 (.24)	-1.02 (.24)
	p=.003	p=.002	p<.001	p<.001
Picture 4	-.84 (.32)	-.85 (.35)	-.85 (.24)	-.85 (.24)
	p=.008	p=.016	p<.001	p<.001
OT	.031 (.74)	-1.35 (.43)	-.75 (.40)	-1.35 (.53)
	p=.97	p=.002	p=.064	p=.011
Contraceptive		.12 (.45)		
		p=.79		
Sex			-1.02 (.41)	-1.73 (.58)
			p=.013	p=.003
OT*sex				1.37 (.81)
				p=.088
Constant	6.50 (.55)	8.26 (.49)	7.94 (.39)	8.29 (.44)
	p<.001	p<.001	p<.001	p<.001
Wald chi^2	29.91	42.25	69.37	72.44
N	225	300	525	525

Each individual scored five different picture types which are added to the models as dummy variables. Standard errors are given in parentheses. All p -values are two-tailed. *Oxytocin* is coded 1, placebo coded 0; *sex* is coded 1 for males, 0 for females; *Pictures 1-4* are respectively a wounded arm, a filthy bathroom, spoiled soup, and a wounded toe. Picture 5, depicting rotten teeth, is the comparison group. *Contraceptive* indicates whether the participant was currently taking hormonal contraceptives (coded 1).

Finally, because OT interacts with estrogen, and estrogen levels vary with the menstrual cycle of women, we redid the above regression analyses for female participants (N = 60) while taking into account the stage in their menstrual cycle. Because hormone levels are expected to fluctuate far more in the naturally cycling group, we conduct separate analyses for the group of women on

hormonal contraceptives (n = 40) and the naturally cycling group (n = 20). With respect to health perception, the time in the menstrual cycle has no significant effect on women's evaluations in either group. This variable furthermore does not interact significantly with OT, and its addition in the regression models does not significantly change the results reported earlier in Table 1. With respect to feelings of disgust, the time in the menstrual cycle does have a significant effect for the group that takes hormonal contraceptives (B = 0.027, S.E. = 0.013, p = 0.032). In the group taking hormonal contraceptives, the effect of OT remains significant (B = -1.61 , S.E. = 0.53, p = 0.002) above and beyond the effect of the menstrual cycle. In the naturally cycling group, the significant effect of OT disappears, probably due to the small sample size. The direction of the effect, and the effect size, however, are similar to the group using contraception. (B = -1.07 , S.E. = 0.74, p = 0.15). There is furthermore no interaction effect between OT and menstrual cycle in either group. Full details of these additional regression analyses are reported in Appendix table 3A and 3B.

4. Discussion

The data reported in this study are difficult to conciliate with the hypothesis that OT facilitates pathogen detection in humans based on visual cues. We focused on the visual system because, as primates evolved from a nocturnal to a diurnal lifestyle, the importance of olfactory processing and the vomeronasal system declined extensively, driving the need to gather social information via visual cues (Curley & Keverne, 2005). Unlike in rodents where OT is associated with olfactory-mediated recognition of infected conspecifics, the results of the current study found nothing equivalent in the human visual system. On the contrary, after exogenous OT administration, women in the current study find pictures of pathogen-loaded situations less disgusting, and men evaluate unhealthy (pale-looking) as well as healthy-looking males to be in worse physical health.

Although these data indicate that intra-nasal OT administration does not help an individual to avoid visual cues related to pathogen infection, the nature of the experimental manipulations imposes limitations that call for cautious interpretation of the data and preclude overgeneralizations. It is, for example still possible that OT (as in rodents) would play a role in human pathogen avoidance mechanisms through affecting olfaction. By using a surrogate task that shows pictures of infected individuals, rather than actual infestations, we deprived participants from one of their senses. Even if the predominance of olfaction has waned during human evolution, there are some indications that humans do use odor as a cue for detecting pathogens and sickness (Moshkin et al., 2012; Olsson et al., 2014). Investigating the role of OT herein would be of interest.

Another factor that precludes generalization is that the current study only examined the effect of OT addition (via an intranasal spay). While administering and blocking OT often have opposite effects

(e.g., facilitating and preventing partner preferences in female prairie voles, see Ferguson, Young, and Insel (2002)), hormone addition and removal do not consistently lead to complementary results. In female meadow voles, for example, administering OT enhances partner preferences, while OT antagonists do not block the formation of preferences (Anacker & Beery, 2013). Hence we cannot really predict what the results of this experiment would have been had we looked at OT subtraction. Furthermore, dose-dependent effects of OT have been reported: 24 IU of intranasal OT significantly improves autobiographical memory recall, whereas no such effect was observed for a dose of 48 IU (Cardoso, Orlando, Brown, Joober, & Ellenbogen, 2012), a finding which has been attributed to the possibility that, at high concentrations, OT begins to occupy arginine-vasopressin receptors, which tend to elicit opposing effects (Olff et al., 2013). Finally, improvements in well-established social skills (such as a social recognition task) may be more difficult to detect in response to OT-addition, and does not dismiss the fact that the role of OT might become apparent with OT depletion, as is the case with OT-gene knockout mice (Kavaliers et al., 2004).

We also acknowledge that some of the methodological choices we needed to make may have possibly introduced biases in the experiment. As mentioned in footnote 2, males in this study evaluated same-sex stimuli in the health dimension, while females evaluated opposite sex-stimuli. By analyzing the data for men and women separately, we avoid the confounding effect of this potential bias. While this set up does not allow us to generalize the null finding of 'women categorizing male stimuli' to 'all stimuli,' this does not dismiss the finding that OT differentially affects men and women's health evaluation of male stimuli. Finally, an anonymous reviewer pointed out that the health data may have inadvertently been biased because the referent face and the face which needed to be evaluated were kept on the same side. We opted for this consistency because we did not want to place an additional burden on working memory by having to make repeated mental switches regarding the referent side. This simplified the task, and, while not ideal, it is difficult to fathom an underlying mechanism that would explain why one side would be deemed significantly unhealthier than the other side.[7] Differences in lighting or color rendition cannot be the reason as the curtains were drawn and the artificial lights were dimmed. Furthermore, such a systematic bias can again not explain the sex differences that emerged. Despite these experimental limitations, three solid conclusions can be drawn from these data.

[7]If there was a side bias in ratings, this would also be apparent when evaluating the neutral faces. For each of the five facial identities, we tested if the ratings for the neutral face differed significantly from 0 (the assigned score of the referent face). This was the case for only one out of the five faces, which was given a higher health rating. In addition, the mean ratings for the neutral faces were three times below zero, and two times above zero, which is not sufficient evidence to substantiate that people would consistently rate the face on one side to be more or less healthy than the other side.

First, OT affects how men perceive health on the basis of facial color. Consistent with prior research (Stephen, Smith, et al., 2009) the data show that health ratings typically go up with facial redness – an indication of blood perfusion and physiological well-being. What we additionally find is that, regardless of whether faces are pale or red, males (but not females) given exogenous OT perceive faces of other men as less healthy compared to males in the placebo condition. Thus, men given OT tend to perceive health cues of other (unknown) male faces differently, but it is unlikely to be related to pathogen detection. While red faces could possibly be interpreted as more feverish and hence indicative of a pathogen load, men in the OT condition still rated red faces as more healthy, just less so than in the placebo condition (Figure 1). Why OT lowers health perceptions of healthy - as well as unhealthy - looking individuals is a puzzle that cannot, at this point, be adequately explained by either the current data or other findings from the literature.

The second conclusion is that, for females (but not males), OT appears to inhibit feelings of disgust to visual cues of pathogen-infested conditions. This finding seems to contradict a role of OT in pathogen avoidance for women, as OT actually impedes their natural repulsion towards filth and contamination. As a straightforward explanation, it appears that the lower disgust feelings could simply be a consequence of OT's general anxiolytic effect, which has been reported numerous times in the literature (Churchland & Winkielman, 2012; Kemp & Guastella, 2011). Women given exogenous OT would then become less worried about infection. But the same argument could equally well apply to men. In fact, only for men does OT appear to decrease amygdala reactivity to negative emotional stimuli, and the opposite may in fact occur in women (Domes et al., 2010; Lischke et al., 2012). While there are recent indications that OT may reduce anxiety through different channels, more research will be needed to understand the exact mechanisms by which men and women differentially respond to stressful (and disgusting) stimuli.

The final conclusion drawn from these data is that there are clear sex differences in the effect of OT on the perception of health and sickness cues. Although sex differences in the functions of OT have long been speculated based on the animal literature and the well-known OT-estrogen interactions (Bos et al., 2012; Kubzansky, Mendes, Appleton, Block, & Adler, 2012), there are to date relatively few experimental studies that allow us to infer exactly how uniform the effects of OT are among men and women (see also Olff et al. (2013), who review the moderating effects of gender and individual differences with respect to OT functions). We propose two avenues to explore in future research. First, there should be further inquiry into how common sexual dimorphism in perception is when people are given exogenous OT. Second, studying how the effect of exogenous OT is related to the saliency of gender-dependent decision contexts would give insights into how computational reasoning varies between the sexes. If the evolution of sociality posed different challenges on males

and females, they would become differentially sensitive to gender-specific contexts or cues. In combination with a very distinct neuroendocrine system, this may have led to current differences in male and female behaviors, decisions, and attitudes.

References

Anacker, A. M. J., & Beery, A. K. (2013). Life in groups: The roles of oxytocin in mammalian sociality. *Frontiers in Behavioral Neuroscience, 7*. doi: 10.3389/fnbeh.2013.00185

Bartz, J. A., Zaki, J., Bolger, N., & Ochsner, K. N. (2011). Social effects of oxytocin in humans: Context and person matter. *Trends in Cognitive Sciences, 15*(7), 301-309. doi: 10.1016/j.tics.2011.05.002

Bos, P. A., Panksepp, J., Bluthe, R. M., & van Honk, J. (2012). Acute effects of steroid hormones and neuropeptides on human social-emotional behavior: A review of single administration studies. *Frontiers in Neuroendocrinology, 33*(1), 17-35. doi: 10.1016/j.yfrne.2011.01.002

Broad, K. D., Curley, J. P., & Keverne, E. B. (2006). Mother-infant bonding and the evolution of mammalian social relationships. *Philosophical Transactions of the Royal Society B-Biological Sciences, 361*(1476), 2199-2214. doi: 10.1098/rstb.2006.1940

Campbell, A. (2008). Attachment, aggression and affiliation: The role of oxytocin in female social behavior. *Biological Psychology, 77*(1), 1-10. doi: 10.1016/j.biopsycho.2007.09.001

Cardoso, C., Orlando, M. A., Brown, C. A., Joober, R., & Ellenbogen, M. A. (2012). Intranasal oxytocin promotes the recall of specific autobiographical memories: Dose-dependant effects and moderation by depressive symptoms. *Biological Psychiatry, 71*(8), 276S-277S.

Choleris, E., Pfaff, D. W., & Kavaliers, M. (2013). *Oxytocin, vasopressin and related peptides in the regulation of behaviour*. Cambridge: Cambridge University Press.

Churchland, P. S., & Winkielman, P. (2012). Modulating social behavior with oxytocin: How does it work? What does it mean? *Hormones and Behavior, 61*(3), 392-399. doi: 10.1016/j.yhbeh.2011.12.003

Curley, J. P., & Keverne, E. B. (2005). Genes, brains and mammalian social bonds. *Trends in Ecology & Evolution, 20*(10), 561-567. doi: 10.1016/j.tree.2005.05.018

De Dreu, C. K. W. (2012). Oxytocin modulates cooperation within and competition between groups: An integrative review and research agenda. *Hormones and Behavior, 61*(3), 419-428. doi: 10.1016/j.yhbeh.2011.12.009

Declerck, C. H., Boone, C., & Kiyonari, T. (2010). Oxytocin and cooperation under conditions of uncertainty: The modulating role of incentives and social information. *Hormones and Behavior, 57*(3), 368-374. doi: 10.1016/j.yhbeh.2010.01.006

Domes, G., Lischke, A., Berger, C., Grossmann, A., Hauenstein, K., Heinrichs, M., & Herpertz, S. C. (2010). Effects of intranasal oxytocin on emotional face processing in women. *Psychoneuroendocrinology, 35*(1), 83-93. doi: 10.1016/j.psyneuen.2009.06.016

Ferguson, J. N., Young, L. J., & Insel, T. R. (2002). The neuroendocrine basis of social recognition. *Frontiers in Neuroendocrinology, 23*(2), 200-224. doi: 10.1006/frne.2002.0229

Fischer-Shofty, M., Shamay-Tsoory, S. G., Harari, H., & Levkovitz, Y. (2010). The effect of intranasal administration of oxytocin on fear recognition. *Neuropsychologia, 48*(1), 179-184. doi: 10.1016/j.neuropsychologia.2009.09.003

Guastella, A. J., & MacLeod, C. (2012). A critical review of the influence of oxytocin nasal spray on social cognition in humans: Evidence and future directions. *Hormones and Behavior, 61*(3), 410-418. doi: 10.1016/j.yhbeh.2012.01.002

Kavaliers, M., & Choleris, E. (2011). Sociality, pathogen avoidance, and the neuropeptides oxytocin and arginine vasopressin. *Psychological Science, 22*(11), 1367-1374. doi: 10.1177/0956797611420576

Kavaliers, M., Choleris, E., Agmo, A., & Pfaff, D. W. (2004). Olfactory-mediated parasite recognition and avoidance: Linking genes to behavior. *Hormones and Behavior, 46*(3), 272-283. doi: 10.1016/j.yhbeh.2004.03.005

Kemp, A. H., & Guastella, A. J. (2011). The role of oxytocin in human affect: A novel hypothesis. *Current Directions in Psychological Science, 20*(4), 222-231. doi: 10.1177/0963721411417547

Kubzansky, L. D., Mendes, W. B., Appleton, A. A., Block, J., & Adler, G. K. (2012). A heartfelt response: Oxytocin effects on response to social stress in men and women. *Biological Psychology, 90*(1), 1-9. doi: 10.1016/j.biopsycho.2012.02.010

Lischke, A., Gamer, M., Berger, C., Grossmann, A., Hauenstein, K., Heinrichs, M., . . . Domes, G. (2012). Oxytocin increases amygdala reactivity to threatening scenes in females. *Psychoneuroendocrinology, 37*(9), 1431-1438. doi: 10.1016/j.psyneuen.2012.01.011

Meyer-Lindenberg, A., Domes, G., Kirsch, P., & Heinrichs, M. (2011). Oxytocin and vasopressin in the human brain: Social neuropeptides for translational medicine. *Nature Reviews Neuroscience, 12*(9), 524-538.

Mikolajczak, M., Gross, J. J., Lane, A., Corneille, O., de Timary, P., & Luminet, O. (2010). Oxytocin makes people trusting, not gullible. *Psychological Science, 21*(8), 1072-1074. doi: 10.1177/0956797610377343

Moshkin, M., Litvinova, N., Litvinova, E. A., Bedareva, A., Lutsyuk, A., & Gerlinskaya, L. (2012). Scent recognition of infected status in humans. *Journal of Sexual Medicine, 9*(12), 3211-3218. doi: 10.1111/j.1743-6109.2011.02562.x

Oaten, M., Stevenson, R. J., & Case, T. I. (2009). Disgust as a disease-avoidance mechanism. *Psychological Bulletin, 135*(2), 303-321. doi: 10.1037/a0014823

Olff, M., Frijling, J. L., Kubzansky, L. D., Bradley, B., Ellenbogen, M. A., Cardoso, C., . . . van Zuiden, M. (2013). The role of oxytocin in social bonding, stress regulation and mental health: An update on the moderating effects of context and interindividual differences. *Psychoneuroendocrinology, 38*(9), 1883-1894. doi: 10.1016/j.psyneuen.2013.06.019

Olsson, M. J., Lundstrom, J. N., Kimball, B. A., Gordon, A. R., Karshikoff, B., Hosseini, N., . . . Lekander, M. (2014). The scent of disease human body odor contains an early chemosensory cue of sickness. *Psychological Science, 25*(3), 817-823. doi: 10.1177/0956797613515681

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087-11092. doi: 10.1073/pnas.0805664105

Stephen, I. D., Coetzee, V., Law Smith, M., & Perrett, D. I. (2009). Skin blood perfusion and oxygenation colour affect perceived human health. *PLoS ONE, 4*(4), e5083. doi: 10.1371/journal.pone.0005083

Stephen, I. D., Smith, M. J. L., Stirrat, M. R., & Perrett, D. I. (2009). Facial skin coloration affects perceived health of human faces. *International Journal of Primatology, 30*(6), 845-857. doi: 10.1007/s10764-009-9380-z

Tybur, J. M., Lieberman, D., Kurzban, R., & DeScioli, P. (2013). Disgust: Evolved function and structure. *Psychological Review, 120*(1), 65-84. doi: 10.1037/a0030778

Appendix

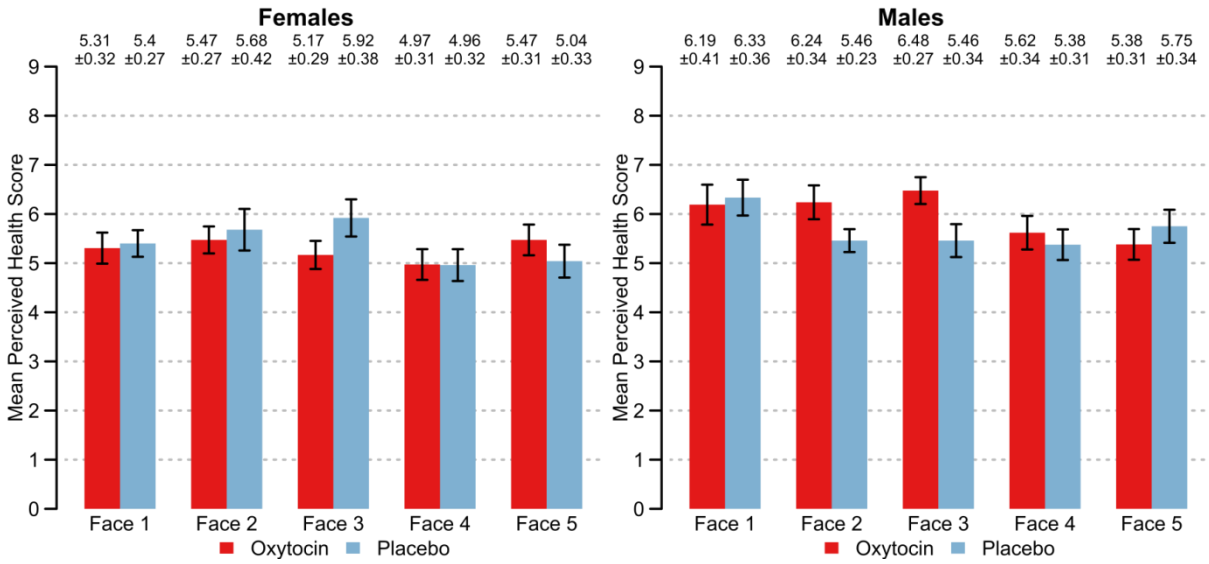

Fig. 1. Perceived health of each of five neutral (computer-generated) Caucasian faces. Error bars represent standard error. *Left*: average health score of female participants. *Right*: average health score of male participants

Table 1. Coefficients of regression models showing the effect of OT and sex on health scores (averaged over the five faces) assigned to neutral faces.

	Model 1		**Model 2**	
Oxytocin (OT)	.052 (.21)	p=.81	-.14 (.28)	p=.61
Sex	.51 (.21)	p=.020	.28 (.31)	p=.37
OT*sex			.45 (.43)	p=.30
Constant	5.29 (.19)	p<.001	5.4 (.22)	p<.001
R^2		.052		.062
N		105		105

Oxytocin and males are coded 1; placebo and females are coded 0. Standard errors are given in parentheses. All p-values are two-tailed.

Table 2. Unstandardized coefficients of random effects GLS regressions (clustered per individual) showing the effect of OT and other predictors on perceived health.

	Model I **Males only**	**Model II** **Males only**	**Model III** **Females only**	**Model IV** **Females only**
Oxytocin (OT)	-.45 (.16)	-.39 (.20)	.17 (.12)	.19 (.17)
	p=.004	p=.052	p=.14	p=.26
Contraceptive			-.026 (.13)	-.026 (.13)
			p=.84	p=.84
Neutral score	.069 (.082)	.069 (.082)	.042 (.052)	.042 (.052)
	p=.40	p=.40	p=.43	p=.43
Pale	-1.86 (.12)	-1.80 (.17)	-2.16 (.12)	-2.14 (.18)
	p<.001	p<.001	p<.001	p<.001
OT*Pale		-.12 (.25)		-.032 (.24)
		p=.63		p=.89
Constant	-.47 (.48)	-.44 (.48	-.49 (.33)	-.48 (.33)
	p=.33	p=.36	p=.14	p=.15
Wald Chi ^{2}	231.00	230.47	349.06	348.12
N	270	270	360	360

The data is in panel form with each individual assigning 7 health scores corresponding to the different grades of redness. These scores represent the average of five different facial identities. *Pale* represents a dummy indicating whether a face is paler (coded 1) or more red (coded 0) than the standard, neutral face. *Neutral score* is the average health score each individual assigned to neutral faces in the preliminary task. *Contraceptive* indicates whether the participant was currently taking hormonal contraceptives (coded 1); *oxytocin* is coded 1, placebo coded 0. Standard errors are given in parentheses. All p-values are two-tailed.

Tables 3A and 3B. The effect of OT on (A) perceived health and (B) disgust scores for female participants, controlling for the time in their menstrual cycle. One participant who reported an unusually long (221 day) cycle while not taking hormonal contraceptives, was excluded from the analyses.

Table 3A. Unstandardized coefficients of random effects GLS regressions (clustered per individual) showing the effect of OT and other predictors on perceived health.

	Model I	**Model II**	**Model III**	**Model IV**
Menstrual cycle	.0025 (.0099)	-.0024 (.012)	-.00072 (.0031)	.0030 (.0047)
	p=.80	p=.85	p=.82	p=.53
Oxytocin (OT)	.21 (.28)	.066 (.35)	.036 (.14)	.17 (.19)
	p=.44	p=.85	p=.80	p=.37
Redness	.59 (.043)	.59 (.043)	.52 (.026)	.52 (.026)
	p<.001	p<.001	p<.001	p<.001
Neutral score	.090 (.11)	.076 (.11)	-.040 (.063)	-.031 (.063)
	p=.41	p=.50	p=.53	p=.62
OT*		.014 (.021)		-.0066 (.0063)
Menstrual cycle		p=.49		p=.30
Constant	-.86 (.57)	-.71 (.62)	-.053 (.37)	-.17 (.39)
	p=.13	p=.25	p=.89	p=.66
Wald Chi ^{2}	191.51	191.89	391.16	392.25
N	140	140	280	280

The data is in panel form with each individual assigning 7 health scores corresponding to the different grades of redness. These scores represent the average of five different facial identities. *Redness* represents a continuous variable ranging from -3 (extreme pale) to +3 (extreme red); *Neutral score* is the average health score each individual assigned to the neutral faces in the preliminary task; *oxytocin* is coded 1, placebo coded 0. *Menstrual cycle* refers to the number of days elapsed since the first day of the last menses. Model I and II test the subjects who are not taking hormonal contraceptives (cycling naturally) while Model III and IV incorporate only those who are currently taking contraceptives. Standard errors are given in parentheses. All p-values are two-tailed.

Table 3B. Unstandardized coefficients of GLS regressions with random effects on panel data (clustered on each individuals) showing the effect of oxytocin on disgust scores.

	Model I	**Model II**	**Model III**	**Model IV**
Menstrual cycle	-.0088 (.031)	-.015 (.039)	.027 (.013)	.018 (.020)
	p=.78	p=.71	p=.032	p=.35
picture 1	-2.6 (.51)	-2.6 (.51)	-1.48 (.46)	-1.48 (.46)
	p<.001	p<.001	p=.001	p=.001
picture 2	-2.25 (.51)	-2.25 (.51)	-1.15 (.46)	-1.15 (.46)
	p<.001	p<.001	p=.012	p=.012
picture 3	-.75 (.51)	-.75 (.51)	-1.23 (.46)	-1.23 (.46)
	p=.14	p=.14	p=.008	p=.008
picture 4	-1.1 (.51)	-1.1 (.51)	-.73 (.46)	-.73 (.46)
	p=.031	p=.031	p=.11	p=.11
OT	-1.07 (.74)	-1.26 (1.08)	-1.61 (.53)	-1.92 (0.74)
	p=.15	p=0.24	p=.002	p=0.010
OT*		.018 (.067)		.016 (.026)
Menstrual cycle		p=.80		p=.54
Constant	8.49 (0.81)	8.60 (0.93)	7.84 (.56)	8.01 (.63)
	p<.001	p<.001	p<.001	p<.001
Wald chi^2	37.63	37.58	25.84	25.99
N	100	100	200	200

Each individual scored five different picture types which are added as dummy variables. *Oxytocin* is coded 1, placebo coded 0. *Pictures 1-4* are respectively a wounded arm, a filthy bathroom, spoiled soup, and a wounded toe. Picture 5, depicting rotten teeth, is the comparison group. *Menstrual cycle* refers to the number of days elapsed since the first day of the last menses. Model I and II test the subjects who are not taking hormonal contraceptives (cycling naturally) while Model III and IV incorporate only those who are currently taking contraceptives. Standard errors are given in parentheses. All p -values are two-tailed.

Chapter 3: A functional MRI study on how oxytocin affects decision making in social dilemmas: cooperate as long as it pays off, aggress only when you think you can win

This chapter is submitted to *Hormones and Behavior* and is currently under review.

