1   **A causal role for estradiol in human reinforcement learning**

2

3   **Authors**

4   Sebastijan Veselic[1,2,3]*, Gerhard Jocham[4], Christian Gausterer[5], Bernhard Wagner[6], Miriam

5   Ernhoefer-Reßler[6], Rupert Lanzenberger[7], Claus Lamm[1,8], Christoph Eisenegger[1] & Annabel

6   Losecaat Vermeer[1]*

7

8   **Affiliations**

9   [1] Neuropsychopharmacology and Biopsychology Unit, Department of Cognition, Emotion,

10   and Methods in Psychology, Faculty of Psychology, University of Vienna, Austria

11   [2] Department of Clinical and Movement Neurosciences, University College London, London,

12   UK

13   [3] Wellcome Centre for Human Neuroimaging, University College London, London, UK

14   [4] Biological Psychology of Decision Making, Institute of Experimental Psychology, Heinrich

15   Heine University Düsseldorf, Germany

16   [5] FDZ-Forensisches DNA Zentrallabor GmbH, Medical University of Vienna, Austria

17   [6] Laboratory for Chromatographic & Spectrometric Analysis, FH JOANNEUM, Graz, Austria

18   [7] Department of Psychiatry and Psychotherapy, Medical University of Vienna, Vienna,

19   Austria,

20   [8] Vienna Cognitive Science Hub, University of Vienna, Austria

21

22

23   (*) corresponding authors: Sebastijan Veselic (*sebastijan.veselic.18@ucl.ac.uk)* and
24   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, while human work has suggested that high estradiol levels are associated with altered 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 the striatal dopaminergic pathway.

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

Introduction

Learning which action 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 estrogen [2].

Estrogens are a class of steroid hormones important for healthy development, with estradiol being the most prevalent and potent form [3,4]. Estradiol has gained traction as a compound that may impact human reward processing by amplifying dopamine signalling via the D1 receptor [2]. The evidence for this hypothesis comes from two lines of work. From human work, we know that fluctuations in circulating estradiol levels are correlated with differences in midbrain BOLD responses, a key area where dopamine is released [5–7]. Furthermore, a recent administration study in women showed that increased salivary estradiol levels led to larger reward prediction errors in the nucleus accumbens [8]. From rodent work, we know that manipulating estradiol levels affects the striatal dopamine system in various ways that are best characterised as a net increase in overall dopamine signalling predominantly via the D1 receptor [9–14].

If increased estradiol levels in humans influence reward processing by increasing reward sensitivity through amplified dopamine signalling via the D1 receptor, then it should be possible to relate its effects to that of dopamine in RL [1,15–17]. Within this field, dopamine agonists and antagonists are often administered to better understand the mechanistic role of dopamine in RL [18–24]. Complementing it with pharmacogenetics, where the effect of genetic variation on the administered drug is considered, has refined our understanding of their relationship [19,22,25–27]. That is, causally manipulating dopamine levels in humans affects performance in RL tasks[27] with these effects depending on individual differences in baseline dopamine levels [18].

Two key polymorphisms that lead to variation in baseline dopamine levels by impacting dopamine synthesis capacity and transmission are the dopamine transporter (DAT1) and

2

69   catechol-O-methyltransferase (COMT) gene [19,24,25]. Variation in the VNTR polymorphism of

70   DAT1 and the val[158]met polymorphism of COMT correlates with performance on working

71   memory and RL in humans [19,24,25,28,29]. Therefore, if estradiol influences reward processing

72   through dopaminergic mechanisms, then variation in the DAT1 and COMT genotype should

73   moderate the effect of estradiol. Furthermore, we would predict it should be moderated by

74   traits related to human reward sensitivity such as the ones measured by the BIS/BAS

75   questionnaire [30–33].

76   Despite abundant evidence from rodent research showing a clear relationship between

77   estradiol and dopamine, evidence in humans have been less conclusive. Namely, it has been

78   shown that high endogenous estradiol levels were associated with increased [5,7] as well as

79   decreased [34] performance on a variety of cognitive tasks that may have recruited different

80   neural mechanisms. Thus, it remains unclear whether administering estradiol influences

81   reward processing as would be expected by amplified D1-receptor signalling. Furthermore,

82   although previous work on humans provided important insights, it had certain shortcomings

83   such as small sample sizes, correlational study designs, and a lack of accounting for baseline

84   differences in dopamine; for exceptions see [8,34,35] [25,36]. All studies thus far have also focussed

85   on female samples who have, on average, higher endogenous levels of estradiol compared

86   to men [37]. These aspects are important for being able to establish a more precise role of

87   estradiol in human reward processing which, therefore, remains an open question (for review

88   see [2]).

