A causal role for estradiol in human reinforcement learning

Authors

Sebastijan Veselic[1,2,3]*, Gerhard Jocham[4], Christian Gausterer[5], Bernhard Wagner[6], Miriam Ernhoefer-Reßler[6], Rupert Lanzenberger[7], Claus Lamm[1,8], Christoph Eisenegger[1] & Annabel Losecaat Vermeer[1]*

Affiliations

[1] Neuropsychopharmacology and Biopsychology Unit, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Austria

[2] Department of Clinical and Movement Neurosciences, University College London, London, UK

[3] Wellcome Centre for Human Neuroimaging, University College London, London, UK

[4] Biological Psychology of Decision Making, Institute of Experimental Psychology, Heinrich Heine University Düsseldorf, Germany

[5] FDZ-Forensisches DNA Zentrallabor GmbH, Medical University of Vienna, Austria

[6] Laboratory for Chromatographic & Spectrometric Analysis, FH JOANNEUM, Graz, Austria

[7] Department of Psychiatry and Psychotherapy, Medical University of Vienna, Vienna, Austria,

[8] Vienna Cognitive Science Hub, University of Vienna, Austria

(*) corresponding authors: Sebastijan Veselic (*sebastijan.veselic.18@ucl.ac.uk)* and Annabel Losecaat Vermeer (annabel.losecaat.vermeer@univie.ac.at)

Abstract

The sex hormone estrogen is hypothesized to play a key role in human cognition via its interactions with the dopaminergic system. Work in rodents has shown that estrogen's most potent form, estradiol, impacts striatal dopamine functioning predominately via increased D1-receptor signalling and correlational evidence in humans has suggested high estradiol levels alter reward sensitivity. Here, we addressed two fundamental questions: 1) whether estradiol causally alters reward sensitivity in men, and 2) whether this effect of estradiol is moderated by individual variation in polymorphisms of dopaminergic genes. To test this, we performed a double-blind placebo-controlled administration study in which hundred men received either a single dose of estradiol (2 mg) or placebo. We found that estradiol administration increased reward sensitivity, which was moderated by baseline dopamine. This was observed in choice behaviour and increased learning rates. These results confirm a causal role of estradiol in reinforcement learning in men that is moderated by striatal and prefrontal dopaminergic pathways.

Keywords: *Estradiol, reward processing, reinforcement learning, DAT1, COMT, Estrogen receptor*

Introduction

Learning which actions to select based on whether the outcome of that action is rewarded or not is a fundamental capacity required for adaptive behaviour. One neuromodulator that has long been linked to this capacity, known as reinforcement learning (RL), is dopamine [1]. More recently, an additional biological substrate that has been suggested to influence RL via dopaminergic mechanisms is the steroid hormone estrogen [2].

Estrogens are a class of steroid hormones important for healthy development in mammals, with estradiol being the most prevalent and potent form [3,4]. Previous human studies implicated estradiol in several cognitive processes with mixed findings in terms of its exact role (for reviews see [2,5]). One recent hypothesis has been that estradiol may specifically impact human reward processing by amplifying dopamine signalling via one of its receptors (i.e. the D1 receptor) [2]. For example, human neuroimaging work has revealed that fluctuations in estradiol levels are correlated with increased reward sensitivity, as documented by an increased BOLD response in the midbrain [6–8]. Similarly, rodent literature has shown that manipulation of estradiol levels affect the striatal dopamine system in various ways, with a net increase in overall dopamine signalling predominantly via the D1 receptor [9–14]. Besides the observed role of estradiol in the striatal dopamine system, it has a hypothesised connection to dopamine in the prefrontal cortex as well. Namely, estradiol metabolites decrease the activity of Catechol-O-methyltransferase (COMT), an enzyme responsible for approximately 60 percent of dopamine degradation in the prefrontal cortex and approximately 15 percent in the striatum [2,15]. Correspondingly, one correlational study previously observed that the association between endogenous estradiol levels and working memory performance is moderated by polymorphisms of the COMT gene [16].

Dopamine's role in reward processing and learning has been well studied using RL tasks and has been formalized with the reward prediction error hypothesis [1,17–19]. A canonical approach to investigate the causal role of dopamine in reward processing is to employ a double-blind placebo-controlled administration protocol using dopamine agonists and

antagonists, respectively [20–26]. Extending this approach through pharmacogenetics, which is the interaction between administered drugs and genetic variation, has enabled a better understanding of how genetic variation modulates dopamine availability and how the latter influences reward processing and cognition more generally [16,21,24,27,28].

This line of work has shown that causal manipulation of dopamine levels in humans affects performance in reinforcement learning [28], and that these effects can depend on individual differences in baseline dopamine levels [20]. Crucially, such individual differences arise from polymorphisms of dopamine-related genes impacting dopamine synthesis capacity and transmission [16,21,26]. For example, the COMT and dopamine transporter (DAT1) gene have polymorphisms that correlate with differences in performance on working memory and reinforcement learning tasks [16,21,26,29,30]. These polymorphisms are the val^{158}met polymorphism of COMT (i.e. the Val/Val, Met/Val, and Met/Met genotypes that are each associated with increasingly higher levels of prefrontal dopamine) and VNTR polymorphism of DAT1 (i.e. the 9/10 and 10/10 genotypes are associated with high and low striatal dopamine, respectively).

Despite abundant evidence from rodent research and work in humans showing the relation between estradiol, dopamine, and human cognition, results so far have been contradictory in terms of estradiol's effects. Namely, it has been shown that high endogenous estradiol levels increased [6,8] as well as decreased [31] performance on a variety of cognitive tasks. Although previous work on humans provided important insights, these were mostly based on correlations (for exceptions see [6,31,32]), small sample sizes (for exception see [32]), and additionally did not explicitly focus on the importance of baseline differences in dopamine (for exceptions see [16,33]). Therefore, the precise role of estradiol in human reward processing remains unclear (for review see [2]).

The aim of the present study was to investigate whether estradiol causally affects reward processing in a probabilistic RL task by employing a pharmacogenetic approach (Fig. 1A). The task required subjects to choose between two options on each trial in order to maximize their earnings. The probability of reward of both options was determined by two independent random Gaussian walks while the reward size was constant across trials (Fig.

1B). A constant reward size allowed us to isolate estradiol's influence on choice behaviour as a function of receiving versus not receiving a reward on each trial. This allowed for a more precise examination how estradiol influences reward processing. We further investigated whether an effect on reward sensitivity was moderated by individuals' baseline dopamine, as indexed through genetic variation in COMT and DAT1. Our main hypothesis was that estradiol administration would increase reward sensitivity which would be observed through increased choice reactivity. We further predicted that an increase in reward sensitivity would be observed in increased Q-learning learning rates, indicative of higher learning. Finally, we predicted that the behavioural and computational effects would uniquely depend on polymorphisms of both COMT and DAT1, as observed in previous work [21,27].

To detect differences at the level of individual genetic variants, we used a sample size (N = 100) in line with previous recommendations in the field [2]. Our sample was pre-screened and matched for key physiological characteristics, behavioural and cognitive traits and states that could have impacted RL behaviour (see Supplementary Materials). Moreover, we aimed at providing a more conclusive and precise account of a dopamine-dependent basis of action through excluding several other mechanistic explanations, which have so far been unaddressed. These included polymorphisms of androgen and estrogen receptors, together with a polymorphism influencing the enzyme aromatase that is responsible for the conversion of androgens to estradiol. These mechanisms are important because previous work has shown that administering estradiol also increases free circulating androgen levels, which are known to be converted to estradiol through aromatase [34] (see also Supplementary Materials).

In brief, we have found that estradiol administration increased reward sensitivity as compared to placebo administration. This was observed in choice behaviour and increased learning rates. Furthermore, we observed that the interaction between estradiol administration and dopamine-related genes predicted choice, in line with predictions from previous work reviewed here. Finally, we have observed several effects related to staying and switching behaviour that depended not only on striatal but also prefrontal baseline dopamine levels. Taken together, the described effects are consistent with the hypothesis that estradiol acts by

amplifying dopamine signalling via the D1 receptor and extend this by showing that the effects of estradiol are moderated by differences in prefrontal dopaminergic functioning as well.

Results

Both treatment groups (estradiol and placebo) were matched on several key characteristics. These included age, height, visceral, and abdominal fat, BMI, and individual traits and states that can impact RL behaviour, including working memory, self-reported impulsivity, behavioural inhibition and approach, and mood. As a manipulation check of our administration protocol, estradiol concentrations were significantly elevated in subjects who had received estradiol compared to placebo after (W = 1545, 95% CI [0.03, 1.87], p < .05), but not before administration (baseline: W = 1498, 95% CI [-0.05, 1.03], p = .09) and subjects' beliefs about whether they had received estradiol or placebo did not correlate with the actual received drug (r = 0.02, p = .82; for further details on group characteristics, matching, and manipulation checks see Supplementary Materials).

First, we investigated our hypothesis that estradiol administration would alter reward sensitivity, which we expected to observe through a systematic difference in choice behaviour across trials compared to placebo. We quantified this systematic difference by computing the cumulative difference in the probability of choosing option A across trials in both groups. This cumulative difference was then compared to a null distribution demonstrating what would be expected by chance (see Methods and materials). Similarly, we looked at the percentage of trials on which estradiol caused a significant difference in the chosen option compared to placebo. Moreover, we looked at whether these differences in choice behaviour also reflected improved task performance. Secondly, we tested our hypothesis that the effect of estradiol administration on choice behaviour would interact with genetic variation of COMT and DAT1. This was followed by a more detailed examination of whether these interactive effects would be observed in the amount of switching and staying behaviour, and choice autocorrelation throughout the task. Finally, we formalized these differences in behaviour within a

reinforcement learning framework that allowed us to exclude the possibility that choice differences were due to more stochastic responding, but instead were due to higher learning rates, indicative of higher weighing of more recent relative to old task-relevant information.

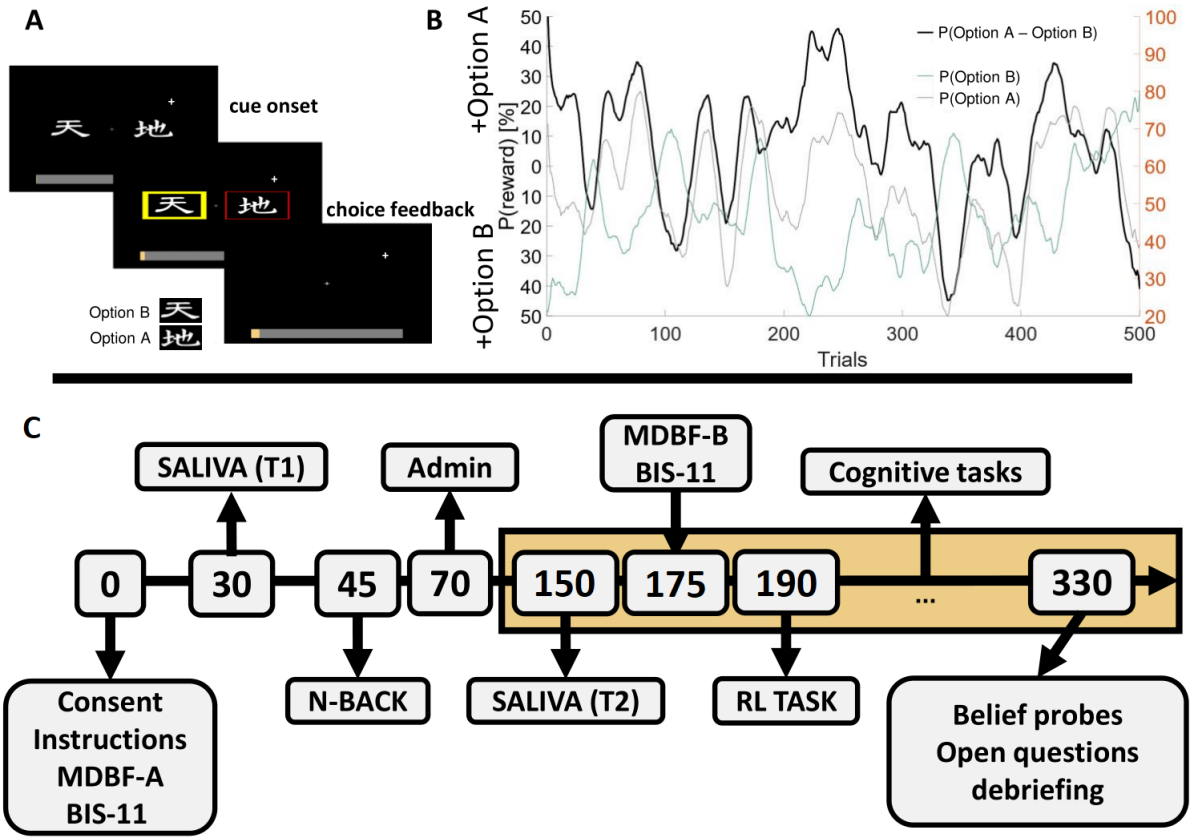


Fig. 1 A) Outline of a trial of the RL task. Each trial started by the presentation of two options (henceforth option A and option B). Subjects were required to choose one of these options. After they made a choice, subjects were presented with feedback, with the chosen option indicated by a thicker frame and the not chosen option by a thinner frame. A yellow frame indicated the rewarded option, whereas a red frame indicated the unrewarded option. Importantly, both options A and B could yield a reward or no reward on the same trial. **B)** The probability of reward upon choice for each option (green and gray lines), which were determined by two independent random Gaussian walks, with the probability shown in percent on the right y-axis in orange. The black line shows the relative probability of reward for one option over the other, which corresponds to the difference in reward probability for option A and option B. On trials where the black line is reaching the top half of the y-axis, option A was more rewarding, and vice versa. **C)** The timeline of the test session. Values in brackets denote minutes from the onset of the test session. We first collected consent and questionnaire data, which was followed by a baseline saliva sample (T1) and the N-BACK task. After administration of estradiol or placebo, subjects were required to rest for two hours before we collected the second saliva sample (T2) and assessed subjects' mood and impulsivity via questionnaires. The RL task began 120 minutes post-administration. This was followed by three other cognitive tasks that are not the focus of the current paper. At the end of the test session, we probed subjects' beliefs about the drug, the experiment, and debriefed them.

Estradiol administration alters choice reactivity

Our first hypothesis was that estradiol administration would increase reward sensitivity. By reward sensitivity, we refer to a systematic difference in the chosen option across trials between the estradiol and the placebo group. Since reward sizes were constant, the only difference across trials was whether a reward was received or not following choice. An effect of estradiol on reward sensitivity would therefore be observed if the difference between the option that each group chose on average across trials would be higher than would be expected by chance. We investigated whether such a difference in choice behaviour exists in two complementary ways. We first computed the probability for each group to select option A vs. option B across trials, subtracted the two group traces from each other (Fig. 2A) and plotted the cumulative choice difference across trials (Fig. 2B).

Under the hypothesis that estradiol systematically influenced choice behaviour, the cumulative difference in the expected chosen option should have exceeded the one obtained from a null distribution. Specifically, in the null distribution choice behaviour was decoupled from the actual treatment (i.e. estradiol vs. placebo) and revealed what degree of cumulative choice difference would be expected by chance or random assignment of treatment (see Methods and materials). Indeed, we observed that the cumulative difference in the expected chosen option between the estradiol and placebo group started to exceed the 100th percentile of a null distribution (Fig. 2B) ($M_{\text{last trial}}$ = 53.48 %, $z_{\text{last trial}}$ = 8.44, p < .001, threshold value for 99.9th percentile of null distribution: 46.20 %). This cumulative choice difference between the estradiol and placebo group remained significant when we collapsed it across time (Fig. 2C), which is demonstrated as the mean and the standard error of the mean remaining above the 99.9th percentile threshold of a null distribution (M = 25.72 ±0.69%, z = 5.80, p < .001, threshold value for 99.9th percentile of null distribution: M = 21.02 %) (see Methods and materials). Both results showed that estradiol administration (vs. placebo) led to systematic differences in subjects' choice.

Secondly, we tested the percentage of trials on which there was a statistically significant difference between the groups in choice behaviour. To test this, we performed a

two-sample proportion z-test on each trial, where we statistically compared the proportion of subjects choosing option A between both groups. We observed that estradiol administration (vs. placebo) led to a statistically significant difference in the proportion of subjects choosing option A vs. option B on 7.6 % of trials (black dots in Fig. 2A). In other words, estradiol administration caused subjects to choose a different option on 7.6 % of trials as compared to placebo. We performed family-wise error control similarly to above (see Methods and materials). For this, we decoupled the responses from the treatment and tested whether this percentage would have been obtained in a null distribution with random allocation of groups. This comparison showed that the change in how groups responded to the rewarding options on 7.6 % of trials exceeded the threshold value of a null distribution ($z = 5.37$, $p < .001$, threshold value for 99.9^{th} percentile of null distribution: 6.4 %).