Bruno Lambert, Carolyn H. Declerck, Christophe Boone, Paul M. Parizel

Abstract

We investigate if the neuropeptide oxytocin (OT), known to moderate social behaviour, influences strategic decision making in social dilemmas by facilitating the integration of incentives and social cues. Participants (N = 29) played two economic games with different incentive structures in the fMRI scanner after receiving OT or placebo (following a double blind, within-subject design). Pictures of angry or neutral faces (the social cues) were displayed alongside the game matrices. Consistent with *a priori* hypotheses based on the modulatory role of OT in mesolimbic dopaminergic brain regions, the results indicate that, compared to placebo, OT significantly increases the activation of the nucleus accumbens during an assurance (coordination) game that rewards mutual cooperation. This increases appetitive motivation so that cooperative behaviour is facilitated for risk averse individuals. OT also significantly attenuates the amygdala, thereby reducing the orienting response to social cues. The corresponding change in behaviour is only apparent in the chicken (or anti-coordination) game, where aggression is incentivized but fatal if the partner also aggresses. Because of this ambiguity, decision making can be improved by additional information, and OT steers decisions in the chicken game in accordance with the valence of the facial cue: aggress when face is neutral; retreat when it is angry. Through its combined influence on amygdala and nucleus accumbens, OT improves the selection of a cooperative or aggressive strategy in function of the best match between the incentives of the game and the social cues present in the decision environment.

1. Introduction

The ability to cooperate *and* aggress is a thumbprint of human nature. But deciding on one or the other when individual interests conflict is not trivial and requires the ability to integrate multiple sources of information at once. How do we do this?

The neuropeptide oxytocin (OT), best known for its hormonal functions facilitating labour and lactation, may have an equally important role as a neurotransmitter in the brain regulating many aspects of human social behaviour that extend beyond reproduction and parental care. Experimental

research with intranasal OT administration has revealed a number of ways by which the oxytocinergic system might affect the outcome of social interactions. First, OT stimulates social approach behaviours and strengthens bonds, supposedly because it links social interactions to feelings of reward (Campbell, 2008; Depue & Morrone-Strupinsky, 2005). This is supported by animal research showing that OT stimulates dopamine release in the nucleus accumbens (Dolen, Darvishzadeh, Huang, & Malenka, 2013), and by neuroimaging studies with humans revealing nucleus accumbens activation when breastfeeding mothers see pictures of their child (Strathearn, Fonagy, Amico, & Montague, 2009), or when males in a relationship see pictures of their partners (Scheele et al., 2013). Second, OT is well known to attenuate amygdala activation (Domes, Heinrichs, Glaescher, et al., 2007; Kirsch et al., 2005; Labuschagne et al., 2010), removing social anxiety and reducing fear (Heinrichs, von Dawans, & Domes, 2009; Neumann & Landgraf, 2012), which in turn further facilitates social approach (Kemp & Guastella, 2011). Finally, OT may increase the saliency of social cues (Shamay-Tsoory, 2010), which has been corroborated by findings that OT improves performance on theory of mind tasks (Domes, Heinrichs, Michel, Berger, & Herpertz, 2007), increases gaze to the eye region of the face (Guastella, Mitchell, & Dadds, 2008), and facilitates finding angry faces in crowds (Guastella, Carson, Dadds, Mitchell, & Cox, 2009), judging the trustworthiness of faces (Lambert, Declerck, & Boone, 2014), or selecting suitable partners in intergroup conflict (De Dreu, Greer, Handgraaf, Shalvi, & Van Kleef, 2012).

However, when it comes to predicting the particular social behaviours resulting from intranasal OT administration, experimental research has been contradictory: OT can either decrease (Preckel, Scheele, Kendrick, Maier, & Hurlemann, 2014) or increase social distance (Scheele et al., 2012), encourage defensive behaviours (De Dreu, Scholte, van Winden, & Ridderinkhof, 2015; Striepens et al., 2012) as well as aggression (Campbell, 2008; De Wall et al., 2014), and it can lead to cooperation (Arueti et al., 2013) as well as competition (De Dreu, 2012). Taken together, these studies indicate that OT is a moderator of behaviour and that it can have prosocial as well as antisocial consequences, depending on the context and individual differences (Bartz, Zaki, Bolger, & Ochsner, 2011; Declerck, Boone, & Kiyonari, 2010; Olff et al., 2013).

To fully grasp the metafunctionality of OT, we need a better understanding of *how* context matters, and what the effects of OT are when there are different sources of information competing for attention. Most experimental studies so far have examined the effects of OT on isolated functions such as perception, social memory, or trust, without considering additional motivational processes that drive moment-to-moment judgments and decision making. Human social behaviour is complex, and key decisions – such as whether to cooperate with or aggress against a stranger- depend on at least two different types of evaluations: first, are there appetitive or aversive stimuli that *incentivize*

us to prefer one behaviour over another? And second, are there other salient stimuli that are *informative* about the potential outcome of the behaviour one is about to choose, irrespective of the particular incentives embedded in the decision context. For example, when working in teams, it may be in one's best interest to select a partner with complementary skills (an appetitive stimulus which promises synergy), but in order to prevent being the target of free-riding, it is wise to also consider the potential partner's reputation as a cue of his trustworthiness.

The purpose of the current experiment is to test the proposition that OT helps to integrate these different types of stimuli (incentives and informative social cues) and thereby facilitates ecologically sound decision making in social contexts. Specifically, we propose that OT improves heuristic social decision making, in such a way that we are more likely to select the behaviour (cooperate or aggress) that provides the best match between the incentives and the social cues present in the decision context.

The reasoning behind this proposition is based on OT's interaction with the mesolimbic dopaminergic system (including the ventral tegmental area, amygdala, nucleus accumbens and hippocampus). Recent theories (Love, 2014; Shamay-Tsoory & Abu-Akel, 2015) outline how oxytocinergic neuromodulation of this system can affect *both* incentive motivation *and* the allocation of attention. This is because dopaminergic neurons, originating from two distinct populations in the ventral tegmental area, respond to different properties of stimuli, coding, on the one hand, the *motivational value* of signals in the nucleus accumbens (by strengthening excitatory potentials for appetitive signals, and inhibiting aversive ones) and, on the other hand, the *salience* of signals by activating or attenuating the amygdala (Bromberg-Martin, Matsumoto, & Hikosaka, 2010). Through its neuromodulatory properties in the mesocorticolimbic system, OT could bias decision making by means of a two-step process: it influences incentive motivation by strengthening dopaminergic coding of appetitive signals in the nucleus accumbens (see for example Romero-Fernandez, Borroto-Escuela, Agnati, and Fuxe (2013)); at the same time, OT-dopamine interactions in the amygdala affect how salient cues (having either a positive or a negative valence) become incorporated in the decision process. Amygdala activation has been associated with vigilance, heightening the orienting response to new, threatening, or unexpected sensory information (Davis & Whalen, 2001). By deactivating the amygdala and strengthening the dopaminergic signal, OT has been proposed to facilitate attention re-orienting (Shamay-Tsoory & Abu-Akel, 2015). Because dopamine neurotransmission is primarily involved in focused attention and action readiness (rather than orienting), OT-dopamine interactions in the amygdala may improve information processing by more accurately responding to the valence

of social cue (Tucker & Williamson, 1984). This is a putative mechanism to explain how OT could link the perception of salient social cues to ecologically sound behaviour.

We set up an experiment in which we test if OT can improve the ecological accuracy of complex social decision making through its effects on the nucleus accumbens and amygdala, depending on, respectively, the incentivizing context (appetitive or aversive) and the valence of peripheral social cues (alerting versus neutral cues). To do so, we rely on two economic games to simulate either win-win, or win-lose situations that offer real monetary incentives (see figure 1). In an assurance game (AG) the highest pay-off can be obtained by coordinating with one's partner, making cooperation appetitive. The optimal choice in the AG is to cooperate when you believe your partner will cooperate. You also know that your partner has the same incentive to cooperate, which reduces the need to pay attention to other social cues. In contrast, in the chicken game (CG), known as an anti-coordination game, more money can be earned by choosing an aggressive strategy. Here aggression is appetitive, but it is also aversive because all is lost if the partner also aggresses. Thus in the chicken game, the optimal choice is to aggress if you believe you can outcompete your partner, and to back off otherwise. To manipulate the valence of alerting social cues, pictures of angry or neutral faces were displayed on the screen alongside the pay-off matrix of the games, so as to suggest a threatening or safe decision environment. Participants (n = 29) each made 80 decisions in the fMRI scanner after they received OT or placebo (following a double blind, within-subject factor design), in two experimental sessions scheduled one month apart.

Thus, the main goal of the study is to test how OT affects neuromodulation of the nucleus accumbens and the amygdala, two regions of the mesolimbic dopaminergic system that are crucial with respect to incentive-driven behaviour. With respect to the fMRI data, we formulate two specific hypotheses: First, in the AG (the context that rewards mutual cooperation and hence stimulates a win-win outcome), we hypothesize that OT will increase appetitive motivation, and that this will be associated with increased activity in the nucleus accumbens. Conversely, in the more ambiguous CG (signalling a win-lose outcome), aggression is the most desirable outcome, but it is also very risky. Given this ambivalent decision context, OT is expected to inhibit appetitive behaviour and deactivate the nucleus accumbens, relative to the AG. In this case, decision making can be improved by additionally relying on salient social cues. The second hypothesis is that this occurs through a reduction of the orienting response in the amygdala. We therefore expect that OT will attenuate amygdala activation, and that this should be especially apparent when cues are threatening.

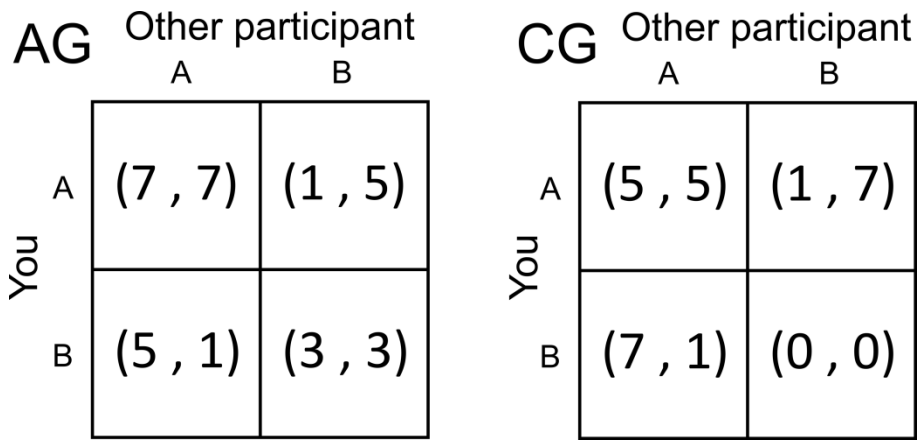

Fig. 1. Example of a pay-off matrix for an assurance game (AG, left panel) and a chicken game (CG, right panel). Each cell depicts the pay-off for yourself and the other participant based on the combination of the chosen options (the first number in the bracket is your outcome, the second one that for the other participant). In an AG, the best outcome is for both to choose A (the cooperative option), yielding 7 euro's each. There is no temptation to earn more by deviating from this choice. This contrasts with the CG, where the highest pay-off is attained when you chose B (the aggressive option) and the other participant choses A (yielding 7 euro's for yourself and 1 for the other). However, if the other also aggresses, both lose everything.

With respect to behaviour, we again formulate two hypotheses. First, by inhibiting the orienting response in the amygdala, OT would allow for more accurate encoding of the valence of the cues. This would lead to cue-appropriate behaviour in the CG: if the partner looks too angry, the best decision rule is to back off, especially if losing is fatal. However, if the partner appears neutral and there is a monetary incentive to win, aggressing is the better course of action. This is not expected in the AG where the strong appetitive motivation makes the valence of social cues irrelevant. However, the increased appetitive motivation (with OT) does not necessarily have to translate into increased cooperative behavior, because decision making in the AG is still not without risk. In a large-scale experiment (Declerck et al., 2010) intranasal OT administration boosted cooperative decisions (relative to placebo) in an AG, but only if participants had met before and already established contact. In the current experiment, we do not manipulate previous contact, but it is still possible that the effect of OT in an AG would depend on individual differences. Already we know that OT increases cooperation especially for anxious individuals (De Dreu, 2012), and trust for cautious individuals (Lambert et al., 2014). Therefore, the second behavioral hypothesis is that OT will increase cooperative decisions in the AG especially for cautious, or risk averse, individuals.

2. Methods

2.1 Participants

Thirty female students (average age = 24 year; S.D. = 2.8 years) were recruited via e-mail invitation, in which the study was described as an investigation of the effect of a synthetic hormone on economic decision making. This sample size was selected out of practical considerations and is similar to that of other OT-fMRI studies (Scheele et al., 2012; Scheele et al., 2013; Strathearn et al., 2009). We limit this study to female participants to reduce noise that might result from OT-gender interactions. All participants were right-handed (assessed with a validated questionnaire (Oldfield, 1971)), heterosexual, and were screened to make sure they met all the safety criteria for MRI examination and OT administration. They attended two experimental sessions during which OT or placebo were administered following a within-subject design, with the order of the treatments randomly determined for each participant. The sessions were scheduled 28 days apart (or a multiple thereof), to increase the likelihood that the participant was twice in the same phase of her menstrual cycle. Data from one participant who did not attend both sessions were removed from subsequent analyses. All procedures were approved by the University's commission of ethics. Debriefing occurred at the conclusion of the study by contacting participants by email and referring them to a website where the intent, results, and procedures of the experiment were fully explained.

2.2 Procedures

Upon arrival at the imaging centre, participants signed an informed consent document (first session only). We checked with a questionnaire that none of the participants suffered from a common cold and that they had not used any interfering substances (nicotine, alcohol for 12 hours prior to the experiment, or other drugs (other than anti-conception)) for 24 hours prior to the experiment. Next they participated in a lottery game to assess their level of risk aversion (first session only). The lottery was adapted from the Eckel-Grossman Risk Task and consisted of choosing one of six possible lotteries that varied in their stakes and probability of winning (Dave, Eckel, Johnson, & Rojas, 2010). Participants were told they could earn between € 0.20 and € 7.00. At the end of the session, the lottery was carried out by coin flip and the participants were payed accordingly. Average earnings from this lottery were € 3.46.

Participants self-administrated a nasal spray under supervision of an assistant: three puffs per nostril with one minute in-between puffs. Following a double-blind random design, neither the assistant, nor the participant knew if the spray contained the placebo or 24 IU OT (Syntocinon, Novartis). The placebo was prepared by the hospital pharmacist. The procedures and timing of OT administration

were the same as in previous studies using intranasal OT administration (Declerck et al., 2010; Lambert et al., 2014).

Participants waited 35 minutes before the fMRI scan (Baumgartner, Heinrichs, Vonlanthen, Fischbacher, & Fehr, 2008; Kirsch et al., 2005) during which time they received instructions on the social dilemma games (see appendix). They were told that they were playing with real partners whose responses had been recorded in advance. The points they earned in each trial (based on the combination of their and the matched partner's choices) were to be exchanged for money at the end of the experiment (1 point = € 0.01; average earnings per session: € 32.85 + a show-up fee of € 10.00). To make sure all participants understood the games and the consequences of their choices, they answered a series of control questions before starting the actual experiment.

While in the scanner, participants played 80 social dilemma games, 40 AG and 40 CG, grouped in 4 rounds of 20 games with a small pause in between each round (60s). Appendix figure 1 shows a trial as used in the experiment. Each round comprised two blocks of five randomly ordered games of the same type (AG or CG). To avoid boredom, 20 different matrices (10 AG and 10 CG) were used. To avoid ceiling effects and to make sure the different matrices elicited sufficient variation in behavioural responses, we selected them based on a pilot study.

Each game matrix was accompanied by a picture of either a neutral (NE) or angry (AN) face (the social cues, see Appendix figure 2). There were a total of 30 different angry and 32 different neutral faces, shown in random order with the restriction that there were an equal number of neutral and angry faces per round and that there were no more than three faces of the same expression displayed in consecution. The pictures were drawn from the Karolinska Directed Emotional Faces database and pretested to make sure that anger and neutral expressions were unambiguous.

The decision (option A or B) for each of the 80 matrices was recorded by pressing a response button held in the right hand (Lumia model LU400, Cedrus, CA, USA). Stimulus presentation and response logging was conducted with the Presentation® software (Neurobehavioral Systems, Inc, Albany, CA, USA).

2.3 Image acquisition and analysis

Images were collected with a 3 Tesla Siemens Trio scanner and an 8-channel head coil (for details and pre-processing, see appendix).

Two general linear models (GLMs) were created per participant (one for each session). The blood oxygen level dependent (BOLD) signal was the dependent variable. The event of interest was the

decision phase, defined as the time interval between the appearance of the slide depicting the game matrix and the participant's response (Appendix figure 1). We created 4 regressors, one for each of the possible combinations of stimuli (CG and angry face; CG and neutral face; AG and angry face; AG and neutral face) and convolved them with the haemodynamic response function. Six movement parameters were added to account for head movement in six dimensions.

For each contrast we computed beta-values (averaged over the regions of interest) for each person to represent the relative effect of a particular contrast on the BOLD signal in that region. Region of interest (ROI) analysis is a useful tool to discern patterns of activity in complex designs with multiple factors. This way we limit statistical testing to regions that are functionally defined based on prior knowledge. By furthermore computing the average beta-value over the entire region, we avoid the problems associated with multiple, voxelwise comparisons (Poldrack, 2007). We compared with two-tailed t-tests if these average beta-values differ from zero and if they differ between the OT and placebo condition. Regions of interests were defined based on Hammers et al. (2003) maximum probability atlas of the human brain.

3. Results

3.1 fMRI analysis

Before testing hypotheses, we conduct an exploratory whole brain analysis. We look at (i) the main effect of OT in all trials (contrasts: $Trial_{[OT>P]}$-$Baseline_{[OT>P]}$ and $Trial_{[OT<P]}$-$Baseline_{[OT<P]}$), (ii) the interaction effect of OT with type of game ($AG_{[OT-P]}$-$CG_{[OT-P]}$) and (iii) the interaction with the type of cue, ($AN_{[P-OT]}$-$NE_{[P-OT]}$). No voxels survived FWE correction (see Appendix table 1), and here we report only the results of subsequent region of interest (ROI) analyses conducted on the left and right nucleus accumbens and amygdala to test the hypotheses set forth in the introduction. These ROI analyses are conducted by comparing the estimators of the *average* brain activation levels (β's) between conditions. Averaging β over an entire region is justified (and the preferred method) because we study sets of functionally dependent voxels (defined as ROI's by probabilistic atlases of macroscopic anatomy) for which we have a-priori hypotheses with respect to their level of activation between conditions (Poldrack, 2007).

To the test if OT affects appetitive motivation in the nucleus accumbens, we compare the effect of OT versus placebo in the contrast AG – CG. Figure 2 shows that there is no difference in nucleus accumbens activation level between the two games in the placebo condition (β's do not differ from zero), and that OT increases nucleus accumbens activation level in the AG relative to the CG (suggesting that approach behaviour is incentivized in the AG). This is statistically significant in the

left hemisphere (p = 0.026, right hemisphere p = 0.054; two-tailed t-test). Furthermore, in the right hemisphere, β in the OT condition is significantly greater than β in the placebo condition (right hemisphere p = 0.014, left hemisphere p = 0.094; two-tailed t-test). Given that we know from a meta-analysis covering 206 studies that the nucleus accumbens is involved in encoding the value of a reward and linking it to the decision outcome (Bartra, McGuire, & Kable, 2013), these results suggest that OT facilitates the encoding of appetitive signals (embedded in the AG) by linking them to the expectation of reward.

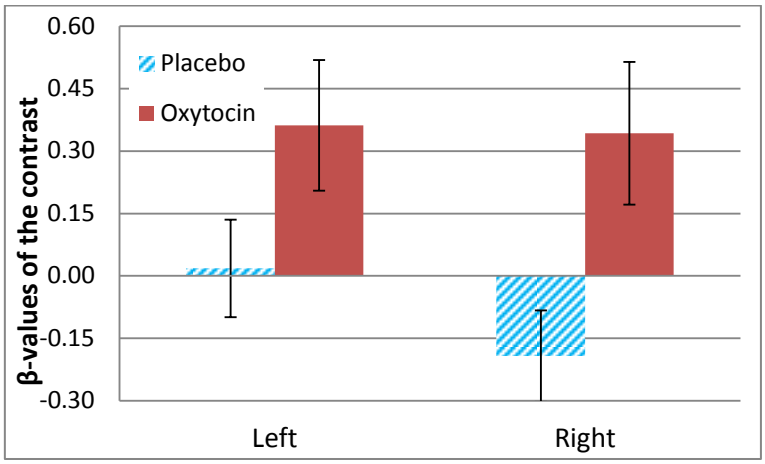

Fig. 2. β-values of the contrast AG-CG on nucleus accumbens activity for the placebo and OT condition. Error bars represent standard error of the mean, N = 26.

Next we investigated if OT inhibits the orienting response in the amygdala, which we presume would heighten focused attention to the cue (the second fMRI hypothesis). We first test if OT has a general anxiolytic effect on the amygdala in all trials (contrasted to baseline, see Figure 3). As expected, β in the right amygdala is indeed significantly reduced in the OT condition compared to the placebo condition (left p = 0.19, right p = 0.049; two-tailed t-test). Note that all β's in this contrast are negative, which is consistent with the proposition that the increased attention needed to perform the experimental tasks is associated with a decrease in amygdala activation. The data show that this is substantially more pronounced in the OT condition. To test if OT reduces the orienting response we furthermore test its effect on the contrast AN – NE. Figure 4 shows that, as expected, angry faces elicit an orienting response in the placebo condition, shown by an increase in amygdala activation relative to neutral eyes. This trend is apparent in the right hemisphere and significant in the left one (left p = 0.022, right p = 0.096; two-tailed t-test). OT attenuates this response, with β's that no longer differ from 0 (left p = 0.93 right p = 0.65; difference between β placebo and β OT: left p = 0.093, right p = 0.287; two-tailed t-test). These data are consistent with the well-known anxiogenic response of the amygdala to threatening cues (Kirsch et al., 2005).

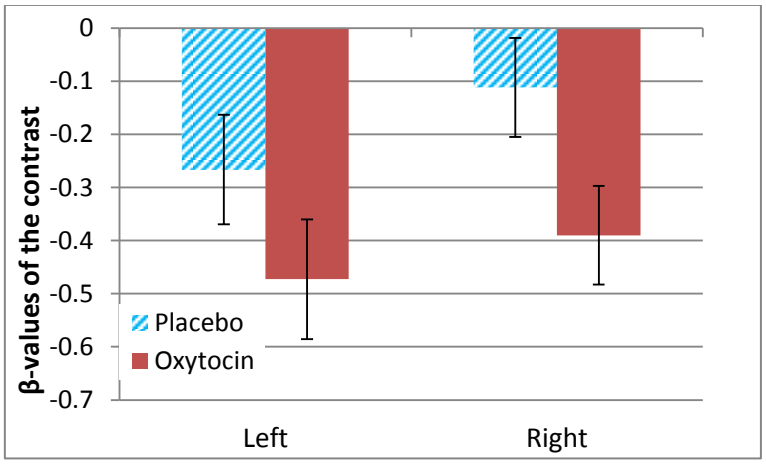

Fig. 3. β-values of the contrast trial-baseline on amygdala activity for the placebo and OT condition. Error bars represent the standard error of the mean, N = 26.

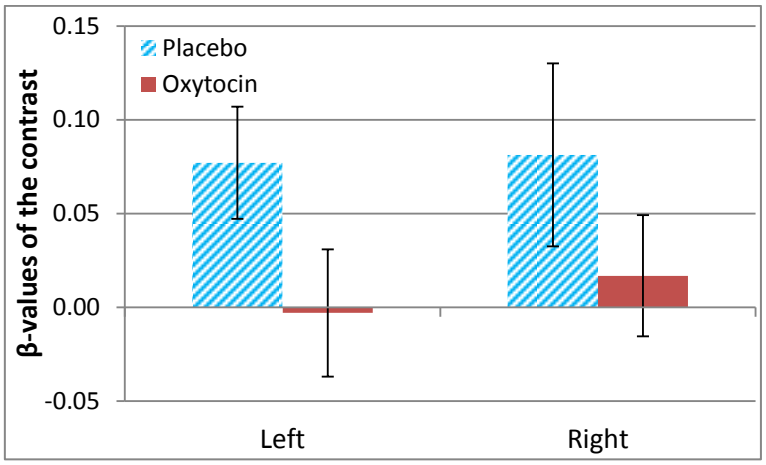

Fig. 4. β-values of the contrast AN-NE on amygdala activity for the placebo and OT condition. Error bars represent the standard error of the mean, n = 26.

3.2 Behavioural data

Next we investigate how these changes in brain activation correspond to changes in behaviour. We test how OT and social cues (AN versus NE) affect behavioural choices in the AG and the CG separately, using logistic regressions with random intercept models in STATA (xtlogit command). This analysis allows one to take unobserved heterogeneity among individuals into account, assuming that these random intercepts are uncorrelated with the effects of the measured independent variables. We use panel data of 80 trials clustered on 29 participants, and we report the effect sizes of the coefficients (odds ratio's), robust standard errors clustered on participants, and the p-values of two-tailed significance tests (table 1). The dependent variable is the decision in each trial (the cooperative decision in an AG or the aggressive decision in a CG = 1; 0 otherwise). The independent variables are the cues that were displayed on the decision screen (neutral faces = 1; angry faces = 0), a risk

aversion score (0 = not risk averse, 5 = very risk averse), and whether the participant had inhaled OT (= 1) or a placebo (= 0) for any particular trial. Risk aversion is added to the models because we know that behaviour in the social dilemma games is highly affected by the willingness to take risks (Cabon-Dhersin & Etchart-Vincent, 2012; Ng & Au, 2016), and that also the effect of OT is dependent on individual differences in caution (Declerck, Boone, & Kiyonari, 2013).

