l-DOPA and oxytocin influence the neurocomputational mechanisms of self-benefitting and prosocial reinforcement learning.

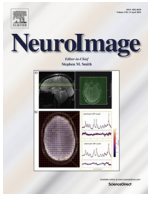
Jansen M, Lockwood PL, Cutler J, de Bruijn ERA

NeuroImage 270 (2023) 119983

Contents lists available at ScienceDirect

NeuroImage

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



Social Neuroscience

L-DOPA and oxytocin influence the neurocomputational mechanisms of self-benefitting and prosocial reinforcement learning

Myrthe Jansen[a,b,*], Patricia L. Lockwood[c,d,e], Jo Cutler[c,d,e], Ellen R.A. de Bruijn[a,b]

[a] Department of Clinical Psychology, Leiden University, the Netherlands
[b] Leiden Institute for Brain and Cognition (LIBC), Leiden, the Netherlands
[c] Centre for Human Brain Health, School of Psychology, University of Birmingham, Birmingham, UK
[d] Institute for Mental Health, School of Psychology, University of Birmingham, Birmingham, UK
[e] Centre for Developmental Science, School of Psychology, University of Birmingham, UK

ARTICLE INFO

Keywords:
Reinforcement learning
Prediction error
Dopamine
Oxytocin
Prosocial behavior

ABSTRACT

Humans learn through reinforcement, particularly when outcomes are unexpected. Recent research suggests similar mechanisms drive how we learn to benefit other people, that is, how we learn to be prosocial. Yet the neurochemical mechanisms underlying such prosocial computations remain poorly understood. Here, we investigated whether pharmacological manipulation of oxytocin and dopamine influence the neurocomputational mechanisms underlying self-benefitting and prosocial reinforcement learning. Using a double-blind placebo-controlled crossover design, we administered intranasal oxytocin (24 IU), dopamine precursor L-DOPA (100 mg + 25 mg carbidopa), or placebo over three sessions. Participants performed a probabilistic reinforcement learning task with potential rewards for themselves, another participant, or no one, during functional magnetic resonance imaging. Computational models of reinforcement learning were used to calculate prediction errors (PEs) and learning rates. Participants behavior was best explained by a model with different learning rates for each recipient, but these were unaffected by either drug. On the neural level, however, both drugs blunted PE signaling in the ventral striatum and led to negative signaling of PEs in the anterior mid-cingulate cortex, dorsolateral prefrontal cortex, inferior parietal gyrus, and precentral gyrus, compared to placebo, and regardless of recipient. Oxytocin (versus placebo) administration was additionally associated with opposing tracking of self-benefitting versus prosocial PEs in dorsal anterior cingulate cortex, insula and superior temporal gyrus. These findings suggest that both L-DOPA and oxytocin induce a context-independent shift from positive towards negative tracking of PEs during learning. Moreover, oxytocin may have opposing effects on PE signaling when learning to benefit oneself versus another.

1. Introduction

Learning about the consequences of our actions is critical for successful and adaptive functioning as it enables us to maximize rewards. According to reinforcement learning theory (RLT), we learn to form associations between actions and outcomes through prediction errors (PEs), which signal the discrepancy between expected and experienced outcomes (Rescorla et al., 1972; Sutton and Barto, 2018). PEs drive learning by updating our expectations of the value of a particular action or stimulus. It does so together with the learning rate, which quantifies the degree to which new information updates the expected value by scaling the PE. The feasibility of RLT as a framework for understanding brain-behavior associations is supported by a wealth of research indicating that midbrain dopamine (DA) neurons code PEs, showing increased firing after better-than-expected outcomes (positive PEs) and reduced firing when outcomes are worse than expected (negative PEs) (Schultz, 2016). RLT therefore provides an important framework for understanding learning based on reinforcers, and points to a critical role of the dopaminergic system in learning.

Neuroimaging studies have used RLT to identify the brain regions in which blood-oxygen-level-dependent (BOLD) responses correlate with PEs (Fouragnan et al., 2018; Garrison et al., 2013). These studies have shown responsiveness to positive (versus negative) PEs in regions that are associated with reward processing, such as the ventral striatum (VS) and ventromedial prefrontal cortex (vmPFC). In contrast, responsiveness to negative (versus positive) PEs has been associated with activity in networks thought to be linked to salience- and performance monitoring and known to be involved in the regulation of alertness and switch-

* Corresponding author at: Department of Clinical Psychology, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, the Netherlands.
E-mail address: m.jansen@fsw.leidenuniv.nl (M. Jansen).

https://doi.org/10.1016/j.neuroimage.2023.119983.
Received 21 November 2022; Received in revised form 3 February 2023; Accepted 23 February 2023

ing behavior, such as the anterior mid-cingulate cortex (aMCC), pre-supplementary motor area (pre-SMA), anterior insula (AI), dorsomedial prefrontal cortex (dmPFC), dorsolateral prefrontal cortex (dlPFC), inferior parietal gyrus (IPG), amygdala and thalamus (Fouragnan et al., 2018). Additionally, a comparable network of brain regions that includes the aMCC, dorsal mid-cingulate cortex (dMCC), AI, precentral gyrus, IPG, dorsal striatum, and midbrain has been implicated in the coding of 'surprise' PEs, which signal the unexpectedness of outcomes, independent of their valence (Fouragnan et al., 2018), and there is evidence that these 'unsigned' or 'absolute' PEs are also encoded by (specific) DA neurons (see e.g., Diederen and Fletcher, 2021).

A wide range of pharmacological studies on reinforcement learning and performance monitoring indicate that PE coding in above regions is modulated by DA (Barnes et al., 2014; De Bruijn, Hulstijn, Verkes, Ruigt, and Sabbe, 2004, 2006; Diederen et al., 2017; Forster et al., 2017; Jocham et al., 2011, 2014; Pessiglione et al., 2006; Santesso et al., 2009; Spronk et al., 2016; Zirnheld et al., 2004). Most studies point towards enhanced positive and negative PE signaling after administration of DA agonists (Barnes et al., 2014; De Bruijn et al., 2004; De Bruijn, Hulstijn, Verkes, Ruigt, and Sabbe, 2005a; Pessiglione et al., 2006; Santesso et al., 2009; Spronk et al., 2016) and reduced signaling after DA antagonists (De Bruijn et al., 2006; Diederen et al., 2017; Jocham et al., 2011, 2014; Pessiglione et al., 2006; Zirnheld et al., 2004). Pharmacological alterations of (learning) performance have been observed as well (Diederen et al., 2017; Guitart-Masip et al., 2014; Pessiglione et al., 2006; Pizzagalli et al., 2008; Santesso et al., 2009; Vo et al., 2018, 2016; Zirnheld et al., 2004), suggesting that pharmacological stimulation of DA may impact how we learn to obtain rewards by enhancing neural PE encoding in brain regions associated with salience and reward processing.

Importantly, to behave in a socially adaptive manner we do not only need to learn how to obtain rewards for ourselves but also how to benefit others, since actions that intend to benefit others, also referred to as prosocial behaviors, are crucial for social functioning and the formation and maintenance of reciprocal social relationships (Carlo, 2013). Prior research suggests that this so-called "prosocial learning" relies on the same reinforcement algorithms as individual or self-benefitting learning (Cutler et al., 2021; Lockwood et al., 2016; Martins et al., 2022; Westhoff et al., 2021). However, whereas the encoding of PEs during both types of learning involved the VS, only prosocial PEs were found to be encoded in the subgenual anterior cingulate cortex (sgACC), with stronger responses in this latter brain region for those with higher levels of empathy, suggesting social specificity on the implementation level (Lockwood et al., 2016; Martins et al., 2022 but see; Westhoff et al., 2021).

Nevertheless, how PE encoding during prosocial learning is affected by DA is less clear. Research indicates that social rewards and prediction errors activate the same DA-innervated regions (e.g., nucleus accumbens and ventral tegmental area) as non-social rewards do (see e.g., Kohls et al., 2012; Manduca et al., 2021). Studies in rodents have additionally shown that DA neurons appear to also code social reward prediction errors (Solié et al., 2022), suggesting that DA may impact PE encoding and learning in a domain-general way. However, other studies have shown, for example, that administration of the DA precursor L-DOPA increased selfish behavior in healthy men in an economic bargaining game (Pedroni et al., 2014) and reduced the hyperaltruistic tendency of preferring harming oneself over harming others in a harm aversion task (Crockett et al., 2015). Conversely, blocking DA transmission using amisulpride during an interpersonal decision reduced prosocial behavior in women and selfish behavior in men (Soutschek et al., 2017). This suggests that upregulating DA might also stimulate a self-serving bias where personal outcomes become more salient compared to the rewards of others, at least in men.

Importantly, a recent study by Martins et al. (2022) demonstrated that oxytocin (OT), a neuropeptide implicated in many aspects of social behavior (see e.g., Y. Ma et al., 2016), may specifically modulate prosocial learning and its neurocomputational mechanisms. They reported dose-dependent effects of intranasal OT administration on PE encoding in the midbrain and sgACC specifically during prosocial learning, with a low dose (9 IU) of OT increasing- and higher doses (18 and 36 IU) of OT decreasing PE encoding in these regions. The low OT dose also prevented a decreased in prosocial performance over time. Another study comparing self-benefitting with prosocial learning found that a 24-IU dose of OT specifically reduced self-benefitting performance (Liao et al., 2021). Our own lab recently demonstrated OT-induced enhancements of early electrophysiological brain responses to errors thought to reflect negative PEs only in a social context (de Bruijn, Ruissen, and Radke, 2017) whereas another study found that OT increased amplitudes of a later electrophysiological brain component associated with error awareness and motivation in response to social feedback and enhanced learning from positive social feedback (Zhuang et al., 2021). These finding are congruent with the social salience hypothesis, which proposes that OT specifically targets social processes by altering the salience of social cues through its interaction with the brain's dopaminergic system (Shamay-Tsoory and Abu-Akel, 2016). There is indeed substantial evidence that OT acts on the mesolimbic DA system to mediate social rewards (see e.g., Borland et al., 2019; Hung et al., 2017). For example, research in mice indicates that OT facilitates the release of DA during social interaction (Hung et al., 2017). Importantly, however, the majority of studies reporting links between DA and OT were performed in preclinical populations and further determination of this association in humans is critically required. Here, we will address this gap by manipulating OT and DA levels in a single study within the same subjects. With this, we can investigate potential neurocomputational overlap in the effects of these neurochemicals for the first time.

Hence, the current study aimed to improve our understanding of the role of OT and DA in self-benefitting and prosocial reinforcement learning. Disentangling the neurochemical mechanisms that support learning in a prosocial context and its underlying neural computations is important, given that disruptions in reinforcement learning (Maia and Frank, 2011) and prosocial behavior (e.g., Mayer et al., 2018; Walsh et al., 2021) are key transdiagnostic factors. Studying these mechanisms may therefore provide more insights into clinical conditions and may lead to the identification of targets for interventions. Using a double blind, cross-over design, healthy male adults ($N = 30$) were administered DA precursor L-DOPA, OT, or placebo across three sessions. In each session (Fig. 1A), they performed a probabilistic reinforcement learning task (Fig. 1B) while in the MRI scanner. In this task, participants were asked to choose between two abstract symbols. One of these symbols was associated with a high probability of obtaining reward (75%), while the other symbol was associated with a low reward probability (25%). These probabilities were not known to the participants but had to be learned through trial and error. Importantly, this task was performed in three different conditions where they could gain rewards for themselves (self-benefitting learning), an anonymous other participant (prosocial learning), or no one (non-social control condition). We then applied computational reinforcement learning models to the behavior during the task to obtain relevant learning parameters and trial-by-trial PE computations. We focused on the 'signed' PE, which incorporate both the valence (better- versus worse-than-expected) and the degree of unexpectedness of outcomes, and assessed both positive and negative correlations of this parametric PE with BOLD responses (Lockwood and Klein-Flügge, 2021). Based on the literature discussed above, we hypothesized that L-DOPA would enhance learning and PE encoding in those regions known to be implicated in processing reward (positive) as well as salience (negative/surprise) PEs (Fouragnan et al., 2018) either in a domain-general or a self-serving manner whereas OT would specifically enhance learning and PE encoding in these regions in a prosocial context. We performed region-of-interest (ROI) analyses focusing on the VS, sgACC and midbrain given previous research implicating these regions in self-benefitting or prosocial learning (Lockwood et al., 2016; Martins et al., 2022; Westhoff et al., 2021) and additionally performed