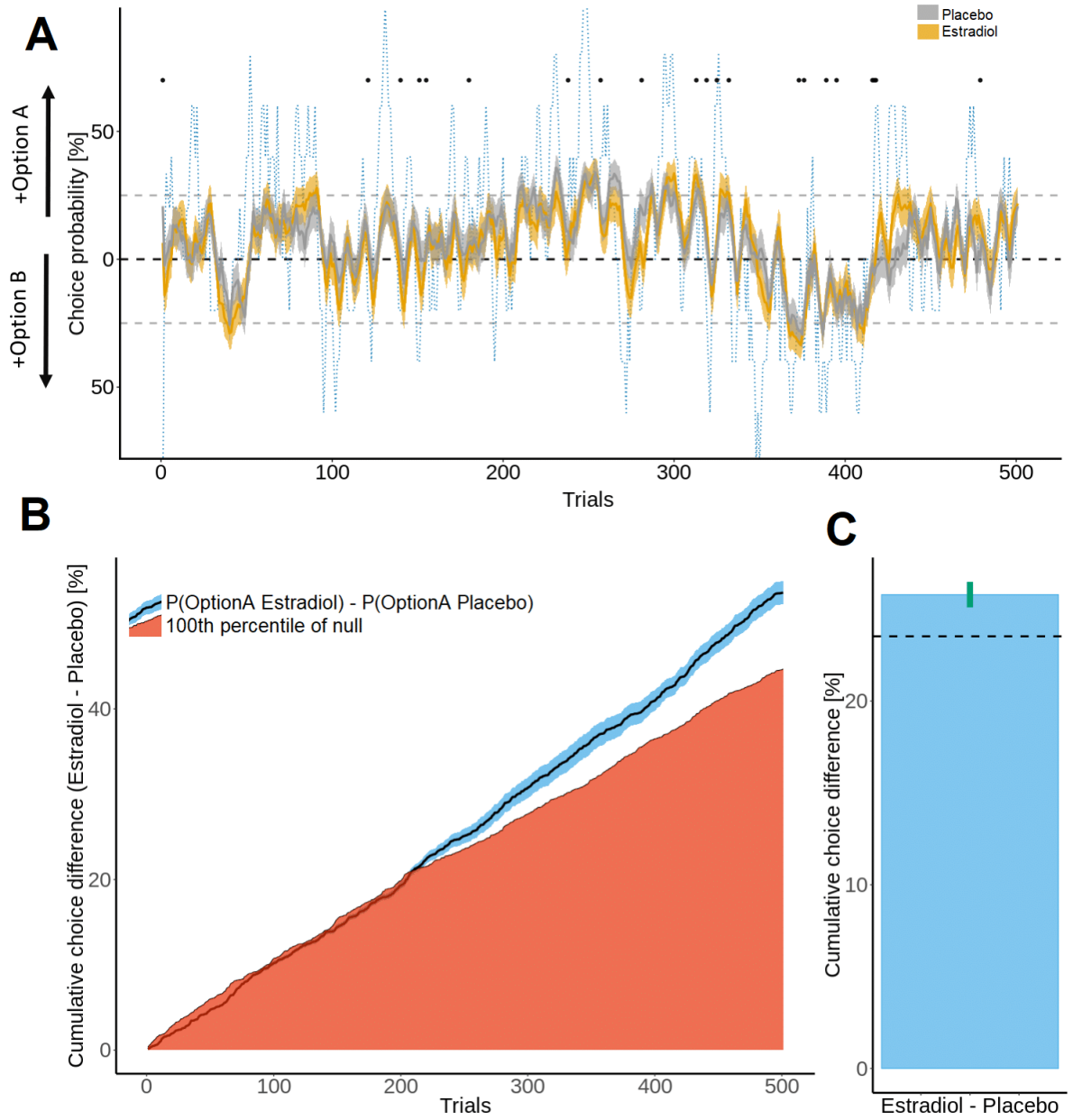


Fig. 2 A) Relative choice probability for choosing option A (top of y-axis) vs. choosing option B (bottom of y-axis) for the estradiol (orange) and placebo (gray) group. Solid thick lines represent trial mean, shaded areas around the thick lines denote standard errors of the mean. The blue dotted line denotes the relative reward probability which was computed from the probability of option A (top of y-axis) minus probability of option B (bottom of y-axis). Horizontal gray dotted lines represent where subjects were on average 25% more likely to select option A (upper line) or option B (lower line). All time-series traces were smoothed with a 5-trial moving average for visual purposes. The black dots indicate trials where there was a statistically significant difference ($p < .05$) between the estradiol and placebo group. The number of significant trials was compared to a null distribution (see Methods and materials). **B**) Cumulative choice difference between the estradiol and placebo group over trials compared to a 100th percentile null distribution. The thick black line is the difference between the orange and gray lines presented in figure **A**, and the blue shaded area is the corresponding difference between the standard errors in **A**. The dark orange area denotes the space in which differences are not significant. Conversely, separation between the lines indicate statistical significance. **C)** Mean cumulative choice difference between the estradiol and placebo group collapsed across trials. The dashed line represents the mean cumulative choice difference of the 100th percentile of the null distribution. Error bars indicate standard error of the mean.

DAT1 genotype marginally moderates the effects of estradiol on accuracy

Following the observed systematic choice difference between both groups, we investigated whether this was reflected in group differences in accuracy (i.e. whether the estradiol group chose the option with higher probability of reward compared to the placebo group). In a comparison of choice accuracy (Fig. 3A), we observed that subjects with exogenously elevated estradiol were not more accurate compared to subjects with placebo ($M_{Estradiol}$ = 57.30 ±6.91, $M_{Placebo}$ = 56.80 ±7.09, $t_{(97.94)}$ = 0.36, *95% CI* [-3.28, 2.28], p = .72, d = 0.07), and responded equally fast ($M_{Estradiol}$ = 0.61 sec ±0.11, $M_{Placebo}$ = 0.62 *sec* ±0.09, $t_{(95.55)}$ = 0.46, p = .65, d = 0.09).

However, based on previous work that showed interactive effects between cognitive performance and dopamine-related genes [20,35], we had hypothesized that the effect of estradiol on accuracy may depend on individual differences in baseline striatal dopamine (indexed with DAT1 polymorphism: 9/10 and 10/10 genotypes are associated with high and low striatal dopamine, respectively). Similarly, we predicted that the effects of estradiol may depend on differences in prefrontal dopamine (indexed with the COMT polymorphism, as Met/Met, Met/Val, and Val/Val genotypes are associated with high, medium, and low prefrontal dopamine, respectively). A general linear model revealed a trend towards an interaction between drug administration and DAT1 genotype on accuracy ($F_{(1, 69)}$ = 3.69, p = .06, Ω^2 = 0.03, Fig. 3B), while controlling for covariates (see Methods and materials). Following up this trend, pairwise comparisons revealed that estradiol administration increased accuracy in subjects with the 9/10 genotype (i.e. high striatal dopamine levels; M = 60.00 ±5.36) compared to those with a 10/10 genotype (i.e. low striatal dopamine levels; 10/10 DAT1, M = 56.00 ±6.51; $t_{(39.60)}$ = 2.14, *95% CI* [0.21, 7.63], p = .04, d = 0.61), but not for the placebo group (9/10 genotype: M = 57.21 ±6.60; 10/10 genotype: M = 56.75 ±6.34; $t_{(31.02)}$ = 0.22, *95% CI* [-3.74, 4.66], p = .82, d = 0.06). Subjects with the 9/10 genotype in the estradiol group were not more

accurate compared to subjects with the 9/10 genotype in the placebo group ($t_{(29.04)}$ = 1.33, *95% CI* [-1.48, 7.00], p = .19, d = 0.45) nor when comparing the groups with the 10/10 genotype $t_{(40.22)}$ = 1.82, *95% CI* [-0.36, 6.80], p = .08, d = 0.61).

Repeating the same analysis for the COMT genotype revealed no interaction between drug administration and COMT on accuracy ($F_{(2, 79)}$ = 1.76, p = .18, Ω^2 = 0.02, Fig. 3C).

In sum, estradiol administration increased reward sensitivity. We observed this in terms of a cumulative difference in the expected chosen option between the estradiol and the placebo group, both across trials and collapsed across trials. Furthermore, on a subset of trials we found a significant difference in the proportion of subjects from the estradiol group compared to placebo group who chose option A. This systematic difference in how subjects responded throughout the task was not reflected in increased accuracy across both groups. However, in line with our hypothesis we found a significant interaction between genetic variation of DAT1 polymorphism and drug administration, such that in the estradiol group subjects with a 9/10 DAT1 genotype showed an improved accuracy (by trend) relative to those with a 10/10 DAT1 genotype, with no such difference in the placebo group. No such interaction was observed for the COMT polymorphism, indicating that estradiol mainly acted on striatal rather than prefrontal dopamine signaling in terms of its effects on task accuracy.

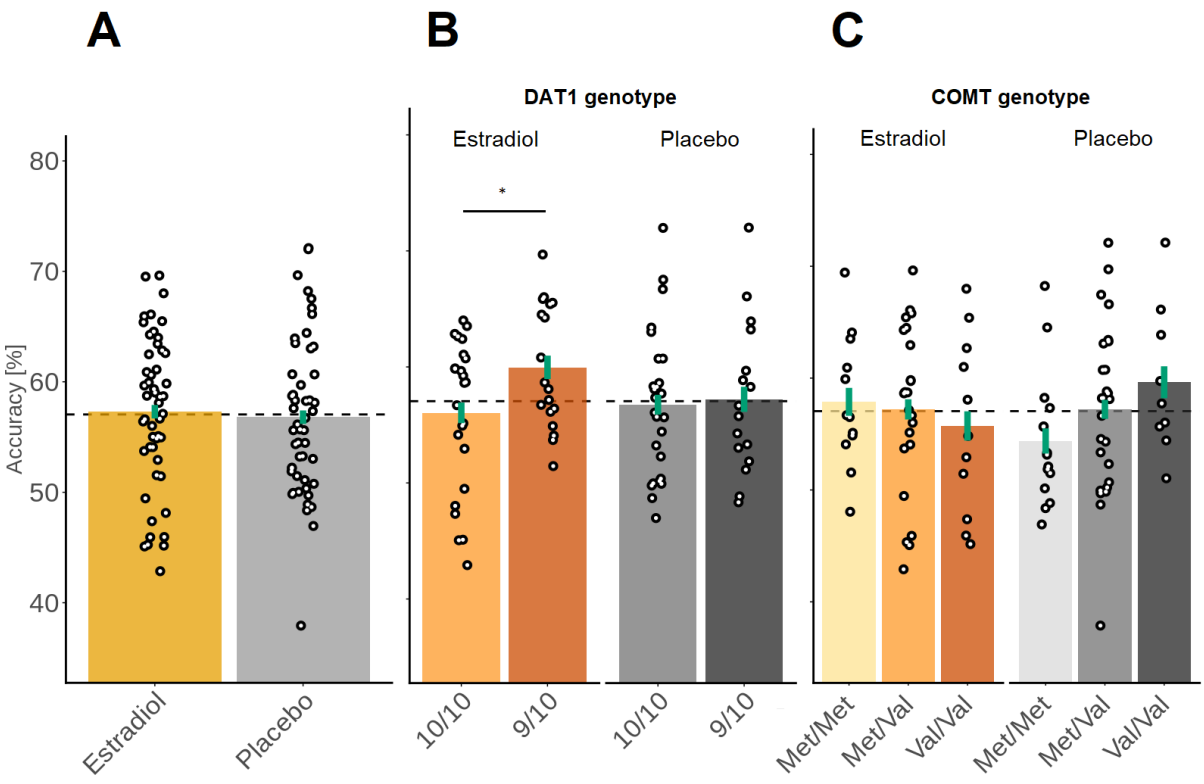


Fig. 3 A) Mean accuracy split according to drug administration. **B**) Mean accuracy split according to drug administration and DAT1 polymorphism. **C**) Mean accuracy split according to drug administration and COMT polymorphism. Green error bars are standard errors of the mean. Dots represent individual subjects. The horizontal black dotted line represents grand mean performance collapsed across groups to show the relative change for individual subgroups. * $p < .05$.

The effect of estradiol administration on choice behaviour is moderated by polymorphisms of both COMT and DAT1

To directly test whether the effect of estradiol administration on choice behaviour is moderated by polymorphisms of dopamine-related genes (e.g. COMT, DAT1), and whether individual variability in these effects may be a contributing factor to the observed effects, we used generalized linear mixed models. We tested whether the interaction between drug, polymorphism (COMT or DAT1), and trial are a significant predictor of choice behaviour (i.e. reward sensitivity).

We predicted a significant interaction due to the observed differences in cumulative choice behaviour described above. Based on the inverted U-shape dopamine hypothesis [35], we predicted that estradiol administration would upregulate reward sensitivity in subjects with

low prefrontal dopaminergic activity (i.e. Val/Val) but would not, or would even impair it, in those with high prefrontal dopaminergic activity (i.e. Met/Met). The model predicted that exogenously elevated estradiol in subjects with a Met/Val (β = 0.20 ± 0.04, *95% CI* [0.11, 0.28], z = 4.56, p < .001) and Val/Val genotype (β = 0.37 ± 0.06, *95% CI* [0.26, 0.48], z = 6.99, p < .001) were more likely to select option A as trials progressed (Fig. 1B, see also Fig. S2 and Fig. S7. Supplementary Materials) – which was the more rewarding option throughout the task (percent trials rewarded: $M_{optionA}$ = 53.70%, $M_{optionB}$ = 42.91%).

Similarly, we predicted that estradiol should indirectly increase striatal dopamine levels, leading to higher reward prediction errors. Based on this, we expected that subjects with the 9/10 genotype (i.e. high striatal dopamine) would select the more rewarding option (i.e. higher value option) more often and less so for subjects with the 10/10 genotype (i.e. low striatal dopamine). This was supported by model predictions showing that that subjects with the 10/10 genotype with placebo (β = -0.12 ± 0.04, *95% CI* [-0.04, -0.20], z = -3.03, p < .01) were the most likely to select the lower valued option A throughout task progression, while estradiol administration dampened this slope in subjects with the same 10/10 genotype (see Fig. S7. Supplementary Materials). Results from both generalized linear mixed effects models showed that once individual variation was considered, the effect of estradiol administration on choice behaviour across trials was moderated by striatal (DAT1) and prefrontal (COMT) polymorphisms (see Fig. S7. Supplementary Materials for model predictions).

Increased reward sensitivity is observed in increased learning rates

Given our observation that estradiol increased reward sensitivity and our hypothesis that the mechanistic explanation for a cumulative choice difference in this task may underlie increased striatal and prefrontal dopamine levels, we predicted that estradiol would enhance the learning of reward probabilities. In a RL framework this would be reflected in increased learning rates. The learning rates represent latent variables dictating one's weighing of recent in comparison

to older information. To test this, we estimated the learning rate by fitting several Q-learning models (see Methods and materials). To test this, we estimated learning rates by fitting several Q-learning models (see Methods and materials). The best model (model 2, leave one out information criterion (LOOIC) = 60179, Fig. 4A) included separate learning rates for each option, a temperature parameter, and an irreducible noise parameter. The model predicted choice behaviour above chance ($t_{(99)}$ = 13.95, *95% CI* [0.64, 0.68], p < .001, Fig. 4B, see also Fig. S8. Supplementary Materials) and did not perform better for either group ($M_{Estradiol}$ = 66.26 % ±10.77, $M_{Placebo}$ = 64.90 % ±11.85; $t_{(97.115)}$ = 0.76, *95% CI* [-0.03, 0.06], p = .45).

Our main hypothesis was that if estradiol increases available striatal and prefrontal dopamine concentrations [2,5], then the behavioural differences in choice over time (Fig. 2A) would be captured in the learning rates. We have found that estradiol administration increased the learning rate for both options compared to placebo ($\alpha_{optionB}$: $M_{Estradiol}$ = 0.27 ±0.16, $M_{Placebo}$ = 0.17 ±0.13, $t_{(85.36)}$ = 4.47, *95% CI* [0.08, 0.21], p < .001, d = 0.9; $\alpha_{optionA}$: $M_{Estradiol}$ = 0.26 ±0.19, $M_{Placebo}$ = = 0.12 ±0.13, $t_{(92.13)}$ = 3.42, *95% CI* [0.04, 0.16], p < .001, d = 0.69, Fig. 4C). We expected that estradiol would affect both learning rates in the same direction due to their intrinsic correlation arising from the fact that both capture the same behaviour (r = 0.84, p <.001). However, contrary to our expectations, the observed main effect of estradiol was not moderated by either polymorphisms of DAT1 or COMT (COMT: $\alpha_{OptionB}$: $F_{(2, 81)}$ = 0.37, p = .69; $\alpha_{OptionA}$: $F_{(2, 72)}$ = 0.29, p = .75; DAT1: $\alpha_{OptionB}$: $F_{(1, 71)}$ = 0.02, p = .89, $\alpha_{OptionA}$: $F_{(1, 71)}$ = 0.03, p = .86).

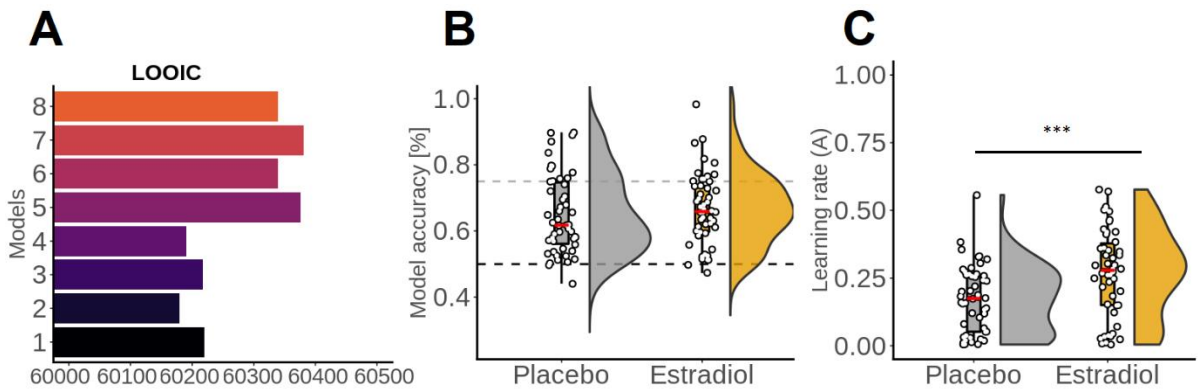


Fig. 4 A) Leave one out information criterion (LOOIC) value for all employed models. Lower LOOIC indicates better model fit – model two was selected as the best model. **B)** The overall model accuracy collapsed over time obtained from the posterior predictive density (see Supplementary Materials) shown for both groups separately. Individual dots represent subjects. The red bar represents the median, the box plot represents the 75% middle most data points, with the whiskers representing 1.95*IQR. **C)** Learning rates by drug treatment. The estradiol group (in orange) had higher learning rates compared to the placebo group (in gray).

In sum, the estradiol group had higher learning rates compared to the placebo group but we observed no moderation of the polymorphisms of both COMT and DAT1 on the model parameters.

Altered reward sensitivity is driven by differences in the number of stay-switch decisions and moderated by COMT and DAT1 genotype.

Finally, to more precisely understand the observed difference in choice behaviour between treatment groups and dopamine-related genes, we tested whether this difference could be attributed to differences in staying and switching behaviour, commonly studied in this field [26,36].

Based on our expectation that estradiol would increase striatal dopamine levels, and through that increase reward prediction errors, we predicted that estradiol administration would enhance staying behaviour moderated by DAT1, but not by COMT polymorphism. As a measure of staying, we computed how many trials subjects chose the same option on average if they were previously rewarded for that option (see Fig. 5). Overall, estradiol administration did not increase the number of stay choices (M = 1.70 ±0.03) compared to placebo (M = 1.65 ±0.04; $t_{(97.91)}$ = 1.07, *95% CI* [-0.15, 0.04], p = .29, d = 0.21).