89   The aim of the present pharmacogenetic study was to address these gaps and

90   investigate whether estradiol administration increases reward sensitivity in men and whether

91   this effect is moderated by baseline differences in dopamine. To test this, we used a

92   probabilistic RL task (Fig. 1A) where subjects had to choose between two options on each trial

93   in order to maximize their earnings. Moreover, we aimed at providing a more conclusive and

94   precise account of a dopamine-dependent basis of action through excluding several other

95   candidate explanations (i.e. gene polymorphisms) that could have given rise to the obtained

96   results and have so far been unaddressed in the literature (see Supplementary Materials).

3

97   We hypothesized that estradiol administration would influence reward processing by

98   increasing reward sensitivity observed through subjects' choice behaviour [8]. We further

99   predicted this would be revealed through computational modelling. Finally, we predicted that

100   the behavioural and computational effects would be moderated by the DAT1 polymorphism,

101   as observed in previous dopamine administration work [19,26].

102

Methods and materials

Subjects

105   We tested one hundred healthy young men between 19 and 34 years (M_{age} = 24.86, SD =

106   3.53) with a body mass index (BMI) between 19.3 and 31.5 (M = 24.45, SD = 2.86). Our

107   screening procedure was based on previous work where pharmacokinetic data for a single 2

108   mg estradiol dose in topical form was obtained [38]. Our sample size was in line with previous

109   recommendations for the field [2]. The study procedure was performed in accordance with the

110   Declaration of Helsinki and approved by the Medical Ethics Committee of the University of

111   Vienna (1918/2015).

Measurement Instruments

Questionnaires

114   We assessed self-reported mood (German Multidimensial Mood State Questionnaire;

115   MDBF[39]), individuals' impulsiveness (Barratt Impulsiveness Scale; BIS-11[40]), and behavioural

116   activation and inhibition (BIS/BAS[41]). Both BIS/BAS and BIS-11 scores have been previously

117   found to correlate with reward learning [42–45] with the BIS/BAS specifically being related to

118   reward sensitivity [30,31]. In addition, we probed subjects' beliefs about estradiol (e.g. whether

119   they believed they received estradiol or a placebo, how certain they were about their belief,

120   and whether they had noticed any subjective changes).

121

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).

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.

Experimental Tasks

Reinforcement Learning: We used a well-established probabilistic reinforcement learning task with two options [17]. The task consisted of 500 trials, with a 10 second pause after the first 250 trials. There was no choice time-out. The two options had independently varying reward probabilities generated by Gaussian random walks (Fig. 1). Importantly, both options could be correct (i.e. rewarding – yellow frame) or incorrect (i.e. non-rewarding – reward frame) on any trial. Subjects received feedback for the chosen (thick frame) and unchosen (thin frame) option. Each correct choice was rewarded with 5 eurocents and added to their cumulative balance. Subjects also saw a bar fill up as they chose the correct option. Once the bar filled up, a 1 € coin was presented next to the bar indicating they had added 1 € to their cumulative balance.

Working memory capacity: We used an adapted version of the standard N-BACK task [25] with four conditions in total (0-BACK, 1-BACK, 2- BACK, 3-BACK). Each condition block had 20 trials which included 20% target, 65% nontarget, and 15% lure trials. Subjects were

5

148   presented with a sequence of letters one-by-one. For each letter, they had to decide if the current

149   letter was the same as the one presented *N* trials ago by pressing "R", in case it was not the

150   same they had to press "O".

Procedure
151

152   Subjects screened for BMI, a history of psychiatric disorders, concurrent involvement in other

153   psychopharmacological studies, and chronic physical injuries, arrived to the lab on two

154   separate days. On the first session, at 4.00 pm, we collected subjects' responses on a battery

155   of questionnaires, their DNA with a buccal swab for genotyping, and measured their BMI and

156   body fat using a body composition monitor (Omron BF51).

157   On the second session (Fig. 1, timeline) after obtaining consent, subjects filled out the

158   MDBF-A, BIS-11, and BIS/BAS. Twenty minutes after arrival, we obtained the first saliva

159   sample (T1) to assess baseline hormone concentrations, followed by the N-BACK task. They

160   were randomly assigned estradiol or placebo in a double-blind manner and self-administered

161   a topical transparent gel containing either 2 mg of estradiol or a placebo. We waited two hours

162   to allow estradiol levels to peak based on a previously established procedure [38]. Fifteen

163   minutes prior to the behavioural testing, subjects filled out MDBF-B, BIS-11, and provided a

164   second saliva sample (T2).

165   The first behavioural task was the probabilistic reinforcement learning task, followed

166   by three other tasks that were not the focus of this publication. After the behavioural testing,

167   we probed subjects' beliefs about the treatment and the tasks. At the end of the study, each

168   subject was paid in accordance to their performance.