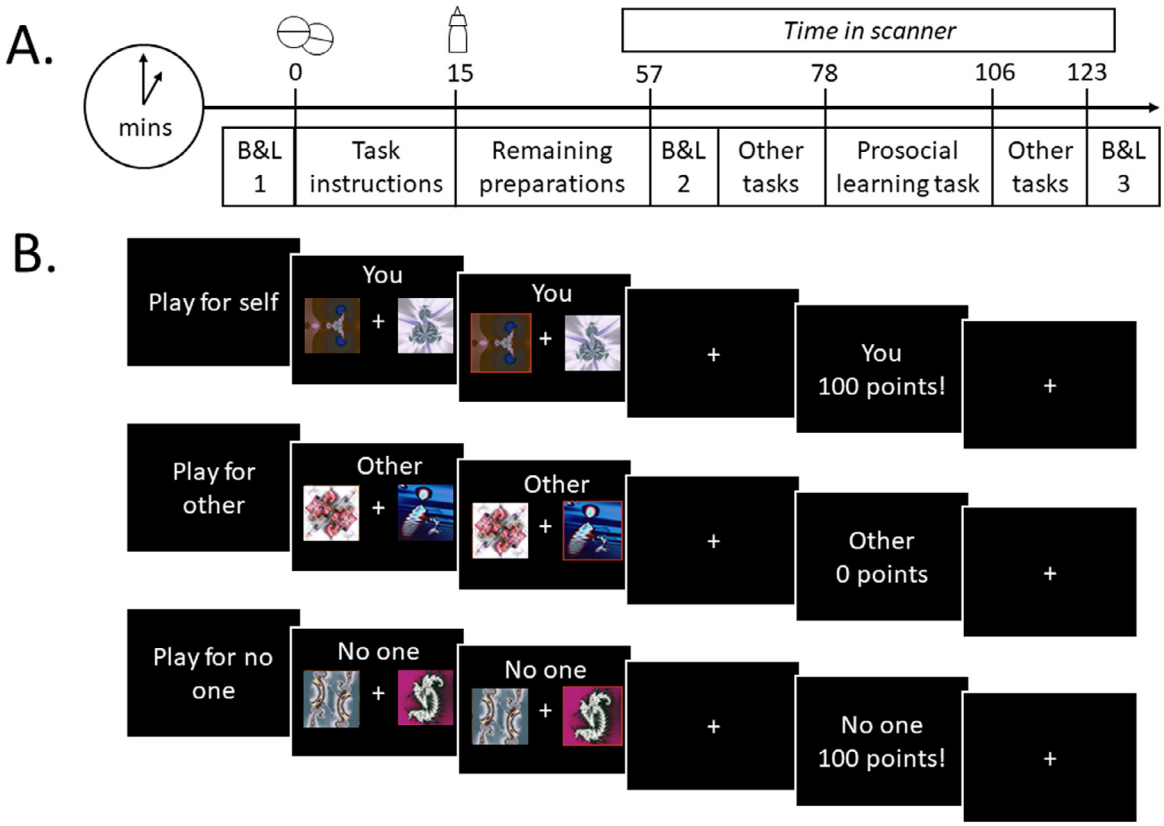


Fig. 1. Overview of the time table for each session (A) and the prosocial learning task (B). (A) During each session participants received a tablet followed by a nasal spray exactly 15 min later (either a placebo tablet followed by a placebo spray, a placebo tablet followed by a spray containing oxytocin, or a L-DOPA tablet followed by placebo spray). On average 78 min (*SD* = 3.4) after administration of the tablet and 63 min (*SD* = 3.4) after administration of the nasal spray, participants performed the prosocial learning task in the scanner. Subjective drug effects (alertness, mood and anxiety) were assessed using the Bond and Lader (1974) mood rating scale (B&L) right before administration of the tablet (T1), approximately one hour later (T2), and at the end of the visit (T3). (B) In this probabilistic reinforcement learning task, participants had to learn the probability that abstract symbols were rewarded to gain points. At the start of each block, a text screen (2000 ms) indicated whether participants performed that block for their own outcomes, for the outcomes of an anonymous other participant, or for no one. Then, each trial started with the presentation of the two symbols (cue), for a maximum of 3000 ms. When a button was pressed, the selected option was highlighted (for 3500 ms – the reaction time to the cue), or when no response was given, the words "Too late" appeared for 500 ms. Then, after a 2000 ms fixation cross, the outcome (100 vs. 0 points) was presented for 1000 ms. Each trial ended with a jittered fixation cross (2000 – 4000 ms). During the cue and outcome screen, the recipient condition was always indicated by the words "you / other / no one". The task was presented using E-prime 3.0 software (Psychology Software Tools, Pittsburgh, PA).

whole-brain analyses to examine potential drug effects in the broad range of regions implicated in coding salience PEs (Fouragnan et al., 2018).

2. Materials and methods

2.1. Participants

Based on a medium effect size (f = 0.25) found in previous studies investigating similar pharmacological manipulations (e.g., Soutschek et al., 2017; Vo et al., 2018, 2016), with 85% statistical power and a correlation between repeated measures of 0.5, an a priori power calculation showed that a sample size of 26 would be sufficient to detect effects. Taking into account potential drop-out (10%), 30 healthy, right-handed males between age 18 and 35 (M = 22.8, SD = 3.6) with a good command of the Dutch language were included in this study. We chose to only include males in order to avoid menstrual cycle-dependent interactions between the dopaminergic system and gonadal steroids (e.g., Jocham et al., 2011, 2014). Exclusion criteria were as follows: history of cardiovascular, endocrine, psychiatric, neurological or hematological disease, counter-indications to MRI, history of medication or drugs within 1 month prior to the start of the study with the exception of occasional use of paracetamol, previous experience of allergic reaction upon administration of a drug, participation in another drug study within 3 months preceding participation in the current study, intake of more than 3 units of alcohol a day, smoking more than 5 cigarettes a day and scoring > 8 on the anxiety or depression subscale of the Hospital Anxiety and Depression Scale (Spinhoven et al., 1997). Participants were asked to refrain from using caffeine, alcohol and smoking 24 h prior to drug administration and to refrain from eating and drinking (except water) 1.5 h prior to arrival at the laboratory. Participants received a monetary compensation of €140 after completion of the experiment. The study was approved by the medical ethical committee of the Leiden University Medical center (P19.031) and was conducted in accordance with the latest version of the declaration of Helsinki.

2.2. Procedure

Participants were recruited using advertisements on social media including Facebook and via the Leiden University Research Participation System (SONA). Participants indicating their willingness to participate in the study received the information letter of the study by email and filled out an online screening questionnaire in Qualtrics. Participants meeting inclusion criteria were assessed for MRI contraindications by telephone. Each participant visited the LUMC three times, each visit separated by at least one week (mean- / mode- / max interval = 13.6 / 7 / 46 days) to ensure complete washout of the drugs (Jocham et al., 2014). All visits took place between 9:00 AM and 5:30 PM, with the time of day for

the visits always being the same for each session of a participant (maximum time difference: 1.5 h). Using a double-blind cross-over design, participants received a tablet containing either L-DOPA (100 mg in combination with 25 mg carbidopa) or a placebo tablet containing no active ingredients, followed by a nasal spray containing either oxytocin (24 intranasal units / one puff of 0.37 ml per nostril) or a placebo (chlorobutanol nasal spray). This dosage of L-DOPA has been shown to reach maximum concentration in plasma after approximately 50 min and has an elimination half-life of approximately 80–90 min (Nyholm et al., 2012). Hence, we expected that drug effects would be sustained during the task, which started approximately 78 min after administration of the tablet (see Fig. 1A for a timetable). Evidence for effects of intranasal OT administration on central levels of OT has mostly been indirect or stem from animal work (Quintana et al., 2021). While many studies test effects of OT already after 20 to 60 min after administration, research has also suggested that a 24-IU dose of OT may take up to 75 min to increase cerebrospinal fluid concentrations of OT (Striepens et al., 2013). Given that a recent study found that the most robust effects on amygdala reactivity were observed between 45 and 70 min after 24 IU of OT (Spengler et al., 2017), we aimed to start within this timeframe. Participants performed the prosocial learning task (see Experimental Task) in the MRI scanner by using fiber-optic buttons that were attached to their legs. We also assessed subjective effects of the drugs at three different timepoint during the session (see Fig. 1A, Supplemental results and Table S6). The order of drug administration across visits was randomized by the Clinical Pharmacy of the LUMC so that both researchers and participants were blind to the drug condition. While in the scanner, respiration and heart rate of participants was continuously monitored using a breathing belt and finger clip peripheral pulse unit. After the task and at the end of every session, participants were asked to complete some statements measuring subjective responses to the task (see Supplementary Results and Table S5). Besides the prosocial learning task, we assessed a resting-state task, a performance-monitoring paradigm, and a task measuring working memory, which will be reported in separate papers. The procedure was the same for each session.

2.3. Experimental task

The prosocial learning task (Lockwood et al., 2016) is a probabilistic reinforcement learning task, whereby participants completed a series of trials on which they were asked to choose between two abstract symbols (Fig. 1B). One of these symbols was associated with a high probability of obtaining reward (75%), while the other symbol was associated with a low probability of reward (25%). These probabilities were not known to the participants but they had to be learned through trial and error. Participants performed the task in three different conditions. They were informed that during the task, they could earn points that would be converted to money at the end of the study, whereby they would either play for their own monetary bonus (self-benefitting learning), the bonus of another participant (prosocial learning) or neither's bonus (non-social control condition). The order of these conditions were counterbalanced between participants. They were additionally informed that the other participant would be randomly chosen and would remain anonymous, and that this participant in his turn played for someone else's bonus, in order to prevent feelings of reciprocity. Central to our manipulation, none of the participants reported disbelief in the other person at debriefing. Even though the confederate was not present in the lab, previous work has shown that it is beliefs that are important in evoking neural activity and behavior related to social decision-making (reviewed in Lockwood et al., 2020). At the end of the final visit, participants were informed that, for ethical reasons, they would not receive a monetary bonus based on task performance, but would instead receive a fixed bonus of 10 euros.

Each symbol pair was presented for 16 trials, with the location (left vs. right) of the symbols randomized across trials, after which a new symbol pair was presented in a new block. In total, participants performed 3 blocks of 16 trials per recipient condition, resulting in a total of 144 trials per session. The order of the conditions were pseudorandomized, with the same condition never appearing twice in a row.

2.4. Computational modeling

In line with previous studies (Cutler et al., 2021; Lockwood et al., 2016; Martins et al., 2022; Westhoff et al., 2021), we modelled learning during the prosocial learning task using a standard reinforcement learning model (Rescorla et al., 1972). The Rescorla-Wagner rule (Eq. (1)) states that individuals form an expectation of the value of an action or stimulus (i) for each future trial ($t+1$) as a function of the current expected value $Q_t(i)$ and the prediction error δ_t, which is the difference between the current expectation and the actually received outcome R_t (where 1 is reward and 0 is no reward). The prediction error is scaled by the learning rate α (bounded between 0 and 1), which determines to what extent the prediction error on the current trials is used to update the expected value. This means that lower learning rate indicate that new information has less influence on the expected value.

$$Q_{t+1}(i) = Q_t(i) + a * \underbrace{\left[R_t - Q_t(i)\right]}_{Prediction\ error\ \delta_t} \tag{1}$$

The softmax choice rule provides a model for how people use these expected values to guide their decisions by quantifying the relation between the expected value of stimulus or action i and the probability of choosing this stimulus or action on trial t. The model (Eq. (2)) assumes that people usually choose the option with the highest expected value, but occasionally explore other options or make a mistake by selecting an option with a lower value. The degree of this exploration or noisiness is determined by the temperature parameter β, where higher β values indicate a larger degree of randomness or exploration in choices, whereas lower β values indicate that the actions or stimulus with the highest expected value is chosen more often.

$$p_t\left[\left(i|Q_t(i)\right)\right] = \frac{e^{(Q_t(i)/\beta)}}{\sum_{i'} e^{(Q_t(i')/\beta)}} \tag{2}$$

2.5. Model fitting

Model fitting was done using MATLAB 2019b (The MathWorks Inc). The model was fitted across the drug sessions, which provides a more conservative comparison compared to performing model fitting separately for each session. An iterative maximum a posteriori (MAP) approach was used, which consists of first estimating the model parameters α and β for each participant using maximum likelihood estimation (MLE), and then estimating the parameters again using priors. This approach is less susceptible to outliers compared to using a single step MLE. Group-level Gaussians were initialized as uninformative priors with means of 0.1 (plus some added noise) and variance of 100. During the expectation, model parameters (α and β) were estimated for each participant using the MLE approach, calculating the log-likelihood of the subject's series of choices given the model. We then computed the maximum posterior probability estimate, given the observed choices and given the prior computed from the group-level Gaussian, and recomputed the Gaussian distribution over parameters during the maximization step. We repeated expectation and maximization steps iteratively until convergence of the posterior likelihood summed over the group or a maximum of 800 steps. Convergence was defined as a change in posterior likelihood <0.001 from one iteration to the next. Note that bounded free parameters were transformed from the Gaussian space into the native model space via appropriate link functions (e.g., a sigmoid function in the case of the learning rates) to ensure accurate parameter estimation near the bounds. The detailed code used for model fitting can be found on the Open Science Framework: https://doi.org/10.17605/OSF.IO/9XZDH.