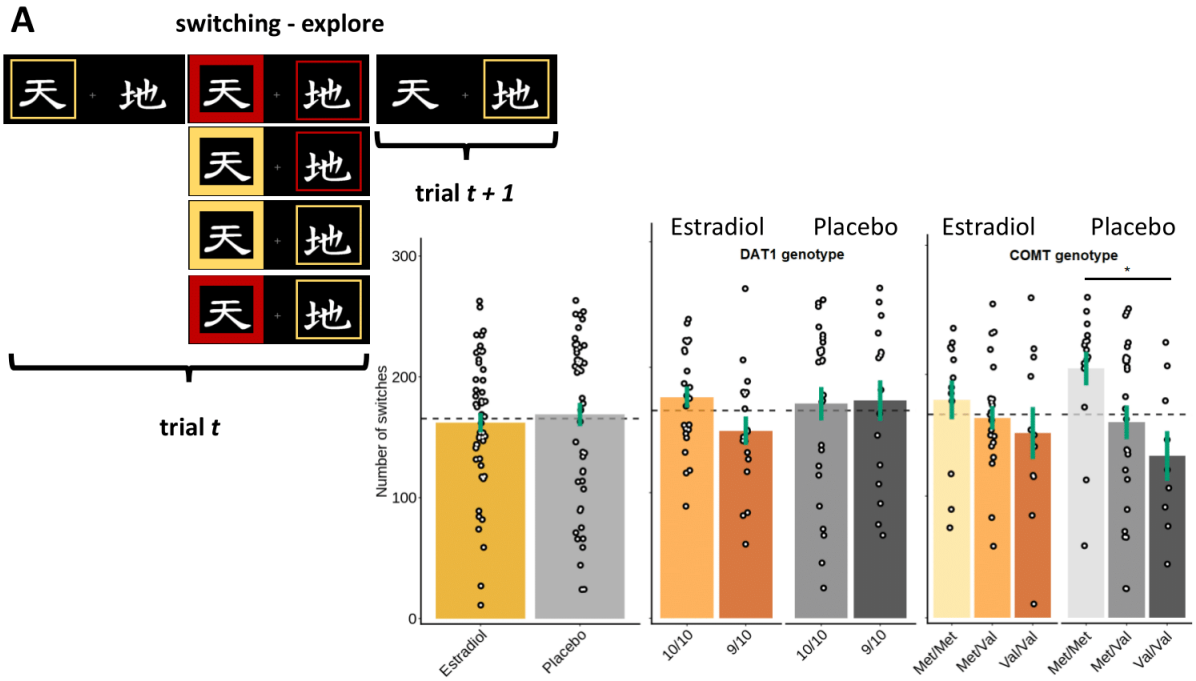
However, estradiol administration in 9/10 DAT1 genotype subjects, who were more accurate compared to subjects who received estradiol and had the 10/10 genotype, also chose the same option on more trials on average after being rewarded for their choice (M = 1.79 ± 0.18; Fig. 5B). This was observed compared to subjects with placebo who had the 9/10 genotype (M = 1.63 ± 0.22; $t_{(29.05)}$ = 2.33, *95% CI* [0.02, 0.3], p = .03, d = 0.41; see also Supplementary Materials Fig. S6), and compared to subjects who had the 10/10 genotype

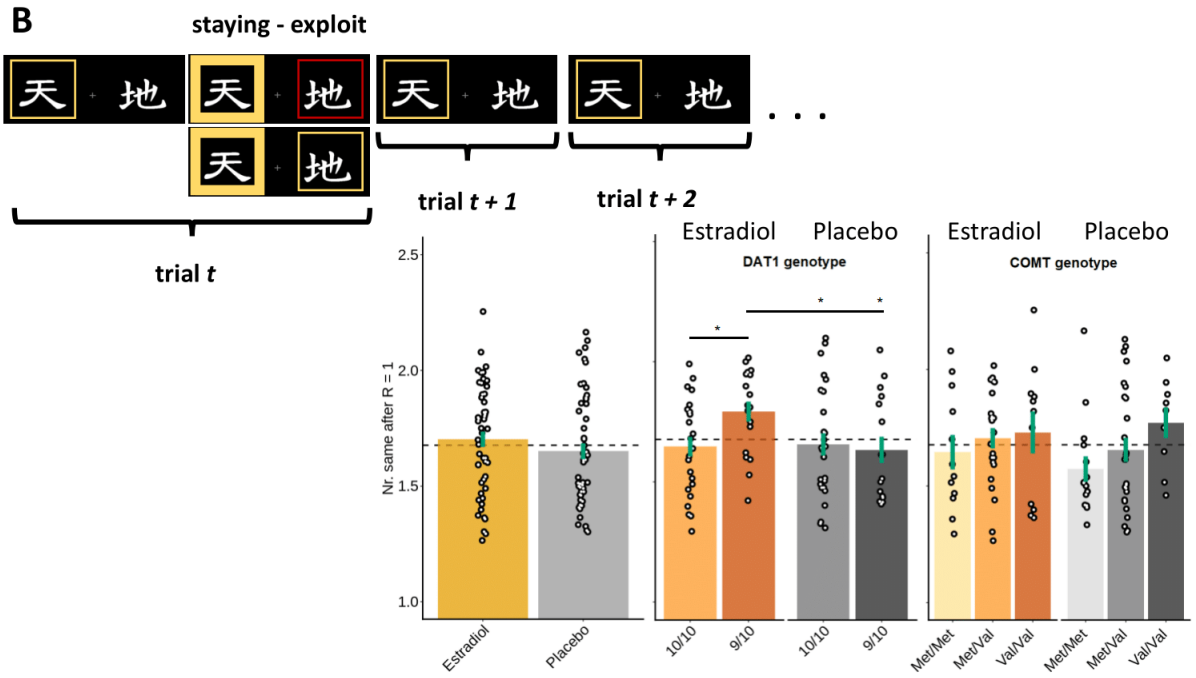
(placebo: $t_{(41.61)}$ = 2.22, *95% CI* [0.01, 0.27], p = .03, d = 0.41; estradiol: ($t_{(38..86)}$ = 2.49, *95% CI* [0.03, 0.27], p = .02, d = 0.64). In other words, the increase in accuracy by exogenously elevated estradiol in individuals with a 9/10 genotype was reflected in increased staying with options for which they were previously rewarded. This is consistent with previous work showing increased striatal prediction errors following dopamine precursor administration [28].

Furthermore, because estradiol administration likely results in increased prefrontal dopamine levels through downregulating COMT enzyme activity, we predicted that the interaction between estradiol administration and COMT polymorphism would be predictive of switching behaviour [26]. As a measure of switching, we assessed the number of times the option chosen on trial t was different from the one chosen at trial $t + 1$ (i.e. a switch), irrespective of the choice outcome on trial t (see Fig. 5). Estradiol administration did not significantly influence switch decisions (M = 162.12 ±56.31) compared to placebo (M = 168.82 ±68.13; $t_{(94.64)}$ = 0.54, *95% CI* [-18.12, 31.51], p = .59, d = 0.11). However, we observed a significant interaction of estradiol administration by COMT genotype ($F_{(2, 80)}$ = 3.22, p = .05, Ω^2 = 0.04, Fig. 5A). The interaction showed that subjects with placebo and a Val/Val genotype (i.e. low prefrontal dopamine availability) switched less often (β = -84.07±33.69, p = .02) compared to all other groups. As predicted by the inverted U-shaped relationship between prefrontal dopamine levels and behaviour [35], Val/Val placebo subjects (Val/Val: M = 132.33 ±61.40) switched less compared to Met/Met placebo subjects (i.e. associated with high prefrontal dopamine availability; Met/Met: M = 204.27 ±53.52, $t_{(15.10)}$ = 2.91, *95% CI* [19.25, 124.54], p = .01, d = 1.46). For the estradiol group, this difference was not present (Val/Val: M = 151.09 ±70.85; Met/Met: M = 178.5 ±55.34; $t_{(18.96)}$ = 1.03, *95% CI* [-28.28, 83.10], p = .32, d = 0.44). That is, estradiol administration attenuated naturally occurring differences in switching behaviour found in subjects with the Met/Met and Val/Val genotypes that are associated with high and low prefrontal dopamine levels, respectively.

Crucially, the effects reported for accuracy, staying, and switching were not explained by other mechanistic explanations (i.e. other genetic polymorphisms assessed here), such as

those related to androgen receptor functioning, androgen to estrogen conversion, or estrogen receptor functioning (see Supplementary Materials), indicating that the observed results are moderated specifically by dopamine-related genes. Furthermore, in Supplementary Materials we further show that the observed differences in staying and switching can be also characterised as differences in choice autocorrelation, and choice autocorrelation as a function of previous reward. In brief, the estradiol group overall exhibited less choice autocorrelation compared to the placebo group, showing that previous responses had a weakened effect on future choices, with these differences being more pronounced based on DAT and COMT polymorphisms.





407

408 **Fig. 5** Stay and switch behaviour split according to drug administration and the COMT and DAT1
409 polymorphisms. **A)** Switching behaviour, measured as the total number of trials on which subjects chose
410 different options on trial *t* and trial *t* + *1,* independent of the choice outcome on trial *t*. **B)** Staying
411 behaviour, measured as the average number of trials the same option was selected when that choice
412 was previously rewarded. In both plots, each dot represents a subject, the green error bar represents
413 standard error of the mean.

414

415

416

Discussion

In this study we examined the causal effects of estradiol on reward processing in human males. A body of previous rodent causal and human correlational work has suggested a role of estradiol in cognition, and reward processing more specifically, via dopamine-related mechanisms [5,7,8,10,12,14,16,37]. However, it remained an open question whether the effect of estradiol administration would be observable in choice behaviour of healthy young men and whether this would be moderated by individual variation in DAT1 and COMT. By employing a pharmacogenetic approach with a probabilistic RL task, we have shown that exogenously elevated estradiol altered various aspects ofchoice behaviour related to reward processing. Moreover, we have shown that effects related to accuracy, staying, and switching were moderated by striatal (DAT1) and prefrontal (COMT) dopamine-related genes, but not by other candidate genes that we tested.

Firstly, we confirmed the hypothesis that estradiol administration increases reward sensitivity in healthy young men as observed through increased choice reactivity. More specifically, we found that the cumulative difference in the expected chosen option between both groups was higher than what would be expected by chance. This was the case when we compared choice behaviour across trials and when we collapsed across trials. When we further quantified the difference by looking at the percentage of trials on which the estradiol group choose a different option compared to the placebo group, we observed they chose differently on a statistically significant subset of trials that was above chance.

In addition to these analyses, we aimed to account for individual variability in the strength of a potential effect on choice behaviour [38,39]. Using two separate generalized linear mixed models, we observed that both models predicted choice, showing that the effect of estradiol on choice behaviour over time was moderated by baseline striatal (DAT1 – first model) and prefrontal dopamine levels (COMT – second model). Overall, these results replicate previous correlational[7,8] neuroimaging work and preliminary evidence from a

pharmacological study on a small sample of women at menopause (*N* = 13) showing changes in BOLD signal due to reward-related information in conditions of high vs. low estradiol conditions[6]. Furthermore, our results show, for the first time, that causal exogenous alteration of estradiol leads to differences in choice behaviour on a reinforcement learning task via striatal and prefrontal modulation.

However, these differences in choice behaviour alone would not yet clearly establish whether estradiol acted by amplifying dopamine D1 receptor signalling[2]. To test this more directly, we investigated whether the observed choice differences resulted in differences in accuracy on our task, as expected by previous work using dopamine precursor administration [28]. We also tested whether this would be moderated by the DAT1 polymorphism [27,33]. This revealed a trending interaction (*p* = .06) between estradiol administration and the DAT1 polymorphism on accuracy. Specifically, a pairwise comparison showed that estradiol administration significantly increased accuracy in subjects with the 9/10 genotype (i.e. high striatal dopamine), but only compared to subjects with the 10/10 genotype. This effect had a medium effect size (Cohen's *d* = 0.61).

While the observed effect on accuracy is in line with our hypothesis and predictions of estradiol amplifying striatal dopamine D1 receptor signalling[2], our findings should be considered as preliminary and warrant replication in future pharmacogenetic administration studies using larger samples per cell. Of note is that previous research, in which striatal dopamine levels were increased exogenously, showed a deterioration in accuracy in their experimental group with the 9/10 genotype, but an improvement in those with the 10/10 genotype [27]. One possible explanation for this contrast with our results is that our administered estradiol dosage most likely acted akin to a "low dosage" of a dopamine precursor. Namely, dopamine precursor administration has been previously shown to impact behaviour in a dose-dependent manner [40,41]. This interpretation is also supported by a recent administration study on a small sample of women (*N* = 34) where 12 mg of estradiol (i.e. 6 times our dose) decreased working memory performance[31], which was interpreted as an overstimulation of

dopaminergic transmission. Similar support comes from a study on hippocampal activity in women following dose-dependent estradiol administration[32].

Overall, our finding of a subtle effect on accuracy following estradiol administration when including the genetic DAT1 polymorphism converges with previous work. That is, previous work has interpreted diverging results in the direction of increased and decreased performance in high estradiol conditions due to different baseline dopamine levels [8,37,42,43]. Our results provide empirical evidence for these previous claims by showing that the effect of estradiol on reward processing is better understood when taking baseline dopamine levels into account. Furthermore, they show that investigating whether and how chronic estradiol administration alters reward processing in humans, dependent on one's genotype, may yield important and novel insights for both basic science as well as clinical practice.

To better understand what drove the effect on accuracy and choice reactivity, we computed metrics of switching and staying behaviour that are commonly investigated in such tasks [26,36] and performed choice autocorrelation analyses. Both metrics showed that within the estradiol group, the subtle difference in accuracy between the DAT1 polymorphisms was reflected by increased staying behaviour. Namely, subjects with a 9/10 genotype chose the same option on more trials, on average, if they were previously rewarded for that choice, compared to the other subgroups. In addition to finding a weak effect through mechanisms of striatal dopamine, we have observed that the interaction between drug administration and COMT predicted switching behaviour. Specifically, estradiol administration attenuated naturally occurring differences in switching observed in the placebo group. While Val/Val placebo subjects switched least and significantly less compared to Met/Met placebo subjects, this difference disappeared in the estradiol group. The switching and staying effects depending on both the COMT and DAT1 polymorphisms were also supported by analyses of choice autocorrelation, and revealed a comparable pattern to the one described, but also enabled us to better understand the effect of choices several trials ago on the current choice.

Finding an effect on switching that is moderated by individual variation in COMT provides, for the first time, evidence for the hypothesis that estradiol has a causal role in frontal dopamine-mediaton[16]. This likely happens due to the inhibition of COMT activity through estradiol metabolites that leads to increased dopamine availability [5,13]. This means that estradiol does not interact only with striatal dopamine levels but also with frontal dopamine levels. Our effect establishes a set of causal findings for future work to replicate and build upon.

Finally, because we predicted that increased reward sensitivity would occur due to larger striatal reward prediction errors because of estradiol administration, we hypothesised that this would be reflected in increased learning rates, as compared to placebo. This demonstrates that estradiol increased the weight of new information relative to old information. These effects are consistent with predictions by previous imaging work who found increased reward sensitivity in high estradiol conditions[6]. This is furthermore supported by decreased choice autocorrelation in subjects who were administered with estradiol compared to placebo. However, the effect of estradiol on learning rates was not moderated by the COMT or DAT1 polymorphism, in contrast to our predictions. Similarly to our interpretation for accuracy and staying behaviour reported below, it is likely that our sample size was not sufficiently large to detect difference at the polymorphism subgroup level. To the best of our knowledge, this is the first examination of behavioural differences through computational modelling as a function of estradiol administration and polymorphisms of dopamine genes in men. The results provide grounding for future work that may benefit by incorporating a computational approach to elucidate the observable behavioural changes following estradiol administration. Framing behavioural effects through a computational framework would allow future work to compare their findings with our work and findings about other hormones, e.g. testosterone [44,45].

Through the behavioural and genetic measures we collected, we were also able to exclude a substantial number of other candidate mechanisms that could have driven some of the effects we observed. These include androgen receptor functioning (polyglutamine (CAG)

and polyglycine (GGN) repeats), differences in androgen to estrogen conversion (CYP 19A1) or estrogen receptor functioning (ERα, ERβ), which have been previously unaddressed (see Supplementary Materials). Furthermore, we were able to exclude confounding measures that are known to influence estradiol metabolism upon administration such as changes in self-reported mood and attention due to drug administration, individual differences in impulsiveness, behavioural approach and inhibition, working memory performance assessed via an N-back task, salivary cortisol levels, and differences in body measurements (weight, height, BMI, abdominal and visceral fat) (see Supplementary Materials).

The current study also encountered some limitations. Based on these, we provide recommendations for future work employing pharmacogenetics with estradiol.

The first is related to increasing sample size. While our sample size was approximately twice as large compared to most previous work [8,16,31,33,37,43,46,47], for one exception see [32]) and in line with suggestions for the field [2], we suspect that we were underpowered to detect all effects of interest at COMT and DAT1 polymorphism level. The reason for this is that we observed several "trend-level" p-values ($p < 0.1$) for which we had strong theoretical predictions. Specifically, this refers to not finding a clear interaction between estradiol administration with both DAT1 and COMT on accuracy [27,33]. In addition, we would have predicted an interaction between estradiol administration with DAT1 on staying, because of increased accuracy. The interaction with COMT on switching behaviour similarly needs replication. Moreover, because previous administration studies did not compute behavioural effect sizes that could have served as a basis for our current work, except of the general recommendation in [2], it was difficult to estimate the minimal viable sample size. Due to general power issues in this field of research, larger sample sizes are required and starting to be used also in other psychoneuroendocrinological work [44,45].

The second recommendation relates to the type of reinforcement learning task used. For future research we would suggest using a reversal learning task [20,48] with parametrically changing reward probability contingencies. Based on our findings, we predict that such a task

could elucidate more clearly the effect of estradiol on behaviour. Namely, the trials where we observed the clearest effect (e.g. trials around 400) is where the largest probability reversals happened. If our prediction is true, it should also more clearly show improved learning and accuracy compared to the Gaussian random walks employed here. An alternative idea for future work is to use the two-step task [49] which would enable to further disentangle both model-free and model-based behaviour and reveal how variation in COMT and DAT1 moderates the influence of administration. We would predict estradiol to have similar effects as found by other work using dopamine precursors where administration increased model-based learning [50].

Our third recommendation is related to dose-dependent effects of estradiol administration. In [31], the authors concluded they may have elicited overstimulation (12 mg) of dopaminergic transmission, while our results (2 mg) show similarity to a low dose of a dopamine precursor due to contrasting results with [27]**.** An extension through a dose-dependent investigation of choice behaviour would show whether this is true for reward processing similarly to dose-dependent observations in [40,41] and further contribute to the understanding of estradiol in relation to the inverted U-shape hypothesis [35].

The final recommendation is to include additional genotypes that may moderate the influence of estradiol on behaviour (e.g. the Taq1A variant in the dopamine D2 receptor gene). This would enable to better disentangle the contribution of different dopamine-related genes [21,24]. Alternatively, neurochemical positron emission tomography as in [20] with estradiol administration would provide a better understanding at the level of receptor binding and show to which degree these effects relate to dopaminergic circuitry in prefrontal and striatal regions.

In conclusion, we have shown that estradiol causally influences choice behaviour by altering reward processing. The observed effects were specifically moderated by frontal (COMT) and striatal (DAT) dopamine-related genes but not estrogen and androgen-related genes (CAG, GGN, CYP 19A1, ERα, ERβ). Our results converge with experimental evidence from rodent work that showed amplified striatal dopamine D1 signalling in high estradiol conditions. Moreover, they confirm the prediction that estradiol has a role in frontal dopamine

577 signalling through the COMT polymorphism [5,13,16]. Finally, our behavioural results were
578 supported by computational modelling showing that estradiol causally increased learning
579 rates, supporting the hypothesis that increased reward prediction errors may have driven the
580 increased reward sensitivity

581 In sum, our study shows the importance of using more complex research designs that
582 are supported by causal work from animal models and correlational human studies. Combining
583 predictions from both and augmenting the hypotheses with pharmacogenetics allows us to
584 elucidate the interactions between hormones, neurotransmitter systems, and cognition, both
585 on a mechanistic, behavioural, and computational level. Such an approach has important
586 implications for a better understanding of the biology and neuroscience of human cognition
587 that is moderated by genes in both health and disorder.