Table 1. Random effects logistic regression using panel data clustered on 29 individuals.

	Assurance game				**Chicken game**			
	OR	p	OR	p	OR	p	OR	p
OT	0.99 (0.12)	0.900	0.18 (0.08)	<0.001	1.17 (0.11)	0.102	0.70 (0.24)	0.310
Cue	1.38 (0.16)	0.007	1.31 (0.22)	0.107	1.57 (0.15)	<0.001	1.31 (0.18)	0.052
Risk Aversion	1.08 (0.46)	0.848	0.86 (0.37)	0.731	0.56 (0.12)	0.006	0.53 (0.12)	0.004
OT*Cue			1.11 (0.26)	0.666			1.44 (0.28)	0.058
OT*Risk Avers.			1.56 (0.18)	<0.001			1.09 (0.09)	0.315
Constant	3.48 (5.82)	0.455	8.71 (14.89)	0.206	5.69 (4.85)	0.041	7.38 (6.43)	0.022
Wald chi^2	7.43		22.62		31.61		36.32	
N	29 (2319)		29 (2319)		29 (2315)		29 (2315)	

In the assurance game, the dependent variable is the cooperative choice, while in the chicken game, the dependent variable is the aggressive choice. Odds ratios (OR) are reported with robust standard error in parentheses. For N, the number in parentheses denotes the decisions made by all 29 individuals.

From table 1 we can infer that, in the AG, the game where the best outcome can be achieved by increased appetitive motivation, OT does not exert a main effect on behaviour (odds ratio OR = 0.99, SE = 0.12, p = 0.900). The likelihood of cooperation in the AG does not depend on how risk averse the participant is (OR = 1.08, SE = 0.46, p = 0.848). The type of cue, however, does exert a significant effect, with neutral cues increasing the chance that a participant will cooperate (OR = 1.38, SE = 0.16, p = 0.007). There is no interactive effect between OT and cue (OR = 1.12, SE = 0.26, p = 0.666), but consistent with the hypothesis set forth in the introduction, we find that OT does interact significantly with risk aversion (OR = 1.56, SE = 0.18, $p < 0.001$). This significance level is well below the Bonferroni adjusted corrected p-value for multiple comparisons, which, with 5 tests (see table 1), would yield an $\alpha = 0.05/5$ or 0.01. To examine the interaction, we have plotted the proportion of cooperation in Figure 5. Note that the effect of OT is opposite for low versus high risk aversion. Figure 6 shows the marginal effects of OT on cooperative behaviour, illustrating that this effect increases with risk aversion. The finding that OT does not have a main effect in the AG, but that its effect is contingent on risk aversion, is consistent with the results of previous studies that indicate

that OT's effect on cooperative behaviour is dependent on context- and/or individual differences (Bartz et al., 2011; Declerck et al., 2010).

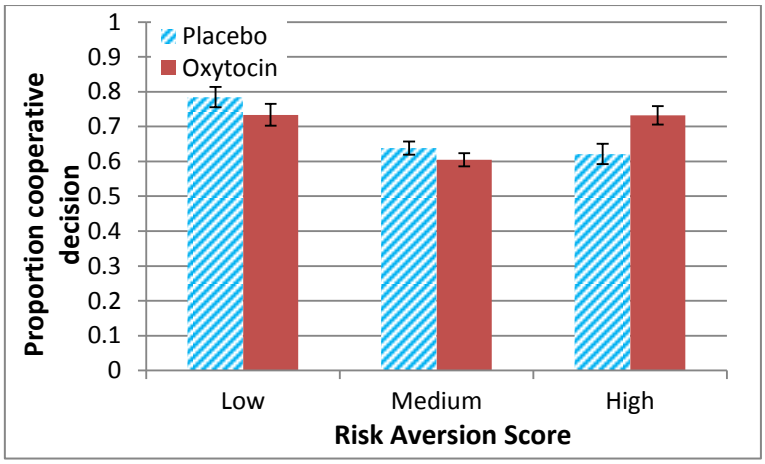

Fig. 5. Proportion of cooperative decisions in an assurance game. Error bars represent the standard error of the mean, N = 29.

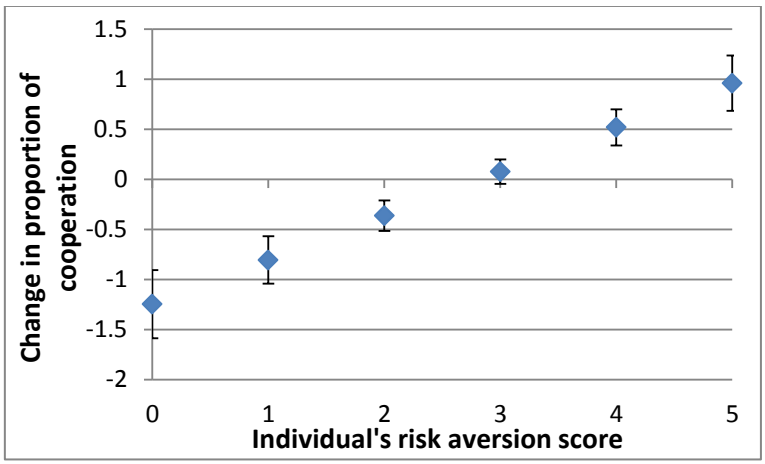

Fig. 6. The marginal effect of OT on cooperation in an AG depending on an individual's risk aversion score. Error bars represent the standard error, N = 2319 decisions, clustered on 29 participants.

Table 1 also shows the results for the CG. Again, there is no main effect of OT on aggressive decisions (OR = 1.17, SE = 0.11, p = 0.102). The presence of a neutral cue (rather than an angry one) increases aggression (OR = 1.57, SE = 0.15 , $p < 0.001$). Risk aversion significantly decreases the likelihood of an aggressive decision (OR = 0.56, SE = 0.12, p = 0.006), but it does not interact with OT (OR = 1.09, SE = 0.09, p = 0.315). Finally, we note that our a-priori hypothesis, namely that reaching an aggressive decision in the CG depends on the interaction of OT and the type of cue, holds at the p = 0.058 level (SE = 0.28), which exceeds the conventional threshold for significance at $\alpha = 0.05$. To examine the interaction, we conducted post-hoc logistic regressions in the CG for angry and neutral faces separately. The odds ratios reveal that participants who received OT relative to placebo are 0.33%

(SE = 0.14, p = 0.826) less likely to aggress when confronted with angry faces, and 40% (SE = 0.19, p = 0.012) more likely to aggress when confronted with neutral faces. Figure 7 illustrates this interaction.

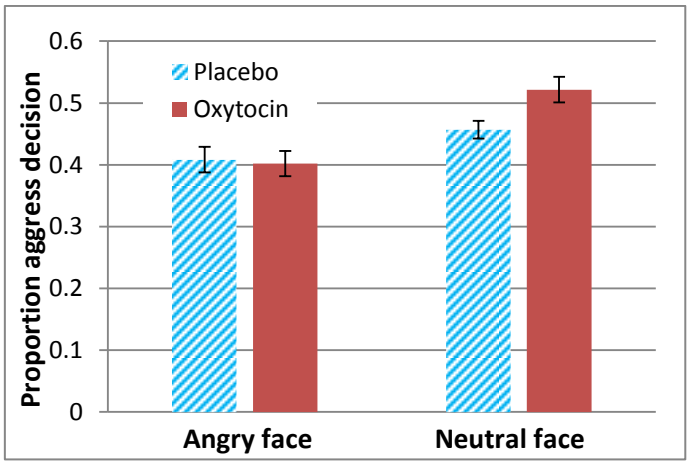

Fig. 7. Proportion of aggress decisions in a chicken game. Error bars represent the standard error of the mean, N = 29.

Because the effect of OT might be influenced by oestrogen levels, we repeated these analyses but included a dummy variable indicating whether the participant used hormonal anticonception (n = 21). This yields very similar odds ratios (OR) and roughly the same level of significance for all effects tested (see Appendix table 2).

To summarize, combining the fMRI data with the behavioural results suggest that the effect of OT in the nucleus accumbens fit with the a priori hypothesis that OT increases appetitive motivation in the AG by activating the nucleus accumbens. This helps risk averse individuals to increase their levels of cooperation in the AG in order to achieve the most desirable outcome (i.e. the outcome with the greatest pay-off). The data also supports the hypothesis that OT decreases the orienting response in the amygdala. We find that OT significantly attenuates amygdala activation in all trials, regardless of incentives or cues. We also see that OT abolishes the orienting response in the amygdala activation in response to angry faces: the significant increase in amygdala activation elicited in the placebo condition is eliminated. In fact, in the OT condition the amygdala responds to neither angry, nor neutral trials, suggesting that the amygdala does not interfere with processing the valence of the cue at a cortical level. We propose that this attenuated amygdala response allows the valence of the cue to be more accurately interpreted in those decision contexts where cues matter. Consistently, we see that OT facilitates the aggressive response in the CG (the response with the best pay-off), on condition that partner faces are neutral. In the AG, the increase in appetitive motivation overrules the need to pay attention to cues.

4. Discussion

Previous research on the effects of intranasal OT administration has spurred several hypotheses with respect to its role in regulating social behaviour: OT encourages social approach (Kemp & Guastella, 2011), it enhances the saliency of social cues (Shamay-Tsoory & Abu-Akel, 2015), and it is anxiolytic (Heinrichs et al., 2009; Neumann & Landgraf, 2012). Starting from a neurophysiological premise that OT increases mesolimbic dopaminergic activity in the amygdala and nucleus accumbens, we set up a multi-factorial experiment which allows us to disentangle some of the boundary conditions that determine when each of these three hypotheses is most likely to hold, thereby trying to making sense of some of the inconsistent results that have recently clouded the OT literature.

Consistent with the *social approach* hypothesis, the data show that OT enhances nucleus accumbens activity in a decision context that is conducive to cooperation because the greatest reward can be obtained by coordinating with one's partner in an AG. Because this decision is not without risk, the marginal effect of OT on approach behaviour is greater for risk averse individuals. This is in line with previous research showing that OT facilitates cooperative decision making in an AG, but that it may depend on other factors, such as having met the partner before (Declerck et al., 2010), or individual differences (De Dreu, 2012; Declerck et al., 2013). The data are also consistent with the proposition (based on animal studies) that OT may strengthen the dopaminergically driven incentive motivation in the nucleus accumbens (Love, 2014; Shamay-Tsoory & Abu-Akel, 2015).

When the context is ambivalent (in a CG), OT activation of the nucleus accumbens does not occur. In that case, behaviour may be improved through neuromodulation of the amygdala, the site where vigilance and attention re-orientation takes place (Davis & Whalen, 2001). If OT is to *enhance the saliency of social cues,* the data we present suggest that it may do so by a *general* anxiolytic effect, whereby also the orienting response is attenuated (i.e. OT counteracts the increased amygdala activation in response to angry faces), thereby shifting attention to interpreting the valence of the cue.

We note that not all previous studies report decreased amygdala activation following intranasal OT administration. Lischke et al. (2012) report increased amygdala for females confronted with angry faces under the fMRI. However, the experimental paradigm is very different, in that the females in the study by Lischke et al. (2012) were only asked to *rate* the valence of the faces, *without repercussion on actual decision making.* In the current study, the valence of the facial expression matters with respect to the decision that needs to be made. If OT strengthens the dopaminergic signal (Love, 2014; Shamay-Tsoory & Abu-Akel, 2015), thereby inducing a shift from vigilance to focused attention and motor readiness (Belujon & Grace, 2015; Tucker & Williamson, 1984) then the

valence of the facial expression needs to be decoupled from the emotion it elicits, as valence becomes a computational component in a decision process where the outcome also depends on momentary incentives. This may explain why, we see that, in the placebo condition, social cues elicit significant changes in amygdala activation with little change in behaviour, while in the OT condition, the attenuated amygdala response is accompanied by increased aggression when cues are neutral, but not when they are threatening.

Improved accuracy in the interpretation of salient social cues has also been shown in previous research: after OT administration, participants are more rational when they couple reward (in the form of a monetary pay-off) with faces (Evans, Shergill, & Averbeck, 2010), and they are more able to accurately rate the likability of faces, regardless of whether or not the participants had previously been conditioned to fear them (Petrovic, Kalisch, Singer, & Dolan, 2008). Together with the results of the current study (showing significant *overall* reduced amygdala activation in the OT condition), the improved interpretation of salient social cues appears to be the result of a generalized *anxiolytic effect* that removes the automatic response to an unexpected and potentially fearful stimulus.

The proposition that OT improves information processing in the brain by a generalized suppression of fear responses has been tested in the rat hippocampus. Here oxytocin agonists were found to selectively increase activity in fast-spiking interneurons which, at the same time, reduced background noise. The result was an overall inhibitory effect of OT on the hippocampus which made for a greater signal-to-noise ratio and hence a better information transfer (Owen et al., 2013). As the hippocampus is also known to play an important role in human fear conditioning (Alvarez, Biggs, Chen, Pine, & Grillon, 2008), we tested if it was affected by OT in the current study. This additional analysis reveals an activation pattern in the hippocampus that is similar to the amygdala (see Appendix figure 3). In all trials (relative to baseline), OT significantly deactivates the hippocampus (left hemisphere $p = 0.048$, right hemisphere, $p = 0.043$; two-tailed t-test), while this is not the case in the placebo condition. The difference in β's between the OT and placebo conditions do not, however, differ statistically significantly, and these results should therefore be treated as preliminary and meriting additional research.

Combining the behavioural findings with, on the one hand the excitatory effects of OT in the nucleus accumbens, and on the other hand the reports of inhibitory effects of OT in the hippocampus and amygdala, suggests a meta-functionality: rather than having a direct role in improving social cognition, OT neuromodulation in the dopaminergic mesolimbic system allows for a better integration of multiple sources of information, leading to ecologically sound decision making. By matching the incentive value of an ongoing social interaction with the valence of a facial cue, choice

behaviour becomes more fine-tuned. Hence OT would improve social approach, reduce anxiety, and heighten the saliency of social cues when the decision environment demands that incentives and social information become integrated.

If one of the functions of OT neuromodulation is to facilitate the integration of perceptual and context-related cues, then administering exogenous OT is not expected to lead to a one-to-one relation between cue and behaviour, and the contradictory findings that have been reported in the literature make more sense. OT has been shown to heighten attention towards happy faces (Domes et al., 2013; Guastella, Mitchell, & Mathews, 2008; Marsh, Yu, Pine, & Blair, 2010), but also to angry (Guastella et al., 2009) and untrustworthy ones (Lambert et al., 2014). Based on the results of the current study we propose that these different effects of OT are due to an experimental design where the perception of cues is decoupled from decision-making. When participants are not motivated by an expected decision outcome, amygdala modulation may still occur independently from activity in the nucleus accumbens. OT in this case may enhance the perceptual sensitivity to salient cues, even if they have no ecological relevance.

Other research with exogenous OT has investigated the effects of altering the decision context without providing additional perceptual social cues (Baumgartner et al., 2008; De Dreu et al., 2015). These experiments report reduced amygdala activation associated with an increase in prosocial response. De Dreu et al. (2015) and Baumgartner et al. (2008) show that the anxiolytic effect of OT in these cases increased social approach irrespective of whether or not it was the best strategy, which may be due to an experimental design lacking perceptual information.

We note that intranasal OT administration studies have recently been under a lot of scrutiny. A first criticism is that OT rarely has a main effect on behaviour, but that it often interacts with one or more moderators (Lane, Luminet, Nave, & Mikolajczak, 2016; Nave, Camerer, & McCullough, 2015). The current study corroborates this, but this should not pose a problem if it turns out that the role of OT neurotransmission truly lies in facilitating the integration of different sources of information, and that it does so by enhancing dopaminergic activity. More research is needed to test this hypothesis, described earlier by Love (2014) and Shamay-Tsoory and Abu-Akel (2015). This study is a first step in that direction. While we realize that the sample size of the current study is small and the statistical power limited, the statistically significant results reported here offer promising insights into the mechanism by which OT may exert a meta-function in regulating behaviour.

A second, more important problem with OT administration studies is that there exists no conclusive theory on how intranasal OT reaches the brain (Leng & Ludwig, 2015). From the nasal cavity, there are at least two possible paths OT could take, as described by Quintana, Alvares, Hickie, and

Guastella (2015). First, intranasal OT can be absorbed into systematic circulation via the blood capillaries permeating the membranes of the nasal cavity. While OT does not cross the blood-brain barrier easily, elevated plasma OT can lead to supraphysiological doses in the cerebrospinal fluid (CSF) of animals (Kang & Park, 2000; Mens, Witter, & Van Wimersma Greidanus, 1983) and recent studies have also shown that intranasally administered OT can affect the CSF concentration in both animals (Rault, 2016) and humans (Striepens et al., 2013). The human studies have hardly been replicated, however, and a more likely path by which intranasal OT might reach the brain is via the olfactory sensory neurons to the olfactory bulb, which projects to the amygdala and hippocampus. A recent study (Quintana et al., 2016), comparing the effects of intravenous and intranasal OT found, similar to the current study, a decrease in amygdala activation in response to angry faces, but only in the intranasal OT condition, implying that a neural path from the nasal cavity to the brain is more likely than a systemic one.

5. Conclusions

In summary, the current study investigates the effect of intranasal OT in an experimental paradigm that combines an incentivized decision making environment with the perception of salient social cues. The behavioural and fMRI results suggest a possible meta-function of OT: by enhancing incentive motivation in the nucleus accumbens, and attenuating the orienting response in the amygdala, OT improves coordination in a cooperative decision environment, while it allows for a more deliberate interpretation of salient cues in a competitive decision environment. This in turn allows for ecologically sound choices that maximize return without compromising safety.

References

Alvarez, R. P., Biggs, A., Chen, G., Pine, D. S., & Grillon, C. (2008). Contextual fear conditioning in humans: Cortical-hippocampal and amygdala contributions. *Journal of Neuroscience, 28*(24), 6211-6219. doi: 10.1523/jneurosci.1246-08.2008

Arueti, M., Perach-Barzilay, N., Tsoory, M. M., Berger, B., Getter, N., & Shamay-Tsoory, S. G. (2013). When two become one: The role of oxytocin in interpersonal coordination and cooperation. *Journal of Cognitive Neuroscience, 25*(9), 1418-1427. doi: 10.1162/jocn_a_00400

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage, 76*, 412-427. doi: 10.1016/j.neuroimage.2013.02.063

Bartz, J. A., Zaki, J., Bolger, N., & Ochsner, K. N. (2011). Social effects of oxytocin in humans: Context and person matter. *Trends in Cognitive Sciences, 15*(7), 301-309. doi: 10.1016/j.tics.2011.05.002

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron, 58*(4), 639-650. doi: 10.1016/j.neuron.2008.04.009

Belujon, P., & Grace, A. A. (2015). Regulation of dopamine system responsivity and its adaptive and pathological response to stress. *Proceedings of the Royal Society B-Biological Sciences, 282*(1805). doi: 10.1098/rspb.2014.2516

Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010). Dopamine in motivational control: Rewarding, aversive, and alerting. *Neuron, 68*(5), 815-834. doi: 10.1016/j.neuron.2010.11.022

Cabon-Dhersin, M.-L., & Etchart-Vincent, N. (2012). The puzzle of cooperation in a game of chicken: an experimental study. *Theory and Decision, 72*(1), 65-87. doi: 10.1007/s11238-010-9220-9

Campbell, A. (2008). Attachment, aggression and affiliation: The role of oxytocin in female social behavior. *Biological Psychology, 77*(1), 1-10. doi: 10.1016/j.biopsycho.2007.09.001

Dave, C., Eckel, C., Johnson, C., & Rojas, C. (2010). Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty, 41*(3), 219-243. doi: 10.1007/s11166-010-9103-z

Davis, M., & Whalen, P. J. (2001). The amygdala: Vigilance and emotion. *Molecular Psychiatry, 6*(1), 13-34. doi: 10.1038/sj.mp.4000812

De Dreu, C. K. W. (2012). Oxytocin modulates cooperation within and competition between groups: An integrative review and research agenda. *Hormones and Behavior, 61*(3), 419-428. doi: 10.1016/j.yhbeh.2011.12.009

De Dreu, C. K. W., Greer, L. L., Handgraaf, M. J. J., Shalvi, S., & Van Kleef, G. A. (2012). Oxytocin modulates selection of allies in intergroup conflict. *Proceedings of the Royal Society B-Biological Sciences, 279*(1731), 1150-1154. doi: 10.1098/rspb.2011.1444

De Dreu, C. K. W., Scholte, H. S., van Winden, F., & Ridderinkhof, K. R. (2015). Oxytocin tempers calculated greed but not impulsive defense in predator-prey contests. *Social Cognitive and Affective Neuroscience, 10*(5), 721-728. doi: 10.1093/scan/nsu109

De Wall, C. N., Gillath, O., Pressman, S. D., Black, L. L., Bartz, J. A., Moskovitz, J., & Stetler, D. A. (2014). When the love hormone leads to violence: Oxytocin increases intimate partner violence inclinations among high trait aggressive people. *Social Psychological and Personality Science, 5*(6), 691-697. doi: 10.1177/1948550613516876

Declerck, C. H., Boone, C., & Kiyonari, T. (2010). Oxytocin and cooperation under conditions of uncertainty: The modulating role of incentives and social information. *Hormones and Behavior, 57*(3), 368-374. doi: 10.1016/j.yhbeh.2010.01.006

Declerck, C. H., Boone, C., & Kiyonari, T. (2013). The effect of oxytocin on cooperation in a prisoner's dilemma depends on the social context and a person's social value orientation. *Social Cognitive and Affective Neuroscience*, DOI: 10.1093/scan/nst1040.