169
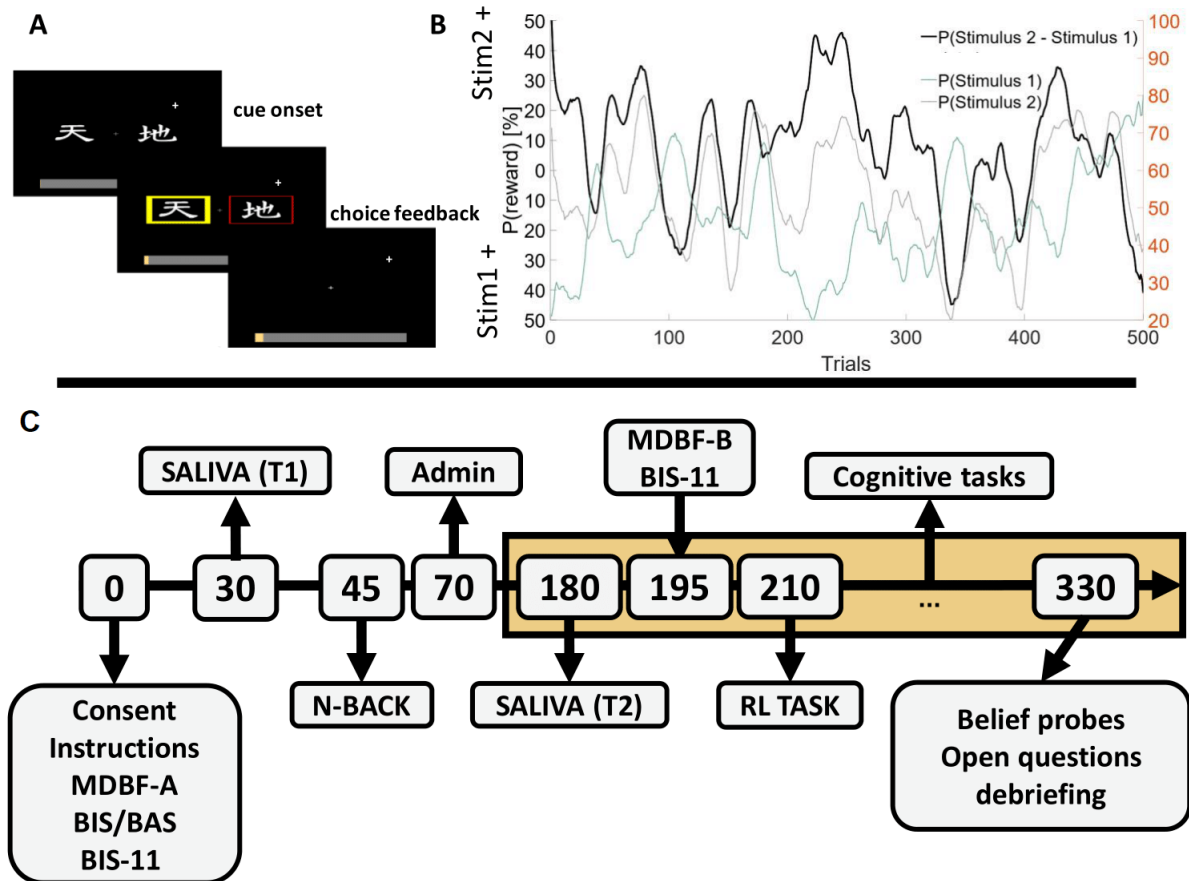
Fig. 1 A) Outline of a trial of the RL task. Each trial started by the presentation of two options (henceforth option A and option B). Subjects were required to choose one of these options. After they made a choice, subjects were presented with feedback, with the chosen option indicated by a thicker 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.

Statistical analysis

Behavioural analysis

To determine the effect of estradiol on choice behaviour, we examined the cumulative difference in the options subjects chose and the percentage of trials on which there was a significant difference in the chosen option. As a measure of family-wise error control, we used permutation testing and Fisher-z-transformations.

We also computed choice accuracy because we predicted that if estradiol would influence reward processing by increasing reward sensitivity, this would result in increased accuracy and be moderated by differences in striatal dopamine functioning (i.e. through the DAT1 genotype) [26]. Due to the same reasoning we computed how many trials subjects would pick the same option if they were rewarded for that option on trial t (staying). Accuracy and staying were statistically evaluated with general linear models.

In addition to the above metrics, we computed subjects' choice autocorrelation. With this, we tested our hypothesis that increased reward prediction errors would, by definition, upregulate saliency of recent events [8]. Behaviourally, this would mean subjects' recent choices should influence what they will choose next more. We computed the relative contribution of choices made from $n - 1$ to $n - 7$ trials back (lags) on the current choice. We then compared the relative contribution of all previous choices (pure choice autocorrelation) and specifically choices when they were rewarded (choice autocorrelation as a function of reward). We then performed independent samples t-tests on individual lags that were z-scored within subject to assess statistical significance.

Computational modelling

To account for behaviour within a computational framework and relate our findings to the field exploring dopamine and its role in RL, we used computational modelling. We fit a series of Q-learning models with softmax choice rules. The winning model included a learning rate for positive and negative prediction errors, a temperature parameter, and an irreducible noise

221   parameter (see Supplementary materials for model selection and parameter recovery). Our

222   main hypothesis was that if estradiol would increase reward sensitivity, that should be

223   captured by the learning rate, reflecting how strongly new information will be weighed and

224   incorporated into the subjects' subjective values.