Table 1
Overview of candidate computational models and model selection criteria. We tested four different variations of the classical Rescorla Wagner model which differed in whether learning rates and beta parameters were fitted separately for the recipient conditions or not. All models were pooled across drug conditions. Our model selection procedure was based on three criteria: the integrated Bayesian Information Criteria (iBIC, lower is better), choice probability (higher is better) (R^2) and the exceedance probability of each model (higher is better). The $3\alpha 1\beta$ model (2) was the winning model according to all three criteria.

Model	Learning rate (α)	Beta (β)	iBIC	Choice probability (R^2)	Exceedance probability
1	A	β	10,837	0.701	0.38
2	α_{self}, α_{other}, $\alpha_{no\ one}$	β	10,713	0.724	0.62
3	α_{self}, $\alpha_{not\text{-}self}$	β	10,820	0.713	0.00
4	α_{self}, α_{other}, $\alpha_{no\ one}$	β_{self}, β_{other}, $\beta_{no\ one}$	10,960	0.721	0.00

2.6. Model comparison

In line with previous work (Cutler et al., 2021; Lockwood et al., 2016), we compared four different learning models, which differed in whether parameters α and β were modelled separately for each recipient condition or not. For model comparison, we calculated the Laplace approximation of the log model evidence (more positive values indicating better model fit) and submitted these to a random-effects analysis using the spm_BMS routine from SPM12 (http://www.fil.ion.ucl.ac.uk/spm/software/spm12/). This generates the exceedance probability: the posterior probability that each model is the most likely of the model set in the population (higher is better, over 0.95 indicates strong evidence in favor of a model). For the models of real participant data, we also calculated the integrated Bayesian Information Criteria (iBIC, Huys et al., 2011; Wittmann et al., 2020) (lower is better) and R^2 as additional measures of model fit. To calculate the model R^2, we extracted the choice probabilities generated for each participant on each trial from the winning model. We then took the squared median choice probability across participants. Table 1 displays the different models and the iBIC, exceedance probability and R^2 for each model. Model 2 (3α, 1β) was the winning model according to all three criteria. There was no significant difference in median choice probability of this model between the placebo ($R^2 = 0.67$), L-DOPA ($R^2 = 0.84$) and oxytocin ($R^2 = 0.72$) sessions, Zs < 1.4, Ps > 0.2. Table S1 describes model fit and comparisons for each drug session separately.

2.7. Simulation experiments

Model identifiability was established in a previous study using the exact same task and models (Cutler et al., 2021). This study also showed strong parameter recovery using 1296 simulated participants will all combinations of alpha and beta values: 0, 0.2, 0.4, 0.6, 0.8 and 1 (see Figure S1A for confusion matrix), for details see Cutler et al. (2021). Here, we additionally demonstrated recoverability of the parameter estimates using 90 synthetic participants with parameter values drawn randomly. Strong Pearson's correlations were obtained between the true simulated and fitted parameter values (all rs > 0.62, all Ps < 0.001, see Figure S1B for confusion matrix), suggesting our experiment was well suited to estimate the model's parameters. A simulation experiment plotting 3000 synthetic learning rates against performance also indicated an optimal learning rate of approximately 0.55 in this task (Cutler et al., 2021). As in this previous study, the range of α values for our participants was below this peak, such that higher learning rates are associated with better performance. In line with this, we observed positive relations between learning rates from our winning model and performance (see Supplemental results).

2.8. Behavioral analysis

All data analyses were performed in Rstudio version 1.3.959 (Team, 2020). One-sample t-tests were used to investigate whether participants selected the option with higher probability of being rewarded above chance (0.5) in each drug and recipient condition separately. The learning rates were analyzed with linear mixed models (LMMs) using the lme4 package (Bates et al., 2014) using drug (placebo, L-DOPA, oxytocin) and recipient (self, other, no-one) as fixed effects. Beta parameters were analyzed with an LMM including only the fixed effect of drug. All mixed models contained random intercepts for participants to account for dependency in the data. The random-effects structure for each model was determined according to the procedure described in (Bates et al., 2015), which consists of 1) fitting the maximal random-effects structure or, if the maximal model does not converge or is degenerate, fitting a reduced zero-correlation parameter model, 2) removing random effects estimated at zero or close to zero that do not result in a significant loss of goodness of fit according to a likelihood-ratio test, and 3) extending the model with correlation parameters again (only) if this improves model fit. The final random-effects structure of each model can be found in the analysis script on https://doi.org/10.17605/OSF.IO/9XZDH after publication. All categorical predictor variables were deviation coded. Post-hoc tests were performed using the emmeans package (Lenth et al., 2018). All post-hoc comparisons were corrected for multiple testing using a false discovery rate (FDR) at $P < 0.05$ (Benjamini and Hochberg, 1995). Correlations between learning rates and task performance were computed using the Pearson correlation coefficient.

To quantify the evidence for or against the learning rate effects, we additionally computed Bayes factors (BFs) using a Bayesian repeated measures ANOVA (JASP, 2020) with default priors. BFs represents the probability ratio of observed data under one model versus another and thus provides an index of the relative strength of evidence for the null or alternative hypothesis (Marsman and Wagenmakers, 2017). BFs > 1 and < 1 favor the alternative hypothesis and null hypothesis, respectively. BFs between 0.33–3 are considered anecdotal evidence indicating data insensitivity. BFs between 3-10, 10–30 and 30–100 or between 0.33–0.1, 0.1–0.3, and 0.03–0.01 are considered moderate, strong and very strong evidence for the alternative or null hypothesis, respectively. For significant interaction effects we report the matched models BFinclusion (Clyde et al., 2011), which compares all models with a particular factor to equivalent models without that factor. One participant showed an outlying learning rate (Z-score of 5.07) for the "Placebo – No one" condition, hence for this specific participant this condition was removed from the LMM model and the participant was removed entirely for the Bayesian model.

2.9. fMRI data acquisition and preprocessing

MRI data was acquired at the LUMC using a 3.0T Philips scanner with a 32-channel head coil. 3DT1 structural images were acquired using 155 slices (FOV 195.8 × 250 × 170.5 mm, voxel size=1.1 mm^3, 0 mm slice gap, matrix size = 228 × 177, flip angle = 8°) with a TR of 7.9 ms and a TE of 3.5 ms. Functional scans were acquired in three separate runs (12 min each) using 40 transverse slices in descending order (FOV 220 × 220 × 120.7, matrix size = 80 × 77, voxel size = 2.75 mm^3, slice gap = 0.275 mm, flip angle = 80°) with a TR of 2200 ms and a TE of 30 ms. The first two volumes of each run were discarded to allow for equilibration of T1 saturation effects. Head motion was restricted using foam inserts.

Imaging data were preprocessed and analyzed using SPM12 (Wellcome Trust center for Neuroimaging, University College London). Preprocessing was performed for each session separately and consisted of the following steps: slice-time correction, correction for field-strength inhomogeneity's using b0 field maps, unwarping and realignment, coregistration to subject-specific structural images, segmentation, normalization to MNI space using the DARTEL toolbox (Ashburner, 2007) and smoothing using an 8-mm full width half maximum isotropic Gaussian kernel.

2.10. fMRI analysis and ROI selection

On the first level, we defined a general linear model for each drug session for each participant with separate regressors for the cue and outcome onsets in each condition (cue_self, cue_other, cue_noone, outcome _self, outcome _other, outcome_noone) and for each run. We added parametric modulators to each of these regressors based on our winning computational model. The prediction error δ_t, was added as a parametric modulator at the time of outcome presentation. Additionally, the expected value $Q_t(i)$ (EV) was added at the time of cue presentation for comparability with previous work (Lockwood et al., 2016; Martins et al., 2022; Westhoff et al., 2021). However, as we did not have specific hypotheses about this, we focused our analyses exclusively on the PE. For completeness and comparability with previous work, we show the main effects of expected value per recipient condition after placebo in Table S3. In line with prior work (Martins et al., 2022), the EV and PE were mean-centered at the session level and estimated using the average alpha estimates across all participants and drug conditions, for each recipient condition separately. Additionally, a regressor for the onsets of the instructions at the start of each block was modelled. For missed responses, a separate regressor for missed trials was included as well. We included the 6 realignment parameters for each run to capture residual effects of head motion and additionally included censor regressors (Siegel et al., 2014) for volumes with more than 1 mm scan-to-scan motion or more than 5 mm absolute motion. All events were modelled with zero duration. We excluded two participants from the fMRI analysis for having to censor more than 5% of the volumes in at least two out of the three sessions. Additionally, we excluded a single drug session for two other participants based on this same criteria, resulting in a final sample included in the MRI analysis of $N = 28$ (of which $n = 2$ with only 2 out of 3 sessions).

First-level contrast maps for each condition were submitted to a flexible factorial model with 9 levels, where we computed main and interaction effects for Drug and Recipient using t-contrasts. We created a single region of interest (ROI) mask consisting of the bilateral VS (Harvard-Oxford Atlas), sgACC (Brodmann areas 25 and s24 from the Anatomy Toolbox) and the midbrain (including both the ventral tegmental area and substantia nigra, derived from a high-resolution atlas of subcortical structures (Pauli et al., 2018), see also Martins et al. (2022). Whole brain effects are reported at $P < 0.05$ family-wise error (FWE) corrected at the voxel level and effects in ROIs at $P < 0.05$ FWE-small volume corrected (SVC). One sample- and paired t-tests were used to test extracted cluster estimates.

3. Results

3.1. Behavioral findings

3.1.1. Behavior is best explained by a model with separate learning rates for each recipient condition

We applied computational modeling to quantify learning. In line with previous work (Cutler et al., 2021; Lockwood et al., 2016), we fitted four different reinforcement learning models which differed in whether the key learning parameters of these models were fitted separately for each recipient condition or not. The learning rate (α) represents the speed by which estimates of reward values are updated, whereas the temperature parameter (β) reflects the extent to which participant's choices are deterministic versus more random or exploratory. Bayesian model comparison showed that the winning model included different learning rate parameters for each recipient condition and one single temperature parameter across conditions ($3\alpha1\beta$: αSelf, αOther and αNo one, one β, see Table 1 and 'Methods' for details of the computational model, model fitting and model comparison). Model identifiability and recoverability of the estimated parameters were established in a previous study using the exact same task and model(s) (see Cutler et al., 2021).

Participants learn at a higher rate for others than for self and no one. A linear mixed model (LMM) including the factors recipient and drugs revealed that, independent of the drug condition, learning rates were significantly higher when playing for the other participant compared to when playing for Self ($b = 0.0038$, $SE = 0.0010$, $t = 3.69$, $P < 0.001$, Fig. 2B) and No one ($b = 0.0104$, $SE = 0.0037$, $t = 2.80$, $P < 0.001$), while the difference between playing for Self and No one did not reach significance ($b = -0.0066$, $SE = 0.0035$, $t = -1.88$, $P = 0.07$). A Bayesian repeated measures (rm) ANOVA including the same factors revealed very strong and strong evidence, respectively, for a difference between Other and Self ($BF_{10} = 42.3$) and between Other and No one ($BF_{10} = 11.3$), while there was no conclusive evidence for or against a difference between the Self and No one condition ($BF_{10} = 0.928$).

3.1.2. L-DOPA and oxytocin do not significantly impact learning rates

The learning rate LMM revealed no main or interaction effects involving drugs (all $ts < 1.24$, $Ps > 0.22$). In line with this, the Bayesian rm ANOVA revealed moderate evidence against the presence of an effect for both L-DOPA ($BF_{10} = 0.186$) and OT ($BF_{10} = 0.179$), and also moderate evidence against an interaction between drug and recipient ($BF_{inclusion} = 0.124$).

Analysis of the Beta parameters (Fig. 2C) also did not reveal any drug effects ($ts < 1.47$, $Ps > 0.15$). A Bayesian rm ANOVA on the Beta parameters indicated moderate evidence against an effect of L-DOPA ($BF_{10} = 0.228$) while there was only anecdotal evidence against a difference between OT and placebo ($BF_{10} = 0.578$).

The absence of significant drug effects were not due to a lack of learning in any drug condition (see Fig. 2A & Supplementary Results), non-significant or non-positive associations between performance and learning rates (Table S2 & Supplementary Results), or reaching maximum performance early in each block (see Supplementary Results). Adding session to the model did also not significantly alter effects (see Supplementary Results). Analysis results for choice performance can be found in the Supplementary Results.