Depue, R. A., & Morrone-Strupinsky, J. V. (2005). A neurobehavioral model of affiliative bonding: implications for conceptualizing a human trait of affiliation. *Behavioral and Brain Sciences, 28*(03), 313-350. doi: 10.1017/S0140525X05000063

Dolen, G., Darvishzadeh, A., Huang, K. W., & Malenka, R. C. (2013). Social reward requires coordinated activity of nucleus accumbens oxytocin and serotonin. *Nature, 501*(7466), 179-184. doi: 10.1038/nature12518

Domes, G., Heinrichs, M., Glaescher, J., Buechel, C., Braus, D. F., & Herpertz, S. C. (2007). Oxytocin attenuates amygdala responses to emotional faces regardless of valence. *Biological Psychiatry, 62*(10), 1187-1190. doi: 10.1016/j.biopsych.2007.03.025

Domes, G., Heinrichs, M., Michel, A., Berger, C., & Herpertz, S. C. (2007). Oxytocin improves "mind-reading" in humans. *Biological Psychiatry, 61*(6), 731-733. doi: 10.1016/j.biopsych.2006.07.015

Domes, G., Sibold, M., Schulze, L., Lischke, A., Herpertz, S. C., & Heinrichs, M. (2013). Intranasal oxytocin increases covert attention to positive social cues. *Psychological Medicine, 43*(08), 1747-1753. doi: 10.1017/S0033291712002565

Evans, S., Shergill, S. S., & Averbeck, B. B. (2010). Oxytocin decreases aversion to angry faces in an associative learning task. *Neuropsychopharmacology, 35*(13), 2502-2509. doi: 10.1038/npp.2010.110

Guastella, A. J., Carson, D. S., Dadds, M. R., Mitchell, P. B., & Cox, R. E. (2009). Does oxytocin influence the early detection of angry and happy faces? *Psychoneuroendocrinology, 34*(2), 220-225. doi: 10.1016/j.psyneuen.2008.09.001

Guastella, A. J., Mitchell, P. B., & Dadds, M. R. (2008). Oxytocin increases gaze to the eye region of human faces. *Biological Psychiatry, 63*(1), 3-5. doi: 10.1016/j.biopsych.2007.06.026

Guastella, A. J., Mitchell, P. B., & Mathews, F. (2008). Oxytocin enhances the encoding of positive social memories in humans. *Biological Psychiatry, 64*(3), 256-258. doi: 10.1016/j.biopsych.2008.02.008

Hammers, A., Allom, R., Koepp, M. J., Free, S. L., Myers, R., Lemieux, L., . . . Duncan, J. S. (2003). Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Human Brain Mapping, 19*(4), 224-247. doi: 10.1002/hbm.10123

Heinrichs, M., von Dawans, B., & Domes, G. (2009). Oxytocin, vasopressin, and human social behavior. *Frontiers in Neuroendocrinology, 30*(4), 548-557. doi: 10.1016/j.yfrne.2009.05.005

Kang, Y.-S., & Park, J.-H. (2000). Brain uptake and the analgesic effect of oxytocin— Its usefulness as an analgesic agent. *Archives of Pharmacal Research, 23*(4), 391. doi: 10.1007/bf02975453

Kemp, A. H., & Guastella, A. J. (2011). The role of oxytocin in human affect: A novel hypothesis. *Current Directions in Psychological Science, 20*(4), 222-231. doi: 10.1177/0963721411417547

Kirsch, P., Esslinger, C., Chen, Q., Mier, D., Lis, S., Siddhanti, S., . . . Meyer-Lindenberg, A. (2005). Oxytocin modulates neural circuitry for social cognition and fear in humans. *The Journal of Neuroscience, 25*(49), 11489-11493. doi: 10.1523/jneurosci.3984-05.2005

Labuschagne, I., Phan, K. L., Wood, A., Angstadt, M., Chua, P., Heinrichs, M., . . . Nathan, P. J. (2010). Oxytocin attenuates amygdala reactivity to fear in generalized social anxiety disorder. *Neuropsychopharmacology, 35*(12), 2403-2413. doi: 10.1038/npp.2010.123

Lambert, B., Declerck, C. H., & Boone, C. (2014). Oxytocin does not make a face appear more trustworthy but improves the accuracy of trustworthiness judgments. *Psychoneuroendocrinology, 40*(0), 60-68. doi: 10.1016/j.psyneuen.2013.10.015

Lane, A., Luminet, O., Nave, G., & Mikolajczak, M. (2016). Is there a publication bias in behavioural intranasal oxytocin research on humans? Opening the file drawer of one laboratory. *Journal of Neuroendocrinology, 28*(4). doi: 10.1111/jne.12384

Leng, G., & Ludwig, M. (2015). Intranasal oxytocin: Myths and delusions. *Biological Psychiatry*. doi: 10.1016/j.biopsych.2015.05.003

Lischke, A., Gamer, M., Berger, C., Grossmann, A., Hauenstein, K., Heinrichs, M., . . . Domes, G. (2012). Oxytocin increases amygdala reactivity to threatening scenes in females. *Psychoneuroendocrinology, 37*(9), 1431-1438. doi: 10.1016/j.psyneuen.2012.01.011

Love, T. M. (2014). Oxytocin, motivation and the role of dopamine. *Pharmacology, Biochemistry, and Behavior, 119*, 49-60. doi: 10.1016/j.pbb.2013.06.011

Marsh, A. A., Yu, H. H., Pine, D. S., & Blair, R. J. R. (2010). Oxytocin improves specific recognition of positive facial expressions. *Psychopharmacology, 209*(3), 225-232. doi: 10.1007/s00213-010-1780-4

Mens, W. B. J., Witter, A., & Van Wimersma Greidanus, T. B. (1983). Penetration of neurohypophyseal hormones from plasma into cerebrospinal fluid (CSF): Half-times of disappearance of these neuropeptides from CSF. *Brain Research, 262*(1), 143-149. doi: 10.1016/0006-8993(83)90478-X

Nave, G., Camerer, C., & McCullough, M. (2015). Does oxytocin increase trust in humans? A critical review of research. *Perspectives on Psychological Science, 10*(6), 772-789. doi: 10.1177/1745691615600138

Neumann, I. D., & Landgraf, R. (2012). Balance of brain oxytocin and vasopressin: Implications for anxiety, depression, and social behaviors. *Trends in Neurosciences, 35*(11), 649-659. doi: 10.1016/j.tins.2012.08.004

Ng, G. T. T., & Au, W. T. (2016). Expectation and cooperation in prisoner's dilemmas: The moderating role of game riskiness. *Psychonomic Bulletin & Review, 23*(2), 353-360. doi: 10.3758/s13423-015-0911-7

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113. doi: 10.1016/0028-3932(71)90067-4

Olff, M., Frijling, J. L., Kubzansky, L. D., Bradley, B., Ellenbogen, M. A., Cardoso, C., . . . van Zuiden, M. (2013). The role of oxytocin in social bonding, stress regulation and mental health: An update on the moderating effects of context and interindividual differences. *Psychoneuroendocrinology, 38*(9), 1883-1894. doi: 10.1016/j.psyneuen.2013.06.019

Owen, S. F., Tuncdemir, S. N., Bader, P. L., Tirko, N. N., Fishell, G., & Tsien, R. W. (2013). Oxytocin enhances hippocampal spike transmission by modulating fast-spiking interneurons. *Nature, 500*(7463), 458-462. doi: 10.1038/nature12330

Petrovic, P., Kalisch, R., Singer, T., & Dolan, R. J. (2008). Oxytocin attenuates affective evaluations of conditioned faces and amygdala activity. *The Journal of Neuroscience, 28*(26), 6607-6615. doi: 10.1523/jneurosci.4572-07.2008

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience, 2*(1), 67-70. doi: 10.1093/scan/nsm006

Preckel, K., Scheele, D., Kendrick, K. M., Maier, W., & Hurlemann, R. (2014). Oxytocin facilitates social approach behavior in women. *Frontiers in Behavioral Neuroscience, 8*(191). doi: 10.3389/fnbeh.2014.00191

Quintana, D. S., Alvares, G. A., Hickie, I. B., & Guastella, A. J. (2015). Do delivery routes of intranasally administered oxytocin account for observed effects on social cognition and behavior? A two-level model. *Neuroscience and Biobehavioral Reviews, 49*, 182-192. doi: 10.1016/j.neubiorev.2014.12.011

Quintana, D. S., Westlye, L. T., Alnæs, D., Rustan, Ø. G., Kaufmann, T., Smerud, K. T., . . . Andreassen, O. A. (2016). Low dose intranasal oxytocin delivered with breath powered device dampens amygdala response to emotional stimuli: A peripheral effect-controlled within-subjects randomized dose-response fMRI trial. *Psychoneuroendocrinology, 69*, 180-188. doi: 10.1016/j.psyneuen.2016.04.010

Rault, J.-L. (2016). Effects of positive and negative human contacts and intranasal oxytocin on cerebrospinal fluid oxytocin. *Psychoneuroendocrinology, 69*, 60-66. doi: 10.1016/j.psyneuen.2016.03.015

Romero-Fernandez, W., Borroto-Escuela, D. O., Agnati, L. F., & Fuxe, K. (2013). Evidence for the existence of dopamine d2-oxytocin receptor heteromers in the ventral and dorsal striatum with facilitatory receptor-receptor interactions. *Molecular Psychiatry, 18*(8), 849-850. doi: 10.1038/mp.2012.103

Scheele, D., Striepens, N., Gunturkun, O., Deutschlander, S., Maier, W., Kendrick, K. M., & Hurlemann, R. (2012). Oxytocin modulates social distance between males and females. *Journal of Neuroscience, 32*(46), 16074-16079. doi: 10.1523/jneurosci.2755-12.2012

Scheele, D., Wille, A., Kendrick, K. M., Stoffel-Wagner, B., Becker, B., Güntürkün, O., . . . Hurlemann, R. (2013). Oxytocin enhances brain reward system responses in men viewing the face of their female partner. *Proceedings of the National Academy of Sciences, 110*(50), 20308-20313. doi: 10.1073/pnas.1314190110

Shamay-Tsoory, S. G. (2010). Oxytocin, social salience, and social approach. *Biological Psychiatry, 67*(6), e35. doi: 10.1016/j.biopsych.2009.11.020

Shamay-Tsoory, S. G., & Abu-Akel, A. (2015). The social salience hypothesis of oxytocin. *Biological Psychiatry, 79*(3), 194-202. doi: 10.1016/j.biopsych.2015.07.020

Strathearn, L., Fonagy, P., Amico, J., & Montague, P. R. (2009). Adult attachment predicts maternal brain and oxytocin response to infant cues. *Neuropsychopharmacology, 34*(13), 2655-2666. doi: 10.1038/npp.2009.103

Striepens, N., Kendrick, K. M., Hanking, V., Landgraf, R., Wüllner, U., Maier, W., & Hurlemann, R. (2013). Elevated cerebrospinal fluid and blood concentrations of oxytocin following its intranasal administration in humans. *Scientific Reports, 3*, 3440. doi: 10.1038/srep03440

Striepens, N., Scheele, D., Kendrick, K. M., Becker, B., Schäfer, L., Schwalba, K., . . . Hurlemann, R. (2012). Oxytocin facilitates protective responses to aversive social stimuli in males. *Proceedings of the National Academy of Sciences, 109*(44), 18144-18149. doi: 10.1073/pnas.1208852109

Tucker, D. M., & Williamson, P. A. (1984). Asymmetric neural control systems in human self-regulation. *Psychological Review, 91*(2), 185-215.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., . . . Joliot, M. (2002). Automated anatomical labeling of activations in Spm using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage, 15*(1), 273-289. doi: 10.1006/nimg.2001.0978

Appendix

Instructions for the social dilemma games

The experiment comprises 4 rounds (lasting about 5 minutes per round), and in each round you will be shown 20 images, depicting a pay-off matrix of the game you are to play accompanied by a picture of the other player. The game is therefore INTERACTIVE!

In between the presentation of the images you will see a fixation cross on which you are to focus your attention for a few seconds before the next image appears.

In between the rounds there is a short break (1 minute). Please lie as still as possible, also during the break. The entire experiment (including the first ten minutes during which you will only see a black screen) will last half an hour.

The other player:

The profit you make in each game depends on the combination of your decision and the decision of the other player. For each game you are paired with a different player randomly appointed to you out of a sample of 32 women who have played the games before. For practical reasons they could not all be present here in the scanning room. Therefore their decisions have been recorded in advance. You will never be paired with the same person twice in one single round. Possibly, a person may reappear in a next round, but her decision is made independently from your previous interaction with her. Your decisions (and those of the other players) remain anonymous.

Economic games:

In each game, you can earn laboratory money, corresponding to the numbers in the cells of the pay-off matrix. At the end of the experiment your earnings will be changed into “real” money. When you return for the second experimental session, you will receive feedback on the results of the games and you will be paid the sum of money you made. It is expected that you will earn anywhere between €30,00 and €114,00, depending on your decisions and those of the other player.

The profit in each game depends on your decision and what your partner decides. The pay-off matrix for each game offers you the option to choose A or B. You choose A by pressing the left button and B by pressing the right button. You have maximum 17 seconds to decide. After 15 seconds you will see an hour glass appear on the screen to tell you that your time is almost up. When you see this, you need to decide right away. If you do not make a decision on time, it will not count towards your earnings at the end.

As mentioned earlier, each game partner has already made her choice, but you will not be told what this choice was. There are subsequently 4 possible scenario's.

1. You both choose A;
2. You choose A, and the other player chooses B;
3. You choose B, and the other player chooses A;
4. You both choose B.

What you and the other player earn differs for each of these 4 scenarios and depends on the combination of your choices. The earnings are summarized in the matrix below.

		Other Player A	Other Player B
You	A	(6 , 6)	(2 , 4)
	B	(4 , 2)	(3 , 3)

If both of you choose A, then you both earn €6. If you choose A, while the other player chooses B, then you receive €2 and the other €4. If the other player chooses A and you choose B, then you receive €4 and the other €2. If you both choose B, you both earn €3.

This pay-off matrix is just an example. The matrices will, just like the other players, change throughout the experiment. Thus it is important that you stay attentive, and that for each game you make an independent decision.

Control questions

Now we want to check that you understood the nature of the task and how to read the pay-off matrices. Look carefully at the next 6 pay-off matrices and answer the accompanying questions. Note, however, that the matrices that will appear on screen during the experiment will be different from these examples.

fMRI: image acquisition and pre-processing

The scanner was a 3 Tesla Siemens Trio scanner (Siemens, Erlangen Germany). A T1-weighted magnetization-prepared rapid acquisition with gradient echo (MP-RAGE) protocol was used to create anatomical images (256 × 192 matrix, 176 1.0 mm sagittal slices, field of view (FOV) = 192*256 mm). Functional images were acquired using T2*-weighted echo planar imaging (EPI) (repetition time (TR) = 2000 ms, echo time (TE) = 35 ms, 64 × 64 image resolution, FOV = 1344 mm², 30 3 mm slices, voxel size = 3.5 × 3.5 × 3.0 mm). We did not use the fMRI data from 3 participants due to excessive movements in the scanner.

Pre-processing and image analysis of the 52 brain volumes (30 slices with a TR of 2s) was conducted with Matlab (MATLAB and Statistics Toolbox Release 2014a, The MathWorks, Inc., Natick, MA, USA) and the Statistical Parameter Mapping package (SPM12; Wellcome Department of Cognitive Neurology, London, UK). The collected brain volumes were (i) corrected for slice timing, (ii) realigned, (iii) normalized against the Montreal Neurologic Institute template and (iv) spatially smoothed (full width at half maximum = 7 mm) and (v) temporally filtered with a 128 s high-pass filter.

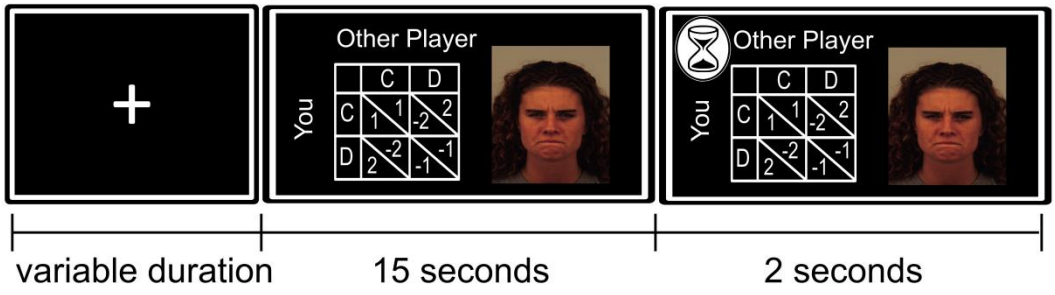

Fig. 1. An example of a trial used in the experiment. The sides on which the matrix and the face were displayed was randomly determined for each trial.

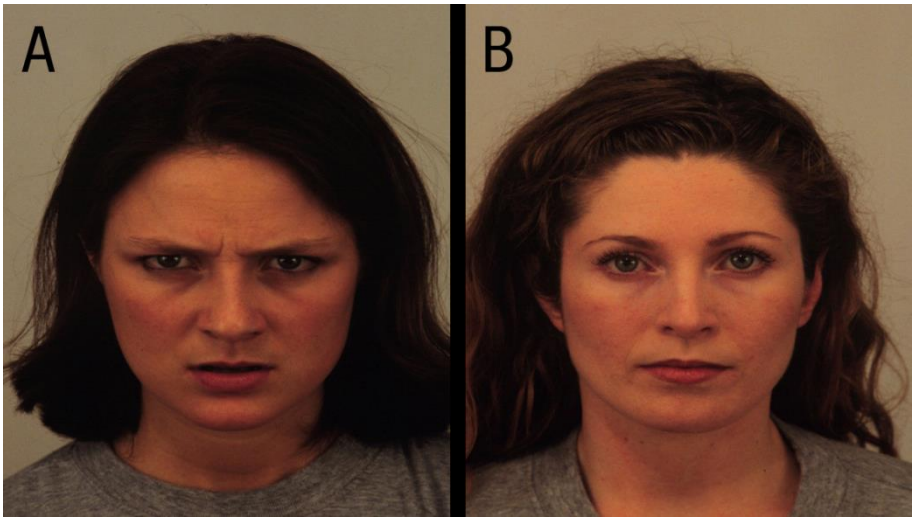

Fig. 2. Example of (A) an angry and (B) neutral face.

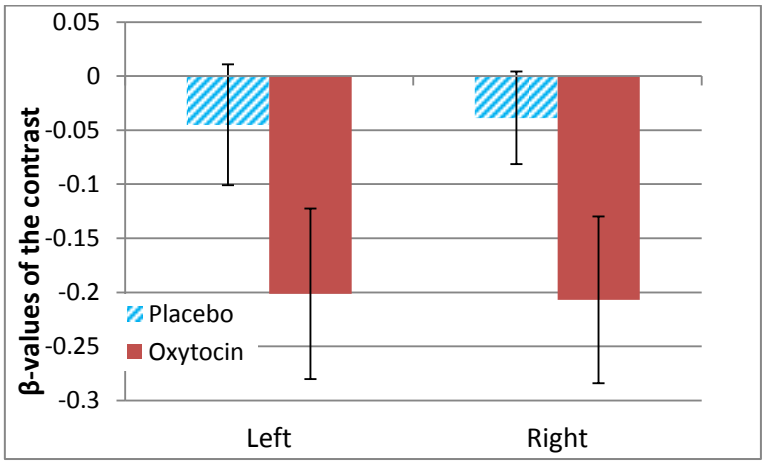

Fig. 3. β-values of the contrast trial - baseline in hippocampus for the placebo and OT condition. Error bars represent the standard error of the mean, N = 26.

Table 1. Clusters of voxels significantly activated (uncorrected p < 0.001) in the whole brain.

Region[a]	*BA*	*Side*	*x*	*y*	*z*	*Size*[b]	*p*[c]
OT - P (only clusters with size ≥ 5)							
No clusters larger than 5 voxels were found							
P - OT (only clusters with size ≥ 5)							
Limbic lobe							
Fusiform		R	38	-28	-18	8	0.997
Hippocampus		R	28	-16	-12	6	0.999
$AG_{[OT-P]}$ - $CG_{[OT-P]}$ (only clusters with size ≥ 25)							
Frontal lobe							
Middle frontal gyrus		L	-26	14	44	29	0.843
Medial frontal gyrus		L	-26	46	8	26	0.883
Superior motor area	6	R	12	10	56	45	0.599
Paracentral lobule	6	R	12	-30	58	30	0.829
Limbic lobe							
Cingulate gyrus		R	10	2	30	37	0.723
Parahippocampus	35	R	18	-20	-18	44	0.614
Temporal lobe							
Superior temporal gyrus		R	44	-62	14	78	0.234
Parietal lobe							
Precuneus		L	-12	-28	60	100	0.121
$AN_{[P-OT]}$ - $NE_{[P-OT]}$ for AG (only clusters with size ≥ 5)							
No clusters larger than 5 voxels were found							
$AN_{[P-OT]}$ - $NE_{[P-OT]}$ for CG (only clusters with size ≥ 5)							
Basal ganglia							
Putamen		L	-22	-2	10	15	0.980

BA = Brodmann area, L = left, R = right. [a]Regions were determined using the AAL atlas (45). [b]Number of statistically significant voxels (voxel size is 2.0 x 2.0 x 2.0 mm). [c]The FWE corrected cluster-level p-value

Table 2. Logistic regression using panel data clustered on 29 individuals.

	Assurance game				**Chicken game**			
	OR	p	OR	p	OR	p	OR	p
OT	0.91 (0.11)	0.463	0.17 (0.10)	0.002	1.10 (0.11)	0.340	0.41 (0.17)	0.028
Cue	1.38 (0.16)	0.007	1.31 (0.22)	0.108	1.57 (0.15)	<0.001	1.31 (0.18)	0.051
Risk Avoidance	1.12 (0.48)	0.779	0.90 (0.39)	0.813	0.57 (0.12)	0.006	0.53 (0.11)	0.003
Anticonception	2.91 (1.39)	0.016	4.68 (2.16)	0.001	2.11 (0.65)	0.015	2.17 (0.71)	0.018
OT*Cue			1.12 (0.27)	0.647			1.44 (0.28)	0.058
OT*Risk Avoid.			1.61 (0.20)	<0.001			1.16 (0.10)	0.097
OT*Antic.			0.71 (0.19)	0.215			1.38 (0.32)	0.164
Constant	0.34 (0.75)	0.830	2.62 (4.62)	0.585	3.10 (2.64)	0.184	4.41 (3.84)	0.088
Wald chi2	13.12		33.25		37.67		45.20	
N	29 (2319)		29 (2319)		29 (2315)		29 (2315)	

In the assurance game, the dependent variable is the cooperative choice, while in the chicken game, the dependent variable is the aggressive choice. Anticonception is included as a control variable. The dependent variables in the random effects logistic regressions are the cooperative choice (= 1; 0 otherwise) in the AG or the aggressive choice (= 1; 0 otherwise) in the CG. Independent variables are the administered OT (= 1; placebo = 0), the cue (neutral faces = 1; angry faces = 0) the risk aversion displayed by the participants (range between not risk averse = 0 to risk averse = 5), and if they use hormonal anticonception (= 1; no = 0). We display the odds ratio's, the standard error in brackets and the p-value (2 sided t-test). For N, the number in parentheses denotes the decisions made by all individuals.

Chapter 4: Trust as commodity: social value orientation affects the neural substrates of learning to cooperate

This chapter has been published as:

Lambert, B., Declerck, C. H., Emonds, G., & Boone, C. (2017). Trust as commodity: social value orientation affects the neural substrates of learning to cooperate. *Social Cognitive and Affective Neuroscience* doi: 10.1093/scan/nsw170

Abstract

Individuals differ in their motives and strategies to cooperate in social dilemmas. These differences are reflected by an individual's social value orientation: proselfs are strategic and motivated to maximize self-interest, while prosocials are more trusting and value fairness. We hypothesize that, when deciding whether or not to cooperate with a random member of a defined group, proselfs, more than prosocials, adapt their decisions based on past experiences: they 'learn' instrumentally to form a base-line expectation of reciprocity. We conducted an fMRI experiment where participants (19 proselfs, 19 prosocials) played 120 sequential prisoner's dilemmas against randomly selected, anonymous and returning partners who cooperated 60% of the time. Results indicate that cooperation levels increased over time, but that the rate of learning was steeper for proselfs than for prosocials. At the neural level, caudate and precuneus activation were more pronounced for proselfs relative to prosocials, indicating a stronger reliance on instrumental learning and self-referencing to update their trust in the cooperative strategy.

1. Introduction

Greed and fear of betrayal are arguably two of the most important motives that impede cooperation in social dilemmas - situations in which there is a characteristic conflict between self- and collective interest. Ever since Pruitt & Kimmel's seminal paper (Pruitt & Kimmel, 1977) researchers have been trying to identify those factors that account for the willingness to cooperate (to overcome greed) and the expectations of reciprocity (to reduce the fear of betrayal).

The willingness to cooperate has been shown to be highly heterogeneous among individuals and reflects a person's social value orientation (SVO; Van Lange, 2000). Individuals with a prosocial value orientation have other-regarding preferences, prefer equal outcomes, and cooperate readily in social dilemmas because they have internalized a moralistic, cooperative norm (Bogaert, Boone, & Declerck, 2008). In contrast, individuals with a proself value orientation have self-regarding

preferences, maximize self-interest in social dilemmas and will therefore defect by default. However, there are many reports that proselfs will cooperate when there are extrinsic incentives that align self-interest with collective interest. This accomplishes a goal transformation whereby greed is no longer an obstacle to cooperation, leading proselfs to strategically cooperate (De Cremer & Van Vugt, 1999). This is, for example, the case when dyadic interactions are repeated, allowing profits for both parties involved to accumulate over time (Axelrod & Hamilton, 1984), when cooperation yields synergy by changing the pay-off structure from a mixed motive dilemma to a coordination task (Boone, Declerck, & Kiyonari, 2010), or when one's reputation is at stake (Declerck, Boone, & Kiyonari, 2014; Simpson & Willer, 2008).

fMRI studies furthermore corroborate that proselfs are more strategic and calculative when they make decisions, while prosocials are more willing to cooperate because they conform to social norms. When participants under the fMRI scanner were playing a series of social dilemma games with different incentive structures (Emonds, Declerck, Boone, Vandervliet, & Parizel, 2011), only proselfs adapted their behavior in accordance with incentives, and this was accompanied by increased activation in the dorsolateral prefrontal cortex, a region implicated in cognitive control and cost/benefit analysis (Miller & Cohen, 2001). Proselfs also showed more activation in the precuneus which is an important region in self-centered cognition (Cavanna & Trimble, 2006; den Ouden, Frith, Frith, & Blakemore, 2005; Kircher et al., 2000) and the posterior superior temporal sulcus (pSTS), a region involved in comparing intentions of self- versus others (Saxe & Wexler, 2005) and hence a crucial element in maximizing pay-offs for self. In contrast, prosocials showed more activation in the anterior portion of the STS, associated with routine, moral judgments (Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006) which suggests that they are more norm-compliant. In addition, when prosocials are treated unfairly, they show more amygdala and nucleus accumbens activity, even when they are under cognitive load (Haruno, Kimura, & Frith, 2014). This further suggests that prosocials engage more in automatic decision making.

The influence of SVO has received less attention with respect to the second factor that determines cooperative decision-making, namely the expectations of reciprocity. According to the triangle hypothesis (reviewed in Bogaert et al. (2008)), proselfs are more likely to assume others are also proselfs, whereas prosocials have a more heterogeneous view of the social world. Accordingly proselfs would generally expect little reciprocity from game partners, while prosocials would rely on trust in order to decide whether or not to cooperate. This is corroborated by an experimental study showing that prosocials are indeed more likely to cooperate in a one-shot social dilemma game when they either have high dispositional trust, or when they have a chance to familiarize themselves with the game partner (Boone et al., 2010). For proselfs, dispositional trust does not matter with respect

to the decision to cooperate, and they are more likely to abuse trust signals if this is to their advantage (Emonds, Declerck, Boone, Seurinck, & Achten, 2014).

However, there are many situations in which it does pay-off to trust partners. Whenever we find ourselves in new and transient groups that impose long-term collaboration, cooperation may be the best strategy leading to the most lucrative outcome, provided that there are sufficient other individuals in the group that are also cooperative. This is the case when we start to work in a new firm, join a sports club, or even purchase on e-bay. In these settings trust is a commodity that makes it possible to maximize self-interest. Such "instrumental trust" could be acquired through reinforcement learning processes by which the base-rate of reciprocity in the group of interacting parties is established. We propose that "learning to trust" in such situations is a strategy by which proselfs will adapt their rate of cooperation to the expected level of reciprocity.

To test this hypothesis, we conduct an fMRI study assessing the neural substrates of decision-making in a sequential prisoner's dilemma (PD) game where a cooperative decision of the first mover is only determined by the expectation of reciprocity and not by greed. We investigate if prosocials and proselfs differ in the underlying mechanism by which they are forming expectations of reciprocity in a transient group of anonymous partners and adapt their level of cooperation accordingly. If proselfs, more than prosocials, update their level of trust based on reinforcement learning, we expect this to be reflected by a relative greater increase in activation of the caudate nucleus, a subcortical region of the brain implicated in instrumental learning (O'Doherty et al., 2004) and updating behavior (Baumgartner, Heinrichs, Vonlanthen, Fischbacher, & Fehr, 2008; King-Casas et al., 2005; Waegeman, Declerck, Boone, Seurinck, & Parizel, 2014).