225

Results

227   Our sample was matched for age, height, visceral, and abdominal fat, BMI, working memory,

228   self-reported impulsivity, behavioural inhibition and approach, and mood. As a manipulation

229   check of our administration protocol, estradiol concentrations were significantly elevated in

230   subjects who had received estradiol compared to placebo after ($W = 1545$, 95% CI [0.03,

231   1.87], $p < .05$), but not before administration (baseline: $W = 1498$, 95% CI [-0.05, 1.03], $p =$

232   .09) and subjects' beliefs about whether they had received estradiol or placebo did not

233   correlate with the actual received drug ($r = 0.02$, $p = .82$; for further details see Supplementary

234   Materials).

Estradiol administration increases reward sensitivity

236   Our first hypothesis was that estradiol administration would increase reward sensitivity.

237   Because reward size was kept constant, the only difference across trials was whether a choice

238   was rewarded or not. Therefore, we predicted that increased reward sensitivity would be

239   reflected in a systematic difference in the options subjects would choose across trials

240   compared to placebo. To test this, we computed the average probability per group to select

241   option A/B (i.e. expected chosen option) on each trial (Fig. 2A), subtracted these traces from

242   each other, and plotted the cumulative choice difference across trials (Fig. 2B).

243   The cumulative choice difference in the expected chosen option between both groups

244   surpassed what would be expected by chance. This was tested by comparing the above

245   results to the 99.9th percentile of a null distribution (see Methods and materials) indicating

246   what would be expected by chance ($M_{\text{last trial}} = 53.48$ %, $z_{\text{last trial}} = 8.44$, $p < .001$, threshold for

9

247   99.9th percentile of the null: 46.20 %, Fig. 2B). The observed cumulative choice difference

248   remained significant when collapsed across time (M = 25.72 ±0.69%, z = 5.80, p < .001,

249   threshold for 99.9th percentile of the null: M = 21.02 %, Fig. 2C). When we traced trials that

250   contributed most to the cumulative difference, we observed a statistically significant difference

251   in what both groups chose on 7.6 % of trials (black dots in Fig. 2A). In other words, estradiol

252   administration caused subjects to choose a different option on 7.6 % of trials as compared to

253   placebo (z = 5.37, p < .001, threshold for 99.9th percentile of null distribution: 6.4 %).

254   However, estradiol administration did not only increase reward sensitivity in choice

255   behaviour as described above. When subjects rated how strongly they observed the changing

256   of reward probabilities of both options throughout the task on a scale from 1 to 100 during

257   debriefing, those receiving estradiol explicitly reported that they observed reward probabilities

258   across trials to change more strongly ($t_{(78.495)}$ = 2.15, *95% CI* = [0.855, 22.61], p = 0.035, d =

259   0.48, BF_{10} = 3.28).
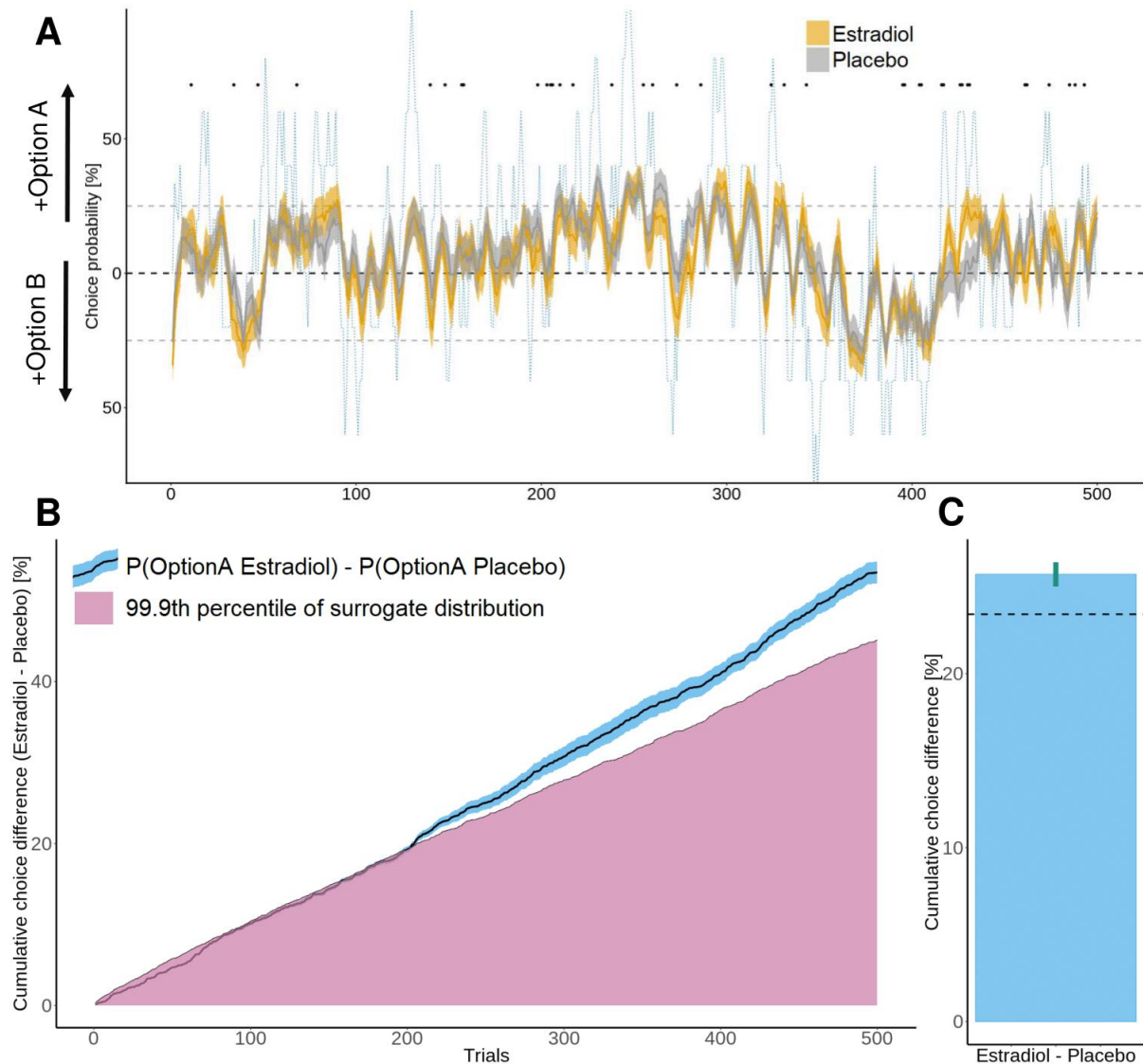
Fig. 2 A) Relative choice probability for choosing option A (top of y-axis) vs. choosing option B (bottom of y-axis) for the estradiol (orange) and placebo (gray) group. Solid thick lines represent trial mean, shaded areas around the thick lines denote standard errors of the mean. The blue dotted line 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 the 99.9th[th] 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.