3.2. fMRI findings

3.2.1. ROIs - L-DOPA and oxytocin both blunt positive signaling of PEs in ventral striatum (independent of recipient)

Next, we examined BOLD responses that correlated with estimated PE magnitude during feedback presentation. Differences and commonalities in neural encoding of prediction errors between recipients after placebo only (for comparison with previous studies) can be found in the Supplementary Results and Table S3. Importantly, testing an additional

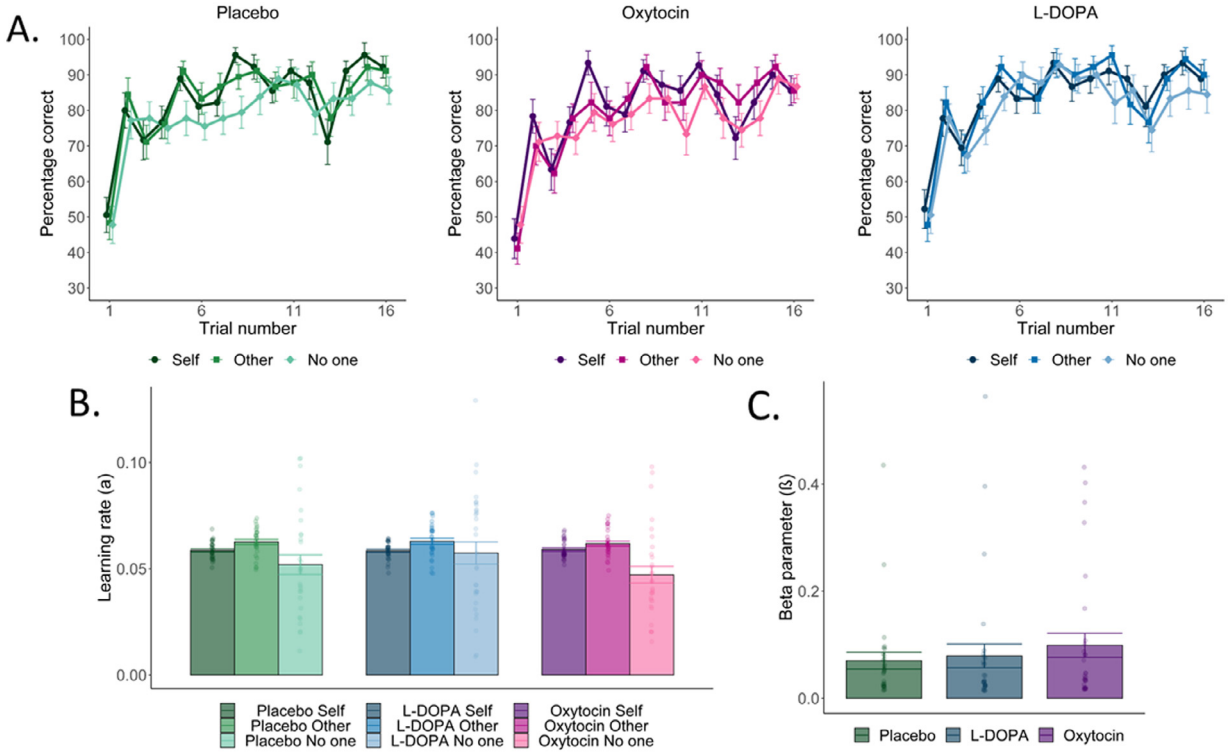


Fig. 2. Choice performance (A), learning rates (B) and beta parameters (C) across conditions. (A) Learning curves for choice behavior in the Self, Other and No one recipients for each drug condition separately showing improved performance for the Self and Other recipients compared to No one, independent of drug condition. Trials are averaged over the three blocks (48 trials in total per recipient presented in three blocks of 16 trials) of the three recipients. Error bars are standard error of the mean. (B) Learning rates (α) from the winning $3\alpha1\beta$ computational model behavior for the Self, Other and No one recipients in each drug condition separately showing higher learning rates when playing for the other compared to self. Error bars display standard error of the mean. (C) Beta estimates (β) from the winning $3\alpha1\beta$ computational model behavior showing no significant differences between drug conditions. Error bars display standard error of the mean.

model with session number as a covariate did not significantly impact any of our fMRI results, confirming that findings do not reflect order effects.

ROI analysis for the main effect of drug (Table S4) revealed a significant effect of L-DOPA (Placebo > L-DOPA) in the left VS ([$x = -14$, $y = 11$, $z = -12$], $k = 10$, $Z = 3.66$, $P = 0.020$ SVC-FWE), which showed significant positive signaling of PEs under placebo ($t(26) = 5.785$, $P < 0.001$) but not after L-DOPA ($t(26) = 1.4$, $P = 0.18$).

ROI analysis also revealed significant main effects of OT (Placebo > OT) in multiple peak clusters in left and right VS ([all $Ps < 0.043$ SVC-FWE, see Table S4 for complete list of peaks). In some of these clusters there was significantly positive signaling after placebo ([$x = -17$, $y = 18$, $z = -2$], $t(26) = 4.61$, $P < 0.001$; [$x = 17$, $y = 20$, $z = -9$], $t(26) = 3.00$, $P = 0.006$; [$x = 6$, $y = 6$, z=−5], $t(26) = 4.59$, $P < 0.001$) but not after OT ($t = -0.91$, $P = 0.37$; $t(27) = -1.33$, $P = 0.20$; and $t = -0.74$, $P = 0.46$, respectively). In contrast, two clusters within the VS mask that lie in the pallidum ([$x = -9$, $y = 3$, $z = -3$], $k = 1$ and [$x = -11$, $y = 5$, $z = -2$], $k = 1$) showed significantly negative signaling after OT ($t(27) = -2.24$, $P = 0.033$; $t(27) = -2.71$, $P = 0.012$, respectively) but not after placebo ($t(26) = 1.59$, $P = 0.12$; $t = 1.37$, $P = 0.18$, respectively).

The reverse contrasts (L-DOPA > Placebo & OT > Placebo) revealed no significant ROI effects and no significant peaks for any of the contrasts were observed in the sgACC or midbrain.

3.2.2. *Whole brain - L-DOPA and oxytocin both lead to negative signaling of PEs in other brain regions (independent of recipient)*

Whole-brain analysis for the contrast Placebo > L-DOPA showed activation in a range of brain regions with peaks in the aMCC extending to the (pre-)SMA, dlPFC, IPG, superior parietal gyrus (SPG), angular gyrus and precentral gyrus (see Table S4 for full list of regions and Fig. 3A). Inspection of the peak coordinates indicates that there was significantly *negative* signaling in each of these clusters after L-DOPA (all $ts > 2.1$, all $Ps < 0.042$). In contrast, after placebo, there was significantly *positive* PE signaling in the dlPFC ([$x = -36$, $y = 32$, $z = 35$], $t(26) = 2.58$, $P = 0.016$) and [$x = 18$, $y = 32$, $z = 32$], $t(26) = 2.06$, $P = 0.049$) whereas there was no significant PE signaling in any of the other regions ($ts < 1.69$, $Ps > 0.10$).

The whole-brain Placebo > OT contrast also revealed significant differences in a wide range of brain regions, including again the aMCC extending to the (pre-)SMA, DLPFC, IPG, SPG, and angular gyrus, as well as the pallidum, caudate, thalamus (anteroventral nucleus) and precuneus (see Table S4 for full list of regions and Fig. 3B). Inspection of the peak coordinates indicates that except for two clusters in the caudate ([$x = 23$, $y = 26$, $z = -9$], $t(27) = -1.36$, $P = 0.19$) and angular gyrus ([$x = 50$, $y = -68$, $z = 33$], $t(27) = -1.82$, $P = 0.081$), there was significant negative signaling after OT in each of these regions ($ts > 2.24$, $Ps < 0.033$). In contrast, after placebo, there was significantly positive signaling of PEs in the caudate ($t(26) = 3.21$, $P = 0.029$), pallidum ([$x = -12$, $y = 0$, $z = 5$], $t(26) = 3.11$, $P = 0.005$), precuneus ([$x = 9$, $y = -56$, $z = 35$], $t(26) = 2.15$, $P = 0.041$), angular gyrus ([$x = 50$, $y = -68$, $z = 33$], $t(26) = 2.45$, $P = 0.021$), and dlPFC ([$x = -29$, $y = 17$, $z = 48$], $t(26) = 2.16$, $P = 0.040$) and no significant signaling in the remaining regions (all $ts < 1.87$, $Ps > 0.072$).

The reverse contrasts (L-DOPA > Placebo & OT > Placebo) revealed no significant whole-brain effects.

3.2.3. *Conjunctions - L-DOPA and oxytocin show overlapping modulatory effects on PE signaling*

A conjunction analysis ([Placebo > L-DOPA]=[Placebo > OT]) within our ROIs revealed significant overlap in signaling for L-DOPA and OT in a peak cluster within the VS ([$x = -14$, $y = 11$, $z = -12$], $k = 10$, $Z = 3.66$, $P = 0.020$ FWE-SVC, see Fig. 3C and Table S4), showing sig-

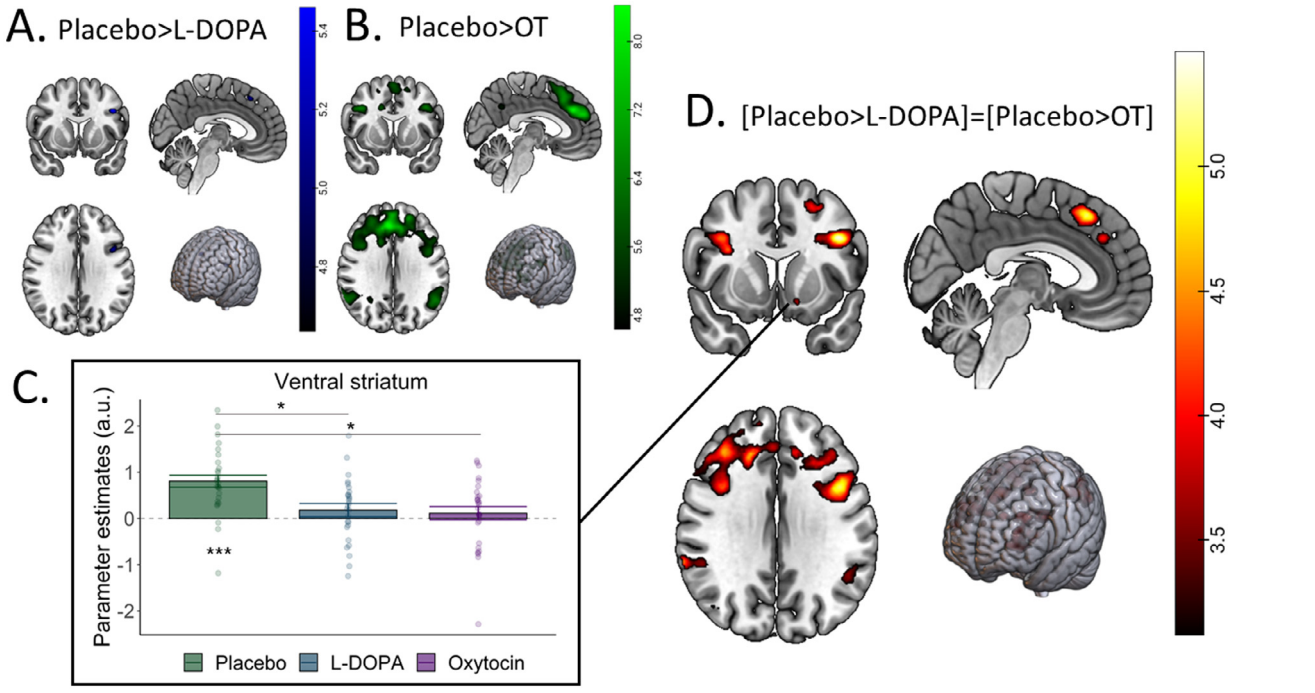


Fig. 3. Whole brain main effects of L-DOPA (A) and Oxytocin (OT) (B), parameter estimates for the ventral striatum conjunction (C) and the whole-brain conjunction of L-DOPA and OT (D). (A) Whole brain activation for the Placebo > L-DOPA contrast, displayed at $P < 0.05$ FWE. The reverse contrast (L-DOPA > Placebo) revealed no significant effects. (B) Whole brain activation for the Placebo > OT contrast, displayed at $P < 0.05$ FWE. The reverse contrast (OT > Placebo) revealed no significant effects. (C) Parameter estimates for the significant conjunction in the ventral striatum [$x = -14$, $y = 11$, $z = -12$]. (D) Whole brain activation for the conjunction analysis ([Placebo > L-DOPA]=[Placebo > OT]) showing overlap in activation for L-DOPA and OT, displayed at $P < 0.001$ uncorrected for illustration purposes. $^{*}P < 0.05$, $^{**}P < 0.01$ $^{***}P < 0.001$. For all whole brain images: $x = 3$, $y = 10$, $z = 30$.

nificantly positive signaling under placebo ($t(27) = 5.79$, $P < 0.001$) but not after L-DOPA ($t(27) = 1.40$, $P = 0.11$) or OT ($t(28) = 0.80$,$P = 0.43$).