The sequential PD lends itself well to studying how expectations of reciprocity are formed. For the first mover there is an incentive to cooperate because he or she can potentially earn more by cooperating, but this is contingent on the second player's decision to reciprocate. Because the decision of the first player is revealed to the second player, the second player cannot fare worse than the first player because a defect decision of the first player is unlikely to be positively reciprocated. This removes greed as a motive for the first player. By repeating the sequential PD interactions within a closed group of anonymous partners that return randomly, we simulate the occurrence of real-life transient groups in which cooperation can emerge because possible future interactions "cast a shadow back on the present and thereby affect the current strategic cooperation" (Axelrod & Hamilton, 1984, p. 12). In this setting it is possible for the first mover to establish the base-rate of reciprocity by relying on instrumental learning processes whereby each instance of positive reciprocation reinforces the presumption that cooperation is paying-off.

From agent-based simulations (Riolo, Cohen, & Axelrod, 2001) and laboratory experiments (Efferson, Lalive, & Fehr, 2008) we know that cooperation can emerge in large pools of anonymous partners that are randomly matched, and typically a stable equilibrium emerges with roughly 60% cooperators (Balliet, Parks, & Joireman, 2009; Bowles & Gintis, 2004; Fischbacher, Gachter, & Fehr, 2001; Kurzban & Houser, 2005). Similarly, the experiment of Kiyonari, Tanida, and Yamagishi (2000) indicated that second movers in a sequential PD reciprocated cooperation 62% of the times, which was much higher than the 38% cooperation in the simultaneously played PD. A plausible interpretation for this jump is that the "reciprocal exchange" nature of the game, which is more salient in the sequential than in the simultaneous PD, provides a strategic incentive to cooperate (Kiyonari et al., 2000), and this would be especially valuable for proself individuals (Simpson, 2004).

In summary the main hypothesis we test in this current study is that, given 60% reciprocation in a repeated and sequentially played PD, "learning to cooperate" will be more pronounced for proselfs, and this will be reflected in the neural mechanisms (especially the caudate nucleus) that substantiate instrumental learning.

2. Methods

2.1 Participants

We recruited participants via e-mail, flyers, and web-based advertisements in which the study was introduced as an investigation of the brain areas that become activated during economic decision-making. We selected thirty-eight participants (22 females, average age = 24.6, S.D. = 4.5) based on (i) a medical screening questionnaire, to make sure the participant met all the safety criteria for MRI examination, and (ii) their SVO, assessed using the decomposed method (Van Lange, Otten, De Bruin, & Joireman, 1997). This measure consists of nine items in which the participant can choose between three distributions of points allocated to oneself and an anonymous partner. Each of these distributions represents a particular social value orientation. We selected only those participants with a consistent prosocial or individualistic orientation (i.e., at least six out of the nine choices in the SVO questionnaire were consistent with that orientation). We refer to the individualistic orientation as "proself."

All procedures were approved by the medical ethics commission of the University of Antwerp. Debriefing occurred at the conclusion of the study by contacting participants by email and referring them to a website where the intent, results, and procedures of the experiment were fully explained.

2.2 Paradigm

Participants played the role of the first mover in a repeated and sequentially played PD with a number of different but returning partners while under the MRI scanner. A pay-off matrix of a PD game played between two people is shown in Figure 1. When the game is played sequentially, the first mover can potentially earn more by cooperating, but this is contingent on the second mover's decision to reciprocate. If the second mover defects, the first mover loses more than if he/she had not cooperated. Thus the participant's intent to cooperate in this game should be a function of the expected reciprocity rate of the partner pool. In this experiment, there were supposedly 25 partners and their decisions were, unbeknownst to the participant, computer-programmed to reciprocate cooperation in 60% of the cases. The program was such that there was never a cooperative decision after the participant chose to defect.

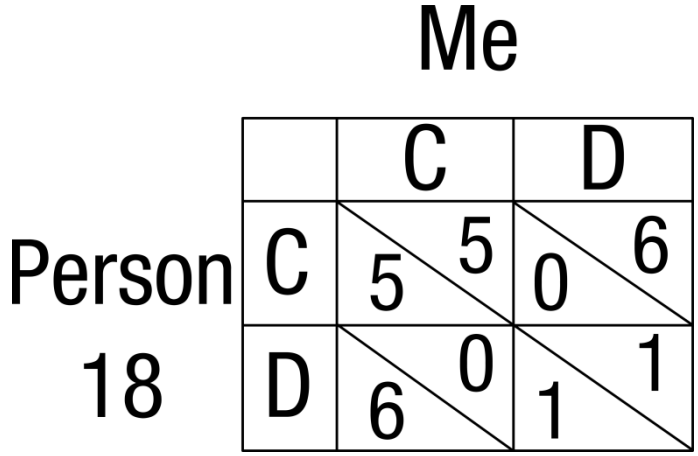

Fig. 1. A PD pay-off matrix as shown in the experiment. The participant can choose between the C or D column. If the participant chooses C (the cooperative decision) 5 points will be earned if the partner subsequently reciprocates by also choosing C (mutual cooperation); 0 points will be earned if the partner were to choose D. If the participant chooses D, the partner response is set to always be D, in which case the participant earns 1 point. In the actual experiment, we changed 'C' and 'D' into 'K' and 'L' to avoid bias.

To increase the participants' believability that they were playing against real partners, they were first shown 25 pictures representing the alleged, anonymous partners. These pictures were obtained from people who previously participated in similar experiments and who approved that their pictures would be used for this purpose. Participants were told that, for each of the 120 trials they were to play under the scanner, they would be randomly matched with one of these 25 partners. Furthermore, they were made clear that their profit earned in each trial would depend upon the combination of their own choice and the choice made by the partner. During the game, the presumed partner was always identified with a number and not with the previously seen picture.

Because these numbers corresponded to 25 possible partners, it was nearly impossible to keep track of returning partners.

Each participant received written instructions explaining that they were to play the first mover in a series of 120 PD games. To make sure all participants understood the pay-offs of the PD correctly, they had to answer three test questions correctly before starting the actual experiment. Once in the scanner, the 120 PD games were played in six rounds of twenty trials each, with a short break after each round. Within a round, participants could not be matched more than once with a certain partner.

The time sequence of the experiment is shown in Figure 2. To avoid boredom, 25 different PD matrices were used, and no matrix is used more than once in a given round. The difference in pay-off between defection or cooperation after a cooperative decision was kept equal for each PD. The values in the game matrix represent points to be exchanged for money at the end of the experiment (1 point = € 0.02; average earnings: € 9.32 + € 10.00 show-up fee). Participants indicated their choice (to cooperate or not) by pressing the corresponding button on the response box held in their right hand (Lumia model LU400, Cedrus, CA, USA). The allotted 8 seconds per trial proved sufficiently long, as the responses were successfully registered in 99% of the cases, and no-one missed more than six out of the 120 decisions. Stimulus presentation and response logging was conducted with the Presentation® software (Neurobehavioral Systems, Inc, Albany, CA, USA).

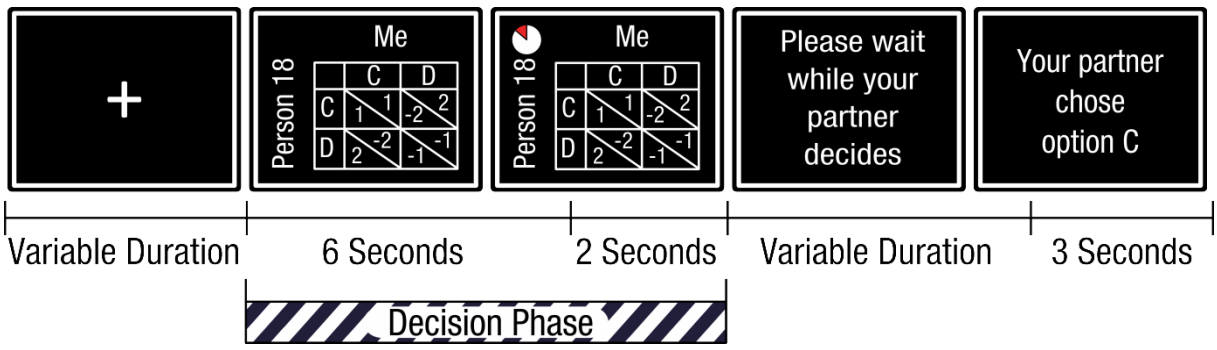

Fig. 2. Time line of the experiment. A fixation cross marks the inter-trial interval which lasts between 1.5 and 4 seconds, after which a screen depicting the PD appears. Six seconds later, an indicator appears and stays on the screen for an additional 2 seconds, signaling the end of the participants' allotted time to make a decision. The decision phase can end earlier when one of the buttons is pressed. In between the decision phase and the feedback phase a screen appears for 2 to 4 seconds asking the participant to wait. The feedback phase starts when the screen revealing the partner response is shown, which lasts 3 seconds.

This experiment was part of a larger study comprising two independent and different experiments using the same participant pool (the results of the first experiment are published elsewhere, see

Emonds et al. (2014)). Because the participants received no feedback after the first experiment, it is highly unlikely that this would have affected the second experiment.

2.3 Modelling the learning effect

To test if participants are learning to trust and update their cooperation rate according to the 60% reciprocity feedback, we adopted two different methods. First we examined how behavior changes *over time* and plot the number of cooperative decisions as a function of trial number. Second, we tested how individuals update their decisions in function of their accumulated experiences in the game, using the *Experience weighted attraction (EWA) learning model* developed by Camerer and Hua Ho (1999). This model (explained in more detail in the Appendix) computes attractions for each strategy that can be used in a game context (i.e., cooperate or defect in the prisoner's dilemma), and this attraction is updated after each trial (see eq. 1 in the Appendix). The attraction for cooperation is calculated by taking the sum of (i) a first term that accounts for previous attractions towards cooperation during the game, weighted by a free discounting parameter that weighs recent trials more heavily than earlier ones, and (ii) a second term that takes the actual or forgone payment of the trial in consideration. This sum is then normalized by dividing it by the discounted number of encounters that have taken place. To investigate how individuals adapt their behavior in function of how much they learned from their previous experience, it is the first term of the EWA equation that is of interest.

2.4 fMRI image acquisition

Images were collected with a 3 Tesla Siemens Trio scanner and an 8-channel head coil (Siemens, Erlangen Germany). A T1-weighted magnetization-prepared rapid acquisition with gradient echo (MP-FRAGE) protocol was used to create anatomical images (256 × 256 matrix, 176 0.9 mm sagittal slices, field of view (FOV) = 220 mm). Functional images were acquired using T2*-weighted echo planar imaging (EPI) (repetition time (TR) = 2400 ms, echo time (TE) = 30 ms, 64 × 64 image resolution, FOV = 224 mm, 39 3 mm slices, voxel size = 3.5 × 3.5 × 3.0 mm). Due to technical problems during the scanning procedure, the results of one participant was lost, leaving a total of 19 proselfs and 18 prosocials for image analysis.

2.5 fMRI data analysis

Image analysis of the 37 brain volumes (39 slices with a TR of 2.4s) was conducted with Matlab (MATLAB and Statistics Toolbox Release 2012a, The MathWorks, Inc., Natick, MA, USA) and the Statistical Parametric Mapping package (SPM8; Wellcome Department of Cognitive Neurology, London, UK).

Three general linear models (GLMs) were created for each participant. In each case, the blood oxygen level dependent (BOLD) signal was the dependent variable. The event of interest was the decision phase, defined as the time interval between the appearance of the slide depicting the PD matrix and the participant's response (Figure 2). We did not differentiate between cooperation or defect decisions because we are interested in revealing the underlying neural correlates of *learning* which depends on previous experiences with both strategies. The actual decision is not important in this setting: one can learn to cooperate *or* defect, depending on the type of feedback that is given.

With the first GLM, we explore which brain regions are involved in the initial decision making process of the first mover in the sequentially played prisoner's dilemma game, and we test if they differ between prosocials and proselfs who we presume have different underlying motives to cooperate. Here we consider only the decisions made during the first round (i.e. first 20 trials), because this is where we expect the difference in the intrinsic motivation to cooperate to differ between prosocials and proselfs. It is also the round in which in the difference in level of cooperation between prosocials and proselfs is the most significant (see results section).

With the second and the third GLM we investigate the neural correlates of learning (and how they differ between prosocials and proselfs), relying on the entire course of the experiment (i.e., the full 120 trials). We model "learning to cooperate" in two different ways: by adding the trial number as parametric modulator (GLM 2), and by adding the learning term of the EWA model (GLM 3; see eq. 3 in the Appendix).

For all three GLMs, all regressors were convolved with the haemodynamic response function. Six movement parameters were added to account for head movement in six dimensions. For each of the three models, we started by conducting a whole brain first level analysis for the contrast [decision – baseline] before conducting a second level analysis which further investigates possible interaction effects with SVO. For subsequent region of interest (ROI) analyses, we relied on independent coordinates derived from the maximum probability atlas of the human brain (Hammers et al., 2003) and from the AAL atlas (Tzourio-Mazoyer et al., 2002).

3. Results

3.1 Behavioural data

For each group (proselfs and prosocials), we plotted the proportion of cooperative decisions in each of the 120 trials (Figure 3). This reveals that proselfs start out cooperating at a lower level than prosocials (proportion of cooperation in first 20 rounds: mean proselfs = 0.50, 95% CI [0.45 0.55]; mean prosocials = 0.62, 95% CI [0.57 0.67], p-value < 0.001). Their behaviors converge during the

experiment (proportion of cooperation in last 20 rounds: mean proselfs = 0.70, 95% CI [0.66 0.75]; mean prosocials = 0.71, 95% CI [0.66 0.76]). Furthermore, the overall proportion of cooperation increases throughout the experiment, and this learning effect is more pronounced for proselfs. To verify this statistically, we estimated a logistic regression model with random effects to account of the unobserved differences among the 38 participants with cooperative decision (coded 1, defect coded 0) in each of the 120 trials as the dependent variable. This corroborates that there is a significant effect of trial number (B = 0.010, 95% CI [0.008 0.013], p-value < 0.000, Wald χ^2 = 79.50) but not of SVO (coded 0 for prosocials and 1 for proselfs). When adding the interaction term SVO * trial number to the regression model, we observe that the increase in cooperative decisions differs significantly for prosocials and proselfs (B = 0.0066, 95% CI [0.002 0.011], p-value = 0.005, Wald χ^2 = 86.07), indicating that indeed proselfs are adapting their behavior over time more so than to prosocials. We repeated this analysis with the EWA estimated probability of making a cooperative decisions instead of the observed decisions, and obtain very similar results (see Appendix figure 1).

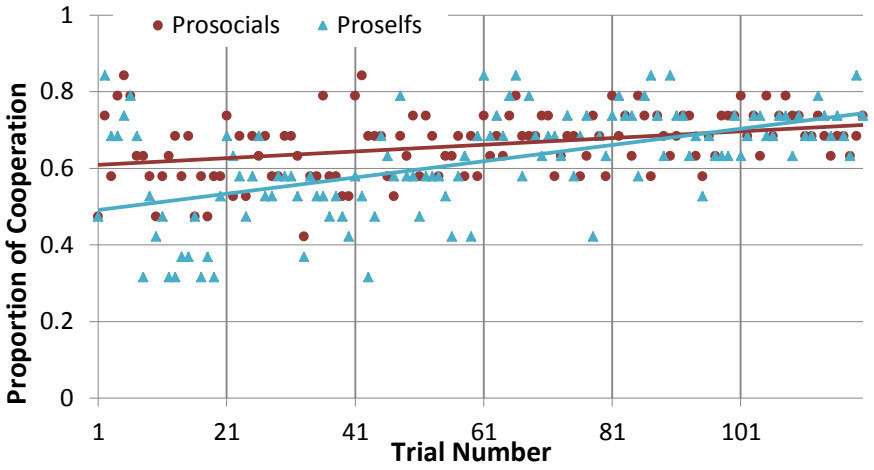

Fig. 3. Proportion cooperative decisions made by the participants for each of the 120 trials, plotted for prosocials and proselfs. Best fit lines, based on least squares, are shown.

If prosocials have internalized a cooperative norm (making them less reliant on instrumental learning), we would expect them to respond more intuitively than proselfs, which would be noticed in shorter response latencies. Therefore we plotted the average reaction times (RT) for proselfs and prosocials (Figure 4). Consistently, we observed that the reaction time is faster for prosocials than proselfs (Mean prosocials = 2.60 s, 95% CI [2.57 s 3.79 s]; Mean proselfs = 3.18 s, 95% CI [2.09 s 3.11 s]). A linear regression with random effects with RT as the dependent variable shows a marginally significant effect of SVO (B = -0.58 , 95% CI [-1.27 0.11], p-value = 0.1), and a significant effect of trial number (B = -0.01, 95% CI [-0.011 -0.0091], p-value < 0.001, Wald χ^2 = 413.20). The interaction term SVO * trial number is also significant (B = 0.0031, 95% CI [0.001 0.005], p-value = 0.002, Wald χ^2 =

424.13), revealing that prosocials respond faster, and proselfs slower, especially early on in the experiment. Thus as time progresses, proselfs show less need to deliberate their choice. This interaction between SVO and time also indicates that the learning effect is greater for proselfs.

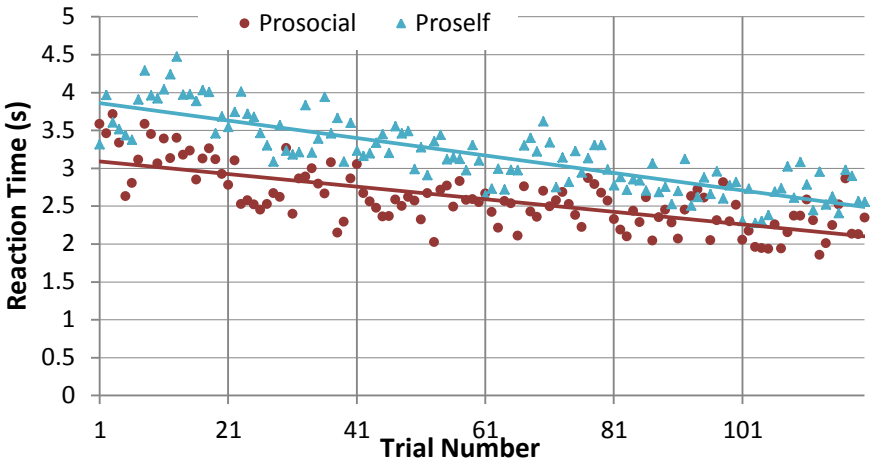

Fig. 4. Average reaction times for prosocials and proselfs throughout the experiment. Lines of best fit are based on least squares.

3.2 Functional MRI data

We first contrasted the decision phase with baseline activation during the first round (average of 20 trials) in the whole brain (using the first GLM). Clusters of significant activation (family wise error [FWE] corrected p-value < 0.05, whole brain) are reported in Table 1. We note that there is significant activation in the precuneus and insula, two regions of great interest in social decision making. Activation of the insula has been reported as a necessary component of overcoming betrayal aversion (Aimone, Houser, & Weber, 2014), while the precuneus is involved in self-referencing (Cavanna & Trimble, 2006). The latter region would be especially important to proselfs (as shown in Emonds et al. (2011)), who we hypothesize are trusting instrumentally, because mutual cooperation is in their best interest. However, a subsequent whole brain analysis contrasting proselfs and prosocials during the decision phase in the first round did not yield any significant differences in brain activation, neither did an ROI on the precuneus (coordinates derived from Tzourio-Mazoyer et al. (2002)) and the insula (coordinates derived from Hammers et al. (2003)).

Table 1. Clusters of voxels significantly activated during the decision phase in the first round (first 20 trials)

Region[a]	***BA***	***Side***	***x***	***y***	***z***	***Size***[b]	***t***	***p***
Frontal								
Middle Frontal Gyrus	-	L	-36	-2	54	133	7.65	<0.001
	6	R	34	2	64	121	8.24	<0.001
Inferior Frontal Gyrus	9	L	-42	2	30	271	8.13	<0.001
Parietal								
Precuneus	-	L	-8	-66	58	3062	10.50	<0.001
Occipital								
Middle occipital Gyrus	-	L	-36	-64	-8	429	8.06	<0.001
Lingual Gyrus	-	R	10	-82	-12	373	8.80	<0.001
Limbic								
Cingulate Gyrus	-	R	6	18	44	372	10.52	<0.001
Sub-lobar								
Insula	-	L	-38	16	2	193	7.84	<0.001
	-	R	34	24	2	346	12.69	<0.001

Summary of all significantly (whole brain FWE corrected p-value < 0.05) activated clusters during the decision phase (coordinates are in MNI space). Only clusters with more than 100 voxels are shown. BA = Brodmann area, L = left, R = right. [a]Regions were determined using the AAL atlas (Tzourio-Mazoyer et al., 2002). [b]Number of statistically significant voxels (voxel size of 2.0 x 2.0 x 2.0 mm).

To investigate learning, the second and third GLM took the decisions of each of the 120 trials into account. In the second GLM, we added the trial number as a parametric modulator and contrasted this with baseline. Results (FWE corrected p-value < 0.05, whole brain) are listed in Table 2. As hypothesized, we observe significant activation in the caudate nucleus, which has previously been identified as one of the most important regions implicated in instrumental learning and updating behavior (Baumgartner et al., 2008; King-Casas et al., 2005; O'Doherty et al., 2004; Waegeman et al., 2014).

Table 2. Clusters of voxels significantly modulated by trial number during the decision phase in the entire experiment (120 trials)

Region[a]	***BA***	***Side***	***x***	***y***	***z***	***Size***[b]	***t***	***p***
Frontal Lobe								
Precentral Gyrus	-	L	-14	-32	72	553	8.43	<0.001
Parietal Lobe								
Precuneus	-	L	-10	-48	62	188	7.39	<0.001
Sub-Gyral	-	R	22	-58	56	927	8.30	<0.001
Limbic Lobe								
Parahippocampal Gyrus	-	L	-24	-44	2	101	7.56	<0.001
Sub-lobar								
Caudate	-	L	-14	20	4	2157	9.65	<0.001
Insula	-	R	32	-22	14	107	8.53	<0.001

Summary of all clusters where activation during the decision phase was significantly (whole brain FWE corrected p-value < 0.05) modulated by trial number (coordinates are in MNI space). Only clusters with more than 100 voxels are shown. BA = Brodmann area, L = left, R = right. [a]Regions were determined using the AAL atlas (Tzourio-Mazoyer et al., 2002). [b]Number of statistically significant voxels (voxel size is 2.0 x 2.0 x 2.0 mm).

When contrasting proselfs versus prosocials in this whole brain contrast, no activation survived correction (FWE corrected p-value < 0.05). We therefore perform ROI-analyses on the caudate (derived from Hammers et al. (2003)), the precuneus, and the insula. Figure 5 shows the β values for the parametric modulation of trial number in these regions (GLM 2), revealing a significant greater effect for proselfs (compared to prosocials) in the former two. Figure 6 projects these significantly activated clusters on a standard brain for both the proself (Figure 6a) and prosocial (Figure 6b) group.

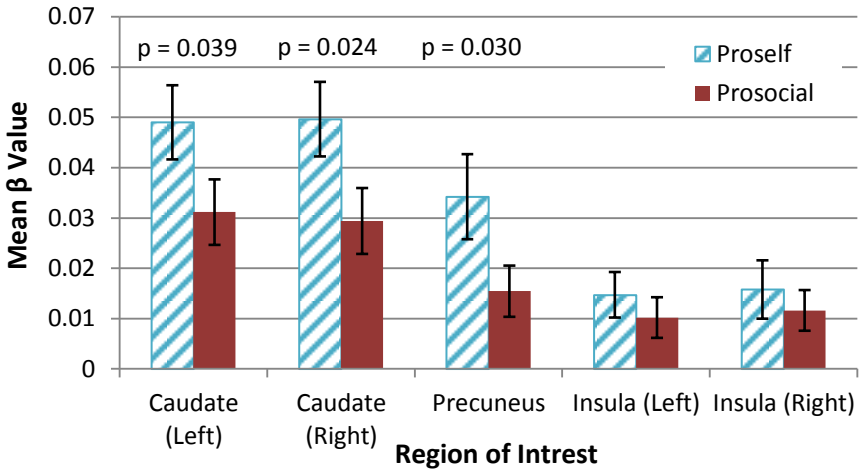

Fig. 5. Parameter estimates for brain activation modulated by trial number during the decision phase in three ROI's. Error bars represent Standard Errors of the Mean. p-values of t-tests reveal significant differences in activation between the prosocial and proself group in the caudate and the precuneus, but not in the insula.

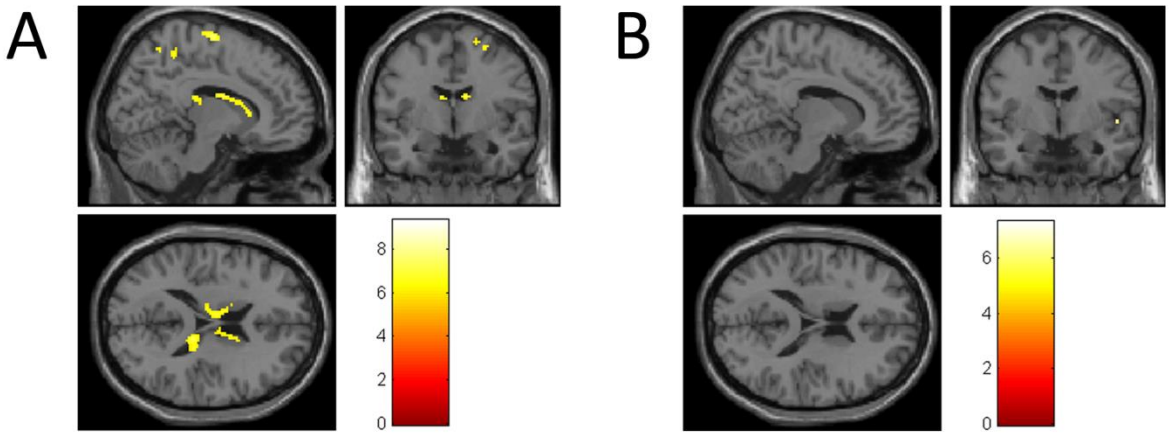

Fig. 6. Whole brain analysis showing activation (in sagittal, coronal and transvers section; x = 12, y = -4, z = 18) modulated by the trial number during the decision phase for (A) 19 proselfs and (B) 18 prosocials. T value cutoff = 5.81 (corresponding to an FWE corrected p-value = 0.05).