11

279   **Estradiol's improvement in accuracy is moderated by DAT1 genotype and**
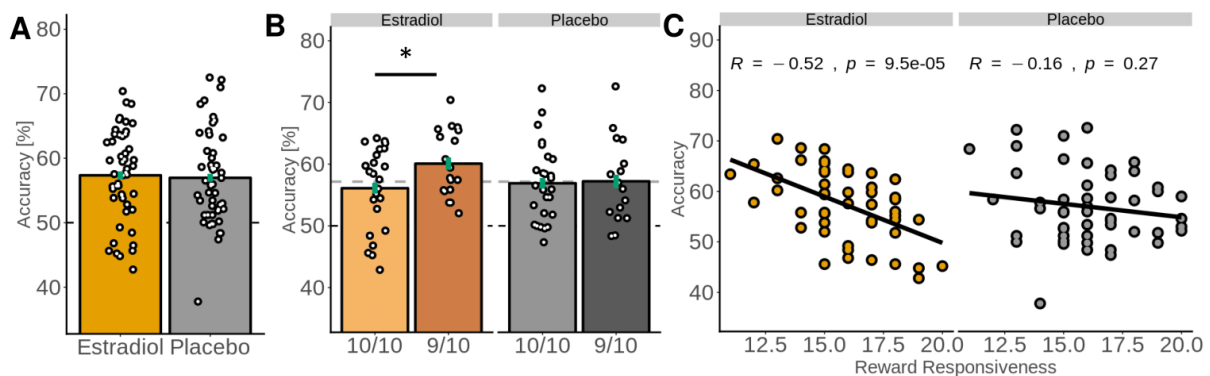
280   **reward responsiveness**

281   Following the observed systematic choice difference between both groups, we investigated

282   whether this was reflected in group differences in accuracy. Both groups were equally accurate

283   ($M_{Estradiol}$ = 57.30 ±6.91, $M_{Placebo}$ = 56.80 ±7.09, $t_{(97.94)}$ = 0.36, p = .72, Fig. 3A), and responded

284   equally fast ($M_{Estradiol}$ = 0.61 sec ±0.11, $M_{Placebo}$ = 0.62 *sec* ±0.09, $t_{(95.55)}$ = 0.46, p = .65).

285   However, we predicted that the effect of estradiol on accuracy will be moderated by

286   individual differences in baseline striatal dopamine, as other work has shown interactive

287   effects between task performance and dopamine-related genes [18,46]. In our study, we used the

288   DAT1 polymorphism as an index of striatal dopamine with the 9/10 and 10/10 genotypes being

289   associated with high and low striatal dopamine, respectively [47].