A conjunction analysis ([Placebo > L-DOPA]=[Placebo > OT]) on the whole brain level additionally revealed significant overlap in negative signaling for L-DOPA and OT in peak clusters within the aMCC extending to the (pre-)SMA, dlPFC, IPG, SPG, angular gyrus and precentral gyrus (all $Ps < 0.042$, see Fig. 3D and Table S4).

The reverse contrast ([L-DOPA > Placebo]=[OT > Placebo]) did not reveal significant effects.

3.2.4. *Opposing signaling of self and prosocial prediction errors after oxytocin (versus placebo) in dorsal anterior cingulate cortex, insula, and superior temporal gyrus*

We computed interaction contrasts to investigate the effect of self-benefitting versus prosocial PE signaling after OT, which revealed no significant effects within our ROIs. However, on the whole-brain level, the contrast (Placebo:-Self[1]-Other[−1], OT:-Self[−1]-Other[1]) revealed significant interactions between OT and recipient in the dACC ([$x = -17$, $y = 33$, $z = 26$], $Z = 4.90$, $k = 4$, $P = 0.011$ FWE), insula ([$x = 44$, $y = 3$, $z = -12$], $Z = 4.76$, $k = 18$, $P = 0.019$ FWE), and the superior temporal gyrus (STG; [$x = -59$, $y = -2$, $z = -11$], $Z = 6.16$, $k = 577$, $P < 0.001$ FWE) extending to the inferior frontal gyrus, pars orbitalis ([$x = -47$, $y = 17$, $z = -12$], $Z = 5.24$, $P = 0.002$ FWE]). Four significant white matter clusters were observed as well but not interpreted (see Table S4). The reverse contrast (Placebo:-Self[−1]-Other[1], OT:-Self[1]-Other[−1]) revealed no significant effects.

Parameter estimates for each of these regions (Fig. 4BCD) showed that after placebo, there was on average positive signaling for Self versus negative signaling for Other. After OT, however, average signaling in the Self condition was negative, while signaling in the Other condition was either less negative (dACC) or positive (insula, STG). Significant differences from zero and between conditions for each region can be found in Fig. 4 and the Supplemental results.

Similar interaction contrasts for L-DOPA (Placebo:-Self[1]-Other[−1], L-DOPA:-Self[−1]-Other[1]; Placebo:-Self[−1]-Other[1], L-DOPA:-Self[1]-Other[−1]) only revealed a significant whole brain-level cluster in the SPG ([$x = 33$, $y = -69$, $z = 56$], $Z = 4.77$, $P = 0.019$ FWE, $k = 10$), see Supplemental results and Figure S2.

3.2.5. *After oxytocin, prosocial prediction error signals in dorsal anterior cingulate cortex, insula and superior temporal gyrus are associated with higher prosocial learning rates*

We next explored whether these differential neural patterns for self-benefitting versus prosocial learning under influence of OT would be related to behavioral differences in the rate of learning. For each of these above regions, we observed significant correlations between learning rates and parameter estimates in the Other condition after OT (dACC: $r(28) = 0.507$, $P = 0.006$; insula: $r(28) = 0.537$, $P = 0.003$; STG: $r(28) = 0.547$, $P = 0.003$, respectively, see Fig. 5A-C) but not after placebo ($rs < 0.205$, $Ps > 0.31$), though the correlations between the difference in prosocial learning rates and parameter estimates between OT versus placebo were not significant ($rs < 0.249$, $Ps > 0.22$). For the dACC and insula we additionally found that after OT, more positive parameter estimates in the Other versus Self condition related to higher prosocial versus self-benefitting learning rates ($r(28) = 0.391$, $P = 0.040$ and $r(28) = 0.420$, $P = 0.026$, respectively, Fig. 5D and E), which was not found after placebo ($rs < 0.203$, $Ps > 0.32$), though again the direct comparisons between placebo and OT of these prosocial versus self-benefitting rates and parameter estimates were not significantly correlated ($rs < 0.063$, $Ps > 0.76$).

4. Discussion

The current study examined the neurochemical mechanisms underlying self-benefitting versus prosocial reinforcement learning by pharmacologically manipulating DA and OT levels in the brain. While we did not observe a modulating impact of L-DOPA and OT on the compu-

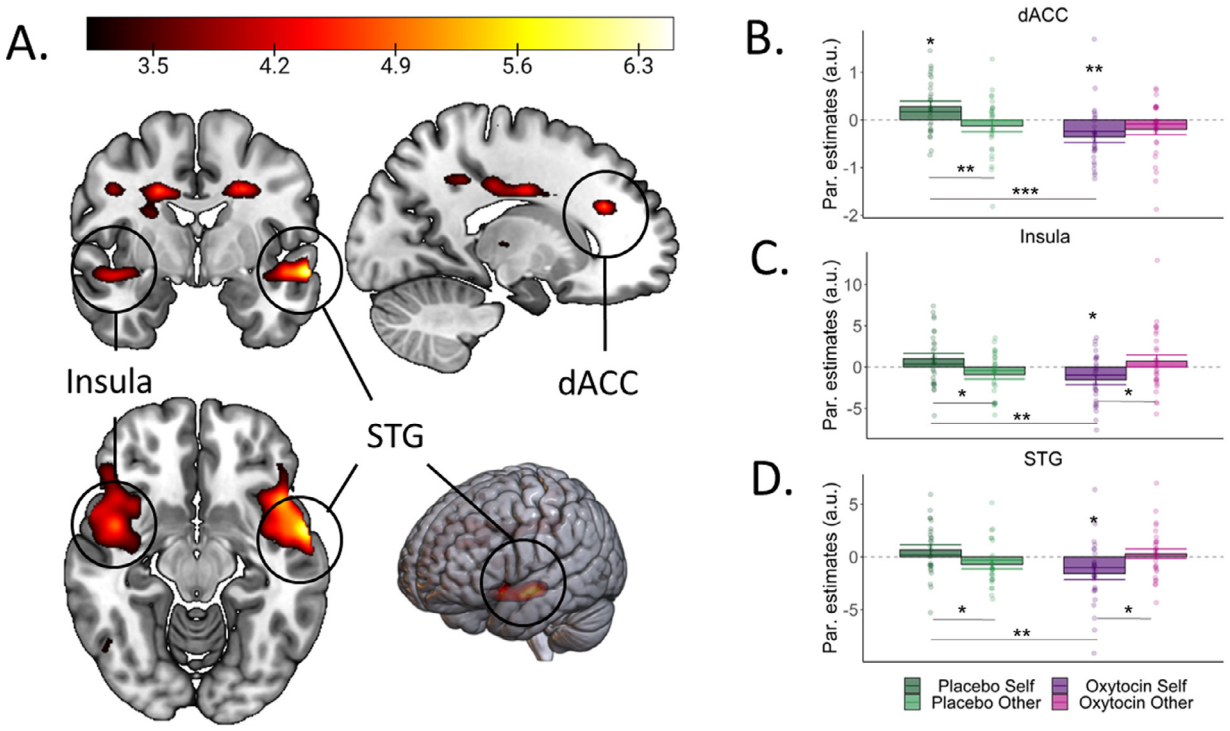


Fig. 4. Whole brain image (A) and extracted parameter estimates in the dACC (B), insula (C) and STG (D) for the interaction between OT and recipient. (A) Whole brain interaction between OT and recipient (Placebo:-Self[1]-Other[−1], OT:-Self[−1]-Other[1]), displayed at $P < 0.001$ uncorrected for illustration purposes, [$x = -17$, $y = -3$, $z = -12$]. The reverse contrast (Placebo:-Self[−1]-Other[1], OT:-Self[1]-Other[−1]), revealed no significant effects. (B) Extracted parameter estimates from the whole-brain interaction in the dorsal anterior cingulate cortex (dACC), [$x = -17$, $y = 33$, $z = 26$], $Z = 4.90$, $k = 4$, $P = 0.011$ FWE. (C) Extracted parameter estimates from the whole-brain interaction in the insula, [$x = 44$, $y = 3$, $z = -12$], $Z = 4.76$, $k = 18$, $P = 0.019$ FWE. (D) Extracted parameter estimates from the whole-brain interaction in the superior temporal gyrus (STG), [$x = -59$, $y = -2$, $z = -11$],], $Z = 6.16$, $k = 577$, $P < 0.001$ FWE. $^{*}P < 0.05$, $^{**}P < 0.01$ $^{***}P < 0.001$.

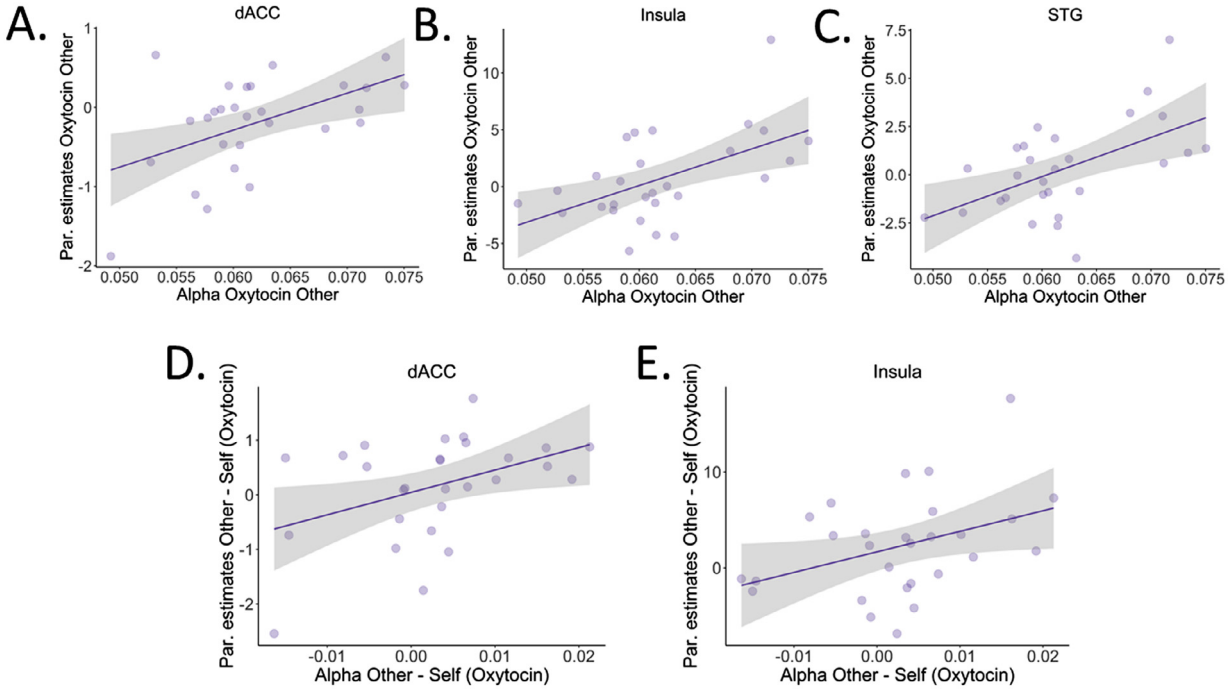


Fig. 5. Scatterplots depicting recipient-specific correlations between extracted parameter estimates and learning rates in dACC, insula and STG after OT. (A) Significant correlation between parameter estimates in the dorsal anterior cingulate cortex (dACC) and learning rates in the prosocial condition under influence of OT, $r(28) = 0.507$, $P = 0.006$. (B) Significant correlation between parameter estimates in the insula and learning rates in the prosocial condition under influence of OT, $r(28) = 0.537$, $P = 0.003$. (C) Significant correlation between parameter estimates in the superior temporal gyrus (STG) and learning rates in the prosocial condition under influence of OT, $r(28) = 0.547$, $P = 0.003$. (D) Significant correlation between parameter estimates in the dACC and learning rates in the prosocial versus self-benefitting condition under influence of OT, $r(28) = 0.420$, $P = 0.026$. (E) Significant correlation between parameter estimates in the insula and learning rates in the prosocial versus self-benefitting condition under influence of OT, $r(28) = 0.391$, $P = 0.040$.

tational parameters or learning performance itself, we show that both drugs widely modulate neural PE encoding during learning. First, our findings indicate that, regardless of the recipient condition, both drugs blunted positive signaling of PEs in the VS. Second, both drugs led to negative signaling of PEs in several regions associated with the encoding of negative or surprise PEs, with conjunction analysis showing substantial overlap after L-DOPA and OT within - amongst others - the aMCC extending to the (pre)-SMA, dlPFC, IPG, and precentral gyrus. Thirdly, we found opposing signaling of self-benefitting and prosocial PEs in dACC, insula and STG after OT (versus placebo). Exploratory analyses also indicated that after OT, more positive encoding of prosocial PEs in these regions related to higher prosocial learning rates.