We repeated the same whole brain and ROI analyses using the third GLM, which includes the learning term derived from the EWA model as a parametric modulator. The whole brain analysis does not reveal any regions that are significantly affected by experience learning. The ROI analyses, however, show that the effect of this parametric modulator on the BOLD signal in the caudate and precuneus differ significantly between prosocials and proselfs, with prosocials showing relatively less activation in these regions than proselfs (see Figure 7). This is consistent with the hypothesis that adopting a cooperative strategy based on previous experience (learning to trust) is less instrumental (i.e. less dependent on reinforcement and self-referencing) for prosocials compared to proselfs. For prosocials, caudate and precuneus neural activity is negatively associated with experience based

attraction towards cooperation. Thus, when deciding to cooperate, they seem to rely less on those regions, given their imprinted preference to seek out mutual cooperation. Updating their attraction to cooperation, conditional on the partner's strategy, is therefore likely to occur by a different mechanism that does not rely as much on either reinforcement learning (involving the caudate nucleus) or self-referencing (involving the precuneus).

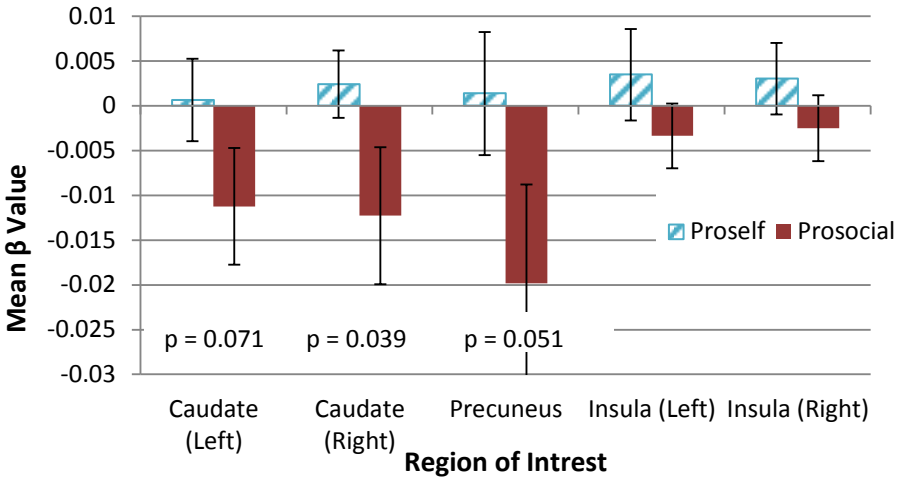

Fig. 7. Parameter estimates for brain activation modulated by the first term of the EWA model during the decision phase in three ROI's. Error bars represent Standard Errors of the Mean. p-values of t-tests reveal significant differences in activation between the prosocial and proself group in the caudate and the precuneus, but not in the insula.

4. Discussion

The results contribute to the body of knowledge indicating that individuals with different social value orientations rely on different strategies to cooperate in social dilemmas. In line with previous findings the behavioral data show that, in a repeated sequential PD, first mover prosocials have an overall greater willingness to cooperate (they start out cooperating at a higher rate and they have faster reaction times). Proselfs, in this setting, are learning to cooperate because it is a strategy that pays off in the long run; their behavior converges with that of prosocials towards the end of the experiment.

The fMRI data shed light on the underlying neural mechanism that may drive some of the difference in behavior between prosocials and proselfs. The initial greater willingness to cooperate of prosocials (relative to proselfs) was, in the current study, not associated with any notable difference in the pattern of brain activation. When it comes to reinforcement-based learning, however, the data are consistent with the hypothesis that proselfs (more than prosocials) are learning to cooperate by adapting their expectations of reciprocity instrumentally, relying on the caudate and precuneus.

As hypothesized, proselfs showed a substantially greater increase in neural activation in the caudate nucleus over time (Figure 5). Furthermore, for proselfs, but not for prosocials, the caudate nucleus remains active when the attraction to the cooperative strategy is updated based on the accumulated (positive and negative) feedback throughout the experiment. The caudate nucleus has previously been shown to be the key structure involved in establishing stimulus-response contingencies (O'Doherty et al., 2004; Packard & Knowlton, 2002). It becomes more active with perseveration and therefore plays an important role in the decision to either update behavior or maintain the status quo (Baumgartner et al., 2008; Smith-Collins et al., 2013; Waegeman et al., 2014). In a repeated dyadic trust game, the caudate nucleus is reported to become increasingly activated as partners are gaining more experience with each other, with peak activity reported at the moment the investor anticipates positive reciprocation (King-Casas et al., 2005). This instrumental trust by which expectations of reciprocity are formed is, in the current study, more pronounced for proselfs. For prosocials in the current experiment, caudate nucleus activation actually decreased with increasing experience (Figure 7), suggesting that trust is not learned instrumentally, but perhaps an intrinsic feature that is adopted automatically from the start.

In addition to the differential activation of the caudate nucleus, the data also reveal that, compared to prosocials, proselfs show a greater increase in precuneus activation as time progresses (Figure 5) and a relative greater precuneus involvement in learning (Figure 7). The precuneus is activated when comparing the outcome of decisions for self and others, and plays an important role in solving social dilemma problems, especially for proselfs who rely more heavily on self-referencing to compute the strategy with the highest pay-off (Emonds et al., 2011). Again, it makes sense that this region would become more active with time, and that activation of this region remains active for proselfs as they gain experience with the cooperative strategy: only at the end of the experiment do the proselfs have sufficient ground to establish positive expectations of reciprocity. At the onset of the experiment, they have no basis for comparison yet, and the role of the precuneus may be less salient (Emonds et al., 2014).

In addition to the caudate and precuneus, the data reveal that the insula also becomes increasingly activated as time progresses, and this is equally true for prosocials as well as proselfs (Table 2). The insula has a well-established role in emotion-processing, and its activation in the context of solving dilemma-type problems points to automatic processing (Kuo, Sjostrom, Chen, Wang, & Huang, 2009). It is also activated when overcoming betrayal aversion (Aimone et al., 2014), which is consistent with the current finding that decision latencies are decreasing with time and, that betrayal aversion is also steadily decreasing as positive expectations of reciprocity are formed. Considering that betrayal

aversion is an important reason why prosocials might chose not to trust, it is not surprising that the increase in insula activation does not differ between prosocials and proselfs (see Figures 5 and 7).

These data, revealing caudate, precuneus, and insula activation in learning to cooperate, are particularly relevant in the light of a recent experiment by Watanabe et al. (2014). These authors scanned participants that were supposedly embedded in a chain of reciprocal donations. The participant had to decide whether or not to donate a sum of money to the next individual in the chain (which was costly for the participant but beneficial for the other). In the first condition, the participant had received money from the previous person in the chain and decided to "pay it forward" to the next one. This activated the caudate nucleus together with the insula, pointing to the involvement of emotions in cooperative decision making. In the second condition, the participant (regardless of what he or she had received previously) decided to donate money to the next person in the chain knowing that this person had a history of donating. In this "reputation-based" condition, the caudate together with the precuneus were activated. Interestingly, we find that these two regions (precuneus and caudate) are more activated in proselfs, which is consistent with their strategic nature: if they are truly establishing the base-rate of reciprocity to solve repeated PD games, it is important that they learn the history of the cooperative behavior of returning partners.

Finally, the current findings also extend the results described by Smith-Collins et al. (2013). These authors investigated how participants make decisions in a trust game where the same and new partners re-appear in consecutive rounds and feedback is given. They report that activation in the caudate is associated with successfully adapting behavior after being confronted with unexpected cooperation or betrayal, i.e. trusting after unexpected cooperation or distrusting after unexpected betrayal. While Smith-Collins et al. (2013) focus on the response towards an *identifiable* partner when group composition is changing, we show that trust-based learning (with increasing caudate activity over time) can also take place based on expectations of the entire group.

We also note limitations to the study. First fMRI data can be influenced by elements that are unrelated to the decision-making process, such as scanner drift (which tends to increase neural activity over time), fatigue, or other effects that are only marginally related to the task. These extraneous factors call for caution in the interpretation of the results, but they cannot explain the differences that we observe between two groups that received the exact same treatment and only differed in personality.

A second limitation pertains to the experimental design and challenges the generalizability of the conclusions. While we decided *a priori* on 60% positive reciprocity (the level needed to sustain mutual cooperation in natural populations, see introduction), this rate also constraints what can be

learned. Prosocials are more likely to cooperate when they sense their partners are cooperative (Kuhlman & Marshello, 1975), so with a 60% cooperation rate, there is little left to be learned for them because their expectation of reciprocity is easily met (i.e., the feedback they receive in early rounds is sufficient to allow their cooperative norm to surface). Hence the current experimental design cannot disentangle whether prosocials are not activating brain regions involved in instrumental learning and self-referencing because they truly differ in the way they integrate newly accumulated information in their utility function, or because their preconceived expectations of reciprocity did not differ substantially from reality. In contrast, proselfs, who are not inclined to cooperate naturally, still have a lot to learn before they can establish that trust is a self-serving strategy, and the data show the involvement of the caudate and precuneus. Hence at least for proselfs, learning to trust is a commodity that they have acquired instrumentally.

An interesting avenue for future studies is to investigate if, and how, prosocials and proselfs differ when they are "unlearning to cooperate," if reciprocity rates were 40% or lower. As mutual cooperation is potentially the most lucrative outcome for the first player in a sequential prisoner's dilemma game, we expect that proselfs would still need to rely on the same instrumental learning and self-reference processes to establish that the reciprocity rate is too poor to adopt a cooperative strategy. The interesting question is how prosocials would adapt their behavior in a pool of negative reciprocators. Being behavior assimilators, they are likely to also adapt to a defect strategy, but we do not know whether this change in behavior can be attributed to the same instrumental learning mechanism as proselfs. Given prosocials' strong egalitarian orientation and retaliatory nature, "unlearning to cooperate" might reveal itself more in a neural signature that updates decision making emotionally rather than instrumentally. This would be consistent with findings that the breaches of fairness that cause punitive actions by prosocials are driven by amygdala and nucleus accumbens activation (Haruno et al., 2014).

In sum, this is the first study to report on the neural basis by which individuals with different social values are forming expectations of reciprocity and thereby learning to cooperate when they experience 60% reciprocation in a transient pool of anonymous partners. While prosocials' decisions based on expectations are more automatic and change little over time, proselfs are learning to trust instrumentally, whereby they activate (more than prosocials) the caudate nucleus and the precuneus.

References

Aimone, J. A., Houser, D., & Weber, B. (2014). Neural signatures of betrayal aversion: An fMRI study of trust. *Proceedings of the Royal Society B: Biological Sciences, 281*(1782). doi: 10.1098/rspb.2013.2127

Axelrod, R., & Hamilton, W. D. (1984). *The evolution of cooperation*. New York: Basic Books.

Balliet, D., Parks, C., & Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes & Intergroup Relations, 12*(4), 533-547. doi: 10.1177/1368430209105040

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron, 58*(4), 639-650. doi: 10.1016/j.neuron.2008.04.009

Bogaert, S., Boone, C., & Declerck, C. (2008). Social value orientation and cooperation in social dilemmas: A review and conceptual model. *British Journal of Social Psychology, 47*, 453-480. doi: 10.1348/014466607x244970

Boone, C., Declerck, C. H., & Kiyonari, T. (2010). Inducing cooperative behavior among proselfs versus prosocials: The moderating role of incentives and trust. *Journal of Conflict Resolution, 54*(5), 799-824. doi: 10.1177/0022002710372329

Borg, J. S., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience*(18), 803-817. doi: 10.1162/jocn.2006.18.5.803

Bowles, S., & Gintis, H. (2004). The evolution of strong reciprocity: Cooperation in heterogeneous populations. *Theoretical Population Biology, 65*(1), 17-28. doi: 10.1016/j.tpb.2003.07.001

Camerer, C., & Hua Ho, T. (1999). Experience-weighted attraction learning in normal form games. *Econometrica, 67*(4), 827-874. doi: 10.1111/1468-0262.00054

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain, 129*, 564-583. doi: 10.1093/brain/awl004

De Cremer, D., & Van Vugt, M. (1999). Social identification effects in social dilemmas: A transformation of motives. *European Journal of Social Psychology, 29*(7), 871-893. doi: 10.1002/(sici)1099-0992(199911)29:7<871::aid-ejsp962>3.0.co;2-i

Declerck, C. H., Boone, C., & Kiyonari, T. (2014). No place to hide: When shame causes proselfs to cooperate. *Journal of Social Psychology, 154*(1), 74-88. doi: 10.1080/00224545.2013.855158

den Ouden, H. E. M., Frith, U., Frith, C., & Blakemore, S. J. (2005). Thinking about intentions. *Neuroimage, 28*(4), 787-796. doi: 10.1016/j.neuroimage.2005.05.001

Efferson, C., Lalive, R., & Fehr, E. (2008). The coevolution of cultural groups and ingroup favoritism. *Science, 321*(5897), 1844-1849. doi: 10.1126/science.1155805

Emonds, G., Declerck, C. H., Boone, C., Seurinck, R., & Achten, R. (2014). Establishing cooperation in a mixed-motive social dilemma. An fMRI study investigating the role of social value orientation and dispositional trust. *Social Neuroscience, 9*(1), 10-22. doi: 10.1080/17470919.2013.858080

Emonds, G., Declerck, C. H., Boone, C., Vandervliet, E., & Parizel, P. M. (2011). Comparing the neural basis of decision making in social dilemmas of people with different social value orientations, a fMRI study. *Journal of Neuroscience Psychology and Economics, 4*(1), 11-24. doi: 10.1037/a0020151

Fischbacher, U., Gachter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters, 71*(3), 397-404. doi: 10.1016/s0165-1765(01)00394-9

Hammers, A., Allom, R., Koepp, M. J., Free, S. L., Myers, R., Lemieux, L., . . . Duncan, J. S. (2003). Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Human Brain Mapping, 19*(4), 224-247. doi: 10.1002/hbm.10123

Haruno, M., Kimura, M., & Frith, C. D. (2014). Activity in the nucleus accumbens and amygdala underlies individual differences in prosocial and individualistic economic choices. *Journal of Cognitive Neuroscience, 26*(8), 1861-1870. doi: 10.1162/jocn_a_00589

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science, 308*(5718), 78-83. doi: 10.1126/science.1108062

Kircher, T. T. J., Senior, C., Phillips, M. L., Benson, P. J., Bullmore, E. T., Brammer, M., . . . David, A. S. (2000). Towards a functional neuroanatomy of self processing: Effects of faces and words. *Cognitive Brain Research, 10*(Article), 133-144. doi: 10.1016/s0926-6410(00)00036-7

Kiyonari, T., Tanida, S., & Yamagishi, T. (2000). Social exchange and reciprocity: Confusion or a heuristic? *Evolution and Human Behavior, 21*(6), 411-427. doi: 10.1016/s1090-5138(00)00055-6

Kuhlman, D. M., & Marshello, A. F. J. (1975). Individual differences in game motivation as moderators of preprogrammed strategy effects in prisoner's dilemma. *Journal of Personality and Social Psychology, 32*(5), 922-931. doi: 10.1037//0022-3514.32.5.922

Kuo, W. J., Sjostrom, T., Chen, Y. P., Wang, Y. H., & Huang, C. Y. (2009). Intuition and deliberation: Two systems for strategizing in the brain. *Science, 324*(5926), 519-522. doi: 10.1126/science.1165598

Kurzban, R., & Houser, D. (2005). Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. *Proceedings of the National Academy of Sciences of the United States of America, 102*(5), 1803-1807. doi: 10.1073/pnas.0408759102

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*(1), 167-202. doi: 10.1146/annurev.neuro.24.1.167

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K. J., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science, 304*(5669), 452-454. doi: 10.1126/science.1094285

Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience, 25*, 563-593. doi: 10.1146/annurev.neuro.25.112701.142937

Pruitt, D. G., & Kimmel, M. J. (1977). Twenty years of experimental gaming: Critique, synthesis, and suggestions for the future. *Annual Review of Psychology, 28*(1), 363-392. doi: 10.1146/annurev.ps.28.020177.002051

Riolo, R. L., Cohen, M. D., & Axelrod, R. (2001). Evolution of cooperation without reciprocity. *Nature, 414*(6862), 441-443. doi: 10.1038/35106555

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia, 43*(10), 1391-1399. doi: 10.1016/j.neuropsychologia.2005.02.013

Simpson, B. (2004). Social values, subjective transformations, and cooperation in social dilemmas. *Social Psychology Quarterly, 67*(4), 385-395. doi: 10.1177/019027250406700404

Simpson, B., & Willer, R. (2008). Altruism and indirect reciprocity: The interaction of person and situation in prosocial behavior. *Social Psychology Quarterly, 71*(1), 37-52.

Smith-Collins, A. P. R., Fiorentini, C., Kessler, E., Boyd, H., Roberts, F., & Skuse, D. H. (2013). Specific neural correlates of successful learning and adaptation during social exchanges. *Social Cognitive and Affective Neuroscience, 8*(8), 887-896. doi: 10.1093/scan/nss079

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., . . . Joliot, M. (2002). Automated anatomical labeling of activations in Spm using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage, 15*(1), 273-289. doi: 10.1006/nimg.2001.0978

Van Lange, P. A. M. (2000). Beyond self-interest: A set of propositions relevant to interpersonal orientations. *European Review of Social Psychology, 11*(1), 297-331. doi: 10.1080/14792772043000068

Van Lange, P. A. M., Otten, W., De Bruin, E. M. N., & Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology, 73*(4), 733-746. doi: 10.1037/0022-3514.73.4.733

Waegeman, A., Declerck, C. H., Boone, C., Seurinck, R., & Parizel, P. M. (2014). Individual differences in behavioral flexibility in a probabilistic reversal learning task: An fMRI study. *Journal of Neuroscience Psychology and Economics, 7*(4), 203-218. doi: 10.1037/npe0000026

Watanabe, T., Takezawa, M., Nakawake, Y., Kunimatsu, A., Yamasue, H., Nakamura, M., . . . Masuda, N. (2014). Two distinct neural mechanisms underlying indirect reciprocity. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1318570111

Appendix

Experience weighted attraction

To model learning by experience and to test if experience-based learning affects which neural correlates are activated when players are playing a repeated sequential prisoner's dilemma game, the experiment made use of the experience weighted attraction model (EWA) developed by Camerer and Hua Ho (1999).

This model computes an attraction for each strategy possible in the game. In a prisoner's dilemma, the two possible strategies are to cooperate (C) or to defect (D). In each subsequent trial, the attractions towards these strategies are updated.

The full equation for the attraction, A_C, towards the cooperative strategy, s_C, in trial t is given by:

eq. 1 $$A_C(t) = \frac{\varphi N(t-1) A_C(t-1) + \left[\delta + (1-\delta) * I\left(s_C, s(t)\right)\right] * \pi\left(s_C, z(t)\right)}{\rho N(t-1) + 1}$$

The first term in the numerator captures what the participant learned from experience in past encounters:

eq. 2 $$\varphi \mathrm{N}(\mathrm{t}-1)\mathrm{A}_{\mathrm{C}}(\mathrm{t}-1)$$

where N(t-1) is the number of "observation-equivalents" referring to past encounters, and A_C(t-1) is the attraction towards the cooperative strategy in round t-1. The factor φ is a discount factor that weights prior experience in earlier trials less than the experience in the most recent trials. This term captures how strongly participants update their preference towards cooperation, based on built–up experience with this strategy, and therefore captures the learning process of the participants.

The second term describes the influence of the pay-off on the attraction in the current trial t:

eq. 3 $$\left[\delta + (1-\delta) * I\left(s_C, s(t)\right)\right] * \pi\left(s_C, z(t)\right)$$

The term $\pi(s_C, z(t))$ is the pay-off received when the participant chooses the cooperative strategy, given the response z(t) by the opponent (which can be cooperate or defect). The term $I\left(s_C, s(t)\right)$ is equal to 1 if the participant chooses to cooperate, and 0 otherwise. Thus, if the participant cooperates, the pay-off term is equal to the actual pay-off and weighted by 1. If the participant defects, attraction to cooperate is affected by the pay-off that was forgone by not cooperating, weighted by δ.

The sum of these two terms is then normalized by dividing the number of previous encounters (the "observation-equivalents", N(t-1)), discounted by the factor ρ.

The probability of cooperation in trial t+1 is defined by the attraction to cooperation vis-à-vis the alternative strategy of defect, A_D:

eq. 4
$$P_C(t+1) = \frac{e^{\lambda A_C(t)}}{e^{\lambda A_C(t)} + e^{\lambda A_D(t)}}$$

in which λ is a scaling factor. To determine the values of the free parameters, we fitted equation 4 to the data using matlab code by den Ouden et al. (2005).

A linear regression model with random effects to account for the differences among participants was estimated, with the estimated probability of cooperation as dependent variable, and trial numbers and SVO as independent variables. This revealed that there is a significant effect of trial number (B = 0.010, 95% CI [0.008 0.011], p < 0.000, Wald χ^2 = 253.63) but not of SVO. Similar to the results reported in the paper, the interaction between trial number and SVO is also significant (B = 0.0006, 95% CI [0.003 0.008], p = 0.005, Wald χ^2 = 275.00).

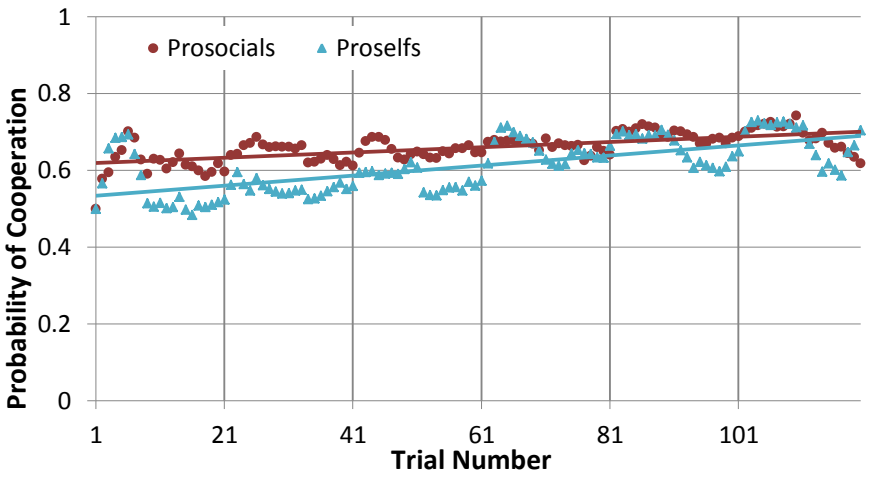

Fig. 1. Estimated probability of cooperative decisions made by proselfs or prosocials for each of the 120 trials. Best fit lines, based on least squares, are shown.

Epilogue

This dissertation has described how cooperative behavior can be established given that social interactions are often permeated by many, sometimes conflicting, sources of information. To arrive at a cooperative decision, an individual evaluates the decision context, computes pay-offs, and matches the expected outcome with his/her own social preferences. By using a combination of behavioral experiments, pharmacological interventions, and neuroimaging, we investigated (i) the role of oxytocin (OT) in evaluating social cues, (ii) how OT affects the way these social cues and pay-offs are integrated in the brain and (iii) the role of social values in establishing trust and cooperation over time.

1. Main contributions

A **first** contribution is that we have gained a better understanding of how OT affects the evaluation of social cues. Contrary to popular believe, OT does not indiscriminately make people trust other people more. Rather, OT seems to heighten caution when this is warranted. Individuals were better able to accurately discriminate between faces that were trustworthy or untrustworthy after intranasal OT administration (chapter 1). When the trustworthiness of a partner was an important piece of information that could improve the decision outcome (as was the case in the anti-coordination game described in chapter 3), OT facilitated the incorporation of this cue in the decision process. In a similar fashion, when asked to evaluate the health of artificial faces, men given exogenous OT became more cautious by responding more negatively to all faces (chapter 2). Women, in contrast, did not respond differently to health cues when given OT, but they became significantly less disgusted by foul situations. This suggests that the modulatory role of OT in evaluating informative cues is gender-specific and that it is not straightforward. It can increase or decrease caution depending on individual differences and on the particular situation.

Evaluating social cues is a first step in the decision to act cooperatively, as these cues may unconsciously bias the decision making process towards or away from cooperation. Especially in strategic social interactions where partners are interdependent, the pay-off from cooperation is uncertain as it will also depends on the behavior of the partner. Taking into account cues that reveal something about the partner can therefore improve the chances of obtaining the expected pay-off. This automatic integration of social cues with strategic incentives to cooperate can be considered a heuristic in decision making: i.e., a fast and frugal shortcut that bypasses rational decision-making but which still leads, on average, to a satisfactory outcome, without having to process an overwhelming amount of information (Gigerenzer & Todd, 1999). The results of chapter 3 suggest

that OT might be involved in this heuristic process as exogenous OT affected the decision to cooperate or compete in economic games, depending on the combination of available cues and the expected pay-offs, which is the **second** contribution of this dissertation. Specifically, we found that OT downregulated the amygdala, corresponding to a reduced orienting response and increased focused attention to the saliency of facial expressions. This is expected to affect behavior especially in an ambiguous context where decision can be improved by paying attention to social cues. Indeed, we found that OT facilitated risky competition with partners displaying neutral, but not angry expressions. When pay-offs incentivized cooperation, OT enhanced nucleus accumbens activity which is known to play a role in linking decision outcomes with the feeling of reward. In this less ambivalent decision context, social cues were less salient and therefore had less influence on decision making.