290   A general linear model revealed an interaction between drug administration and DAT1

291   genotype on accuracy ($F_{(1, 69)}$ = 4.10, p = .047, Ω^2 = 0.037, Fig. 3B) while controlling for

292   covariates (see Supplementary Materials). Pairwise comparisons revealed that estradiol

293   administration increased accuracy in subjects with the 9/10 genotype (i.e. high striatal

294   dopamine levels; M = 60.00 ±5.36) compared to those with a 10/10 genotype (i.e. low striatal

295   dopamine levels; 10/10 DAT1, M = 56.00 ± 6.51; $t_{(39.60)}$ = 2.14, *95% CI* [0.21, 7.63], p = .04, d

296   = 0.66, BF_{10} = 3.16), while this difference did not exist in the placebo group (9/10 genotype: M

297   = 57.21 ±6.60; 10/10 genotype: M = 56.90 ±6.34; $t_{(30.01)}$ = 0.15, p = .88). However, estradiol

298   administration did not improve accuracy in subjects with the 9/10 genotype relative to placebo

299   (9/10: $t_{(28.48)}$ = 1.35, p = .19; 10/10: $t_{(40.11)}$ = 1.79, *95% CI* [-0.41, 6.77], p = .08, d = 0.55, BF_{10}

300   = 1.92).

301   We also predicted that the effect of estradiol would be moderated by subjects' traits

302   related to reward responsiveness such as the one measured by BIS/BAS [30,31,33]. Indeed, when

303   we accounted for subjects' reward responsiveness (Fig 3C), we found that those who received

304   estradiol (vs. placebo) were more accurate (β = 24.17 ±11.14, $F_{(1, 85)}$ = 4.7, p = 0.03, Ω^2 =

305   0.008), which was further moderated by reward responsiveness ($F_{(1, 85)} = 4.6$, $p = 0.03$, $\Omega^2 =$

306   0.036) while controlling for covariates. That is, estradiol enhanced accuracy specifically in

307   subjects who were less reward responsive ($r = -0.52$, $p < .001$) with no such correlation in the

308   placebo group ($r = -0.16$, $p = 0.27$). This result further supports the hypothesis that estradiol

309   administration increased reward sensitivity by enhancing striatal reward prediction errors that

310   increased the saliency of each trial [8]. In our task, this hypothesis would predict more reward

311   responsive subjects to switch between the choice options more frequently. Indeed, we found

312   that within the estradiol group, there was a significant positive correlation between reward

313   responsiveness and switching ($r = 0.37$, $p = 0.009$) that was only trending in subjects who

314   received placebo ($r = 0.26$, $p = 0.07$). Importantly, the degree of switching negatively predicted

315   subjects' task accuracy ($F_{(1, 98)} = 69.38$, $p < .001$, $\Omega^2 = 0.41$) across both groups, explaining

316   why more reward responsive subjects who received estradiol were less accurate.

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

Increased reward sensitivity is observed in increased learning rates

325   Given our observation that estradiol increased reward sensitivity (Fig. 2), the interactive effect

326   with DAT1 and reward responsiveness (Fig. 3), and our hypothesis that the differences in

327   choice behaviour might occur due to upregulated striatal prediction errors, we predicted that

13

328   estradiol would enhance the learning of reward probabilities. In a RL framework this would be

329   reflected in increased learning rates.

330   To compare learning rates we used the maximum a posteriori estimates of the winning

331   model fitted in a hierarchical Bayesian way [48], which included separate learning rates for

332   positive and negative prediction errors, a temperature parameter, and an irreducible noise

333   parameter (see Supplementary Materials for details on model selection and parameter

334   recovery).

335   This model revealed that estradiol administration increased the learning rates for

336   positive and negative prediction errors compared to placebo ($\alpha_{Positive}$: $M_{Estradiol} = 0.47 \pm 0.22$,

337    $M_{Placebo} = 0.37 \pm 0.24$, $t_{(97.51)} = 2.36$, *95% CI* [0.017, 0.19], $p = .02$, $d = 0.47$, $BF_{10} = 4.77$; $\alpha_{Negative}$:

338    $M_{Estradiol} = 0.45 \pm 0.31$, $M_{Placebo} = = 0.27 \pm 0.25$, $t_{(93.55)} = 3.2$, *95% CI* [0.068, 0.29], $p = .002$, $d =$

339   0.64, $BF_{10} = 35.03$, Fig. 4A). We also compared both parameters by computing 95% Highest

340   Density Interval estimates [49] (Fig. 4D) with stronger evidence in favour of the negative learning

341   rate 95% HDI [0.04, 0.39] being higher in subjects who received estradiol compared to the

342   positive learning rate 95% HDI [-0.04, 0.32]. Contrary to our expectations, the observed main

343   effect of estradiol was not moderated by the DAT1 polymorphism in case of either learning

344   rate ($\alpha_{Positive}$: $F_{(1, 80)} = 0.24$, $p = .89$, $\alpha_{Negative}$: $F_{(1, 80)} = 0.12$, $p = .73$).