588

589 ## Acknowledgements

590 The authors would like to thank Christina Faschinger and Isa Krol for their assistance in data
591 collection, Nace Mikus for his help in data collection and analysis suggestions, and Lei Zhang
592 for comments on the final manuscript. The study was supported by the Vienna Science and
593 Technology Fund (WWTF VRG13-007).

594 ## Conflict of interests

595 The authors declare no potential conflicts of interest with respect to the research, authorship,
596 and/or publication of this article. RL received travel grants and/or conference speaker
597 honoraria within the last three years from Shire, Heel, Bruker, and support from Siemens
598 Healthcare regarding clinical research using PET/MR. He is a shareholder of BM Health
599 GmbH since 2019.

600

Methods and materials

Subjects

One hundred healthy young males between 19 and 34 years (M_{age} = 24.86, *SD* = 3.53) participated in the study. We only included men in this study as the employed administration procedure was previously validated on a sample of health young men. Therefore, these results are only representative for the male population and need replication in women as well. All subjects had a body mass index (BMI) between 19.3 and 31.5 (*M* = 24.45, *SD* = 2.86). We screened potential subjects for the presence or a history of psychiatric disorders, self-reported weight and height, concurrent involvement in other studies with pharmacological agents, and presence of a chronic physical injury that might have prevented them from participation in a longer experiment. The short version of e-MINI [51] was used to screen and exclude those who had a non-diagnosed, disclosed, or a diagnosed psychiatric disorder. The screening procedure and the sample size estimate were based on previous work for which we obtained pharmacokinetic data for a single 2 mg estradiol dose in topical form [34]. Subjects were recruited through social media, web portals, and flyers on university premises. All subjects provided written informed consent and were financially compensated for the completion of the experiment (50€) and received an additional maximum bonus of 40€ (range 7€ – 30€) based on their performance in the all the tasks. The procedure described was performed in accordance with the Declaration of Helsinki and approved by the Medical Ethics Committee of the University of Vienna (1918/2015).

Measurement Instruments

Questionnaires

We used a battery of questionnaires to assess self-reported mood (German Multidimensial Mood State Questionnaire; [52], individuals' impulsiveness (Barratt Impulsiveness Scale, BIS-

11; [53]), and reward responsiveness (BIS/BAS; [54]), to test for changes after estradiol administration and ensure there were no interindividual differences between both groups, as previously both BIS/BAS and BIS-11 scores have been found to correlate with reward learning [55–58] (see Supplementary Materials). In addition, we probed subjects' beliefs and confidence about estradiol (e.g. whether they believed they received estradiol or a placebo, how certain they were of this answer, and whether they noticed any changes). This was done to later regress out the potential contribution of beliefs arising, for example, from subjects researching potential side effects of the hormone prior the experiment. Namely, individuals' beliefs about having received the hormone and beliefs about the effects of the hormone on their performance have previously shown to modulate behaviour independent of whether subjects had received the hormone [59].

Hormone concentrations

We collected hormone samples via passive drool and stored them at -30 degrees Celsius. Saliva samples were analyzed for estrone and estradiol using gas chromatography tandem mass spectrometry (GC-MS/MS) and hydrocortisone including testosterone with liquid chromatography tandem mass spectrometry (LCMS/MS) (see Supplementary Materials for details of procedure).

Genotyping

We collected DNA using sterile cotton buccal swabs (Sarstedt AG, Germany) and extracted it by applying the QIAamp DNA Mini kit (Qiagen, Germany). Repeat length polymorphisms (AR(CAG), AR(GGN), DAT1(VNTR), ERα(TA) and ERβ(CA)) were investigated by PCR with fluorescent-dye-labeled primers and capillary electrophoresis. The single base primer extension (SBE) method also known as minisequencing was applied for the typing of single

nucleotide polymorphism (SNP) variants (Val158Met) in the COMT gene (see Supplementary Materials for details of procedure).

Experimental Tasks

For each task, we gave subjects paper instructions including control questions to check whether all subjects understood the instructions. All tasks except for the N-BACK task were monetarily incentivised.

Working memory capacity: We assessed working memory capacity using an adapted version of the standard N-BACK task [16]. In our version we added a 1-BACK condition, creating four conditions in total (i.e. a 0-BACK, 1-BACK, 2- BACK, 3-BACK). One condition block had 20 trials which included 20% target, 65% nontarget, and 15% lure trials. Subjects were presented with a sequence of letters one-by-one. For each letter, they had to decide if the current letter was the same as the one presented *N* trials ago by pressing "R", in case it was not the same they had to press "O". For example, in the 3-back condition, the letter sequence "***A B D A A***" would require subjects to press "R" only to the second occurrence of A, as this was the same letter as the one 3 trials ago. The last A in this example sequence is defined as a lure trial, while the other letters were nontarget trials. Lure trials were present only in the 2-BACK and 3-BACK conditions as in [16], and while lure trials were added to keep the task consistent with their implementation, we did not further analyse them separately as they were not relevant for our question. In total, there were four blocks per condition. Each block was announced by an instruction lasting for 2 sec (Fig. 1A), a fixation cross (1 sec) and a sequence of 20 trials. Each trial was presented for 1 sec with a 1 sec feedback phase and a 1 sec inter-stimulus interval. After every 20 trials, subjects had a 3 sec resting period, before the next block was announced. A lack of response to any cue was considered a miss.

Reinforcement Learning: We employed a probabilistic reinforcement learning task [19] to investigate differences in choice behaviour based on the hypothesized altered reward

677 processing. The task consisted of 500 trials, with a 10 second pause after the first 250 trials.
678 Prior to this, subjects performed 10 practice trials with two initial options, which were changed
679 before the main trials. We did this to avoid carry-over effects from practice to the main task.
680 Throughout the task subjects were exposed to the same set of two options with independently
681 varying reward probabilities. We informed subjects that it was possible that both options could
682 be correct (i.e. rewarding) or incorrect (i.e. non-rewarding) on any given trial, as the reward
683 probability of one option was independent of the other and vice versa. As shown in Fig. 1,
684 each trial included three stages: (1) a cue onset stage (5 sec) where subjects had to decide
685 between the two options and press the corresponding key. If they did not respond within that
686 time frame, they would see a warning message indicating they should respond and try to be
687 faster next time; (2) a choice feedback stage (1 s) where subjects received information about
688 both the chosen (thick frame) and unchosen (thin frame) option (yellow - correct, red - wrong);
689 and (3) an inter-trial interval (*M* = 1.5 s, jittered between 0.9 to 2.1 s). Each correct choice was
690 rewarded with 5 eurocents and added to their cumulative balance. To amplify the association
691 between their performance and earnings, subjects saw a yellow bar filling up incrementally
692 with each correct response. Each time the bar was completely filled, a 1 € coin was presented
693 next to the bar indicating they had gained 1 € to their cumulative balance.

694

695 **Procedure**

696 We asked potential candidates to fill out an online survey with screening questions probing for
697 exclusion criteria described in *Subjects*. Following this, we screened them for the general
698 exclusion criteria. We invited suitable candidates to two separate test sessions. They were
699 scheduled to occur with a maximal difference of one week to prevent major changes in weight
700 and/or other bodily measures.

701 The first session always took place at 4.00 pm. We first provided subjects general
702 information about the study procedure, after which subjects provided written informed consent

and filled out a battery of questionnaires. Moreover, we assessed their height, weight, abdominal, and visceral fat. These metrics were included as they could impact estradiol metabolization, and therefore, we included them as nuisance regressors in our linear models [60,61]. Twenty minutes after arrival, subjects provided a saliva sample. At the end of the session, we obtained a small amount of blood from the finger on a Micro FTA card and a buccal swab for genotyping.

On the second test day (see timeline, Fig. 1, bottom panel) we gave subjects general instructions and information regarding the day. After subjects provided informed consent, they filled out a mood (MDBF-A scale) and impulsiveness (BIS-11) questionnaire. We obtained a first saliva sample (T1, 20 minutes after arrival) to assess baseline hormone concentrations. This was followed by the N-BACK task which we used to assess their baseline working memory performance. Following the N-BACK, subjects applied a topical transparent gel on their chest and shoulders that either contained 2 mg of estradiol (Divigel, Orion Pharma AG, Zug Switzerland) or a placebo. They were randomly assigned estradiol or placebo in a double-blind manner. A male experimenter was present to ensure that the subjects applied the gel correctly. After gel application, we waited for two hours to allow estradiol levels to peak based on our previously established procedure [34]. During this time subjects could read magazines available in the room or books they brought with them. Fifteen minutes prior to the behavioural testing, we required them to fill out a second mood (MDBF-B scale) and impulsiveness (BIS-11) questionnaire followed by a second saliva sample (T2).

The behavioural testing commenced two hours after administration of the drug. The first task was the probabilistic reinforcement learning task which contained a block of practice trials to familiarise subjects with the task setup. After they completed the reinforcement learning task, three other decision-making tasks that were not the focus of this publication followed. After the behavioural testing, we probed subjects' beliefs about the treatment and the tasks. At the end of the study, each participant was paid in accordance to their performance.

730

731 ## Analysis of behaviour

732 ### Statistical analysis of behaviour

733 For the reinforcement learning task, we first looked at the cumulative difference in response
734 proportions between the estradiol and placebo group. That is, we first computed the relative
735 response probability for each group. This value tells us what percentage of subjects from the
736 estradiol/placebo group chose one of the two options (e.g. option A, Fig. 2A). For the relative
737 response probability, we also computed the corresponding standard errors of the mean which
738 gave us a group-level probability and confidence estimate for choosing, e.g. option A, on each
739 trial. We then subtracted the mean and both the lower and upper bound of the standard error
740 of the mean between both groups for each trial. This gave us a difference in the expected
741 chosen option for each trial that reflected how strong the groups differed in the probability of
742 choosing, e.g. option A. Because we were interested in the absolute difference (i.e. we were
743 not interested in the sign of the difference), we took the absolute value on a per trial basis and
744 computed the cumulative choice difference from this which is presented in Fig. 2B.

745 To quantify statistical significance for this metric, on each trial we shuffled the
746 responses of subjects and therefore decoupled labels from responses to build a null
747 distribution that would tell us what kind of difference would be expected by chance. By shuffling
748 responses on each trial, we took a more conservative approach to a permutation test when
749 compared to shuffling responses within and across trials as it preserves systematic variance
750 across trials in terms of subjects' choice. We then generated a null distribution of 2000
751 iterations where for each iteration we computed the cumulative choice difference between two
752 random groups that would be expected by chance. From these cumulative difference traces,
753 we took the 100^{th} percentile of the null distribution for each trial (null distribution in Fig. 2B).
754 This value shows the maximum possible cumulative value that would have been expected by
755 chance (i.e. by two random groups). Therefore, values that exceed this null distribution cannot

be attributable due to chance. Namely, if estradiol administration would not have impacted choice behaviour systematically, then cumulatively the difference between the actual estradiol and placebo group would not surpass the threshold of the null distribution.

We also computed this metric by averaging across trials. This gave us a measure of the average percentage in choice difference that was cumulative across trials. That is, on average, how strongly estradiol influenced choice difference. As above, we also did the same to the corresponding null distribution to observe whether the obtained empirical percentage exceeded the null distribution showing us what would have been expected by chance.

Similarly, we employed two-sample proportion z-tests which tests for whether the proportion of successes from one group is statistically different from the proportion of successes in the other group. These tests were not performed on the relative response probabilities but on the raw responses. That is, we tested whether the number of subjects who chose option A in one group was statistically significantly different from the other group. We repeated this test on every trial to determine on what percentage of trials there was a statistically significant difference between both groups.

As a measure of family-wise error control and to ensure that the values we observed were not due to chance, but due to estradiol administration, we shuffled the responses from subjects for each trial 2000 times and thereby decoupled responses from the labels. This yielded a null distribution that showed on what percentage of trials we could expect to find a statistically significant difference between two random groups with intact response variance across trials. By intact response variance we mean that on some trials, both groups were more likely to select one or the other option. Therefore, if we had also shuffled across trials and subjects, it would have been possible to invoke a larger number of false positives in our null distribution (i.e. lower percentages of trials with a statistically significant difference between both random groups). In short, for each permutation test we obtained a percentage reflecting the number of trials with a statistically significant difference in response proportions between two random groups that would have been obtained by chance.

For all cases where we computed a null distribution, we computed z-scores as measures of standardized effect size, as in [62]. We obtained a z-score by subtracting from the quantity of interest the mean of the null distribution and dividing it by the standard deviation of the null distribution. From this, we were able to use the Fisher-z-transformation to determine statistical significance.

Next, we computed accuracy, defined as the proportion of responses where the option with higher probability of reward was chosen. We collapsed this value across time (Fig. 2C). We computed two additional metrics. The first metric was a measure of switching behaviour; the number of trials where the chosen option on trial *t* and the one chosen at *t + 1* were different. The second metric quantified how many trials on average would subjects stay with the same option on subsequent trials if they were rewarded for the same option on trial *t*. We used this metric as a measure of staying behaviour.

Accuracy, reaction times, switching, and staying were statistically evaluated with general linear models where the first model always included a predictor for drug administration (estradiol, placebo). For all models we subsequently included interaction terms for the polymorphisms of genes of interest. Unless explicitly mentioned in the main result section, all reported linear models regressed out z-scored nuissance regressors. These included cortisol levels following administration, beliefs about the drug (see Belief Probes), and body measurement characteristics (weight, BMI, abdominal and visceral fat). Weight and BMI were summed together to generate a composite score [63] because of their high intrinsic correlation ($r = 0.89$). (See also Supplementary Materials: Selecting linear models). General linear models for accuracy also included z-scored reaction times to control for accuracy-speed trade-offs.

In addition, we analysed choice autocorrelation (see Supplementary Materials: Impact of previous choice on current choice). In brief, for each participant we computed the relative contribution of choices made from $t - 1$ to $t - 7$ trials back (lags) on current choice. The obtained regression weights indicated how strong the relative influence of individual trials on the current choice was. We performed this both for choice as a function of previous choice

(pure choice autocorrelation) and choice as a function of previous rewarded choice (choice autocorrelation as a function of reward). We then performed independent samples Welch *t*-tests on individual lags to assess statistical significance.

To control for the variance of random effects such as subjects themselves, we used generalized linear mixed effects models that do not require data aggregation [39,64]. In two separate sets of analyses, we investigated whether treatment group (estradiol, placebo) interacted with the val^{158}met polymorphism of the COMT gene or with the VNTR polymorphism of the DAT1 gene across trials. We fitted separate models, as the sample size per smallest cell was too small otherwise (Table S6, Supplementary Materials). We ran these models using R (version 3.6.0 R Development Core Team, 2019), with the lme4 package [64]. Our simplest model included only an intercept and a random effects structure which included subject-level intercepts. We used a likelihood ratio test to determine whether including group as a fixed factor improved the model fit. From there we fitted separate models for the VNTR polymorphism of the DAT1 gene and the val^{158}met polymorphism of the COMT gene. In both cases, the starting model had a fixed effect interaction between group (estradiol, placebo) and gene (either COMT or DAT1) and subject-level intercepts as random effects. From this model we incrementally increased the complexity of our model until the most complex one. The most complex model was identical for both the VNTR and val^{158}met polymorphism. The model included a three-way interaction between group (estradiol, placebo), gene (COMT or DAT1) and time (trial number). This was our main measure of interest and the one for which we hypothesized effects – that estradiol administration would differentially influence choice as the task progressed, depending on subjects' genotype. The random effect structure for this model included random intercepts for each subject. All models were estimated using the "nloptwrap" optimizer. Models without convergence or singularity warnings were then compared with likelihood ratio tests. We used BIC [65] to pick the winning model but also inspected their AIC [66] and deviance scores for converging information. Below we report the two winning models; both models were identical, except for the polymorphism:

$$y = choice \sim group * COMT * trial + (1|subject) \quad 1$$

In the case of DAT1, the winning model was:

$$y = choice \sim group * DAT1 * trial + (1|subject) \quad 2$$

Computational modelling

A canonical approach to estimate subjects' learning is afforded by reinforcement learning. To test if subjects in the estradiol group would behave differently compared to the placebo group, because of increased striatal prediction errors, we formalized behaviour within a reinforcement learning framework and fitted several Q-learning models [67] with softmax choice rules:

Q-learning model (equation 3):

$$Q_{t+1}^{A} = Q_t^A + \alpha(R_t^A - Q_t^A) \quad (3)$$

Softmax choice rule (equation 4):

$$p = \frac{1}{1 + e^{\tau(Q_t^A - Q_t^B)}} \quad (4)$$

Where, t is time, A is option A, Q is subjective value, α is the learning rate, R is the obtained reward, and τ is the temperature parameter. Equations 3 and 4 represent our first model (model 1). In Q-learning, the basic idea is that agents learn subjective values for actions in their environment. Subjective values are learned and updated through a value function (Equation 3) following feedback after each action. A teaching signal known as the learning rate-weighted prediction error dictates how strongly the subjective value will be updated on each action. The prediction error corresponds to the difference between the obtained and expected reward (i.e. the subjective value prior to making the new choice). Within this process, the learning rate dictates how heavily new information will be weighted in proportion to previous information about the option, and therefore how strongly the subjective value will change from its current estimate. The softmax equation then yields the probability of selecting

an action given the learning rate and the temperature parameter, which reflects stochasticity of choice behaviour.

By employing computational modelling of this sort, we were able to obtain parameter estimates that quantify the difference in subjects' behaviour which we predicted. Our main hypothesis was that estradiol would increase reward sensitivity which should be captured by the learning rate, but not influence choice stochasticity across trials.

To obtain a more precise account of the effect of estradiol on reward processing, we extended the basic Q-learning model in several ways, as described below.

The first extension (model 2, equation 5a and 5b) allowed for separate learning rates for Q_A and Q_B, because subjects were able to track the outcome of both the chosen and unchosen option.

$$Q_{t+1}^{A} = Q_t^A + \alpha_A(R_t^A - Q_t^A) \quad (5a)$$

$$Q_{t+1}^{B} = Q_t^B + \alpha_B(R_t^B - Q_t^B) \quad (5b)$$

Furthermore, due to reward stochasticity of our n-armed bandit implementation (obtained by a Gaussian random walk – Fig. 1B), we added an additional parameter ξ, representing irreducible noise [68] in our perceptual model (model 3, equation 6):

$$p = \frac{1}{1 + e^{\tau(Q_t^A - Q_t^B)}}(1-\xi) + \frac{\xi}{2} \quad (6)$$

Finally, we added a perseverance parameter λ [69] to the response model (model 4, equation 7):

$$p = \frac{1}{1 + e^{\tau(Q_t^A - Q_t^B + \lambda C)}}(1-\xi) + \frac{\xi}{2} \quad (7)$$

Where C = 1, if the same cue was chosen on trial *n* and trial *n+1*, and C = -1 if the converse was true. In summary, our full model space had separate learning rates for two separate

884 options, a choice stochasticity, and irreducible noise parameter. All other models were
885 reduced cases of this model and all possible combinations of the described free parameters
886 therefore yielded eight models in total for which we estimated parameters. The model fitting
887 was performed using JAGS and the rjags (v 4.9) package in R (v 3.6.0). Each model was run
888 with 5000 samples each with 1000 burn-in samples on three chains. Priors over parameters
889 and hyperparameters were set to default as described in [70]. We computed the leave one out
890 information criterion using the loo package [71] and used this metric to compare the models.
891 Furthermore, we performed Bayesian model comparison by computing the (protected)
892 exceedance probability [72] using the VBA toolbox [73] to determine the best model and compare
893 its congruency with the LOOIC measure. Finally, we extracted the posterior predictive density
894 for each participant as a measure of predictive power of the best model. This was then
895 compared to the actual behaviour as a measure of static (accuracy collapsed across time) and
896 dynamic (accuracy at each trial across subjects) predictive accuracy.