We observed widespread modulatory effects of both drugs on neural PE signaling. Importantly, we found blunted positive signaling of PEs in the VS after L-DOPA administration compared to placebo, independent of the recipient-condition. While we hypothesized that positive signaling in VS would be enhanced rather than attenuated, previous neuroimaging studies investigating DA agonists have been inconsistent. Whereas the few studies administering L-DOPA report enhanced signaling in the VS (Pessiglione et al., 2006), restored VS signaling in old age (Chowdhury et al., 2013) and increased VS activity to rewards (Pleger et al., 2009), several factors may contribute to different findings. For example, Pessiglione et al. (2006) employed a between-subject design with a small sample size ($N = 13$ per group) and only observed an increase in signaling compared to the DA antagonist haloperidol and not compared to placebo, while the study by Chowdhury et al. (2013) observed enhanced VS signaling only in (a subgroup of) older adults, who are more likely to show suboptimal DA functioning than our sample of healthy young adults. Additionally, the VS enhancement reported by Pleger et al. (2009) was based on a somatosensory judgement task and focused on the BOLD response to reward rather than the positive PE directly. Sex differences could play a role here as well, as these previous studies have been conducted in mixed-sex samples, whereas our current study only included male participants, and previous research has for example shown that there are sex differences in striatal DA release (Munro et al., 2006). Importantly, in line with our current results, two other studies using DA agonists methamphetamine (Bernacer et al., 2013) and methylphenidate (Evers et al., 2017) also found reduced PE signaling in VS. The latter authors speculated that methylphenidate might have led to decreased (phasic DA-dependent) PEs through a drug-induced increase in tonic levels of DA (Evers et al., 2017), which can inhibit phasic DA release through increased stimulation of presynaptic DA autoreceptors (Beaulieu and Gainetdinov, 2011) and reduce postsynaptic responses to phasic DA (Jonasson et al., 2014). This explanation could be relevant to L-DOPA as well, since this drug is also known to increase tonic DA levels (Harun et al., 2016). Our results could also be seen to be in line with an "overdose" hypothesis of DA (Cools, 2006), whereby L-DOPA administration may impair PE signaling in the VS due to an overdosing of this already optimally-functioning brain area. However, other studies report no significant effects of DA agonist bromocriptine on PE signaling (Diederen et al., 2017) or processing of rewarding feedback (van der Schaaf et al., 2014) in the VS, and increased signaling after low dose amisulpride, a specific D2 receptor antagonist that is thought to boost D1 signaling (Jocham et al., 2011). These inconsistent results highlight the need for further research on the exact circumstances under which (different types of) DA-stimulating drugs lead to enhanced versus reduced VS encoding of PEs during reward learning.

Interestingly, we observed similar blunting of PEs in the VS after OT administration. This effect was independent of the recipient condition, which is consistent with OT exerting domain-general effects on PE signaling during learning. While most studies report social-specific effects of OT, there is an increasing amount of research indicating that OT may also have effects outside of- or independent of the social domain and serves general approach-avoidance behavior (Harari-Dahan and Bernstein, 2014) or the regulation of allostatic processes (Quintana and Guastella, 2020). OT is known to affect the mid-brain DA circuit, which supports processing of both social and non-social reward, and as such has been proposed to exert both social and non-social behavioral effects (Harari-Dahan and Bernstein, 2014).

Importantly, our design uniquely enabled us to conduct conjunction analyses, which demonstrated overlap in the modulatory effects of L-DOPA and OT within specific voxels in the VS. This fits with a growing number of studies highlighting the extensive interactions between DA and OT (for reviews see e.g., Love, 2014; Shamay-Tsoory and Abu-Akel, 2016). For example, OT receptors are located throughout the mesocorticolimbic dopamine system (Love, 2014). Moreover, animal studies have demonstrated that optogenetic manipulation of OT release induces DA release in the midbrain during social interaction (Hung et al., 2017) whereas optogenetic stimulation of oxytocinergic terminals have also been found to generally modulate midbrain DA activity (Xiao et al., 2017). There is also evidence from a human neuroimaging study that OT administration increases the BOLD signal in the midbrain to (non-social) monetary reward (Mickey et al., 2016). Hence, our analogous findings after OT and L-DOPA suggest that, in humans, administration of OT may induce similar effects through its alteration of DA.

Furthermore, whole-brain analyses revealed that both L-DOPA and OT induced *negative* signaling of PEs in several brain regions, with conjunction analyses showing significant overlap in many brain regions including the aMCC (extending to the [pre]-SMA), dlPFC, IPG, and precentral gyrus, all areas previously implicated in coding negative or surprise PEs (Fouragnan et al., 2018). These overlapping effects of OT and DA on negative signaling of PEs again fit with the notion that OT may exert effects through its impact on DA release (Love, 2014; Shamay-Tsoory and Abu-Akel, 2016). The findings are also line with previous electroencephalography studies indicating that DA stimulation enhances electrophysiological brain responses thought to reflect negative PEs (Barnes et al., 2014; De Bruijn et al., 2004; de Bruijn, Hulstijn, Verkes, Ruigt, and Sabbe, 2005b; Spronk et al., 2016), and originate from the aMCC (Debener et al., 2005). Our results furthermore indicate that, rather than enhancing both positive and negative signaling of PEs, both drugs may in fact induce a (recipient-independent) shift from positive to negative signaling of PEs in the brain. Given the switch from positive signaling in areas associated with reward to negative signaling in areas primarily associated with the coding of negative and surprise PE, the current outcomes might reflect an attentional shift whereby worse-than-expected outcomes become more salient than better-than-expected outcomes. However, since we focused on parametrically signed PEs, we cannot disentangle whether this negative signaling of PEs is more likely to reflect an inverse tracking of better-than-expected outcomes or increased tracking of worse-than-expected outcomes, or both. Notably, previous pharmacological fMRI studies using similar learning paradigms have mainly focused on positive correlations with parametrically signed PEs, whereas negative correlations are either ignored or not observed. Our findings thus additionally highlight the importance of testing for negative correlations when investigating the neural encoding of PEs (Lockwood and Klein-Flügge, 2021). We would also like to acknowledge that there is some debate with regard to the classification of prediction errors. Some researchers argue that to justify that a brain region truly encodes PEs rather than mere outcome valence, BOLD activity should be positively correlated with the outcome and negatively with the expectation term (Behrens et al., 2008; Zhang et al., 2020), which could be followed up in future work.

In addition to domain-general drug effects, we also observed whole-brain interactions between recipient and OT. Administration of OT versus placebo resulted in opposite prosocial versus self-benefitting PE signaling in dACC, insula and STG. Importantly, these regions have previously been associated with the encoding of negative and surprise PEs (Fouragnan et al., 2018), but also form part of the social brain network, with each of these regions implicated in, for example, the representation and/or experience of other's mental states (Bzdok et al., 2012; Lieberman, 2007). Hence, from a social salience perspective (Shamay-Tsoory and Abu-Akel, 2016), the fact that OT modulated activity in these

areas in a recipient-dependent manner, may reflect OT-induced changes in saliency of other- versus self-relevant outcomes. Interestingly and in line with this, exploratory analyses indicated that individual differences in prosocial learning rates correlated with PE signals in these areas after OT, which suggests that the extent to which these specific brain areas are engaged in PE encoding may be predictive of how well individuals learn in a prosocial (versus self-benefitting) context. While our interpretation of neural differences is complicated by the absence of significant behavioral differences, these findings may indicate that OT induces alternative or additional neural mechanisms during learning, with OT specifically inducing a role for these regions in learning for others.

Previous studies using this task have suggested enhanced sgACC signaling in the prosocial versus self-benefitting condition (Lockwood et al., 2016) and recipient-specific modulations of PE signaling in midbrain and sgACC after a medium dose of OT (Martins et al., 2022). We did not observe these effects here. Notably, we also did not observe midbrain involvement when looking at recipient-specific correlations with the PE for the placebo condition separately (see Table S3). These differences between the current and previous work could, however, be explained by the different behavioral learning patterns: crucially, our participants showed comparable levels of prosocial and self-benefitting performance and even enhanced prosocial learning rates already under placebo, in contrast to previous research, showing mainly self-biases in learning performance (Cutler et al., 2021; Liao et al., 2021; Lockwood et al., 2016; Martins et al., 2022; Westhoff et al., 2021). The fact that we failed to show such biases and even observed a bias towards better learning for others, is noteworthy. Possibly, differences in participant characteristics and contextual factors play a role here. For one, in contrast to previous studies, our data was collected in the middle of the COVID outbreak, a time which may have facilitated altruistic behavior through a heightened sense of common identity and emotional connection with others that are experiencing the same challenges (Drury, 2018; Wider et al., 2022). Moreover, most previous studies were carried out in the United Kingdom and there may thus be cultural differences with our native Dutch-speaking participants (Cutler et al., 2021; Lockwood et al., 2016; Martins et al., 2022). The only other study conducted in the Netherlands consisted of both males and females and concerned a younger age group (Westhoff et al., 2021), while previous work suggests that self-biases in learning tend to decrease as we age (Cutler et al., 2021). Hence, it would be interesting to see if our finding can be replicated in future studies.

Based on what we know about the neurobiology of reinforcement learning in non-human animals (Schultz, 2016) and research in humans (Pessiglione et al., 2006), it may be expected that L-DOPA would facilitate learning rates. Similarly, based on the social salience hypothesis (Shamay-Tsoory and Abu-Akel, 2016), which argues that OT enhances the salience of social versus self-relevant stimuli, and prior findings (Liao et al., 2021; Martins et al., 2022), we expected that OT would specifically facilitate prosocial learning. However, we did not find evidence for behavioral learning effects of either drug. Several explanations for this may be put forward. First of all, it should be acknowledged that the lack of significant behavioral findings in the presence of neural differences could be due to insufficient sensitivity of our learning paradigm or analyses to detect behavioral effects. Importantly, the task was perceived as relatively easy (see Table S5) and accuracy rates were high (see Fig. 2A), which together with the within-subject design could have contributed to ceiling effects. In line with our null-findings, a recent review of the literature investigating the impact of DA agonists on self-benefitting learning indicated very mixed results (for a comprehensive review see Webber et al., 2021), with L-DOPA enhancing (e.g., Pessiglione et al., 2006), decreasing (e.g., Guitart-Masip et al., 2014; Pizzagalli et al., 2008; Vo et al., 2016) or not impacting (Weis et al., 2013; Wunderlich et al., 2012) learning rates and/or performance. These inconsistencies have been suggested to stem from the fact that healthy adults already show optimal DA levels, meaning that further enhancement of DA is unlikely to lead to improvements (Webber et al., 2021). Some studies also indicate that individual differences in baseline levels of DA (e.g., Clatworthy et al., 2009; Cools et al., 2009; Mueller et al., 2014) determine how DA affects learning, and may need to be taken into account. Furthermore, while we also speculated about a potential self-serving bias of L-DOPA on performance based on previous studies observing modulatory effects on decisions regarding self- versus other-benefitting outcomes (Crockett et al., 2015; Pedroni et al., 2014; Soutschek et al., 2017), it is important to note that these studies involved a trade-off where others' outcomes influenced own rewards. In this respect, it may be interesting for future research to study how L-DOPA affects learning for others when this comes at personal costs.

Additionally, while we expected that OT would specifically facilitate prosocial learning, we noted that in our study, learning performance and rates were already similar and higher, respectively, when playing for the other versus oneself in the placebo condition. Hence, unlike previous studies reporting OT modulations of prosocial versus self-benefitting learning (Liao et al., 2021; Martins et al., 2022), our sample appeared already highly motivated to play for others to begin with. Also in contrast with our study, Liao et al. (2021) employed a between-subject design in both males and females from a different cultural background, all factors that could modulate OT effects (Borland et al., 2019; Xu et al., 2017). While our task contained three blocks of trials in each recipient condition, Martins et al. (2022) observed preserving effects of OT on prosocial performance only in a fourth block, which could suggest that their OT effect related to the increased efforts required to sustain performance. Moreover, this effect was only observed for a low dose (9 IU), and not higher doses (18 and 36 IU). Interestingly, a recent account proposes that OT is an allostatic hormone that helps maintain stability through changing environments, and predicts that OT may specifically improve reversal- rather than stable learning, regardless of whether the stimuli are social or non-social (Quintana and Guastella, 2020). Hence, it may be worthwhile for future studies to explore drug effects on learning tasks that require higher effort or cognitive demands, using different doses.