345   Finally, to provide construct validation for the obtained parameters, we correlated them

346   with the BIS/BAS Drive subscale measuring motivation for goal-directed behaviour [40]. We

347   found that both positive and negative learning rates were weakly correlated in subjects who

348   received estradiol ($\alpha_{Positive}$: r = 0.3, p = 0.03, $\alpha_{Negative}$: r = 27, p = 0.06) but not who received

349   placebo ($\alpha_{Positive}$: r = -0.10, p = 0.47, $\alpha_{Negative}$: r = -.09, p = 0.53).

14
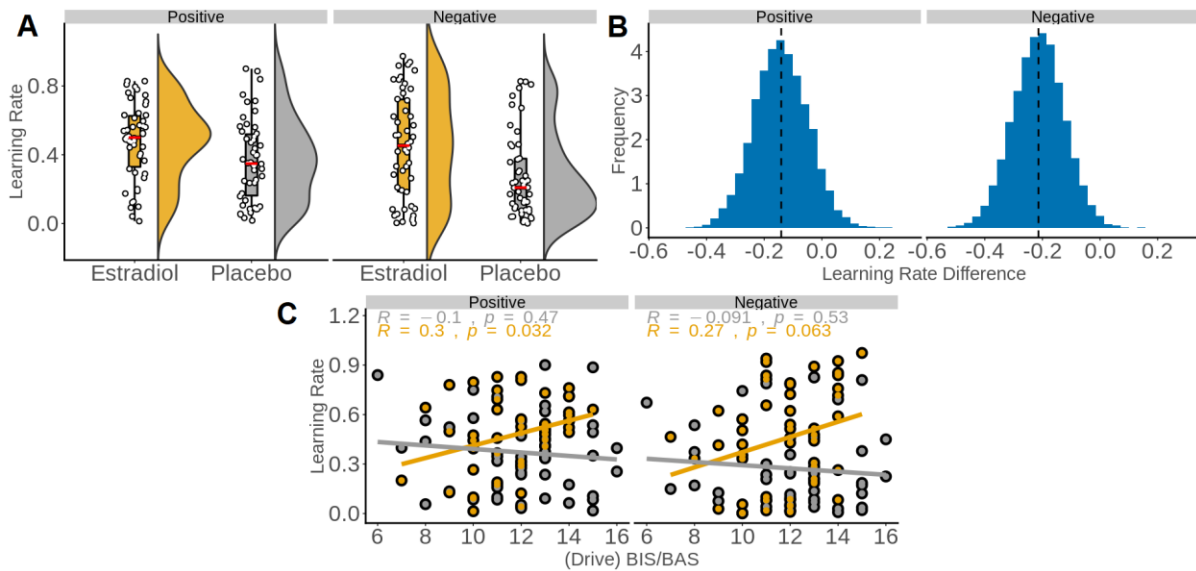
Fig. 4 A) Learning rates by drug treatment. Individual dots represent subjects. The red bar represents the median, the box plot represents 75% of data, with the whiskers representing 1.95*IQR. **B)** 95% highest density interval (HDI) estimates for learning rate differences between both groups. **C)** Correlation between both learning rates and the drive subscale of BIS/BAS measuring motivation for goal-directed behaviour.

Increased reward sensitivity is driven by differences in stay decisions, choice autocorrelation, and is moderated by the DAT1 genotype

Finally, to more precisely understand the observed difference in choice behaviour between both groups and DAT1 genotype, we tested whether these differences could be attributed to differences in staying behaviour and choice autocorrelation [24,50].

As would be predicted by increased accuracy in the 9/10 DAT genotype subjects, on average these subjects also chose the same option more often after being rewarded for this option before ($M = 1.79 \pm 0.18$; Fig. 5A). This was true when comparing them to subjects with placebo who had the 9/10 genotype ($M = 1.63 \pm 0.22$; $t_{(29.05)} = 2.33$, *95% CI* [0.02, 0.3], $p = .03$, $d = 0.80$, $BF_{10} = 5.01$), and when comparing them to subjects who had the 10/10 genotype (placebo: $t_{(41.61)} = 2.22$, *95% CI* [0.01, 0.27], $p = .03$, $d = 0.66$, $BF_{10} = 3.34$; estradiol: ($t_{(38..86)} = 2.49$, *95% CI* [0.03, 0.27], $p = .02$, $d = 0.64$, $BF_{10} = 5.95$). In contrast, this difference was not observed between both drug groups when not accounting for differences in DAT1 genotype ($t_{(97.94)} = 1.10$, $p = 0.28$). In other words, the increase in accuracy by exogenously elevated

15

371   estradiol in individuals with a 9/10 genotype was reflected in increased decisions to stay with

372   options for which they were previously rewarded.

373   We then extended this metric by examining how subjects' previous choice impacted

374   their current choice. We observed that estradiol administration caused subjects to weigh their

375   choice on the previous trial (n - 1) more heavily compared to placebo (Fig. 5B), both when that

376   choice was rewarded ($t_{(97)}$ = 2.61, *95% CI* [0.073, 0.538], *p* = 0.01, *d* = 0.52, BF_{10} = 8.07) and

377   irrespective of the outcome ($t_{(97)}$ = 2.21, *95% CI* [0.034, 0.639], *p* = .03, *d* = 0.44, BF_{10} = 3.49).