897

898 ## References

899 1. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward.
900 *Science (80-.).* **275**, 1593–1599 (1997).

901 2. Diekhof, E. K. Estradiol and the reward system in humans. *Curr. Opin. Behav. Sci.* **23**,
902 58–64 (2018).

903 3. Saldanha, C. J., Remage-Healey, L. & Schlinger, B. A. Synaptocrine signaling:
904 Steroid synthesis and action at the synapse. *Endocr. Rev.* **32**, 532–549 (2011).

905 4. Luine, V. N. Estradiol and cognitive function: Past, present and future. *Horm. Behav.*
906 **66**, 602–618 (2014).

907 5. Colzato, L. S. & Hommel, B. Effects of estrogen on higher-order cognitive functions in
908 unstressed human females may depend on individual variation in dopamine baseline

levels. *Front. Neurosci.* **8**, 65 (2014).

6. Thomas, J., Météreau, E., Déchaud, H., Pugeat, M. & Dreher, J. Hormonal treatment increases the response of the reward system at the menopause transition : A counterbalanced randomized placebo-controlled fMRI study. *Psychoneuroendocrinology* **50**, 167–180 (2014).

7. Dreher, J. *et al.* Menstrual cycle phase modulates reward-related neural function in women. *PNAS* **104**, 2465–2470 (2007).

8. Diekhof, E. K. & Ratnayake, M. Menstrual cycle phase modulates reward sensitivity and performance monitoring in young women: Preliminary fMRI evidence. *Neuropsychologia* **84**, 70–80 (2016).

9. Lévesque, D. & Di Paolo, T. Rapid conversion of high into low striatal D2-dopamine receptor agonist binding states after an acute physiological dose of 17β-estradiol. *Neurosci. Lett.* **88**, 113–118 (1988).

10. Becker, J. B. Gender Differences in Dopaminergic Function in Striatum and Nucleus Accumbens. *Pharmacol. Biochem. Behav.* **64**, 803–812 (1999).

11. Becker, J. B. Direct effect of 17β-estradiol on striatum: Sex differences in dopamine release. *Synapse* **5**, 157–164 (1990).

12. Pasqualini, C., Olivier, V., Guibert, B., Frain, O. & Leviel, V. Acute Stimulatory Effect of Estradiol on Striatal Dopamine Synthesis. *J. Neurochem.* **65**, 1651–1657 (1995).

13. Ball, P., Knuppen, R., Haupt, M. & Breuer, H. Interactions Between Estrogens and Cateehol Amines III. Studies on the Methylation of Catechol Estrogens, Catechol Amines and other Catechols by the Catechol-O-Methyltransferase of Human Liver. *J. Clin. Endocrinol. Metab.* **34**, 736–746 (1972).

14. Yoest, K. E., Quigley, J. A. & Becker, J. B. Rapid effects of ovarian hormones in dorsal striatum and nucleus accumbens. *Horm. Behav.* **104**, 119–129 (2018).

15. Männistö, P. T. & Kaakkola, S. Catechol-O-methyltransferase (COMT): Biochemistry, molecular biology, pharmacology, and clinical efficacy of the new selective COMT inhibitors. *Pharmacol. Rev.* **51**, 593–628 (1999).

16. Jacobs, E. & D'Esposito, M. Estrogen Shapes Dopamine-Dependent Cognitive Processes: Implications for Women's Health. *J. Neurosci.* **31**, 5286–5293 (2011).

17. Schultz, W., Stauffer, R. W. & Lak, A. The phasic dopamine signal maturing: from reward via behavioural activation to formal economic utility. *Curr. Opin. Neurobiol.* **43**, 139–148 (2017).

18. Glimcher, P. W. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *PNAS* **108**, 15647–54 (2011).

19. Frank, M. J., Seeberger, L. C. & O'Reilly, R. C. By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. *Science (80-.).* **306**, 1940–1943 (2004).

20. Cools, R. *et al.* Striatal Dopamine Predicts Outcome-Specific Reversal Learning and Its Sensitivity to Dopaminergic Drug Administration. *J. Neurosci.* **29**, 1538–1543 (2009).

21. den Ouden, H. E. M. *et al.* Dissociable Effects of Dopamine and Serotonin on Reversal Learning. *Neuron* **80**, 1090–1100 (2013).

22. Jocham, G., Klein, T. A. & Ullsperger, M. Dopamine-Mediated Reinforcement Learning Signals in the Striatum and Ventromedial Prefrontal Cortex Underlie Value-Based Choices. *J. Neurosci.* **31**, 1606–1613 (2011).

23. Jocham, G., Klein, T. a & Ullsperger, M. Differential Modulation of Reinforcement Learning by D2 Dopamine and NMDA Glutamate Receptor Antagonism. *J. Neurosci.* **34**, 13151–13162 (2014).

24. Eisenegger, C. *et al.* Role of dopamine D2 receptors in human reinforcement learning. *Neuropsychopharmacology* **39**, 2366–2375 (2014).

959 25. Swart, J. C. *et al.* Catecholaminergic challenge uncovers distinct Pavlovian and
960 instrumental mechanisms of motivated (in)action. *Elife* **6**, 1–36 (2017).

961 26. Frank, M. J. *et al.* Genetic triple dissociation reveals multiple roles for dopamine in
962 reinforcement learning. *PNAS* **104**, 16311–16316 (2007).

963 27. Eisenegger, C. *et al.* DAT1 Polymorphism Determines L-DOPA Effects on Learning
964 about Others ' Prosociality. *PLoS One* **8**, e67820 (2013).

965 28. Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J. & Frith, C. D. Dopamine-
966 dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*
967 **442**, 1042–1045 (2006).

968 29. Doll, B. B., Bath, K. G., Daw, N. D. & Frank, M. J. Variability in dopamine genes
969 dissociates model-based and model-free reinforcement learning. *J. Neurosci.* **36**,
970 1211–1222 (2016).

971 30. Nemmi, F. *et al.* Interaction between striatal volume and DAT1 polymorphism predicts
972 working memory development during adolescence. *Dev. Cogn. Neurosci.* **30**, 191–
973 199 (2018).

974 31. Sommer, T. *et al.* Effects of the experimental administration of oral estrogen on
975 prefrontal functions in healthy young women. *Psychopharmacology (Berl).* **235**, 3465–
976 3477 (2018).

977 32. Bayer, J., Gläscher, J., Finsterbusch, J., Schulte, L. H. & Sommer, T. Linear and
978 inverted U-shaped dose-response functions describe estrogen effects on
979 hippocampal activity in young women. *Nat. Commun.* **9**, 1–12 (2018).

980 33. Jakob, K., Ehrentreich, H., Holtfrerich, S. K. C., Reimers, L. & Diekhof, E. K. DAT1-
981 genotype and menstrual cycle, but not hormonal contraception, modulate
982 reinforcement learning: Preliminary evidence. *Front. Endocrinol. (Lausanne).* **9**, 60
983 (2018).

34. Eisenegger, C., von Eckardstein, A., Fehr, E. & von Eckardstein, S. Pharmacokinetics of testosterone and estradiol gel preparations in healthy young men. *Psychoneuroendocrinology* **38**, 171–178 (2013).

35. Cools, R. & D'Esposito, M. Inverted-U-shaped dopamine actions on human working memory and cognitive control. *Biol. Psychiatry* **69**, 113–125 (2011).

36. Daw, N. D., Doherty, J. P. O., Dayan, P., Seymour, B. & Dolan, R. J. Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876–879 (2006).

37. Reimers, L., Büchel, C. & Diekhof, E. K. How to be patient. The ability to wait for a reward depends on menstrual cycle phase and feedback-related activity. *Front. Neurosci.* **8**, 401 (2014).

38. Baayen, R. H., Davidson, D. J. & Bates, D. M. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* **59**, 390–412 (2008).

39. Bolker, B. M. *et al.* Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution* vol. 24 127–135 (2009).

40. Chowdhury, R. *et al.* Dopamine restores reward prediction errors in old age. *Nat. Neurosci.* **16**, 648–653 (2013).

41. Chowdhury, R., Guitart-Masip, M., Bunzeck, N., Dolan, R. J. & Düzel, E. Dopamine modulates episodic memory persistence in old age. *J. Neurosci.* **32**, 14193–14204 (2012).

42. Smith, C. T., Sierra, Y., Oppler, S. H. & Boettiger, C. A. Ovarian Cycle Effects on Immediate Reward Selection Bias in Humans: A Role for Estradiol. *J. Neurosci.* **34**, 5468–5476 (2014).

43. Diekhof, E. K. Be quick about it. Endogenous estradiol level, menstrual cycle phase and trait impulsiveness predict impulsive choice in the context of reward acquisition. *Horm. Behav.* **74**, 186–193 (2015).

1009 44. Geniole, S. N. *et al.* Using a Psychopharmacogenetic Approach To Identify the
1010 Pathways Through Which—and the People for Whom—Testosterone Promotes
1011 Aggression. *Psychol. Sci.* **30**, 481–494 (2019).

1012 45. Losecaat Vermeer, A. B. *et al.* Exogenous testosterone increases status-seeking
1013 motivation in men with unstable low social status. *Psychoneuroendocrinology* **113**,
1014 (2020).

1015 46. Colzato, L. S., Hertsig, G. & Wildenberg, van den Hommel, B. Estrogen modulates
1016 inhibitory control in healthy human females: evidence from the stop-signal paradigm.
1017 *Neuroscience* **167**, 709–715 (2010).

1018 47. Colzato, L. S., Pratt, J. & Hommel, B. Estrogen modulates inhibition of return in
1019 healthy human females. *Neuropsychologia* **50**, 98–103 (2012).

1020 48. Schaaf, M. E. Van Der, Fallon, S. J., Huurne, N., Buitelaar, J. & Cools, R. Working
1021 Memory Capacity Predicts Effects of Methylphenidate on Reversal Learning.
1022 *Neuropsychopharmacology* **38**, 2011–2018 (2013).

1023 49. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-based
1024 influences on humans' choices and striatal prediction errors. *Neuron* **69**, 1204–1215
1025 (2011).

1026 50. Wunderlich, K., Smittenaar, P. & Dolan, R. J. Dopamine Enhances Model-Based over
1027 Model-Free Choice Behavior. *Neuron* **75**, 418–424 (2012).

1028 51. Sheehan, D. V. *et al.* The validity of the Mini International Neuropsychiatric Interview
1029 (MINI) according to the SCID-P and its reliability. *Eur. Psychiatry* **12**, 232–241 (1997).

1030 52. Steyer, R., Schwenkmezger, P., Notz, P. & Eid, M. Testtheoretische Analysen des
1031 Mehrdimensionalen Befindlichkeitsfragebogen. *Diagnostica* **40**, 320–328 (1994).

1032 53. Patton, J. H., Stanford, M. S. & Barratt, E. S. Factor structure of the barratt
1033 impulsiveness scale. *J. Clin. Psychol.* **51**, 768–774 (1995).

54. Carver, Charles, S. & White, Teri, L. Behavioral Inhibition, Behavioral Activation, and Affective Responses to Impending Reward and Punishment: The BIS/BAS Scales. *J. Pers. Soc. Psychol.* **67**, 319–333 (1994).

55. Kim, S. H., Yoon, H. S., Kim, H. & Hamann, S. Individual differences in sensitivity to reward and punishment and neural activity during reward and avoidance learning. *Soc. Cogn. Affect. Neurosci.* **10**, 1219–1227 (2014).

56. Sali, A. W., Anderson, B. A. & Yantis, S. Reinforcement learning modulates the stability of cognitive control settings for object selection. *Front. Integr. Neurosci.* **7**, 95 (2013).

57. Unger, K., Heintz, S. & Kray, J. Punishment sensitivity modulates the processing of negative feedback but not error-induced learning. *Front. Hum. Neurosci.* **6**, 186 (2012).

58. Lighthall, N. R., Gorlick, M. A., Schoeke, A., Frank, M. J. & Mather, M. Stress Modulates Reinforcement Learning in Younger and Older Adults. *Psychol Aging* **28**, 35–46 (2013).

59. Eisenegger, C., Naef, M., Snozzi, R., Heinrichs, M. & Fehr, E. Prejudice and truth about the effect of testosterone on human bargaining behaviour. *Nature* **463**, 356–359 (2010).

60. Fishman, J., Boyar, R. M. & Hellman, L. Influence of body weight on estradiol metabolism in young women. *J. Clin. Endocrinol. Metab.* **41**, 989–991 (1975).

61. Schneider, J. *et al.* Effects of Obesity on Estradiol Metabolism: Decreased Formation of Nonuterotropic Metabolites. *J. Clin. Endocrinol. Metab.* **56**, 973–978 (1983).

62. Maidenbaum, S., Miller, J., Stein, J. M. & Jacobs, J. Grid-like hexadirectional modulation of human entorhinal theta oscillations. *PNAS* **115**, 10798–10803 (2018).

63. Aeberli, I., Molinari, L. & Zimmermann, M. B. A composite score combining waist

circumference and body mass index more accurately predicts body fat percentage in 6- to 13-year-old children. *Eur. J. Nutr.* **52**, 247–253 (2013).

64. Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, (2015).

65. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **6**, 461–464 (1978).

66. Akaike, H. A New Look at the Statistical Model Identification. *IEEE Trans. Automat. Contr.* **19**, 716–723 (1974).

67. Watkins, C. J. C. H. & Dayan, P. Q-learning. *Mach. Learn.* **8**, 279–292 (1992).

68. de Boer, L. *et al.* Dorsal striatal dopamine D1 receptor availability predicts an instrumental bias in action learning. *PNAS* **116**, 261–270 (2019).

69. Rutledge, R. B. *et al.* Dopaminergic Drugs Modulate Learning Rates and Perseveration in Parkinson ' s Patients in a Dynamic Foraging Task. *J. Neurosci.* **29**, 15104–15114 (2009).

70. Ahn, W.-Y., Haines, N. & Zhang, L. Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package. *Comput. Psychiatry* **1**, 24–57 (2017).

71. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).

72. Rigoux, L., Stephan, K. E., Friston, K. J. & Daunizeau, J. Bayesian model selection for group studies - Revisited. *Neuroimage* **84**, 971–985 (2014).

73. Daunizeau, J., Adam, V. & Rigoux, L. VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. *PLoS Comput. Biol.* **10**, (2014).

74. Longcope, C., Kato, T. & Horton, R. Conversion of blood androgens to estrogens in normal adult men and women. *J. Clin. Invest.* **48**, 2191–2201 (1969).

75. Yaffe, K. *et al.* Androgen receptor CAG repeat polymorphism is associated with cognitive function in older men. *Biol. Psychiatry* **54**, 943–946 (2003).

76. Beyenburg, S. *et al.* Androgen receptor mRNA expression in the human hippocampus. *Neurosci. Lett.* **294**, 25–28 (2000).

77. Kovacs, D. *et al.* The androgen receptor gene polyglycine repeat polymorphism is associated with memory performance in healthy Chinese individuals. *Psychoneuroendocrinology* **34**, 947–952 (2009).

78. Comings, D. E., Chen, C., Wu, S. & Muhleman, D. Association of the androgen receptor gene (AR) with ADHD and conduct disorder. *Neuroreport* **10**, 1589–1592 (1999).

79. Gillies, G. E. & McArthur, S. Estrogen actions in the brain and the basis for differential action in men and women: A case for sex-specific medicines. *Pharmacol. Rev.* **62**, 155–198 (2010).

80. Bayer, J. *et al.* Estrogen and the male hippocampus: Genetic variation in the aromatase gene predicting serum estrogen is associated with hippocampal gray matter volume in men. *Hippocampus* **23**, 117–121 (2013).

81. Kravitz, H. M., Meyer, P. M., Seeman, T. E., Greendale, G. A. & Sowers, M. F. R. Cognitive Functioning and Sex Steroid Hormone Gene Polymorphisms in Women at Midlife. *Am. J. Med.* **119**, 94–102 (2006).

82. Ma, S. L. *et al.* Polymorphisms of the estrogen receptor (ESR1) gene and the risk of Alzheimer's disease in a southern Chinese community. *Int. Psychogeriatrics* **21**, 977–986 (2009).

83. Almey, A., Milner, T. A. & Brake, W. G. Estrogen receptors in the central nervous system and their implication for dopamine-dependent cognition in females. *Horm. Behav.* **74**, 125–138 (2015).

1108

1109

Supplementary Materials:

A causal role for estradiol in human reinforcement learning

Sebastijan Veselic[1,2,3]*, Gerhard Jocham[4], Christian Gausterer[5], Bernhard Wagner[6], Miriam Ernhoefer-Reßler[6], Rupert Lanzenberger[7], Clauss Lamm[1,8], Christoph Eisenegger[1] & Annabel Losecaat Vermeer[1]*

Methods

Genotyping

DNA extraction and quantification

Buccal swabs were collected using sterile cotton swabs (Sarstedt AG, Germany). DNA was extracted from swabs using the QIAamp DNA Mini kit (Qiagen, Hilden, Germany) and eluted in a final volume of 50 µL of QIAamp buffer AE (Qiagen). Human nuclear DNA was quantified using the Applied Biosystems (AB) 7500 real-time PCR instrument (Thermo Fisher Scientific, Waltham, MA) and the Quantifiler Human Plus quantification Kit (AB) following manufacturer's recommendations.

Typing of repeat length polymorphisms

Genomic DNA fragments that contain polymorphic repeat sequences were amplified in two separate reactions: i.e. a multiplex PCR (simultaneously targeting AR(CAG)n, DAT1 VNTR, Erα(TA)n and Erβ(CA)n) and a singleplex PCR (targeting solely AR(GGN)n), respectively.

The multiplex PCR was performed using 5 ng template DNA in a reaction mix (total volume of 25 µL) consisting of 1 × GeneAmp PCR buffer (AB), 0.25 mM each dNTP, 2.5 units AmpliTaq Gold polymerase (AB) and target specific primers (AR(CAG), DAT1, ERα and Erβ; including 5'-fluorescent-dye-labeled forward primers; details provided in

Table 1). The following protocol was applied using the Veriti 96-well thermal cycler (AB): 35 cycles at 95 °C for 30 seconds, 55 °C for 1 minute, and 72 °C for 1 minute. Before the first cycle, an initial denaturation (95 °C for 5 minutes) was included, and the last cycle was followed by a final extension step at 72 °C for 45 minutes.