These results suggest that the decision to cooperate or compete in one-shot interactions with strangers is indeed influenced by heuristic processes that take into account salient social cues, at least when they can improve strategic decision making by reducing uncertainty. When social interactions are repeated, however, peripheral, ad hoc social cues become less salient in heuristic processing as the context now provides opportunities for uncertainty reduction by the building up of trust. This is what we investigated in the last chapter of this dissertation, leading to the **third** contribution.

In this final chapter, we investigated the underlying mechanism of trust formation in individuals who are by nature more or less inclined to cooperate. We found, not surprisingly, that individuals relied on experiences in previous encounters with similar partners to *learn* what the rate of reciprocity was. Consistently, the data revealed an increase in cooperation when individuals encountered more and more trials in which, on average, their cooperative behavior was reinforced. This was further corroborated by the fMRI data, which showed an increase in caudate activity over time. The caudate nucleus is a region with a well-established role in instrumental- or reinforcement learning. The correlation between activity in the caudate and learning was greater for proselfs, which led us to conclude that trust for them is more likely a commodity on which they call when it is useful to maximize their own profits. Prosocials, in contrast, put more trust in anonymous others from the onset of the experiment, suggesting that they regarded cooperation as a socially shared norm.

2. Limitations

The contributions of this dissertation with regards to understanding the biological roots of cooperative behavior are the result of a novel, multidisciplinary approach, combining psychological theories with endocrine studies, fMRI, and economic games. However, some of the methods we have

relied on (intranasal OT administration and fMRI) have recently been subject to much scientific debate. Here we address this debate and acknowledge the limitations of our study.

A first limitation of the dissertation research pertains to the interpretation of fMRI data. We are aware that the fMRI studies reported in chapters three and four allow us to correlate the increase or decrease in blood flow (BOLD signal) with behavior during the tasks, but that we cannot infer a cause-and effect relation. It would be wrong to reason backwards and conclude that the brain regions we identified as being activated during the tasks are responsible for the ongoing cognitive processes. This would be reverse inferencing (Poldrack, 2006). One way to avoid reverse inferencing is to start out from theory-driven hypotheses or to formulate hypotheses that built on previous experimental findings. In chapter three, the investigated contrasts and brain regions were selected on the basis of an underlying theory of OT-dopamine interactions (Love, 2014; Shamay-Tsoory & Abu-Akel, 2015). In chapter four, we based our conclusion on a large body of previous research that implicated the caudate nucleus in reinforcement learning (Baumgartner, Heinrichs, Vonlanthen, Fischbacher, & Fehr, 2008; O'Doherty et al., 2004). A second problem with fMRI is double dipping in which the same data set is used for both the selection of regions of interest and the subsequent statistical analysis within these regions (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009). This creates circular reasoning in which statistics are distorted. To avoid this, we always created relevant brain mask for regions of interest analyses by using independent probability maps.

The next major concern that can be raised in this dissertation is the question of how intranasal OT reaches the brain (Leng & Ludwig, 2015). Due to its size (molecular weight: 1007.19 Da), it was thought unlikely to cross the blood-brain-barrier (McEwen, 2004). However, a study by Born et al. (2002) showed that intranasal vasopressin, a cyclic peptide differing from oxytocin in only two amino acids, does appear in the CSF (cerebrospinal fluid). Since then there have been many reports indicating that also OT can cross the blood-CSF barrier. This is true for rodents (Neumann, Maloumby, Beiderbeck, Lukas, & Landgraf, 2013), primates (macaques) (Chang & Platt, 2014; Dal Monte, Noble, Turchi, Cummins, & Averbeck, 2014; Modi, Connor-Stroud, Landgraf, Young, & Parr, 2014), pigs (Rault, 2016), and is indicated in at least one study with humans (Striepens et al., 2013). The authors of this last paper administrated 24 IU of OT to 11 participants and extracted CSF via lumbar puncture after 45, 60 and 75 minutes. They found a significant increase in CSF OT relative to the samples of 4 participants who received placebo, albeit only after 75 minutes (it has to be mentioned that the site of sampling is distant from the relevant brain regions), while a significant increase in plasma OT was observed after 15 minutes. A typical intranasal administration of 24 IU of OT increases the physiological OT concentration in plasma significantly, but only a very small fraction

crosses the blood-CSF-barrier, which is then diluted even more before crossing the CSF-brain-barrier. Whether this small amount is sufficient to affect behavior, is still uncertain (Leng & Ludwig, 2015).

A second, and more probable way that intranasal administrated OT can reach the brain, is via the olfactory sensory neurons or the trigeminal nerve fibres that are present in the nasal cavity. The olfactory sensory neurons lead towards the olfactory bulb, which ultimately projects to the amygdala and hippocampus. The trigeminal nerves converge in the trigeminal ganglion, located in the brainstem. Experiments with labelled insulin-like growth factor have shown that transmission via these pathways is plausible in rats (Thorne, Pronk, Padmanabhan, & Frey Ii, 2004). A recent rigorous study conducted by Quintana et al. (2016), compared the effects of intravenous and intranasal administrated OT. They described how intranasal OT administration attenuates the activation of the amygdala when confronted with the same pictures of angry faces used in our study. Because amygdala modulation was only found in the intranasal OT condition, and not in the intravenously administered OT condition, it corroborates the notion that OT can reach the brain via the nasal cavity, without going through the bloodstream. The time needed to reach and affect the brain in their experiment, is in accordance with the timing used in our research.

While the uptake of intranasal OT by brain tissue remains contested, there is evidence that peripheral OT can also affect behavior. Oxytocin receptors are found within the heart (Gutkowska et al., 1997) and gut (Klein, Tamir, & Welch, 2011) among others. OT has been shown to decrease cardiac activity (Gutkowska & Jankowski, 2012) and to facilitate parasympathetic activity (Quintana, Kemp, Alvares, & Guastella, 2013). In a study by Gamer and Büchel (2012), the authors observed that OT had a phasic effect on heart rate, conditional on the valence of the cues with which the participant was confronted. They interpreted this modulation of valence-dependent parasympathetic response as an indication that peripheral oxytocin facilitates assigning motivational value to facial expressions. Because this study did not include neuroimaging, we do not know if there were significant changes in brain activity occurring simultaneously. The study of Quintana et al. (2016) showed that an increase in peripheral OT (administered intravenously) does not affect the amygdala. We leave it open whether or not the behavioral effects of intranasal OT observed in our studies are caused by a direct uptake from the nasal cavity into the brain (affecting the amygdala), or by additional parasympathetic effects instigated by OT in the periphery.

Nevertheless, even if behavioral effects of intranasal OT are well documented and there are plausible theories with regard to its path into the brain, there is not yet a concrete study pinpointing an exact mechanism of OT uptake. A possible research avenue would be the molecular labeling of administered OT to chart the pharmacokinetics in the body.

Third, there is at this moment no “grand” theory of OT social functions. There are many ad hoc explanations for the often contradictory findings, but these are not yet unified in an overarching framework. There are few, if any, consistent main effects of OT on behavior in most experimental studies, which has tempered the initial view of OT as a cuddle hormone that promotes unconditional bonding (Lane, Luminet, Nave, & Mikolajczak, 2016; Nave, Camerer, & McCullough, 2015). Inspired by Kosfeld’s 2005 paper, reporting that administering 4 IU of OT intranasally (compared to a placebo) increased investments in a trust game, many social scientists have conducted experiments to search for associations between OT and prosocial behaviors. The current state of the art was recently reviewed by Nave et al. (2015), who conclude that the convergent evidence that human trust is caused or related to OT is not robust. The effects of intranasal administration of OT on trust do not seem to replicate well. Furthermore, the review reports that endogenous (plasma) levels of OT correlates poorly with individual differences in trusting behavior, and that there are no consistent associations between specific OT-related genetic polymorphisms and trust.

These conflicting results that have emerged from over a decade of OT research do not necessarily have to be a limitation of the research, but rather could reflect a true property of OT. OT is in the first place a neuromodulator known to interact with many other hormones and neurotransmitters in the central nervous system (e.g., dopamine, serotonin (Mottolese, Redouté, Costes, Le Bars, & Sirigu, 2014), testosterone and estrogen (Bos, Panksepp, Bluthe, & van Honk, 2012; Meyer-Lindenberg, Domes, Kirsch, & Heinrichs, 2011)). In addition, a large body of research has accumulated showing that the effects of intranasal OT on social behavior is context-dependent (summarized in Bartz, Zaki, Bolger, and Ochsner (2011)). The relation between OT and trust, for example, was later shown to depend on the presence of social cues. Declerck, Boone, and Kiyonari (2010) compared the cooperative behavior of players engaged in an anonymous, one-shot assurance game (where a cooperative decision hinges on trusting the other player, see chapter 3), after intranasal OT or placebo administration. When two players were allowed to meet briefly before the experiment, OT (compared to placebo) significantly boosted cooperative decisions. But without this cue, OT significantly increased distrust and decreased cooperation.

For the above reasons, the hypotheses set forth in this dissertation assume a modulatory, rather than a main effect of OT, and are theory-driven, based on current knowledge regarding OT interactions with the mesocorticolimbic dopamine system. The results obtained in this dissertation again corroborate that effects of OT on behavior are context-dependent.

3. Future research and applications.

3.1 Follow-up studies on the meta-function of OT

The three leading theories on the social function of OT describe it as being anxiolytic, facilitating approach and increasing social saliency, all of which can be derived from the overarching proposition by Love (2014) and Shamay-Tsoory and Abu-Akel (2015) that OT enhances mesolimbic dopaminergic activity. Churchland and Winkielman (2012) propose that OT has a broad physiological effect on basal functions (e.g., attenuating generalized anxiety) and that the changes in higher-order cognitive functions and subsequent behavior (e.g., trust) are dependent on cortical afferent neurons to the prefrontal cortex.

To deepen our understanding of how OT could affect the integration of cognition and emotion to affect heuristic processing, the recent study by Owen et al. (2013) on the hippocampus is very relevant. This study shows that OT affects fast-spiking interneurons, which has inhibitory effects on nearby pyramidal cells and suppresses spontaneous firing. The presence of OT suppresses background firing, leading to an enhanced signal-to-noise ratio and which improves accurate transfer to the prefrontal cortex. These results are fascinating in the light of understanding the meta-function of OT and can serve as a general framework from which to derive future specific hypotheses. The study described in chapter three suggests that a similar mechanism may be operating in the pathway linking the amygdala to the nucleus accumbens. The experimental paradigm in chapter four can readily be adapted to further test how OT affect information integration. For example, remembered partners carry more information than anonymous partners and would hence be more salient in the social decision making process. We expect OT to increase the signal in the hippocampus more to remembered partners than to anonymous partners. If this leads to improved information transfer, then the valence of the remembered partners should matter more: positively remembered partners would affect decision making differently from negatively remembered partners.

Context dependent effects of OT on social behavior are furthermore known to be influenced by individual differences. Researchers have begun to pay attention to this, showing significant effects of gender (Lischke et al., 2012), social value orientation (Declerck, Boone, & Kiyonari, 2013), and attachment (Bartz, 2012). This is not surprising considering the many interactions of OT with other hormones (e.g., the sex hormones (Bos et al., 2012), serotonin (Mottolese et al., 2014)). So far little research has addressed the developmental patterns by which OT affects social behavior. Considering that OT exposure and social bonding is probably the greatest during breastfeeding, it is likely that individual differences in OT sensitivity are formed during a sensitive period early in life. Already there are indications that children deprived of adequate care in orphanages have a decreased OT

sensitivity and show deficiencies in their socioemotional development (Wismer Fries, Ziegler, Kurian, Jacoris, & Pollak, 2005). These OT insensitivities caused by maternal love withdrawal stay present in grown-up females (Riem et al., 2013), which in turn affects their caregiving abilities (Strathearn, Fonagy, Amico, & Montague, 2009). A recent experiment investigated the effect of OT on how women respond to the sound of crying children and showed that the administration of intranasal OT led to a reduction in amygdala activation and handgrip force (Riem, Bakermans-Kranenburg, & van Ijzendoorn, 2016). This effect was only significant in women with an insecure attachment representation, who show by default emotional, behavioral, and neural hyperreactivity to crying (Riem, Bakermans-Kranenburg, van Ijzendoorn, Out, & Rombouts, 2012).

Shedding further light upon the nature of this sensitive period and on how the effects of OT unfold during development would help to design treatment programs to alleviate problematic social behavior in young adulthood and to prevent such problems from being passed on to the next generation.

3.2 Practical applications in translational medicine

Future research on the social functions of OT will increasingly contribute to advances in translational medicine. First, the therapeutic value of OT has already been proven useful in treatment of neuropsychiatric disorders (Cochran, Fallon, Hill, & Frazier, 2013). Because of the potential positive effect on social behavior, e.g., facilitating approach, OT has been administered in patients with disorders characterized by defects in social skills, such as autism spectrum disorder and schizophrenia. In both these cases, OT administration led to a relief of symptoms. Patients were able to more accurately identify emotions (i.e., improved social cognition) (Averbeck, Bobin, Evans, & Shergill, 2012; Domes et al., 2013; Evans et al., 2011; Guastella et al., 2010; Pedersen et al., 2011; Woolley et al., 2014) even after some time had passed (Hollander et al., 2007). Underlying these disorders are defects in the dopaminergic system, and it has been speculated that OT administration alleviates the effects of dopamine deficiencies (Gordon et al., 2016; Rosenfeld, Lieberman, & Jarskog, 2011).

For some time now there has been a special interest in OT with respect to its role in autism. For patients with autism or autism spectrum disorder (ASD), OT administration has been shown to lead to increased retention of social information (Hollander et al., 2007), improved social behavior (Andari et al., 2010) and decreased repetitive behavior (Hollander et al., 2002). While past research seemed to insinuate that a deficit in the oxytocinergic system can be causally linked to autism, more recent work shows that similar concentrations of endogenous OT are found in both children with ASD and their unaffected siblings (Parker et al., 2014). However low levels of endogenous OT were indeed

correlated with severe social difficulties, while these symptoms were not as pronounced in those children with higher endogenous OT levels. But this link between social skills and OT levels was present in all children, regardless of an ASD diagnosis. Thus, while not specific to autism, endogenous OT appears to be positively related to social functioning, and a recent study on the use of intranasal OT as a therapeutic means showed promising results (Yatawara, Einfeld, Hickie, Davenport, & Guastella, 2016): for five weeks long, 39 children between 3 and 8 years old diagnosed with autism received either oxytocin or a placebo. The results show that social responsiveness was significantly improved in the OT group. This increase in social skills after OT administration is possibly attributed to the increased brain activation of regions associated with social behavior. Gordon et al. (2013) administrated OT to 17 children with ASD and measured brain activity using fMRI while they judged social and non-social pictures. They found an increase in brain regions associated with reward processing and social salience (i.e., the striatum, middle frontal gyrus, medial PFC, superior temporal sulcus, and premotor cortex).

The use of OT as a therapeutic means has also been investigated for other disorders, such as social anxiety (Guastella, Howard, Dadds, Mitchell, & Carson, 2009; Neumann & Slattery, 2016), mood disorders (Pincus et al., 2010; Scantamburlo, Ansseau, Geenen, & Legros, 2011), and substance abuse (Lee & Weerts, 2016; McGregor & Bowen, 2012; Sarnyai & Kovacs, 2014; Zanos et al., 2017), sometimes with mixed results.

Because of its context dependent effects, the administration of OT for therapeutic ends needs to be thoroughly researched and approached with care. Given that OT mostly moderates behavior and that its effects are not unilateral, therapeutically administered OT could sometimes alleviate, but other times enhance symptoms. For example, in an environment with ambiguous social signals, OT might also increase anxiety, defeating the purpose of therapy in treating anxiety disorder. It is therefore that Shamay-Tsoory and Abu-Akel (2015) and Hurlemann (2017) stress the importance of “strictly controlling the therapeutic context to avoid the risk of unfavorable outcomes” (p. 377).

Another problem that hinders the use of oxytocin as a therapeutic treatment is that there may be negative effects from chronic intake. Huang et al. (2014) showed that male mice given a single intranasal dose of OT increase their number of social interactions with females, decrease their interactions with unfamiliar males, and do not change their behavior towards known males. On the other hand, after a chronic treatment of OT, the number and durations of social interactions with both females and males were reduced. So the long-term effects of OT treatment do not necessarily extend its short-term effect. Furthermore, chronic administration led to an decrease of OT receptors in the brain of these mice. More extensive research will definitely be necessary, especially when

targeting developing children. Researches have shown that chronic administration of intranasal OT to young prairie voles induced a change in partner preferences later on when they were of adult age (Bales et al., 2013).Even a single high dose of OT in infant voles induced a change in vasopressin receptors (Bales et al., 2007). While experimental treatments with children have not yet uncovered consistent negative effects, additional studies on the long-term effects of exogenously administered OT on the human (developing) mind and behavior is necessary before efficient treatment can be developed (Guastella & Hickie, 2016; Young, 2013).

3.3 Societal implications

Studying the determinants of social behavior has societal implications. Many psychiatric conditions, like ASD discussed above, are related to deficits in social functions that hinder cooperation. But also for healthy people, engaging in mutually reciprocal cooperative behaviors is fulfilling and essential for personal as well as societal well-being. No community today would exist without cooperation. Yet, establishing cooperation does not occur automatically because selfish people can gain more by free-riding on the cooperative efforts of others. At the same time, cooperative people are reluctant to do so because they fear exploitation by greedy free-riders. Knowing what motivates people to cooperate makes it possible to design environments in which trust and cooperation thrive.

In this spirit, governments in several countries are initiating ‘nudging’ programs. Nudging is a term that describes the ‘easy and cheap’ subtle manipulations that policy-makers and regulators can make to induce the desired behavior by offering contextual cues that hijack the heuristic processes underlying decision making (Thaler & Sunstein, 2008). Thus a nudge gently biases the decision-maker in the direction of what is considered to be normative, without creating a feeling of being obliged to do so. The advantage is that the desired behavior is implemented voluntarily, and that no external enforcement is necessary. Nudging already has led to positive results in many different domains, such as an increase in organ donations, more hygienic urinals, and an overall healthier lifestyle (Sousa Lourenco, Ciriolo, Rafael Rodrigues Vieira De Almeida, & Troussard, 2016).

Finally, the last chapter of this dissertation also points to the need of considering individual differences in value orientations when designing institutions that promote cooperation. In organizations members are typically heterogeneous with respect to how much they value personal gain over the collective benefits that can be achieved. Proself individuals may prefer to serve themselves first, but they will adapt their behavior to the situation if it is in their advantage to do so. Thus they can learn that cooperating is a viable strategy in an environment that rewards cooperation in the long term. For prosocial individuals cooperation can be achieved by creating a strong norm-sensitive environment. Other research has pointed to the importance of transparency and

punishment because such an environment facilitates the exposure and subsequent discouragement of potential free-riding (Fehr & Gächter, 2000).

By studying the biological roots of cooperation, this dissertation has added to the growing body of knowledge that much of our behavior is driven by values and unconscious computations. It also gives us hints on how different social cues and pay-off structures can steer behavior towards cooperation.

References

Andari, E., Duhamel, J.-R., Zalla, T., Herbrecht, E., Leboyer, M., & Sirigu, A. (2010). Promoting social behavior with oxytocin in high-functioning autism spectrum disorders. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.0910249107

Averbeck, B. B., Bobin, T., Evans, S., & Shergill, S. S. (2012). Emotion recognition and oxytocin in patients with schizophrenia. *Psychological Medicine, 42*(2), 259-266. doi: 10.1017/s0033291711001413

Bales, K. L., Perkeybile, A. M., Conley, O. G., Lee, M. H., Guoynes, C. D., Downing, G. M., . . . Mendoza, S. P. (2013). Chronic intranasal oxytocin causes long-term impairments in partner preference formation in male prairie voles. *Biological Psychiatry, 74*(3), 180-188. doi: 10.1016/j.biopsych.2012.08.025

Bales, K. L., Plotsky, P. M., Young, L. J., Lim, M. M., Grotte, N., Ferrer, E., & Carter, C. S. (2007). Neonatal oxytocin manipulations have long-lasting, sexually dimorphic effects on vasopressin receptors. *Neuroscience, 144*(1), 38-45. doi: 10.1016/j.neuroscience.2006.09.009

Bartz, J. A. (2012). Oxytocin, attachment, betrayal and self-interest: A commentary on "Oxytocin modulates the link between adult attachment and cooperation through reduced betrayal aversion" by Carsten KW De Dreu, Psychoneuroendocrinology, doi:10.1016/j.psyneuen.2011.10.003. *Psychoneuroendocrinology, 37*(7), 1106-1108. doi: 10.1016/j.psyneuen.2012.03.003

Bartz, J. A., Zaki, J., Bolger, N., & Ochsner, K. N. (2011). Social effects of oxytocin in humans: Context and person matter. *Trends in Cognitive Sciences, 15*(7), 301-309. doi: 10.1016/j.tics.2011.05.002

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron, 58*(4), 639-650. doi: 10.1016/j.neuron.2008.04.009

Born, J., Lange, T., Kern, W., McGregor, G. P., Bickel, U., & Fehm, H. L. (2002). Sniffing neuropeptides: A transnasal approach to the human brain. *Nature Neuroscience, 5*(6), 514-516. doi: 10.1038/nn849

Bos, P. A., Panksepp, J., Bluthe, R. M., & van Honk, J. (2012). Acute effects of steroid hormones and neuropeptides on human social-emotional behavior: A review of single administration studies. *Frontiers in Neuroendocrinology, 33*(1), 17-35. doi: 10.1016/j.yfrne.2011.01.002

Chang, S. W. C., & Platt, M. L. (2014). Oxytocin and social cognition in rhesus macaques: Implications for understanding and treating human psychopathology. *Brain Research, 1580*, 57-68. doi: 10.1016/j.brainres.2013.11.006

Churchland, P. S., & Winkielman, P. (2012). Modulating social behavior with oxytocin: How does it work? What does it mean? *Hormones and Behavior, 61*(3), 392-399. doi: 10.1016/j.yhbeh.2011.12.003

Cochran, D., Fallon, D., Hill, M., & Frazier, J. A. (2013). "The role of oxytocin in psychiatric disorders: A review of biological and therapeutic research findings". *Harvard Review of Psychiatry, 21*(5), 219-247. doi: 10.1097/HRP.0b013e3182a75b7d

Dal Monte, O., Noble, P. L., Turchi, J., Cummins, A., & Averbeck, B. B. (2014). CSF and blood oxytocin concentration changes following intranasal delivery in macaque. *PLoS ONE, 9*(8). doi: 10.1371/journal.pone.0103677

Declerck, C. H., Boone, C., & Kiyonari, T. (2010). Oxytocin and cooperation under conditions of uncertainty: The modulating role of incentives and social information. *Hormones and Behavior, 57*(3), 368-374. doi: 10.1016/j.yhbeh.2010.01.006

Declerck, C. H., Boone, C., & Kiyonari, T. (2013). The effect of oxytocin on cooperation in a prisoner's dilemma depends on the social context and a person's social value orientation. *Social Cognitive and Affective Neuroscience*, DOI: 10.1093/scan/nst1040.

Domes, G., Heinrichs, M., Kumbier, E., Grossmann, A., Hauenstein, K., & Herpertz, S. C. (2013). Effects of intranasal oxytocin on the neural basis of face processing in autism spectrum disorder. *Biological Psychiatry, 74*(3), 164-171. doi: 10.1016/j.biopsych.2013.02.007

Evans, S., Shergill, S. S., Chouhan, V., Bristow, E., Collier, T., & Averbeck, B. B. (2011). Patients with schizophrenia show increased aversion to angry faces in an associative learning task. *Psychological Medicine, 41*(7), 1471-1479. doi: 10.1017/s0033291710001960

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *The American Economic Review, 90*(4), 980-994.

Gamer, M., & Büchel, C. (2012). Oxytocin specifically enhances valence-dependent parasympathetic responses. *Psychoneuroendocrinology, 37*(1), 87-93. doi: 10.1016/j.psyneuen.2011.05.007

Gigerenzer, P., & Todd, P. M. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Gordon, I., Jack, A., Pretzsch, C. M., Vander Wyk, B., Leckman, J. F., Feldman, R., & Pelphrey, K. A. (2016). Intranasal oxytocin enhances connectivity in the neural circuitry supporting social motivation and social perception in children with autism. *Scientific Reports, 6*. doi: 10.1038/srep35054

Gordon, I., Vander Wyk, B. C., Bennett, R. H., Cordeaux, C., Lucas, M. V., Eilbott, J. A., . . . Pelphrey, K. A. (2013). Oxytocin enhances brain function in children with autism. *Proceedings of the National Academy of Sciences of the United States of America, 110*(52), 20953-20958. doi: 10.1073/pnas.1312857110

Guastella, A. J., Einfeld, S. L., Gray, K. M., Rinehart, N. J., Tonge, B. J., Lambert, T. J., & Hickie, I. B. (2010). Intranasal oxytocin improves emotion recognition for youth with autism spectrum disorders. *Biological Psychiatry, 67*(7), 692-694. doi: 10.1016/j.biopsych.2009.09.020

Guastella, A. J., & Hickie, I. B. (2016). Oxytocin treatment, circuitry, and autism: A critical review of the literature placing oxytocin into the autism context. *Biological Psychiatry, 79*(3), 234-242. doi: 10.1016/j.biopsych.2015.06.028

Guastella, A. J., Howard, A. L., Dadds, M. R., Mitchell, P., & Carson, D. S. (2009). A randomized controlled trial of intranasal oxytocin as an adjunct to exposure therapy for social anxiety disorder. *Psychoneuroendocrinology, 34*(6), 917-923. doi: 10.1016/j.psyneuen.2009.01.005

Gutkowska, J., & Jankowski, M. (2012). Oxytocin revisited: Its role in cardiovascular regulation. *Journal of Neuroendocrinology, 24*(4), 599-608. doi: 10.1111/j.1365-2826.2011.02235.x

Gutkowska, J., Jankowski, M., Lambert, C., Mukaddam-Daher, S., Zingg, H. H., & McCann, S. M. (1997). Oxytocin releases atrial natriuretic peptide by combining with oxytocin receptors in the heart. *Proceedings of the National Academy of Sciences of the United States of America, 94*(21), 11704-11709.