It has also been proposed that DA agonists may generally enhance exploration over exploitation during learning (Beeler, 2012), which would be reflected in enhanced temperature parameters. Yet, our findings indicate moderate evidence against a difference in beta values between placebo and L-DOPA. Interestingly, recent research shows that L-DOPA may specifically facilitate directed exploration during learning (Chakroun et al., 2020), suggesting that such effects of L-DOPA may only be captured with more complex reinforcement learning models that differentiate between directed and random exploration.

It should be acknowledged that the current study focused on one specific type of prosocial setting. Specifically, learning took place in a private context, where outcomes affected another participant who was in turn responsible for the outcomes of someone else, which allowed us to capture altruistic motivation to benefit others while ruling out effects of reciprocity and reputation, in line with previous work (Cutler et al., 2021; Lockwood et al., 2016; Martins et al., 2022; Westhoff et al., 2021). Of course, in real-life settings learning and prosocial behavior often take place in the presence of others, or require significant personal costs or efforts (Lockwood et al., 2017). However, the advantage of our task was that self and other-benefitting learning could be directly compared as interpretation of neural signals is more complicated in situations where benefitting others come at personal costs (i.e., does a region signal loss to self or benefit to other?). Furthermore, in the current study participants played for an anonymous peer. Prosocial learning may also differ depending on the beneficiary, and L-DOPA and OT may differentially modulate learning when benefitting close others such as friends or family members as compared to those with higher perceived social distance, and these effects may also differ in female versus male participants (e.g., X. Ma et al., 2018). Future research should therefore aim to explore these different kinds of scenarios.

Using a novel design, where both L-DOPA and OT were administered to the same subjects for the first time, we demonstrated that both compounds modulated brain responses underlying self-benefitting and

prosocial learning in a sample of healthy male adults. The effects of DA on PE signaling were primarily domain-general, suggesting that DA facilitation may impact the neural implementation of reinforcement learning independent of whether one is learning for their own benefit or that of another. Interestingly, our results showed that OT also induced domain-general effects. Conjunction analyses revealed widespread overlap in signaling between OT and L-DOPA, with both drugs blunting positive PE signaling in the VS and inducing negative PE signaling in several other regions. This is in line with a growing body of preclinical research highlighting the extensive interactions between DA and OT in the brain. Additionally, we observed recipient-specific effects of OT on PE signaling in the brain, with opposing signaling of self-benefitting and prosocial prediction errors after OT in dACC, insula and STG, with a positive correlation between these areas and prosocial learning rates. This suggests that OT-induced recruitment of these areas specifically supports learning for others. A possible mechanism for OT-induced changes in these areas may be through attention shifts as a result of altered salience of social versus self-relevant stimuli (Shamay-Tsoory and Abu-Akel, 2016). However, social salience cannot explain the currently demonstrated recipient-independent effects of OT, clearly revealing OT-induced domain-general effects on PE encoding during learning as well (Harari-Dahan and Bernstein, 2014; Quintana and Guastella, 2020). Specifically, the neural overlap between OT and L-DOPA suggests that effects of OT may indeed be established through its interactions with DA. The current findings thus reveal the communalities between these neurochemicals for the first time and provide support for their involvement in the neural computations underlying self-benefitting and prosocial reinforcement learning.

Data availability

Data, analysis scripts and materials for this study will be made available on DataverseNL and the Open Science Framework (https://doi.org/10.17605/OSF.IO/9XZDH). Spatially normalized group-level MRI data will be made available on NeuroVault (https://identifiers.org/neurovault.collection:13141).

Credit authorship contribution statement

Myrthe Jansen: Conceptualization, Software, Investigation, Project administration, Formal analysis, Visualization, Writing – original draft, Data curation. **Patricia L. Lockwood:** Methodology, Software, Resources, Writing – review & editing. **Jo Cutler:** Methodology, Software, Resources, Writing – review & editing. **Ellen R.A. de Bruijn:** Conceptualization, Methodology, Funding acquisition, Supervision, Writing – review & editing.

Acknowledgments

This work was supported by a personal grant from the Netherlands Organization for Scientific Research awarded to E. R. A. de Bruijn (NWO; VIDI grant nr. 452-12-005) and by the interdisciplinary research program Social Resilience and Security at Leiden University. We are grateful to all the students involved in the data collection for this study: Francesca Casetta, Neil Schön, Babak Mirheli, Fabien Bijsterbosch, Burak Günem, Emilie Sutter, Simone Härtel, Katarina Solita Puškaš and Eva Goetzke.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2023.119983.

References

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. Neuroimage 38 (1), 95–113.
Barnes, J.J., O'Connell, R.G., Nandam, L.S., Dean, A.J., Bellgrove, M.A, 2014. Monoaminergic modulation of behavioural and electrophysiological indices of error processing. Psychopharmacology (Berl.) 231 (2), 379–392.
Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.
Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
Beaulieu, J.-.M., Gainetdinov, R.R., 2011. The physiology, signaling, and pharmacology of dopamine receptors. Pharmacol. Rev. 63 (1), 182–217.
Beeler, J.A., 2012. Thorndike's law 2.0: dopamine and the regulation of thrift. Front. Neurosci. 6, 116.
Behrens, T.E., Hunt, L.T., Woolrich, M.W., Rushworth, M.F., 2008. Associative learning of social value. Nature 456 (7219), 245–249.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Statist. Soc.: Ser. B (Methodolog.) 57 (1), 289–300.
Bernacer, J., Corlett, P.R., Ramachandra, P., McFarlane, B., Turner, D.C., Clark, L., … Murray, G.K., 2013. Methamphetamine-induced disruption of frontostriatal reward learning signals: relation to psychotic symptoms. Am. J. Psychiatry 170 (11), 1326–1334.
Bond, A., Lader, M., 1974. The use of analogue scales in rating subjective feelings. Br J Med Psychol.
Borland, J.M., Rilling, J.K., Frantz, K.J., Albers, H.E., 2019. Sex-dependent regulation of social reward by oxytocin: an inverted U hypothesis. Neuropsychopharmacology 44 (1), 97–110.
Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A.R., Langner, R., Eickhoff, S.B., 2012. Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. Brain Struct. Funct. 217 (4), 783–796.
Carlo, G., 2013. The development and correlates of prosocial moral behaviors. In: Handbook of Moral Development. Psychology Press, pp. 208–234.
Chakroun, K., Mathar, D., Wiehler, A., Ganzer, F., Peters, J., 2020. Dopaminergic modulation of the exploration/exploitation trade-off in human decision-making. Elife 9, e51260.
Chowdhury, R., Guitart-Masip, M., Lambert, C., Dayan, P., Huys, Q., Düzel, E., Dolan, R.J., 2013. Dopamine restores reward prediction errors in old age. Nat. Neurosci. 16 (5), 648–653.
Clatworthy, P.L., Lewis, S.J., Brichard, L., Hong, Y.T., Izquierdo, D., Clark, L., … Fryer, T.D., 2009. Dopamine release in dissociable striatal subregions predicts the different effects of oral methylphenidate on reversal learning and spatial working memory. J. Neurosci. 29 (15), 4690–4696.
Clyde, M.A., Ghosh, J., Littman, M.L., 2011. Bayesian adaptive sampling for variable selection and model averaging. J. Comput. Graph. Statist. 20 (1), 80–101.
Cools, R., 2006. Dopaminergic modulation of cognitive function-implications for L-DOPA treatment in Parkinson's disease. Neurosci. Biobehav. Rev. 30 (1), 1–23.
Cools, R., Frank, M.J., Gibbs, S.E., Miyakawa, A., Jagust, W., D'Esposito, M, 2009. Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. J. Neurosci. 29 (5), 1538–1543.
Crockett, M.J., Siegel, J.Z., Kurth-Nelson, Z., Ousdal, O.T., Story, G., Frieband, C., … Dolan, R.J., 2015. Dissociable effects of serotonin and dopamine on the valuation of harm in moral decision making. Curr. Biol. 25 (14), 1852–1859.
Cutler, J., Wittmann, M.K., Abdurahman, A., Hargitai, L.D., Drew, D., Husain, M., Lockwood, P.L., 2021. Ageing is associated with disrupted reinforcement learning whilst learning to help others is preserved. Nat. Commun. 12 (1), 1–13.
De Bruijn, E.R., Hulstijn, W., Verkes, R.J., Ruigt, G.S., Sabbe, B.G., 2004. Drug-induced stimulation and suppression of action monitoring in healthy volunteers. Psychopharmacology (Berl.) 177 (1), 151–160.
De Bruijn, E.R., Hulstijn, W., Verkes, R.J., Ruigt, G.S., Sabbe, B.G., 2005a. Altered response evaluation: monitoring of late responses after administration of {d}-amphetamine. J. Psychophysiol. 19 (4), 311.
de Bruijn, E.R., Hulstijn, W., Verkes, R.J., Ruigt, G.S., Sabbe, B.G., 2005b. Altered response evaluation: monitoring of late responses after administration of D-amphetamine. J. Psychophysiol. 19 (4), 311–318.
de Bruijn, E.R., Ruissen, M.I., Radke, S., 2017. Electrophysiological correlates of oxytocin-induced enhancement of social performance monitoring. Soc. Cogn. Affect. Neurosci. 12 (10), 1668–1677.
De Bruijn, E.R., Sabbe, B.G., Hulstijn, W., Ruigt, G.S., Verkes, R.J., 2006. Effects of antipsychotic and antidepressant drugs on action monitoring in healthy volunteers. Brain Res. 1105 (1), 122–129.
Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., Von Cramon, D.Y., Engel, A.K., 2005. Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. J. Neurosci. 25 (50), 11730–11737.
Diederen, K.M., Fletcher, P.C., 2021. Dopamine, prediction error and beyond. Neurosci. 27 (1), 30–46.
Diederen, K.M., Ziauddeen, H., Vestergaard, M.D., Spencer, T., Schultz, W., Fletcher, P.C., 2017. Dopamine modulates adaptive prediction error coding in the human midbrain and striatum. J. Neurosci. 37 (7), 1708–1720.
Drury, J., 2018. The role of social identity processes in mass emergency behaviour: an integrative review. Eur. Rev. Soc. Psychol. 29 (1), 38–81.
Evers, E., Stiers, P., Ramaekers, J., 2017. High reward expectancy during methylphenidate depresses the dopaminergic response to gain and loss. Soc. Cogn. Affect. Neurosci. 12 (2), 311–318.