The singleplex PCR was conducted using 5 ng template DNA in a reaction mix (total volume of 20 µL) containing target specific primers (AR(GGN)n, details provided in Table 1)), 0.5 µL Phire Hot Start II DNA polymerase (Thermo Fisher) in 1 × Phire reaction buffer (Thermo Fisher). Amplification was carried out on the Veriti thermal cycler (AB) and included an initial denaturation step at 98 °C for 30 seconds, followed by 33 cycles of 10 seconds at 98 °C, 30 seconds at 60 °C and 30 seconds at 72 °C. The last cycle was followed by a final extension at 72 °C for 10 minutes.

Aliquots of PCR products were diluted with Hi-Di formamide (AB), mixed with internal lane standard LIZ 600 v.2 (AB) and separated on the ABI 3500 Genetic Analyzer applying standard conditions. The number of repeats predicted by the GeneMapper ID-X software (AB) was in full agreement to the actual repeats determined by direct sequencing of PCR products using the BigDye Terminator Sequencing Kit v3.1 (AB) in selected DNA samples.

Table SI. Panel of loci and primer sets used for the typing of repeat length polymorphisms

Marker	Location[b]	Primer sequence 5'-3' [c]	Dye	Orientation	Conc. (nM)[d]
AR(GGN)n[a]	chrX:67546447-67546603(+)	CCTGGCACACTCTCTTCACA	VIC	forward	625
		*GTTTCT*GGCCGAGTGTAGCCGTAG		reverse	
AR(CAG)n	chrX:67545237-67545434(+)	CGCGAAGTGATCCAGAACC	6-FAM	forward	200
		*GTTTCT*AGAACCATCCTCACCCTGCT		reverse	
DAT1 VNTR	chr5:1393559-1394008(-)	TGTGGTGTAGGGAACGGCCTGAGA	6-FAM	forward	400
		TGTTGGTCTGCAGGCTGCCTGCAT		reverse	
ERα(TA)n	chr6:151806472-151806594(+)	AACTATCCAAGATTATAGACGCATGA	NED	forward	600
		*GTTTCT*AACATGCACACGCACATACA		reverse	
Erβ(CA)n	chr14:64253529-64253650(-)	GTGCTGCGAGCAGAGATA	PET	forward	800
		*GTTTCT*AATGAGTGGGCCTCCCTTAG		reverse	

[a] AR(GGN)n primers only used in singleplex PCR; all other primers combined in a multiplex PCR
[b] Chromosome number and genomic location of targeted sequence (orientation provided in brackets) according to UCSC version hg38 (http://genome.ucsc.edu/)
[c] The non-specific primer tail is underlined in Italics
[d] The final primer concentrations in the reaction mix

Typing of the COMT Val158Met polymorphism

SNaPshot minisequencing was applied for the typing of Val158Met variants in the COMT gene. Therefore, a 177 bp fragment of genomic DNA harbouring the causative single nucleotide polymorphism (SNP rs4680) in its centre was amplified by PCR. The reaction mix comprised 5 ng template DNA, 1 × GeneAmp PCR buffer (AB), 0.25 mM each dNTP, 2.5 units AmpliTaq Gold polymerase (AB) and target specific primers (details provided in Table 2) in a total reaction volume of 25 µL. Thermal cycling was performed applying the Veriti cycler (AB) and conditions as follows: 95 °C for 5 min; 35 cycles of 95 °C for 15 seconds, 59 °C for 30 seconds and 72 °C for 1 minute; final extension at 72 °C for 5 minutes.

1179 **Table S2. Primer set used for PCR of the COMT fragment**

Marker	Location[a]	Primer sequence 5'-3' [b]	Orientation	Conc. (nM)[c]
COMT	chr22:19963623-19963799(+)	GGGCCTACTGTGGCTACTCA	forward	400
		GCCCTTTTTCCAGGTCTGA	reverse	

1180 [a] Chromosome number and genomic location of targeted sequence (orientation provided in brackets)
1181 according to UCSC version hg38 (http://genome.ucsc.edu/).

1182

1183 PCR products were purified from excess primers and dNTPs by ExoSAP-IT (Thermo
1184 Fisher) treatment following manufacturer's recommendations. Minisequencing was
1185 conducted on a Veriti thermal cycler (AB) in a total volume of 10 µL containing 3 µL of
1186 purified PCR product, 5 µL SNaPshot Multiplex Ready Reaction mix (Thermo Fisher)
1187 and 2 µL minisequencing primer (2 µM; details see Table 3). The cycling conditions (25
1188 cycles) were as follows: denaturation at 96 °C for 10 seconds, annealing at 50 °C for 5
1189 seconds and extension at 60 °C for 30 seconds.

1190

1191 **Table S3. Minisequencing primer information**

SNP sequence variation	Location[a]	Primer sequence 5'-3' [b]	Orientation
G>A	chr22:19963728-19963747	*(GATC)*$_4$ GGATGGTGGATTTCGCTGGC	forward

1192 [a] Chromosome number and genomic location of primer binding site (orientation provided in brackets)
1193 according to UCSC version hg38 (http://genome.ucsc.edu/). The targeted SNP rs4680 is located at position
1194 chr22:19963748.
1195 [b] The non-specific primer tail is underlined in Italics
1196

1197 ExoSAP-IT treatment was again applied for the clean-up of the minisequencing
1198 reaction. 5 µl of purified minisequencing reaction product was then mixed with 9.3 µL
1199 Hi-Di formamide (AB) and 0.2 µL of GeneScan-LIZ 120 internal size standard (AB).
1200 After a denaturing step for 5 min at 98 °C followed by cooling to 4 °C the fragments
1201 were separated on an ABI PRISM 310 Genetic Analyzer (AB) with POP4 polymer and
1202 analysed with GeneMapper v3.2 software. Calling of SNP variants based on

minisequencing was in full agreement to results from direct sequencing of PCR products in selected DNA samples.

Hormone concentrations

Quantification of estrone and estradiol in saliva samples was performed with derivatization using pentafluorobenzoyl chloride (PFBCl) and the addition of the isotopically labeled internal standards estrone-d_4 and estradiol-d_5. Organic saliva was reacted with 1.0 mL 1% PFBCl and 0.1 mL pyridine at 60°C for 30 min. The derivatization agents were evaporated, the sample was reconstituted with 0.5 mL NaHCO3 and extracted with 1 mL n-hexane. The organic phase was substituted with 0.2 mL dodecane and subjected to optimized GC-MS/MS analysis using an Agilent 7890 GC with Agilent DB-17ht 15 m x 0.25 mm x 0.15 µm capillary column connected to an Agilent 7010 tandem mass spectrometer operated in MRM mode using negative chemical ionization at 150°C with methane as a reaction gas (40%, 2 mL/min). Method validation was performed using ion transition m/z 464 -> 400 as a quantifier for estrone and m/z 660 -> 596 for estradiol, whereas a LLOQ of 1.92 fg o.c. and 1.94 fg was obtained, respectively.

Quantification of hydrocortisone and testosterone in saliva samples was performed using liquid chromatography tandem mass spectrometry (LCMS/MS), with an Agilent 6460 with electrospray ionization in positive mode coupled to a 1290 UHPLC system. Collision energy was optimized for specific MRM transitions of Hydrocortisone (363.2/121.1 m/z; 363.2/91.1 m/z), Testosterone (289.2/109.1; 289.2/97.1 m/z), 2,3,4-13C3-Hydrocortisone (366.2/124 m/z) and 2,3,4-13C3-Testosterone (292.2/100 m/z). Agilent Poroshell 120 EC-C18 was used for chromatographic separation under reversed phase conditions. The internal standard preparation and internal standard mixture was prepared containing 2,3,4-13C3-Hydrocortisone; 2,3,4-13C3-Testosterone, 2,4,16,16,17-d5-17b-Estradiol and concentration of 5ng/mL each.

Samples were prepared by adding 100 µl internal standards (5 ng/mL) to 500µl plasma or saliva and the steroids were extracted using 4 mL MTBE. After 10 min. overhead shacking, the samples were centrifuged for 5 min. at 3000 rpm and the top MTBE layer was transferred to a test tube. MTBE was evaporated using a centrivap concentrator at 40°C (Labconco). The residual sample was then re-dissolved in methanol and analyzed by LC-MS/MS.

Questionnaires

***Mood*:** To control for a potential confound of mood, tiredness, or alertness from the treatment affecting subjects' performance [24], we assessed participants' self-reported mood before and after administration of the treatment, using the German Multidimensial Mood State Questionnaire ("Der Mehrdimensionale Befindlichskeitfragebogen - MDBF) [52] Both versions of this questionnaire (A and B) contain 12 items with a 5-level Likert scale and three subscales that test for different continuums of mood (Good-Bad [α_{pre} = .81, α_{post} = .77], Awake-Tired [α_{pre} = .84, α_{post} = .87], Calm-Nervous [α_{pre} = .73, α_{post} = .75]).

***Impulsiveness*:** We used the Barratt Impulsiveness Scale (BIS-11; [53] to measure participants' impulsiveness as [43] observed that variations in estradiol levels differentially affected women with low trait as opposed to high trait impulsiveness. BIS-11 is a widely used measure for impulsiveness with 30 items describing common behaviour and preferences related to (non)impulsiveness which individuals have to rate on a 4-point scale (1 - rarely/never, almost always/always - 4). The General Impulsiveness (α_{pre} = .71, α_{post} = .75) factor together with its three second-order factors (Motor Impulsiveness (α_{pre} = .47, α_{post} = .54) Nonplanning Impulsiveness (α_{pre} = .6, α_{post} = .63), Attentional Impulsiveness (α_{pre} = .49, α_{post} = .52) are reported.

***Behavioural inhibition and activation*:** we measured the trait behavioural activation and inhibition with the Behavioural inhibiton/Behavioural Activation Scales (BIS/BAS; [54]. The BAS scale is a 24-item questionnaire answered on a four-level scale (1- very true for me, 4 - very false for me). It is subdivided into Drive (α= .74), Fun Seeking (α= .67), and Reward Responsiveness (α= .6) while the BIS scale (α= .77) is unidimensional. Drive is thought to measure the persistent pursuit of goals (e.g. "I go out of my way to get the things I want"), Fun Seeking: the desire for new rewards and willingness to approach events that would be potentially rewarding (e.g. "I crave excitement and new sensations"), while Reward Responsiveness focuses on positive responses that would occur if a reward is anticipated (e.g. "When I am doing well at something I love to keep doing it"). Finally, the BIS scale measures sensitivity to negative events (e.g. "Criticism or scolding hurts me quite a bit").

Belief probes

In addition, we probed participants' beliefs and confidence about estradiol (e.g. whether they believed they received estradiol or a placebo, how certain they were of this answer, and whether they noticed any changes). This was done to later regress out the potential contribution of beliefs arising, for example, from participants researching potential side effects of the hormone prior the experiment. Namely, individuals' beliefs about having received the hormone and beliefs about the effects of the hormone on their performance have previously shown to modulate behaviour independent of whether participants had received the hormone [59].

Matching of both groups

We compared both treatment groups for age and other bodily characteristics (i.e. BMI, height, weight, visceral, and abdominal fat) and potential differences in self-reported

mood (MDBF), impulsiveness (BIS-11) and reward responsiveness (BIS/BAS) (see Questionnaires, Table S4 and S5). We used two-tailed independent samples Welch t-tests, or Wilcoxon signed-rank test if assumptions of normality were not met, to test whether the groups matched on all variables. To test for mood differences after administration between the treatment groups, we performed an ANCOVA for each of the three subscales of the MDBF questionnaire where we controlled for baseline mood scores. Two-way ANOVAs were further performed on the individual subscales of the BIS-11 questionnaire to investigate whether there was an interaction between the group (estradiol, placebo) and session (pre, post) on impulsiveness.

To compare working memory capacity assessed by the N-BACK task, we analyzed target accuracy, reaction times, and d-prime. We analyzed this with an ANOVA containing the between-subject variable group (estradiol, placebo) and within-subject variable for condition together with an interaction term for group and condition.

Results

Matching of both groups

In the first part of the supplementary results, Table S4 and S5 show that our random assignment was successful as the groups did not differ in any of the measured parameters before (Table S4) administration and as a function of administration (Table S5). However, we did observe the expected change in estradiol metabolite concentrations in the estradiol group, outlined below.

Hormone concentrations

We observed a statistically significant post-administration difference between both groups in log-transformed estradiol concentrations (W = 1545, 95% CI [0.03, 1.87], p < .05) with the estradiol group having higher estradiol metabolite concentration following administration (estradiol: *Mdn* = 41.77 ±531.54), placebo: *Mdn* = 5.55 ±230.23) but not

before (estradiol: *Mdn* = 3.38 ±230.97), placebo: Mdn = 1.89 ±21.92) compared to the placebo group (*W* = 1498, 95% CI [-0.05, 1.03], *p* = .09). We report the median for the values above because even after log-transforming the metabolite concentrations, they were not distributed normally. Because of this a mean would not have been a good measure of central tendency. Importantly, because we have observed high interindividual variance in estradiol concentrations prior to administration, we have reason to believe the obtained metabolite concentrations were contaminated during the handling of the samples following our data collection. Namely, in previous work such baseline variation was not observed despite an identical procedure and dosage with the main difference being that serum levels of estradiol were measured there [34]. Log-transformed estrone and cortisol concentrations after administration were also examined showing no differences between both groups. Estrone: (experimental: *Mdn* = 8.79 ±4226.69), control: *Mdn* = 5.80 ±161.99) (*W* = 1427, 95% CI [-0.17, 1.05], *p* = .16), cortisol: (experimental: *Mdn* = 0.77 ±0.94), control: Mdn = 0.73 ±1.15) (*W* = 1207, 95% CI [-0.31, 0.27], *p* = .90).

Bodily measures and behavioural characteristics

As outlined in Table S4, both the estradiol and placebo group were also matched for their weight, height, BMI, visceral, abdominal fat, and individual sub scales of the BIS/BAS questionnaire (Drive, Reward, Fun-Seeking, Behavioural Inhibition). Similarly, separate one-way ANOVAs revealed no interaction for the four subscales of BIS-11 (Table S5) (General: $F_{(1, 195)} = 0.01$, $p = 0.91$, Attentional: $F_{(1, 195)} = 0.04$, $p = .85$, Motor: $F_{(1, 195)} = 0.59$, $p = .45$, nonplanning: $F_{(1, 195)} = 0.08$, $p = .78$).

Furthermore, we ensured that both the estradiol and placebo group did not differ in pre-existing differences in working memory (Figure S2A, S2B, S2C) in addition to testing whether administration influenced mood (Figure S2D). By doing so we were able to exclude differences in working memory and mood leading to the observed results [27,48]. Separate ANCOVAs for the three subscales (Alertness, Mood, Calmness) of the

MDBF revealed no differences in post-administration (Post) scores between the estradiol and placebo group when controlling for baseline scores (Pre) as a covariate (Mood: $F_{(1, 96)} = 0.30$, $p = 0.58$, $\Omega^2 = 0.08$; Alertness: $F_{(1, 96)} = 1.35$, $p = .25$, $\Omega^2 = 0.01$; Calmness: $F_{(1, 96)} = 1.34$, $p = .25$, $\Omega^2 = 0.01$). Similarly, we observed no interaction between group membership and post-administration score (Mood: $F_{(1, 96)} = 0.06$, $p = .81$, $\Omega^2 = 0.01$; Alertness: $F_{(1, 96)} = 1.88$, $p = .17$, $\Omega^2 = 0.01$; Calmness: $F_{(1, 96)} = 1.55$, $p = .22$, $\Omega^2 = 0.01$).

Table S4: Descriptive statistics by treatment (Estradiol, Placebo).

	Group				
	Estradiol	**Placebo**	***n***	**statistic [95% CI]**	***p***
Age (years)	25.12 (3.63)	24.6 (3.44)	100	1381 [-0.99, 1.49][1]	0.99
BMI	24.54 (2.65)	24.35 (3.08)	99	1286 [-0.99, 1.99][1]	0.99
Height (cm)	181.90 (6.88)	180.40 (5.95)	99	1.16 [-1.07, 4.07]	0.94
Weight (kg)	81.09 (9.66)	79.48 (11.44)	99	0.76 [-2.61, 5.83]	0.99
Visc. Fat (%)	6.20 (2.48)	6.06 (2.90)	98	1248 [-0.99, 1.00][1]	0.99
Abd. Fat (%)	20.66 (5.97)	19.74 (6.32)	98	0.74 [-1.55, 3.38]	0.99
BIS/BAS					
BIS	17.54 (2.80)	18.40 (3.68)	100	-1.32 [-1.86, -0.25]	0.88
Drive	11.72 (1.77)	12.12 (2.32)	100	1074.5	0.91
Reward	11.62 (1.99)	12.46 (2.04)	100	921[1]	0.18
Fun Seeking	15.88 (2.00)	16.08 (2.17)	100	-0.48 [-1.03, 0.63]	0.99

Note: Values in cells denote M, parentheses denote SD. The superscript 1 denotes the Mann-Whitney-Wilcoxon W value. For the remaining group comparisons, two-tailed independent samples Welch *t*-tests were employed. In cases where *n* is not equal to *N* = 100, data was not recorded for that particular variable. p-values are Bonferroni corrected.