Hollander, E., Bartz, J., Chaplin, W., Phillips, A., Sumner, J., Soorya, L., . . . Wasserman, S. (2007). Oxytocin increases retention of social cognition in autism. *Biological Psychiatry, 61*(4), 498-503. doi: 10.1016/j.bipsych.2006.05.030

Hollander, E., Novotny, S., Hanratty, M., Yaffe, R., DeCaria, C. M., Aronowitz, B. R., & Mosovich, S. (2002). Oxytocin infusion reduces repetitive behaviors in adults with autistic and Asperger's disorders. *Neuropsychopharmacology, 28*(1), 193-198.

Huang, H., Michetti, C., Busnelli, M., Managò, F., Sannino, S., Scheggia, D., . . . Papaleo, F. (2014). Chronic and acute intranasal oxytocin produce divergent social effects in mice. *Neuropsychopharmacology, 39*(5), 1102-1114. doi: 10.1038/npp.2013.310

Hurlemann, R. (2017). Oxytocin-augmented psychotherapy: Beware of context. *Neuropsychopharmacology, 42*(1), 377-378. doi: 10.1038/npp.2016.188

Klein, B. Y., Tamir, H., & Welch, M. G. (2011). PI3K/Akt responses to oxytocin stimulation in Caco2BB gut cells. *Journal of Cellular Biochemistry, 112*(11), 3216-3226. doi: 10.1002/jcb.23243

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature, 435*(7042), 673-676. doi: 10.1038/nature03701

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience – The dangers of double dipping. *Nature Neuroscience, 12*(5), 535-540. doi: 10.1038/nn.2303

Lane, A., Luminet, O., Nave, G., & Mikolajczak, M. (2016). Is there a publication bias in behavioural intranasal oxytocin research on humans? Opening the file drawer of one laboratory. *Journal of Neuroendocrinology, 28*(4). doi: 10.1111/jne.12384

Lee, M. R., & Weerts, E. M. (2016). Oxytocin for the treatment of drug and alcohol use disorders. *Behavioural Pharmacology, 27*(8), 640-648. doi: 10.1097/fbp.0000000000000258

Leng, G., & Ludwig, M. (2015). Intranasal oxytocin: Myths and delusions. *Biological Psychiatry*. doi: 10.1016/j.biopsych.2015.05.003

Lischke, A., Gamer, M., Berger, C., Grossmann, A., Hauenstein, K., Heinrichs, M., . . . Domes, G. (2012). Oxytocin increases amygdala reactivity to threatening scenes in females. *Psychoneuroendocrinology, 37*(9), 1431-1438. doi: 10.1016/j.psyneuen.2012.01.011

Love, T. M. (2014). Oxytocin, motivation and the role of dopamine. *Pharmacology, Biochemistry, and Behavior, 119*, 49-60. doi: 10.1016/j.pbb.2013.06.011

McEwen, B. B. (2004). Brain–fluid barriers: Relevance for theoretical controversies regarding vasopressin and oxytocin memory research *Advances in Pharmacology* (Vol. Volume 50, pp. 531-592): Academic Press.

McGregor, I. S., & Bowen, M. T. (2012). Breaking the loop: Oxytocin as a potential treatment for drug addiction. *Hormones and Behavior, 61*(3), 331-339. doi: 10.1016/j.yhbeh.2011.12.001

Meyer-Lindenberg, A., Domes, G., Kirsch, P., & Heinrichs, M. (2011). Oxytocin and vasopressin in the human brain: Social neuropeptides for translational medicine. *Nature Reviews Neuroscience, 12*(9), 524-538.

Modi, M. E., Connor-Stroud, F., Landgraf, R., Young, L. J., & Parr, L. A. (2014). Aerosolized oxytocin increases cerebrospinal fluid oxytocin in rhesus macaques. *Psychoneuroendocrinology, 45*, 49-57. doi: 10.1016/j.psyneuen.2014.02.011

Mottolese, R., Redouté, J., Costes, N., Le Bars, D., & Sirigu, A. (2014). Switching brain serotonin with oxytocin. *Proceedings of the National Academy of Sciences, 111*(23), 8637-8642. doi: 10.1073/pnas.1319810111

Nave, G., Camerer, C., & McCullough, M. (2015). Does oxytocin increase trust in humans? A critical review of research. *Perspectives on Psychological Science, 10*(6), 772-789. doi: 10.1177/1745691615600138

Neumann, I. D., Maloumby, R., Beiderbeck, D. I., Lukas, M., & Landgraf, R. (2013). Increased brain and plasma oxytocin after nasal and peripheral administration in rats and mice. *Psychoneuroendocrinology, 38*(10), 1985-1993. doi: 10.1016/j.psyneuen.2013.03.003

Neumann, I. D., & Slattery, D. A. (2016). Oxytocin in general anxiety and social fear: A translational approach. *Biological Psychiatry, 79*(3), 213-221. doi: 10.1016/j.biopsych.2015.06.004

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K. J., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science, 304*(5669), 452-454. doi: 10.1126/science.1094285

Owen, S. F., Tuncdemir, S. N., Bader, P. L., Tirko, N. N., Fishell, G., & Tsien, R. W. (2013). Oxytocin enhances hippocampal spike transmission by modulating fast-spiking interneurons. *Nature, 500*(7463), 458-462. doi: 10.1038/nature12330

Parker, K. J., Garner, J. P., Libove, R. A., Hyde, S. A., Hornbeak, K. B., Carson, D. S., . . . Hardan, A. Y. (2014). Plasma oxytocin concentrations and OXTR polymorphisms predict social impairments in children with and without autism spectrum disorder. *Proceedings of the National Academy of Sciences, 111*(33), 12258-12263. doi: 10.1073/pnas.1402236111

Pedersen, C. A., Gibson, C. M., Rau, S. W., Salimi, K., Smedley, K. L., Casey, R. L., . . . Penn, D. L. (2011). Intranasal oxytocin reduces psychotic symptoms and improves Theory of Mind and social perception in schizophrenia. *Schizophrenia Research, 132*(1), 50-53. doi: 10.1016/j.schres.2011.07.027

Pincus, D., Kose, S., Arana, A., Johnson, K., Morgan, P. S., Borckardt, J., . . . Nahas, Z. (2010). Inverse effects of oxytocin on attributing mental activity to others in depressed and healthy subjects: A double-blind placebo controlled fMRI study. *Frontiers in Psychiatry, 1*, 134. doi: 10.3389/fpsyt.2010.00134

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences, 10*(2), 59-63. doi: 10.1016/j.tics.2005.12.004

Quintana, D. S., Kemp, A. H., Alvares, G. A., & Guastella, A. J. (2013). A role for autonomic cardiac control in the effects of oxytocin on social behavior and psychiatric illness. *Frontiers in Neuroscience, 7*, 48. doi: 10.3389/fnins.2013.00048

Quintana, D. S., Westlye, L. T., Alnæs, D., Rustan, Ø. G., Kaufmann, T., Smerud, K. T., . . . Andreassen, O. A. (2016). Low dose intranasal oxytocin delivered with breath powered device dampens amygdala response to emotional stimuli: A peripheral effect-controlled within-subjects randomized dose-response fMRI trial. *Psychoneuroendocrinology, 69*, 180-188. doi: 10.1016/j.psyneuen.2016.04.010

Rault, J.-L. (2016). Effects of positive and negative human contacts and intranasal oxytocin on cerebrospinal fluid oxytocin. *Psychoneuroendocrinology, 69*, 60-66. doi: 10.1016/j.psyneuen.2016.03.015

Riem, M. M. E., Bakermans-Kranenburg, M. J., & van Ijzendoorn, M. H. (2016). Intranasal administration of oxytocin modulates behavioral and amygdala responses to infant crying in females with insecure attachment representations. *Attachment & Human Development, 18*(3), 213-234. doi: 10.1080/14616734.2016.1149872

Riem, M. M. E., Bakermans-Kranenburg, M. J., van Ijzendoorn, M. H., Out, D., & Rombouts, S. A. R. B. (2012). Attachment in the brain: Adult attachment representations predict amygdala and behavioral responses to infant crying. *Attachment & Human Development, 14*(6), 533-551. doi: 10.1080/14616734.2012.727252

Riem, M. M. E., van Ijzendoorn, M. H., Tops, M., Boksem, M. A. S., Rombouts, S. A. R. B., & Bakermans-Kranenburg, M. J. (2013). Oxytocin effects on complex brain networks are moderated by experiences of maternal love withdrawal. *European Neuropsychopharmacology, 23*(10), 1288-1295. doi: 10.1016/j.euroneuro.2013.01.011

Rosenfeld, A. J., Lieberman, J. A., & Jarskog, L. F. (2011). Oxytocin, dopamine, and the amygdala: A neurofunctional model of social cognitive deficits in schizophrenia. *Schizophrenia Bulletin, 37*(5), 1077-1087. doi: 10.1093/schbul/sbq015

Sarnyai, Z., & Kovacs, G. L. (2014). Oxytocin in learning and addiction: From early discoveries to the present. *Pharmacology Biochemistry and Behavior, 119*, 3-9. doi: 10.1016/j.pbb.2013.11.019

Scantamburlo, G., Ansseau, M., Geenen, V., & Legros, J. J. (2011). Intranasal oxytocin as an adjunct to escitalopram in major depression. *Journal of Neuropsychiatry and Clinical Neurosciences, 23*(2), E5. doi: 10.1176/appi.neuropsych.23.2.E5 & 10.1176/jnp.23.2.jnpe5

Shamay-Tsoory, S. G., & Abu-Akel, A. (2015). The social salience hypothesis of oxytocin. *Biological Psychiatry, 79*(3), 194-202. doi: 10.1016/j.biopsych.2015.07.020

Sousa Lourenco, J., Ciriolo, E., Rafael Rodrigues Vieira De Almeida, S., & Troussard, X. (2016). *Behavioural insights applied to policy - European report 2016*. Publications Office of the European Union.

Strathearn, L., Fonagy, P., Amico, J., & Montague, P. R. (2009). Adult attachment predicts maternal brain and oxytocin response to infant cues. *Neuropsychopharmacology, 34*(13), 2655-2666. doi: 10.1038/npp.2009.103

Striepens, N., Kendrick, K. M., Hanking, V., Landgraf, R., Wüllner, U., Maier, W., & Hurlemann, R. (2013). Elevated cerebrospinal fluid and blood concentrations of oxytocin following its intranasal administration in humans. *Scientific Reports, 3*, 3440. doi: 10.1038/srep03440

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, Conn.: Yale University Press.

Thorne, R. G., Pronk, G. J., Padmanabhan, V., & Frey Ii, W. H. (2004). Delivery of insulin-like growth factor-I to the rat brain and spinal cord along olfactory and trigeminal pathways following intranasal administration. *Neuroscience, 127*(2), 481-496. doi: 10.1016/j.neuroscience.2004.05.029

Wismer Fries, A. B., Ziegler, T. E., Kurian, J. R., Jacoris, S., & Pollak, S. D. (2005). Early experience in humans is associated with changes in neuropeptides critical for regulating social behavior. *Proceedings of the National Academy of Sciences of the United States of America, 102*(47), 17237-17240. doi: 10.1073/pnas.0504767102

Woolley, J. D., Chuang, B., Lam, O., Lai, W., O'Donovan, A., Rankin, K. P., . . . Vinogradov, S. (2014). Oxytocin administration enhances controlled social cognition in patients with schizophrenia. *Psychoneuroendocrinology, 47*, 116-125. doi: 10.1016/j.psyneuen.2014.04.024

Yatawara, C. J., Einfeld, S. L., Hickie, I. B., Davenport, T. A., & Guastella, A. J. (2016). The effect of oxytocin nasal spray on social interaction deficits observed in young children with autism: A randomized clinical crossover trial. *Molecular Psychiatry, 21*(9), 1225-1231. doi: 10.1038/mp.2015.162

Young, L. J. (2013). When too much of a good thing is bad: Chronic oxytocin, development, and social impairments. *Biological Psychiatry, 74*(3), 160-161. doi: 10.1016/j.biopsych.2013.05.015

Zanos, P., Georgiou, P., Weber, C., Robinson, F., Kouimtsidis, C., Niforooshan, R., & Bailey, A. (2017). Oxytocin and opioid addiction revisited: Old drug, new applications. *British Journal of Pharmacology*. doi: 10.1111/bph.13757

Summary

A defining feature of human social groups is the high level of cooperation, which has made it possible for us to form societies and build economies. But achieving cooperation is not an obvious task because our social groups are highly complex and often interdependent: decisions in social interactions do not only affect ourselves, but also impact others, and vice versa. This makes social decisions strategic: we decide in function of what we anticipate the others will do. To do so quickly and effectively depends on several different factors.

First, humans have a strong self-interest motive driving behavior that maximizes positive outcomes for themselves. This motive can still lead to cooperation when both parties know that they are better off working together. But when resources are scarce and the gain of one person equals the loss of another, the self-interest motive leads to competition. Second, humans also have social motives that prompt behavior that benefits others, sometimes even at a cost to themselves. The tendency to engage in such prosocial behaviors is predicted by an individual's social value orientation (SVO), a psychological trait corresponding to stable preferences in how resources are divided, regardless of incentives. But prosocial behavior in the absence of cooperative incentives makes individuals vulnerable to betrayal. Therefore, an important third factor is the presence of social cues that relay trustworthiness information about the interaction partners.

While the role of these three factors with respect to human cooperation has received much coverage in psychology, economics and decision neuroscience, they have been mostly investigated independently, with insufficient attention to how they interact. The aim of this dissertation is to fill this gap and to study how combinations of these factors are integrated by the brain and how they may contribute to heuristic processes and subsequent (non)cooperative behavior.

In the first two chapters, we study how social cues are evaluated and how the neurohormone oxytocin (OT) affects 'fast and frugal' social judgments. Chapter one describes an experiment in which we test whether OT increases perceived trustworthiness of faces, and/or whether it improves the discriminatory ability of trustworthiness perception. In the second chapter, participants indicated (i) how (un)healthy they judged faces, and (ii) how disgusting they found certain situations. This allowed us to further assess whether OT increases cautious behavior as previously described in the literature.

In the third chapter, we report the results of an experiment in which we combine functional magnetic resonance imaging (fMRI) with intranasal OT administration. We investigate if OT

influences heuristic decision making in economic games that incentivize either cooperation or competition. We additionally include contextual social cues (expressions of neutral or angry faces) that are informative with respect to assessing the trustworthiness/aggressiveness of interaction partners. Thus we combine different sources of social information in an fMRI experiment to investigate the context-dependent neural and behavioral effects of OT.

In the fourth and final chapter, we investigate how cooperation can develop over time as individuals are accumulating more and more trustworthiness information of different partners. With fMRI we identify which brain regions are recruited in the process of learning to cooperate. In this experimental setting individuals can learn from previous encounters, and accumulate more and more trustworthiness information of different partners. We are particularly interested in how individual differences in social value orientation shape how trust is formed over time.

Together, the four chapters of this dissertation reveal how cooperative behavior can be established given that social interactions are often permeated by many, sometimes conflicting, sources of information. To arrive at a cooperative decision, an individual evaluates the decision context, computes pay-offs, and matches the expected outcome with his/her own social preferences. By using a combination of behavioral experiments, pharmacological interventions, and neuroimaging, we shed light on some of the biological roots of cooperation and contribute to a growing body of knowledge that a substantial part of our behavior is driven by values and unconscious computation.

Samenvatting

Het is eigen aan een groep van mensen om onderling te willen samenwerken. Deze eigenschap maakt het ons mogelijk om handel te drijven en samenlevingen te vormen. Het is echter niet vanzelfsprekend om tot samenwerking of coöperatie te komen, omdat sociale middens vaak complex zijn, en we van elkaar afhangen: een beslissing in een sociale interactie heeft niet enkel invloed op onszelf, maar ook op anderen, en visa versa. Daarom is beslissen in een sociale context een strategisch proces: we beslissen in functie van wat we denken dat anderen zullen doen. Om dit beslissingsproces vlot en efficiënt te laten verlopen, laten we ons leiden door verschillende factoren.

Ten eerste zijn mensen sterk gemotiveerd om in hun eigenbelang te handelen. Dit kan wel tot samenwerking leiden, maar enkel indien de interactiepartners denken dat deze samenwerking hun het meeste oplevert. Wanneer grondstoffen schaars zijn, en de winst van de één een verlies voor de ander betekent, kan dit eigenbelang competitief gedrag in de hand werken. Ten tweede zijn mensen ook gevoelig voor sociale belangen, wat kan leiden tot altruïstisch gedrag, zelfs indien dit gepaard gaat met nadelen voor henzelf. De neiging van een persoon om zich op dergelijke pro-sociale manier te gedragen, wordt voorspeld door zijn *social value orientation* (SVO). Deze psychologische maat beschrijft de stabiele voorkeuren van een persoon over hoe grondstoffen verdeeld moeten worden tussen zichzelf en anderen, ongeacht de mogelijke winst. Echter, pro-sociaal gedrag zonder garanties tot samenwerking maakt personen kwetsbaar voor verraad en uitbuiting. Een derde belangrijke factor is daarom de aanwezigheid van sociale informatie die een beeld kan geven van de betrouwbaarheid van de interactiepartner(s).

Hoewel deze drie factoren belangrijk zijn om samenwerking tot stand te laten komen, en ze uitvoerig besproken worden in de psychologie, economie en (beslissings-)neurowetenschappen, worden ze steeds afzonderlijk bestudeerd, zonder te kijken naar hun onderlinge interacties. Deze dissertatie probeert deze leemte op te vullen door te onderzoeken hoe deze factoren in het brein geïntegreerd worden, en hoe deze combinatie van factoren de heuristische processen en daaropvolgend (non-) coöperatief gedrag beïnvloedt.

In de eerste twee hoofstukken bestuderen we hoe sociale signalen geëvalueerd worden en hoe het neurohormoon oxytocine (OT) het sociale oordelen beïnvloedt. Het eerste hoofdstuk beschrijft een experiment waarin we testen of het toedienen van OT aan deelnemers hen de betrouwbaarheid van tegenspelers hoger laat inschatten, en of het hen toelaat om een accurater oordeel van tegenspelers te vormen. In het tweede hoofdstuk interpreteerden de deelnemers de gezondheidstoestand van

tegenspelers en gaven ze hun afkeer over enkele situaties weer. Dit experiment liet ons toe om na te gaan in hoeverre de wetenschappelijke notie dat OT personen aanzet tot voorzichtigheid klopt.

In het derde hoofdstuk geven we de resultaten van een experiment waarin we *functional magnetic resonance imaging* (fMRI) combineren met het toedienen van OT. We onderzoeken of OT een invloed heeft op het nemen van heuristische beslissingen in economische spelen waarin coöperatie of competitie gepromoot wordt. Tijdens dit experiment verschijnen er ook sociale signalen (gezichten met een neutrale of kwade gelaatsuitdrukking) die informatie verschaffen over de betrouwbaarheid van de interactiepartners. Op deze manier combineren we verschillende bronnen van contextuele informatie in een fMRI experiment, en kunnen we de contextafhankelijke effecten van OT op hersenactiviteit en gedrag onderzoeken.

In het vierde en laatste hoofdstuk onderzoeken we hoe samenwerking kan ontwikkelen in de tijd wanneer er continu sociale informatie verzameld wordt. We gebruiken opnieuw fMRI technieken, om die hersenregio's die gebruikt worden om anderen te leren vertrouwen te identificeren. In deze experimentele setting kunnen individuen leren van vorige interacties en steeds meer informatie over de betrouwbaarheid van de partners verzamelen. We zijn bovendien geïnteresseerd in hoe individuele verschillen in SVO de vorming van vertrouwen beïnvloedt.

Samen schetsen deze vier hoofdstukken een beeld over hoe coöperatief gedrag gevormd kan worden, gegeven de verschillende vormen van, soms tegenstrijdige, informatie die beschikbaar zijn tijdens sociale interacties. Om tot een coöperatieve beslissing te komen moet een individu de context beoordelen, de mogelijke winst bereken en de verwachte uitkomst vergelijken met zijn of haar eigen sociale voorkeuren. Door gebruik te maken van een combinatie van gedragsexperimenten, toediening van hormonen en neuro-imaging technieken kunnen we licht werpen op de biologische oorsprong van samenwerking en zodoende bijdragen aan een steeds groter wordende kennis over hoe een belangrijk deel van ons gedrag gedreven wordt door waarden en onbewuste berekeningen.

Academic Curriculum Vitae

Personalia

Bruno Lambert
°20/01/1987, Merksem
Melsele

Education

2012 – present:	PhD student, University of Antwerp, Belgium
2011 – present	Teaching Degree (biology and physics) University of Antwerp, Belgium
2008 – 2010	Master in Bioscience Engineering: Cell and Gene Biotechnology University of Ghent, Belgium
2005 – 2008	Bachelor in Bioscience Engineering University of Antwerp, Belgium

Work experience

2017	Physics Teacher Sint-Lucas KSO, Antwerp, Belgium
2010 – 2011	Research Assistant Department of Bioscience Engineering, University of Antwerp, Belgium

Publications

Lambert, B., Declerck, C. H., Emonds, G., & Boone, C. (2017). Trust as commodity: social value orientation affects the neural substrates of learning to cooperate. Soc Cogn Affect Neurosci. doi: 10.1093/scan/nsw170

Declerck, C. H., Lambert, B., & Boone, C. (2014). Sexual dimorphism in oxytocin responses to health perception and disgust, with implications for theories on pathogen detection. Hormones and Behavior, 65(5), 521-526. doi: 10.1016/j.yhbeh.2014.04.010

Lambert, B., Declerck, C. H., & Boone, C. (2014). Oxytocin does not make a face appear more trustworthy but improves the accuracy of trustworthiness judgments. Psychoneuroendocrinology, 40(0), 60-68. doi: 10.1016/j.psyneuen.2013.10.015

Oral Presentations

Lambert, B., Declerck, C. H., Boone, C. & Parizel P. M.; "Oxytocin fine-tunes decision making in social dilemmas: cooperate if it pays off, but aggress only when you can win!" NeuroPsychoEconomics Conference. June 3rd, 2016. Bonn, Germany

Lambert, B., Declerck, C. H., Emonds, G., & Boone, C.; "Learning to cooperate differs with social value orientation: an fMRI study." Doctoral Day organized by the Department of Applied Economics. November 25th 2015. University of Antwerp, Belgium.

Lambert, B., Declerck, C. H., Emonds, G., & Boone, C.; "The neural foundation of learning to cooperate reflects an individual's social value orientation." NeuroPsychoEconomics Conference. June 18th – 19th, 2015. Copenhagen, Denmark.

Lambert, B., Declerck, C. H., Emonds, G., & Boone, C.; "The neural foundation of learning to cooperate reflects an individual's social value orientation." UZA Research Club. April 22nd, 2015. Antwerp, Belgium.

Lambert, B., Declerck, C. H., Emonds, G., & Boone, C.; "The influence of social value orientation and trust on the processing of negative reciprocity: an fMRI study". 10th NeuroPsychoEconomics Conference. May 30th, 2014. Munich, Germany.

Declerck, C. H., & Lambert, B.; "How did the neuropeptide oxytocin contribute to the evolution of prosocial behavior." Workshop organized by the Human Evolution and Behavior Network (HEBEN). November 24th, 2013. Ghent, Belgium.

Lambert, B., Declerck, C. H., & Boone, C.; "Oxytocin does not make a face appear more trustworthy but improves the accuracy of trustworthiness judgments." NeuroPsychoEconomics Conference. June 7th, 2013.Bonn, Germany.

Poster presentations

Lambert, B., Declerck, C. H., & Boone, C.; "Oxytocin does not make a face appear more trustworthy but improves the accuracy of trustworthiness judgments." Annual Meeting of the Society for Neuroeconomics. September 27th, 2013. Lausanne, Swiss.

Lambert, B., Declerck, C. H., & Boone, C.; "Oxytocin does not make a face appear more trustworthy but improves the accuracy of trustworthiness judgments." Symposium on Stress, Brain and Behavior. September 26th, 2013. Lausanne, Swiss.

Awards

Applied Neuroscience Award for "The neural foundation of learning to cooperate reflects an individual's social value orientation." NeuroPsychoEconomics Conference. June 18th – 19th, 2015. Copenhagen, Denmark.

Runner-up Applied Neuroscience Award for "Oxytocin fine-tunes decision making in social dilemmas: cooperate if it pays off, but aggress only when you can win!" NeuroPsychoEconomics Conference. June 3rd, 2016. Bonn, Germany.