Forster, S.E., Zirnheld, P., Shekhar, A., Steinhauer, S.R., O'Donnell, B.F., Hetrick, W.P, 2017. Event-related potentials reflect impaired temporal interval learning following haloperidol administration. Psychopharmacol. (Berl.) 234 (17), 2545–2562.
Fouragnan, E., Retzler, C., Philiastides, M.G., 2018. Separate neural representations of prediction error valence and surprise: evidence from an fMRI meta-analysis. Hum. Brain. Mapp. 39 (7), 2887–2906.
Garrison, J., Erdeniz, B., Done, J., 2013. Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. Neurosci. Biobehav. Rev. 37 (7), 1297–1310.
Guitart-Masip, M., Economides, M., Huys, Q.J., Frank, M.J., Chowdhury, R., Duzel, E., … Dolan, R.J., 2014. Differential, but not opponent, effects of L-DOPA and citalopram on action learning with reward and punishment. Psychopharmacol. (Berl.) 231 (5), 955–966.
Harari-Dahan, O., Bernstein, A., 2014. A general approach-avoidance hypothesis of oxytocin: accounting for social and non-social effects of oxytocin. Neurosci. Biobehav. Rev. 47, 506–519.
Harun, R., Hare, K.M., Brough, E.M., Munoz, M.J., Grassi, C.M., Torres, G.E., … Wagner, A.K., 2016. Fast-scan cyclic voltammetry demonstrates that L-DOPA produces dose-dependent, regionally selective bimodal effects on striatal dopamine kinetics in vivo. J. Neurochem. 136 (6), 1270–1283.
Hung, L.W., Neuner, S., Polepalli, J.S., Beier, K.T., Wright, M., Walsh, J.J., … Dölen, G., 2017. Gating of social reward by oxytocin in the ventral tegmental area. Science 357 (6358), 1406–1411.
Huys, Q.J., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R.J., Dayan, P., 2011. Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. PLoS Comput. Biol. 7 (4), e1002028.
Jocham, G., Klein, T.A., Ullsperger, M., 2011. Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. J. Neurosci. 31 (5), 1606–1613.
Jocham, G., Klein, T.A., Ullsperger, M., 2014. Differential modulation of reinforcement learning by D2 dopamine and NMDA glutamate receptor antagonism. J. Neurosci. 34 (39), 13151–13162.
Jonasson, L.S., Axelsson, J., Riklund, K., Braver, T.S., Ögren, M., Bäckman, L., Nyberg, L., 2014. Dopamine release in nucleus accumbens during rewarded task switching measured by [11C] raclopride. Neuroimage 99, 357–364.
Kohls, G., Chevallier, C., Troiani, V., Schultz, R.T., 2012. Social 'wanting'dysfunction in autism: neurobiological underpinnings and treatment implications. J. Neurodev. Disord. 4 (1), 1–20.
Lenth, R., Singmann, H., Love, J., Buerkner, P., Herve, M., 2018. Emmeans: estimated marginal means, aka least-squares means. R Pack. Vers. 1 (1), 3.
Liao, Z., Huang, L., Luo, S., 2021. Intranasal oxytocin decreases self-oriented learning. Psychopharmacol. (Berl.) 238 (2), 461–474.
Lieberman, M.D., 2007. Social cognitive neuroscience: a review of core processes. Annu. Rev. Psychol. 58, 259–289.
Lockwood, P.L., Apps, M.A., Chang, S.W., 2020. Is there a 'social'brain? Implementations and algorithms. Trend. Cogn. Sci. (Regul. Ed.) 24 (10), 802–813.
Lockwood, P.L., Apps, M.A., Valton, V., Viding, E., Roiser, J.P., 2016. Neurocomputational mechanisms of prosocial learning and links to empathy. Proceed. Natl. Acad. Sci. 113 (35), 9763–9768.
Lockwood, P.L., Hamonet, M., Zhang, S.H., Ratnavel, A., Salmony, F.U., Husain, M., Apps, M.A., 2017. Prosocial apathy for helping others when effort is required. Nat. Hum. Behav. 1 (7), 1–10.
Lockwood, P.L., Klein-Flügge, M.C., 2021. Computational modelling of social cognition and behaviour—A reinforcement learning primer. Soc. Cogn. Affect. Neurosci. 16 (8), 761–771.
Love, T.M., 2014. Oxytocin, motivation and the role of dopamine. Pharmacol. Biochem. Behav. 119, 49–60.
Ma, X., Zhao, W., Luo, R., Zhou, F., Geng, Y., Xu, L., … Kendrick, K.M., 2018. Sex-and context-dependent effects of oxytocin on social sharing. Neuroimage 183, 62–72.
Ma, Y., Shamay-Tsoory, S., Han, S., Zink, C.F., 2016. Oxytocin and social adaptation: insights from neuroimaging studies of healthy and clinical populations. Trend. Cogn. Sci. (Regul. Ed.) 20 (2), 133–145.
Maia, T.V., Frank, M.J., 2011. From reinforcement learning models to psychiatric and neurological disorders. Nat. Neurosci. 14 (2), 154–162. doi:10.1038/nn.2723.
Manduca, A., Carbone, E., Schiavi, S., Cacchione, C., Buzzelli, V., Campolongo, P., Trezza, V., 2021. The neurochemistry of social reward during development: what have we learned from rodent models? J. Neurochem. 157 (5), 1408–1435.
Marsman, M., Wagenmakers, E.-J., 2017. Bayesian benefits with JASP. Eur. J. Develop. Psychol. 14 (5), 545–555.
Martins, D., Brodmann, K., Veronese, M., Dipasquale, O., Mazibuko, N., Schuschnig, U., … Paloyelis, Y., 2022a. Less is more": a dose-response account of intranasal oxytocin pharmacodynamics in the human brain. Prog. Neurobiol. 211, 102239.
Martins, D., Lockwood, P., Cutler, J., Moran, R., Paloyelis, Y., 2022b. Oxytocin modulates neurocomputational mechanisms underlying prosocial reinforcement learning. Prog. Neurobiol. 213, 102253.
Mayer, S.V., Jusyte, A., Klimecki-Lenz, O.M., Schönenberg, M., 2018. Empathy and altruistic behavior in antisocial violent offenders with psychopathic traits. Psychiatry Res. 269, 625–632.
Mickey, B.J., Heffernan, J., Heisel, C., Peciña, M., Hsu, D.T., Zubieta, J.-.K., Love, T.M., 2016. Oxytocin modulates hemodynamic responses to monetary incentives in humans. Psychopharmacol. (Berl.) 233 (23), 3905–3919.
Mueller, E.M., Burgdorf, C., Chavanon, M.L., Schweiger, D., Hennig, J., Wacker, J., Stemmler, G., 2014. The COMT Val158Met polymorphism regulates the effect of a dopamine antagonist on the feedback-related negativity. Psychophysiology 51 (8), 805–809.
Munro, C.A., McCaul, M.E., Wong, D.F., Oswald, L.M., Zhou, Y., Brasic, J., … Ye, W., 2006. Sex differences in striatal dopamine release in healthy adults. Biol. Psychiatry 59 (10), 966–974.
Nyholm, D., Lewander, T., Gomes-Trolin, C., Bäckström, T., Panagiotidis, G., Ehrnebo, M., … Aquilonius, S.-.M., 2012. Pharmacokinetics of levodopa/carbidopa microtablets versus levodopa/benserazide and levodopa/carbidopa in healthy volunteers. Clin. Neuropharmacol. 35 (3), 111–117.
Pauli, W.M., Nili, A.N., Tyszka, J.M., 2018. A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. Sci. Data 5 (1), 1–13.
Pedroni, A., Eisenegger, C., Hartmann, M.N., Fischbacher, U., Knoch, D., 2014. Dopaminergic stimulation increases selfish behavior in the absence of punishment threat. Psychopharmacol. (Berl.) 231 (1), 135–141.
Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., Frith, C.D., 2006. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature 442 (7106), 1042–1045.
Pizzagalli, D.A., Evins, A.E., Schetter, E.C., Frank, M.J., Pajtas, P.E., Santesso, D.L., Culhane, M., 2008. Single dose of a dopamine agonist impairs reinforcement learning in humans: behavioral evidence from a laboratory-based measure of reward responsiveness. Psychopharmacol. (Berl.) 196 (2), 221–232.
Pleger, B., Ruff, C.C., Blankenburg, F., Klöppel, S., Driver, J., Dolan, R.J., 2009. Influence of dopaminergically mediated reward on somatosensory decision-making. PLoS Biol. 7 (7), e1000164.
Quintana, D.S., Guastella, A.J., 2020. An allostatic theory of oxytocin. Trend. Cogn. Sci. (Regul. Ed.) 24 (7), 515–528.
Quintana, D.S., Lischke, A., Grace, S., Scheele, D., Ma, Y., Becker, B., 2021. Advances in the field of intranasal oxytocin research: lessons learned and future directions for clinical research. Mol. Psychiatry 26 (1), 80–91.
Rescorla, R.A., Wagner, A.R., Black, A.H., & Prokasy, W.F. (1972). Classical conditioning II: current research and theory.
Santesso, D.L., Evins, A.E., Frank, M.J., Schetter, E.C., Bogdan, R., Pizzagalli, D.A., 2009. Single dose of a dopamine agonist impairs reinforcement learning in humans: evidence from event-related potentials and computational modeling of striatal-cortical function. Hum. Brain Mapp. 30 (7), 1963–1976.
Schultz, W., 2016. Dopamine reward prediction error coding. Dialog. Clin Neurosci. 18 (1), 23–32.
Shamay-Tsoory, S.G., Abu-Akel, A., 2016. The social salience hypothesis of oxytocin. Biol. Psychiatry 79 (3), 194–202.
Siegel, J.S., Power, J.D., Dubis, J.W., Vogel, A.C., Church, J.A., Schlaggar, B.L., Petersen, S.E., 2014. Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. Hum. Brain. Mapp. 35 (5), 1981–1996.
Solié, C., Girard, B., Righetti, B., Tapparel, M., Bellone, C., 2022. VTA dopamine neuron activity encodes social interaction and promotes reinforcement learning through social prediction error. Nat. Neurosci. 25 (1), 86–97.
Soutschek, A., Burke, C.J., Raja Beharelle, A., Schreiber, R., Weber, S.C., Karipidis, I.I., … Kalenscher, T., 2017. The dopaminergic reward system underpins gender differences in social preferences. Nat. Hum. Behav. 1 (11), 819–827.
Spengler, F.B., Schultz, J., Scheele, D., Essel, M., Maier, W., Heinrichs, M., Hurlemann, R., 2017. Kinetics and dose dependency of intranasal oxytocin effects on amygdala reactivity. Biol. Psychiatry 82 (12), 885–894.
Spinhoven, P., Ormel, J., Sloekers, P., Kempen, G., Speckens, A.E., van Hemert, A.M., 1997. A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. Psychol. Med. 27 (2), 363–370.
Spronk, D.B., Verkes, R.J., Cools, R., Franke, B., Van Wel, J.H., Ramaekers, J.G., De Bruijn, E.R., 2016. Opposite effects of cannabis and cocaine on performance monitoring. Eur. Neuropsychopharmacol. 26 (7), 1127–1139.
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. Sci. Rep. 3 (1), 1–5.
Sutton, R.S., Barto, A.G., 2018. Reinforcement learning: An introduction. MIT press.
Team, R.C. (2020). R: a language and environment for statistical computing. R Foundation for Statistical Computing.
van der Schaaf, M.E., van Schouwenburg, M.R., Geurts, D.E., Schellekens, A.F., Buitelaar, J.K., Verkes, R.J., Cools, R., 2014. Establishing the dopamine dependency of human striatal signals during reward and punishment reversal learning. Cereb. Cort. 24 (3), 633–642.
Vo, A., Seergobin, K.N., MacDonald, P.A., 2018. Independent effects of age and levodopa on reversal learning in healthy volunteers. Neurobiol. Aging 69, 129–139.
Vo, A., Seergobin, K.N., Morrow, S.A., MacDonald, P.A., 2016. Levodopa impairs probabilistic reversal learning in healthy young adults. Psychopharmacol. (Berl.) 233 (14), 2753–2763.
Walsh, J.J., Christoffel, D.J., Wu, X., Pomrenze, M.B., Malenka, R.C., 2021. Dissecting neural mechanisms of prosocial behaviors. Curr. Opin. Neurobiol. 68, 9–14.
Webber, H.E., Lopez-Gamundi, P., Stamatovich, S.N., de Wit, H., Wardle, M.C., 2021. Using pharmacological manipulations to study the role of dopamine in human reward functioning: a review of studies in healthy adults. Neurosci. Biobehav. Rev. 120, 123–158.
Weis, T., Brechmann, A., Puschmann, S., Thiel, C.M., 2013. Feedback that confirms reward expectation triggers auditory cortex activity. J. Neurophysiol. 110 (8), 1860–1868.
Westhoff, B., Blankenstein, N.E., Schreuders, E., Crone, E.A., van Duijvenvoorde, A.C., 2021. Increased ventromedial prefrontal cortex activity in adolescence benefits prosocial reinforcement learning. Dev. Cogn. Neurosci. 52, 101018.
Wider, W., Lim, M.X., Wong, L.S., Chan, C.K., Maidin, S.S., 2022. Should I Help? Prosocial Behaviour during the COVID-19 Pandemic. Int. J. Environ. Res. Public Health 19 (23), 16084.
Wittmann, M.K., Fouragnan, E., Folloni, D., Klein-Flügge, M.C., Chau, B.K., Khamassi, M., Rushworth, M.F., 2020. Global reward state affects learning and activity in raphe nucleus and anterior insula in monkeys. Nat. Commun. 11 (1), 1–17.

Wunderlich, K., Smittenaar, P., Dolan, R.J., 2012. Dopamine enhances model-based over model-free choice behavior. Neuron 75 (3), 418–424.
Xiao, L., Priest, M.F., Nasenbeny, J., Lu, T., Kozorovitskiy, Y., 2017. Biased oxytocinergic modulation of midbrain dopamine systems. Neuron 95 (2), 368–384 e365.
Xu, X., Yao, S., Xu, L., Geng, Y., Zhao, W., Ma, X., … Kendrick, K.M., 2017. Oxytocin biases men but not women to restore social connections with individuals who socially exclude them. Sci. Rep. 7 (1), 1–10.
Zhang, L., Lengersdorff, L., Mikus, N., Gläscher, J., Lamm, C., 2020. Using reinforcement learning models in social neuroscience: frameworks, pitfalls and suggestions of best practices. Soc. Cogn. Affect. Neurosci. 15 (6), 695–707.
Zhuang, Q., Zhu, S., Yang, X., Zhou, X., Xu, X., Chen, Z., … Yao, S., 2021. Oxytocin-induced facilitation of learning in a probabilistic task is associated with reduced feedback-and error-related negativity potentials. J. Psychopharmacol. 35 (1), 40–49.
Zirnheld, P.J., Carroll, C.A., Kieffaber, P.D., O'donnell, B.F., Shekhar, A., Hetrick, W.P, 2004. Haloperidol impairs learning and error-related negativity in humans. J. Cogn. Neurosci. 16 (6), 1098–1112.