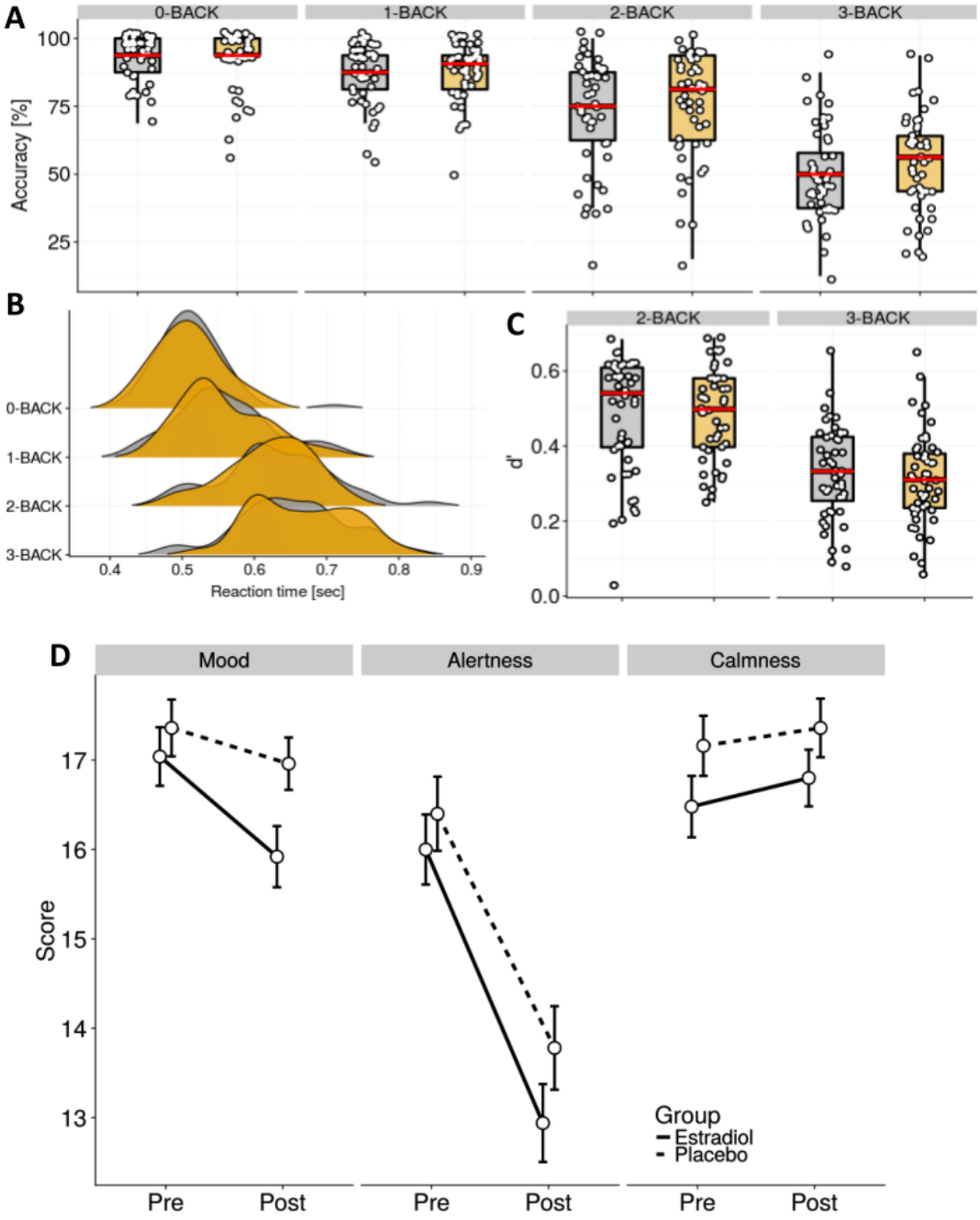


Figure S1. A) Accuracy for individual conditions. The red bar represents the median, the box plot represents the 75% middle most data points, with the whiskers representing 1.95*IQR. Orange depicts the estradiol and gray the placebo group. That is, they represent the division of subjects according to whether they would subsequently be allocated to the estradiol or placebo group. This color convention is used throughout all figures. **B**) shows density plots for reaction time data for individual conditions. **C**) shows d' in the most difficult two conditions (2-BACK, 3-BACK) as there were no false alarms in the 0-BACK and 1-BACK, thus accuracy is reduced to d'. **D**) Average scores prior and post administration for the three subscales of the MDBF.

Furthermore, our working memory (N-BACK) task revealed a comparable picture for accuracy (Figure S2A), reaction times (Figure S2B), and d-prime (Figure S2C). That is, there was no statistically significant difference between the estradiol and placebo group in accuracy, average reaction times, and d-prime. We did observe an expected drop in performance in terms of decreased accuracy (0-BACK: 92.94 ±9.34, 1-BACK: 88.06 ±10.78, 2-BACK: 74.25 ±19.38, 3-BACK: 51.56 ±17.37), and d-prime (2-BACK: 0.48 ±0.14, 3-BACK: 0.32 ±0.12), and increased reaction times (0-BACK: 0.51 ±0.05, 1-BACK: 0.56 ±0.06, 2-BACK: 0.63 ±0.07, 3-BACK: 0.66 ±0.07) as the condition became more difficult (i.e. went from 0-BACK to 3-BACK). Separate linear models were used to compute to check for main effects of drug ($F_{(1, 196)}$ = 2.01, p = .16, Ω^2 = 0.00) and an interactive effect of drug and condition on d-prime ($F_{(1, 196)}$ = 0.82, p = .37, Ω^2 = 0.00). As mentioned above, we also did this for accuracy (main effect of drug: $F_{(1, 392)}$ = 1.07, p = .30, Ω^2 = 0.00; drug*condition interaction: $F_{(3, 392)}$ = 2.30, p = .08, Ω^2 = 0.00), and reaction times (main effect: $F_{(1, 347)}$ = 1.31, p = .25, Ω^2 = 0.00; drug*condition interaction: $F_{(1, 347)}$ = 0.99, p = .39, Ω^2 = 0.00).

Table S5. Descriptive statistics of MDBF and BIS-11 subscales.

		Pre-administration		Post-administration	
	N	**Estradiol**	**Placebo**	**Estradiol**	**Placebo**
MDBF					
Mood	100	17.04 (2.31)	17.36 (2.25)	15.92 (2.41)	16.96 (2.07)
Alertness	100	16.00 (2.77)	16.40 (2.93)	12.94 (3.08)	13.78 (3.30)
Calmness	100	16.48 (2.43)	17.16 (2.38)	16.80 (2.24)	17.36 (2.32)
BIS-11					
General	100	57.86 (7.68)	59.28 (7.45)	59.96 (8.51)	60.46 (7.86)
Motor	100	20.76 (3.00)	21.78 (3.42)	22.02 (3.42)	22.42 (3.65)
Attention	100	14.34 (3.37)	14.14 (2.17)	15.10 (3.30)	14.64 (2.68)
Nonplanning	100	22.76 (3.61)	23.36 (4.29)	22.84 (4.05)	23.40 (4.21)

Note: values in cells denote *M*, values in parentheses denote *SD*.

In summary, both groups were matched on working memory and post-administration mood scores. They were additionally matched for age, height, visceral and abdominal fat, BMI, BIS-BAS, and impulsivity (BIS-11). The estradiol group had

higher estradiol concentrations after but not before administration compared to the placebo group. Importantly, there was no correlation between subjects' belief about whether they had received estradiol or placebo and actually receiving estradiol ($r = 0.02$, $p = .82$), the certainty of that belief and actually receiving estradiol ($r = 0.02$, $p = .82$), or between the reported observed changes and actually receiving estradiol ($r = -0.08$, $p = .42$). This shows that our double-blind procedure worked and that our placebo gel preparation was indistinguishable from the actual drug. Overall, the described results show that our administration procedure was successful and both groups were matched on key traits that could have potentially impacted the observed behaviour. This allowed us to constrain the number of possible alternative explanations of our main results.

Table S6. Frequencies of individual polymorphisms of DAT and COMT genes.

Polymorphism	Group	N
9/10	Estradiol	18
9/10	Placebo	16
10/10	Estradiol	21
10/10	Placebo	26
Val/Val	Estradiol	11
Val/Val	Placebo	9
Met/Val	Estradiol	23
Met/Val	Placebo	26
Met/Met	Estradiol	12
Met/Met	Placebo	15

Note: the split according to both COMT and DAT does not sum to 100 because for a few subjects it was not possible to determine their polymorphism.

Reinforcement learning task

Selecting linear models

For all general linear models assessing interactions described in our results, we started with the simplest model which included our interaction of interest (either drug*COMT or drug*DAT) and regressed out the belief of having received the drug. We considered

this belief as a nuisance regressor because of our previous work showing the impact of beliefs about a hormone on subsequent behaviour [59]. Additional nuisance regressors included bodily measures known to impact estradiol metabolism which we collected: weight, BMI, abdominal and visceral fat [60,61] and post-administration cortisol levels [58]. All linear models were compared with BIC and AIC. Unless stated otherwise in the main text, for all reported results the winning model regressed out cortisol levels following administration, beliefs about having received the drug, the certainty of that belief and whether they had observed any changes in themselves, a composite score of weight and BMI (main text), visceral, and abdominal fat. For general linear models involving accuracy, we also regressed out reaction times to control for accuracy-speed trade-offs. All nuisance regressors were z-scored.

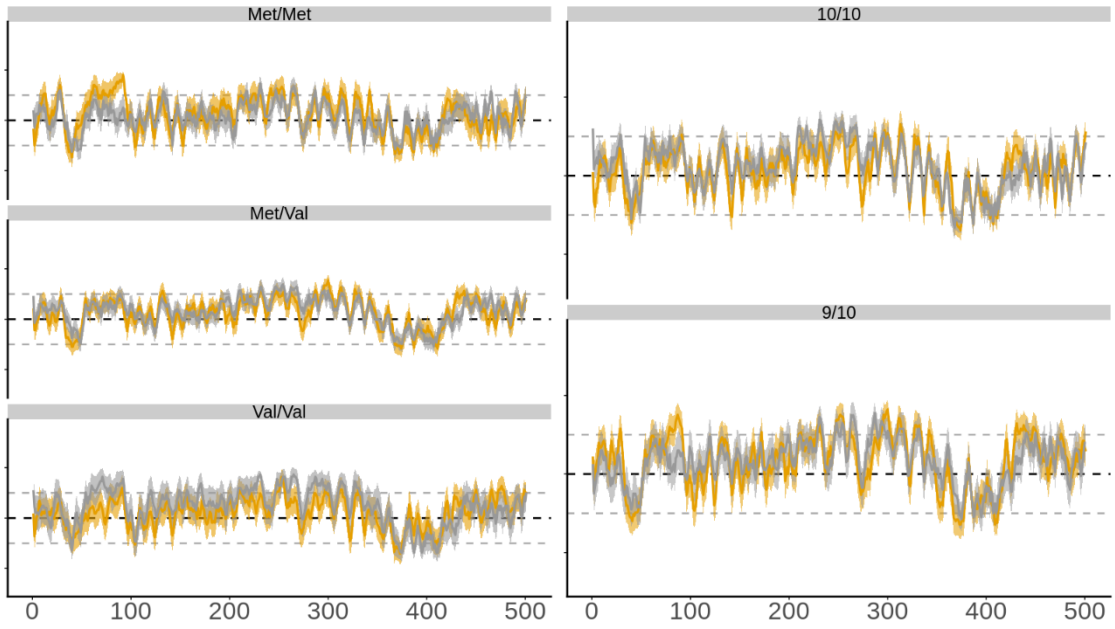


Figure S2. Relative choice probability for choosing option A (top of y-axis) vs. choosing stimulus 1 (bottom of y-axis) for the placebo (gray) and estradiol (orange) group split according to both polymorphisms assessed in the main text: COMT (left panel), DAT (right panel) across trials (1-500). Thick lines represent trial means, shaded areas denote standard error of the means. The blue line in the background denotes the empirical relative reward probability which was computed from the probability of stimulus two being rewarding (top of y-axis) - stimulus one being rewarding (bottom of y-axis). Gray dotted lines represent where participants were on average 25% more likely to select option A (upper line) or stimulus 1 (lower line). All time-series traces are smoothed with a 5-trial moving average for visual purposes.

Figure S2 reveals a differential effect of estradiol administration on choice behaviour that depends on polymorphisms of both COMT and DAT. In the case of the COMT polymorphism this is most clearly visible in the lower left panel. The panel shows that placebo Val/Val subjects exhibited a clear tendency towards stimulus two until trial ~370. After this, they did not reverse back towards choosing it more often despite stimulus two being more rewarding from trial ~420 onwards. This is in contrast with results for subjects with other polymorphisms of COMT and results when subjects were split according to the DAT1 polymorphism. Estradiol Met/Met subjects exhibited choice behaviour more aligned with the reward probability distribution in the beginning at trial ~80 compared to subjects from the placebo group with the same polymorphism. When we then split subjects according DAT1 polymorphism, the estradiol 9/10 subjects can similarly be seen following the reward probability distribution more closely compared to the placebo 9/10.

Model prediction for switching behaviour

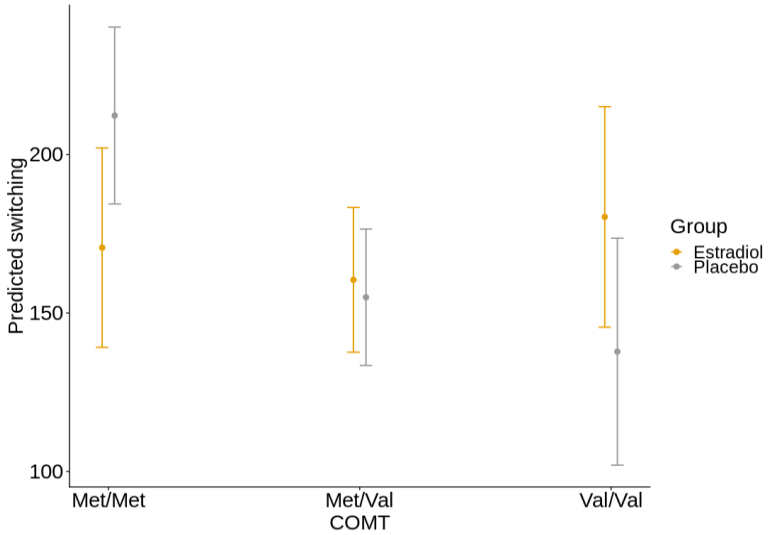


Figure S3. General linear model prediction for switching behaviour (i.e. a change in chosen stimulus on trial *t + 1* from trial *t*, independent of choice outcome on trial *t*). Estradiol administration dampened naturally occurring differences in switching behaviour when subjects were split according to the COMT polymorphism, i.e. whether subjects would switch the stimulus they chose on trial t compared to trial *t + 1* irrespective of choice outcome on trial *t*. Figure S3 shows that our linear model made comparable predictions about this switching behaviour for all three polymorphisms in the estradiol group. In contrast, in the placebo group it predicted a clear linear decrease in switching from the Met/Met genotype (i.e. high prefrontal dopamine) towards the Val/Val genotype (i.e. low prefrontal dopamine).

The role of CYP 19A1, ERα, ERβ, CAG, and GGN

Because the results we report in the main text and the supplementary materials have other mechanistic explanations and/or could have been moderated through other candidate mechanisms, we further analyzed these mechanisms together by providing theoretical motivation for these analyses. We analyzed the candidate mechanisms for both accuracy and reported switching behaviour. Here, we first briefly outline their importance and then summarize the observed results.

It is known that androgens are converted to estrogen [74]. This means that the increase in estrogen levels arises from the conversion process and the administration more directly. Furthermore, variation in the length of two functional polymorphisms (CAG – polyglutamine, and GGN – polyglycine) are known to modulate the functioning of the androgen receptor gene [75]. This is important for two reasons. The first is that our procedure has previously shown to increase circulating testosterone levels which could have raised estradiol levels whilst being moderated by subjects' androgen receptor characteristics [34]. Following from this, previous work has shown that brain regions important for memory and learning contain androgen receptors [76]. Therefore, it could be possible that interindividual differences in both functional polymorphisms could have moderated our observed results due to interindividual variability. For example, greater CAG repeat length has previously been associated with lower scores in different cognitive tests in older men [75]. Similarly, there has been an association between GGN repeats and immediate and delayed logical memory recall as a function of GGN repeat length found in women [77]. Furthermore, longer repeats of both the CAG and GGN polymorphism have been previously associated with different disorders including attentional deficit and hyperactivity disorder, conduct disorder, and oppositional defiant disorder [78]. All described results show a correlation between interindividual variability in androgen receptor functioning and cognitive performance, giving rise to the CAG and GGN polymorphisms being potential candidate mechanisms moderating the observed effect of estradiol on accuracy and switching behaviour. Repeat polymorphism of two

most studied functional polymorphisms in the androgen receptor gene - CAG and GGN - were therefore examined.

Throughout the conversion process from androgens to estrogens, the CYP19A1 gene encodes instructions for aromatase – the enzyme converting androgens to estrogens [79]. The single nucleotide polymorphisms (SNPs) associated with the CYP19A1 gene regulate the metabolism of androgens and mediate brain estrogen activity. Two specific SNPs (rS700518, rs936306) have been previously shown to have a role in cognitive functioning in humans. For example, men with the homozygous AA allele have been shown to have higher estradiol serum levels and greater bilateral posterior hippocampal gray matter volume compared to those homozygous with the GG allele [80]. While other work has shown a differential impact of homozygous CC alleles versus homozygous TT alleles on episodic memory recall in women [81]. Given that our procedure has previously shown to increase circulating testosterone levels and that polymorphisms of the CYP19A1 gene are known to have a role in cognitive functioning, we aimed to exclude the possibility of that driving our observed effects and analyzed both single nucleotide polymorphisms of the CYP19A1 gene.

Once androgens are converted to estrogens, estrogen action is mediated through the known estrogen receptors (ERα, ERβ). Both receptors are widely distributed throughout the brain in regions important for cognitive functioning. So far, it has been shown that ERα is responsible for most of estrogen-related activation. For example, it has been shown that SNPs of ERα are related to Alzheimer's disease and are associated with the likelihood of developing cognitive impairment [82]. We have, therefore, focussed on two particular SNPs of ERα: rs9340799, rs2234693. In contrast, little is known of a potential impact of ERβ. As an exploratory measure, we have included repeats of this receptor in our analysis as well.

Of the described candidates (CAG, GGN, CYP 19A1, ERα, ERβ), no test revealed any effect of interest. There was no interaction between group membership (i.e. estradiol or placebo) and either the SNPs of ERα: rs9340799 ($F_{(2, 84)}$ = 0.66, p =

.52), rs2234693 ($F_{(2, 84)}$ = 0.63, p = .53) in relation to accuracy. Furthermore, the same was true for the interaction between CAG repeats and group membership ($F_{(1, 87)}$ = 0.45, p = .51), GGN repeats and group membership ($F_{(1, 87)}$ = 1.31, p = .26), and SNPs of the CYP19A1 gene and group membership (rs700518 $F_{(2, 84)}$ = 1.84, p = .15, rs936306 $F_{(2, 84)}$ = 0.34, p = .72). In a final examination, we also looked at the repeats of ERβ to determine whether this could have driven any of the observed effects. However, this was not the case for either recorded variant of ERβ (ERβ1: $F_{(1, 87)}$ = 0.02, p = .89, ERβ2: $F_{(1, 87)}$ = 0.00, p = .96).

Identical results were obtained for switching behaviour. While we observed a statistically significant interaction between estradiol administration and the COMT polymorphism, this was not true for any of the other mechanistic explanations. That is, no model showed an interaction between group membership and either of the SNPs of ERα: rs9340799 ($F_{(2, 84)}$ = 2.90, p = .06), rs2234693 ($F_{(2, 84)}$ = 2.88, p = .06), CAG repeats ($F_{(1, 87)}$ = 0.10, p = .76), GGN repeats $F_{(1, 87)}$ = 1.32, p = .25), and SNPs of the CYP19A1 gene (rs700518 $F_{(2, 84)}$ = 1.81, p = .17, rs936306 $F_{(2, 84)}$ = 1.08, p = .35) in relation to switching behaviour. As in the case of accuracy, we also looked at the repeats of ERβ. Again, there was no statistically significant contribution to switching behaviour from this predictor for either recorded variant of ERβ (ERβ1: $F_{(1, 87)}$ = 3.05, p = .08; ERβ2: $F_{(1, 87)}$ = 0.96, p = .33).

We finally repeated the set of analyses for staying behaviour with no effects found. SNPs of ERα: rs9340799 ($F_{(2, 84)}$ = 1.69, p = .19), rs2234693 ($F_{(2, 84)}$ = 1.79, p = .17), CAG repeats ($F_{(1, 87)}$ = 0.38, p = .54), GGN repeats $F_{(1, 87)}$ = 0.30, p = .59), SNPs of the CYP19A1 gene (rs700518 $F_{(2, 84)}$ = 1.27, p = .29, rs936306 $F_{(2, 84)}$ = 0.59, p = .55), and variant of ERβ (ERβ1: $F_{(1, 87)}$ = 1.35, p = .25; ERβ2: $F_{(1, 87)}$ = 0.86, p = .36).

In brief, we have shown that the effects did not depend on overall androgen receptor functioning assessed by investigating the repeat length of two different functional polymorphisms (CAG and GGN). Both polymorphisms were investigated due

to the known conversion process of androgens to estrogen which could have moderated these results [74,80]. We excluded that interindividual variability in the conversion process itself would predict the observed effects, by investigating two polymorphisms of the CYP19A1 gene which plays a key role in converting androgens to estrogens [80,81]. Finally, we excluded the possibility that following the conversion process, the observed effects were a consequence of polymorphisms (ERα) or repeats (ERβ) of known estrogen receptors, given that both are widely distributed throughout the brain, especially in regions of importance for reward processing [83]. All of the described candidates revealed no effect for either accuracy or switching behaviour that are reported above.

Impact of previous choice on current choice

Since we observed a difference in group choice behaviour in Figure S2 in the main results, and that the estradiol and placebo group systematically chose differently on 7.5% of the trials, we ran separate logistic regressions to compute whether this would also be observed in how past choices would affect the current choice. We predicted there would be a difference between the estradiol and placebo group in pure choice autocorrelation (i.e. if I choose option A on trial *t*, is it more likely I will choose it again on trial *t* + *1*) and reward-related autocorrelation (i.e. if I choose option A on trial *t* and it is rewarded, is it more likely I will choose it again on trial *t* + *1*). We further predicted that splitting these two groups according to the DAT1 and COMT polymorphism would show differences depending on the polymorphism.

Information about subjects' choices *n* trials ago was varied from 1 trial to 7 trials ago and used as a regressor to predict current choice. Therefore, in the design matrix we had information about their choice from 7 trials to 1 trial ago. The value 1 meant they repeated their choice, while 0 meant they did not. We first split participants according to the estradiol and placebo group (Figure S4).

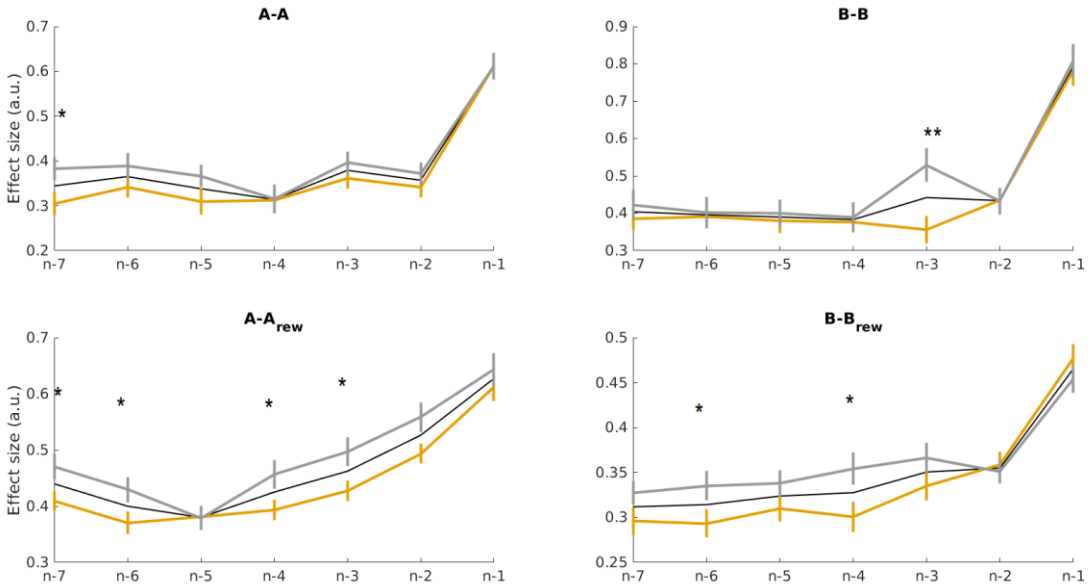


Figure S4. The top panels show pure choice autocorrelation: Choosing A if A was chosen previously or choosing B if B was chosen previously. Bottom panels show reward-related choice autocorrelation: Choosing A if A was previously rewarded or choosing B if B was previously rewarded. The lines show the averaged beta coefficient from the regression. Error bars are standard errors of the mean. Orange line depicts the estradiol group, gray lines depict the placebo group. * $p < .05$, ** $p < .01$.

Contrary to our prediction, the top panel in Figure S4 does not reveal a systematic difference in choice autocorrelation between the estradiol and placebo group. One notable exception is the contribution of the choices made three trials ago where the placebo group was more likely to consider those choices compared to the estradiol group ($p < .01$). However, the bottom panel reveals that the estradiol group had lower reward-related autocorrelation for both options. That is, if they were rewarded for a choice several trials ago, they were less likely to persevere with that choice compared to the placebo group. This is consistent with Figure 2A where the estradiol group followed the reward probability distribution better compared to the placebo group. Figure S4 reveals why that may have been the case; they were less likely to persevere due to information received several trials ago, but not the one that just occurred t – 1 trials ago.

We then further split the same participants according to the COMT (Figure S5) DAT (Figure S6) polymorphisms. We see that the autocorrelation difference for choosing option B three trials ago reported in Figure S4 was driven by the group with the Val/Met genotype specifically. In contrast, the difference between the estradiol and placebo group in terms of reward-related choice autocorrelation was driven by the placebo group with the Val/Val genotype (i.e. low prefrontal dopamine), as seen in the third column. Only in the Val/Val comparison was there a systematic difference between the estradiol and placebo subgroup. This difference disappeared in the other COMT polymorphisms and was also only true for option A. Conversely, in column four a difference between the estradiol and placebo group only became observable in subjects with the Met/Met genotype (i.e. high prefrontal dopamine).

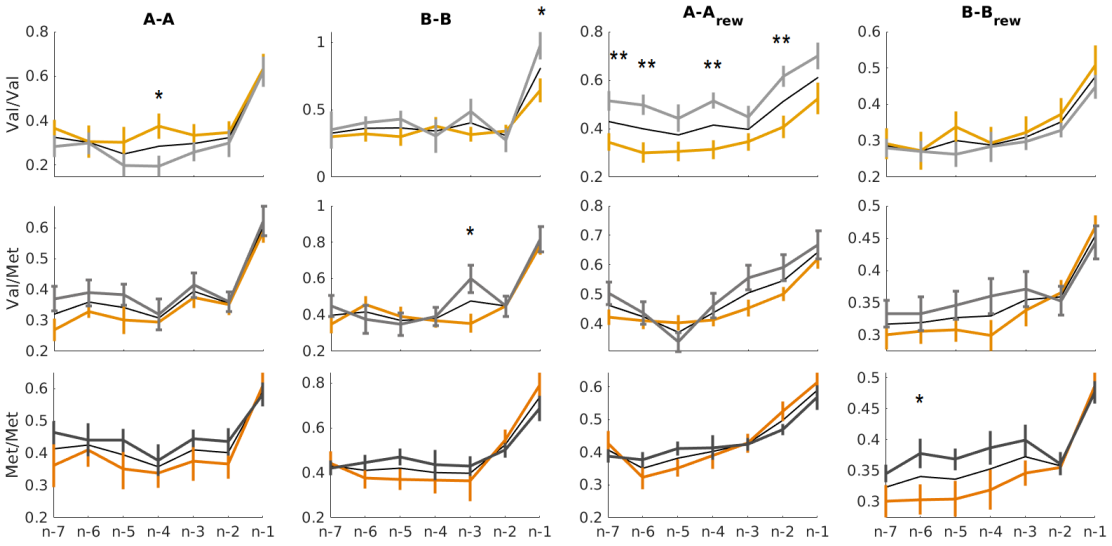


Figure S5. Individual columns show the same as individual panels in Figure S4. Here, they are additionally split according to the COMT polymorphism. The lines show the averaged beta coefficient from the regression. Error bars are standard errors of the mean. Orange line depicts the estradiol group, gray lines depict the placebo group. *$p < .05$, **$p < .01$.

The final split was according to the DAT1 polymorphism. This did not reveal clearly interpretable systematic differences apart from the autocorrelation difference for option B between the estradiol and placebo group being driven by subjects with the 10/10 genotype (i.e. low striatal dopamine) as opposed to subjects with the 9/10

genotype. Similarly, estradiol 10/10 genotype subjects also exhibited lower reward-related autocorrelation compared to the placebo 10/10 genotype subjects. However, this was also present in the 9/10 subjects for both stimuli, indicative of them being more likely to stick with identical choices after being rewarded.

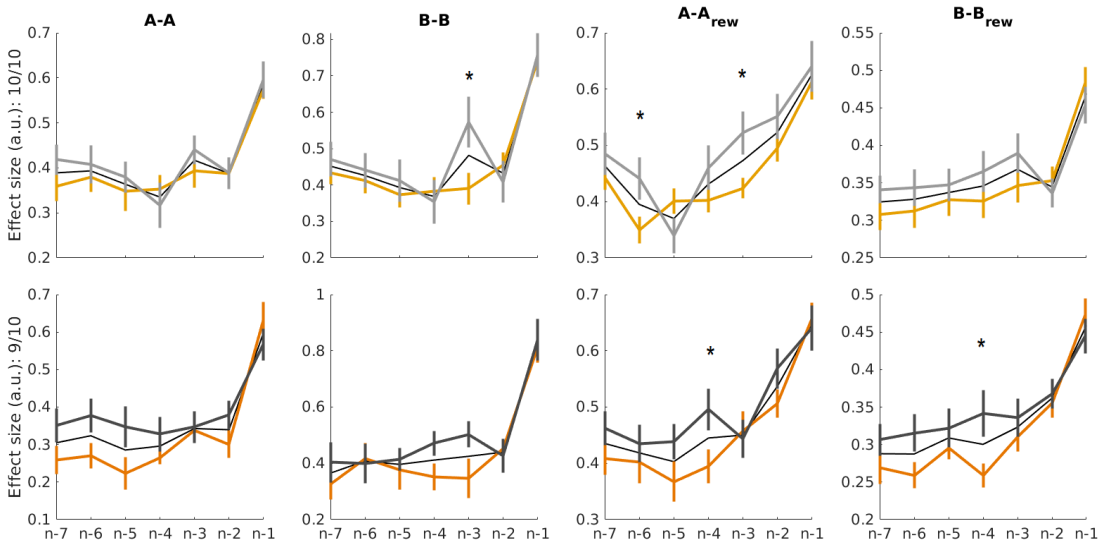


Figure S6. Individual columns show the same as individual panels in Figure S4. Here, they are split according to the VNTR polymorphism of the DAT gene. * $p < .05$, ** $p < .01$.

Generalized linear mixed effects model predictions for choice

Figure S7 reveals strong interactive effects for both the DAT polymorphism with drug over time on choice (A) and the COMT polymorphism (B) with the same model structure. We did not include models that would combine both genotypes as they would have given rise to an insufficient size per smallest cell (Table S6).

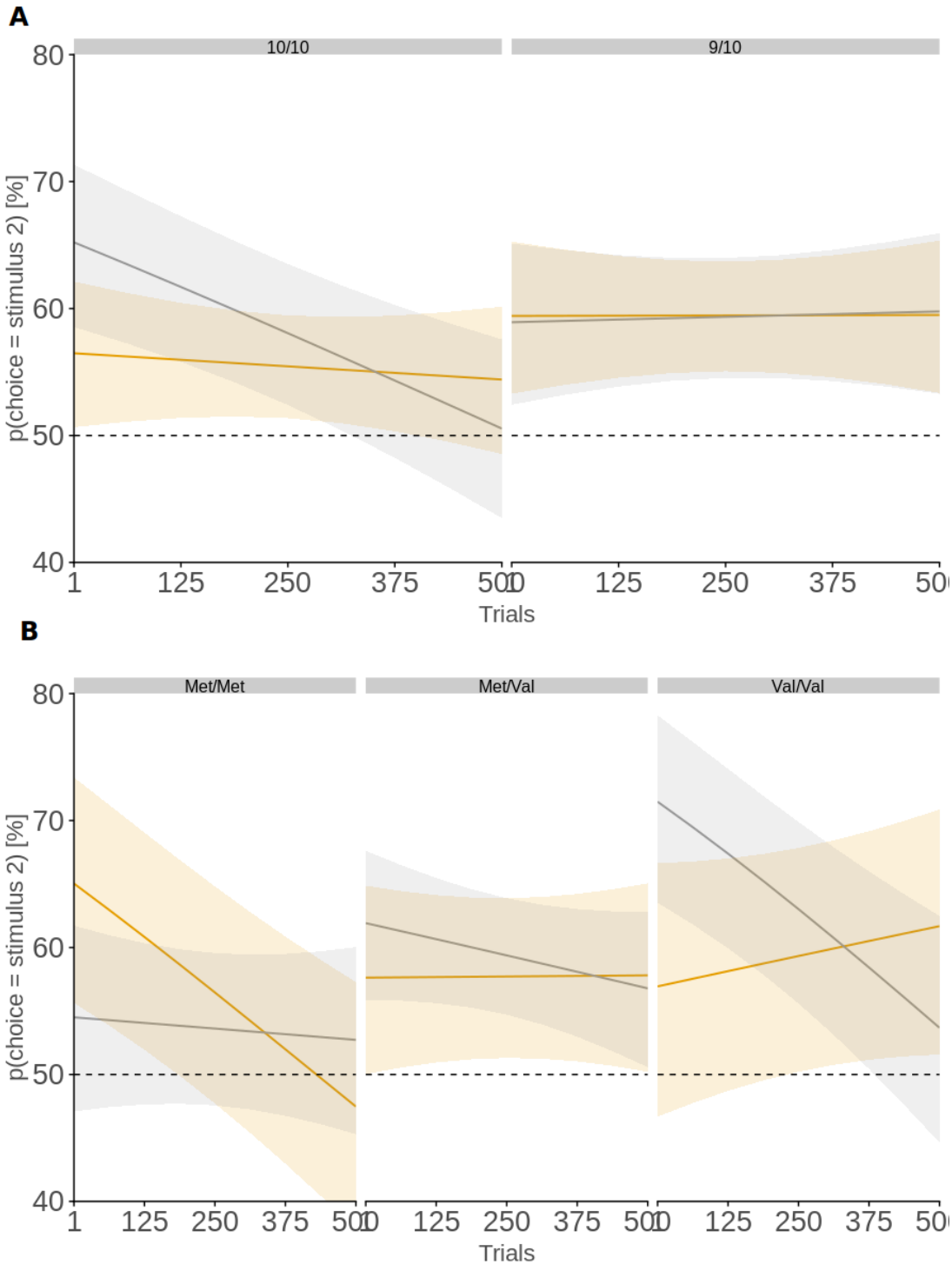


Figure S7. Predictions from winning models of the generalized linear mixed effects models for **A**) the interaction between drug, DAT, and trial on choice, and **B**) the interaction between drug, COMT, and trial on choice.

Formal model comparison

In addition to computing the leave-one-out information criterion to perform model comparison [71] we similarly computed the exceedance probability of the winning model using the VBA toolbox [73]. This value showed a strong preference for the winning model

P(model two) = 98%. Furthermore, we computed protected exceedance probability [72] as an extension which, while yielding an expected decrease in the winning model probability, still favoured model two over other competing models (*P*(model two) = 12.5%). The likely decrease was due to the reinforcement learning task not being optimized to detect behavioural differences between the models tested. However, in all reported models, the latent variable of interest, i.e. the learning rate, remained unaltered. We would therefore expect the increase in learning rates to be present if we were to select the learning rates from models that best fit individual subjects.

Validating model

We further tested the model validity and predictions by computing posterior predictive densities, i.e. what predictions does the model make on a trial by trial basis for subjects with the parameters such as those that were extracted from our participants. Posterior predictive densities showed no difference in a fit between both the estradiol and placebo group and approximated the empirical reward probability distribution (Figure S8A). To quantify this, we then compared model predictions from posterior predictive densities with actual participant behaviour to assess model accuracy collapsed across time (Figure 4B) showing it performed above chance and equally well for both groups. We further compared accuracy on each trial across participants to ensure that there were no unexpected drops in accuracy. This did not happen as the model (Figure S8C) had no discernible drops in performance.

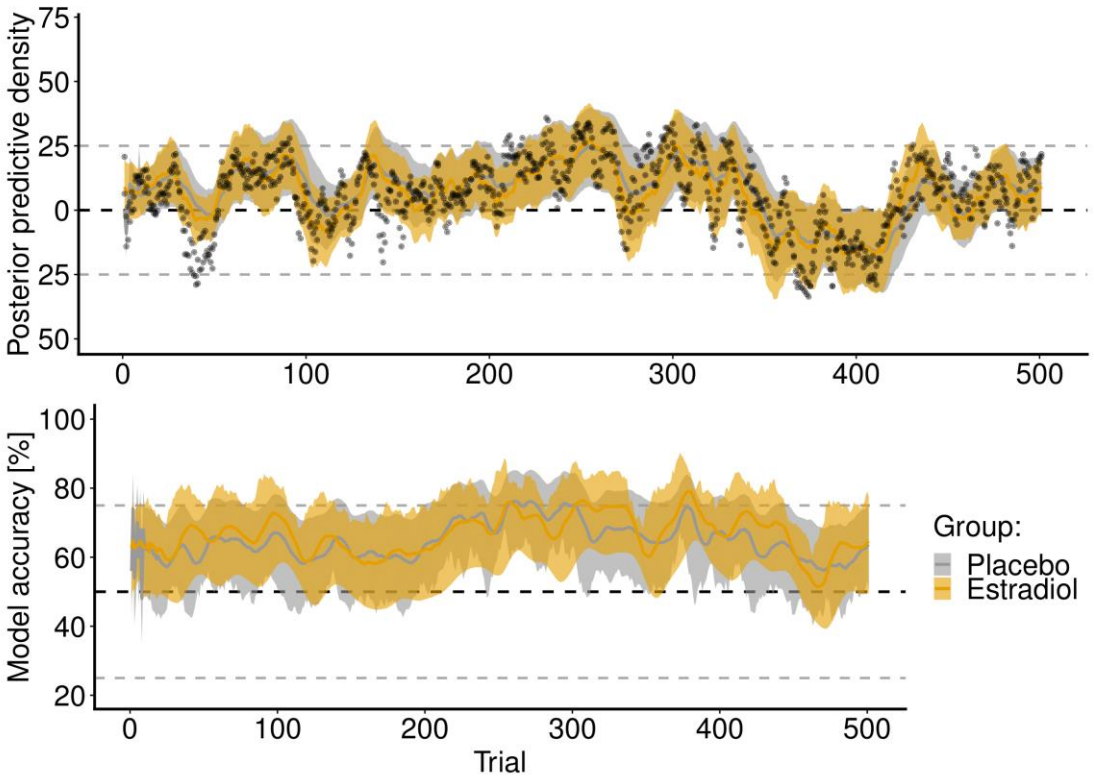


Figure S8. A) Posterior predictive density computed for both groups with overlaid average responses for both groups across trials **B**) Accuracy for both groups obtained from the posterior predictive density for both groups separately.