378   However, we also observed a decrease in the weight of previous rewarded choices occurring

379   more than one trial ago (n − 4) ($t_{(97)}$ = 2.59, *95% CI* [0.064, 0.482], *p* = 0.01, *d* = 0.52, BF_{10} =

380   7.93). In other words, while recent rewarded choices carried more weight, choices that were

381   rewarded further in the past carried less weight, which is in line with having higher learning

382   rates compared to the placebo group.

383   When we further split this analysis according to the DAT1 genotype (Fig. 5C), we

384   observed similar results to the ones reported above. Estradiol administration enhanced the

385   weight of the last choice (n − 1) on the current choice in the 9/10 subgroup ($t_{(32)}$ = 2.12, *95%*

386   *CI* [0.022, 1.027], *p* = .04, *d* = 0.73, BF_{10} = 3.44) while this was not the case in the 10/10

387   subgroup ($t_{(48)}$ = 1.51, *p* = .14).

16
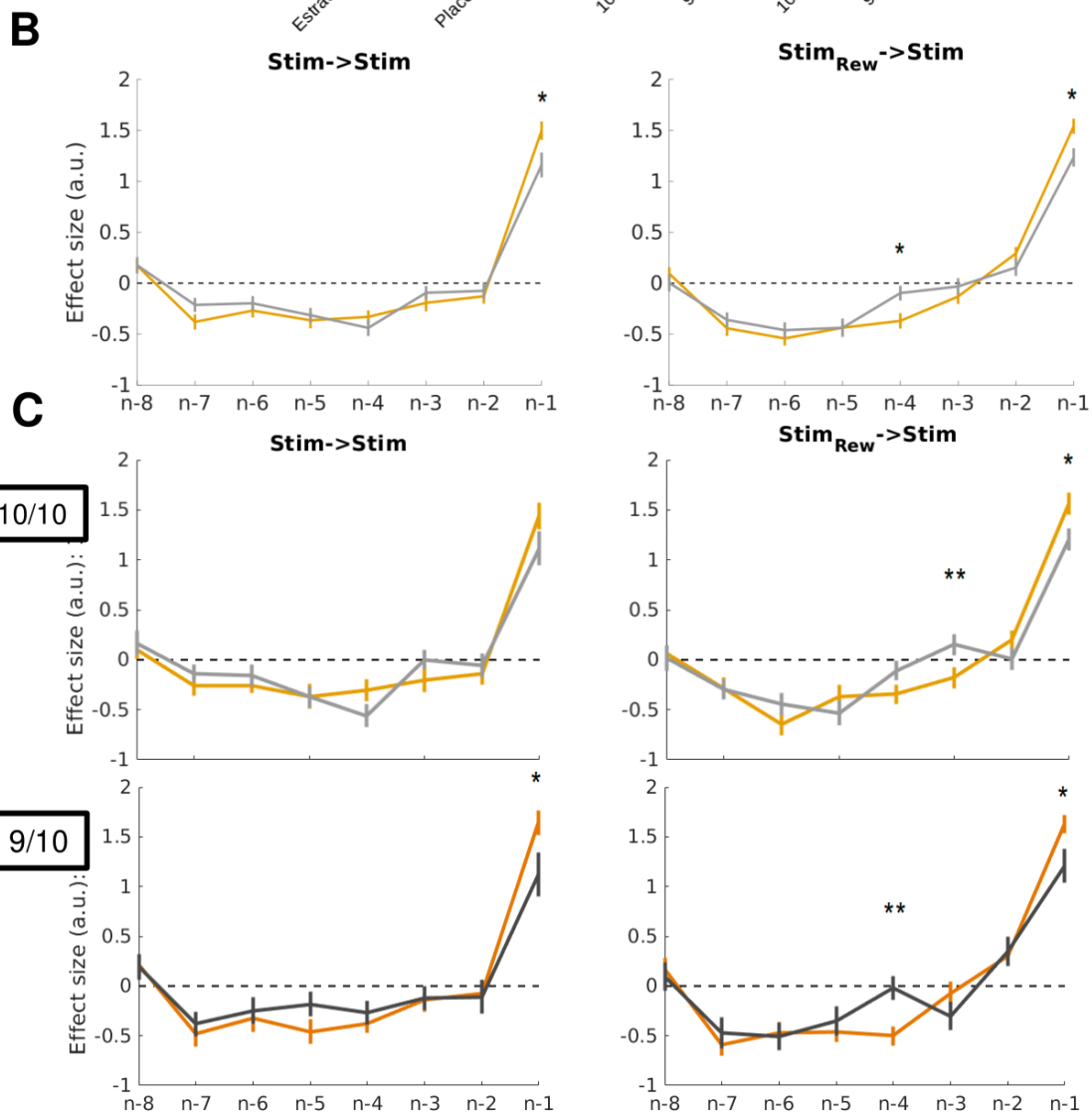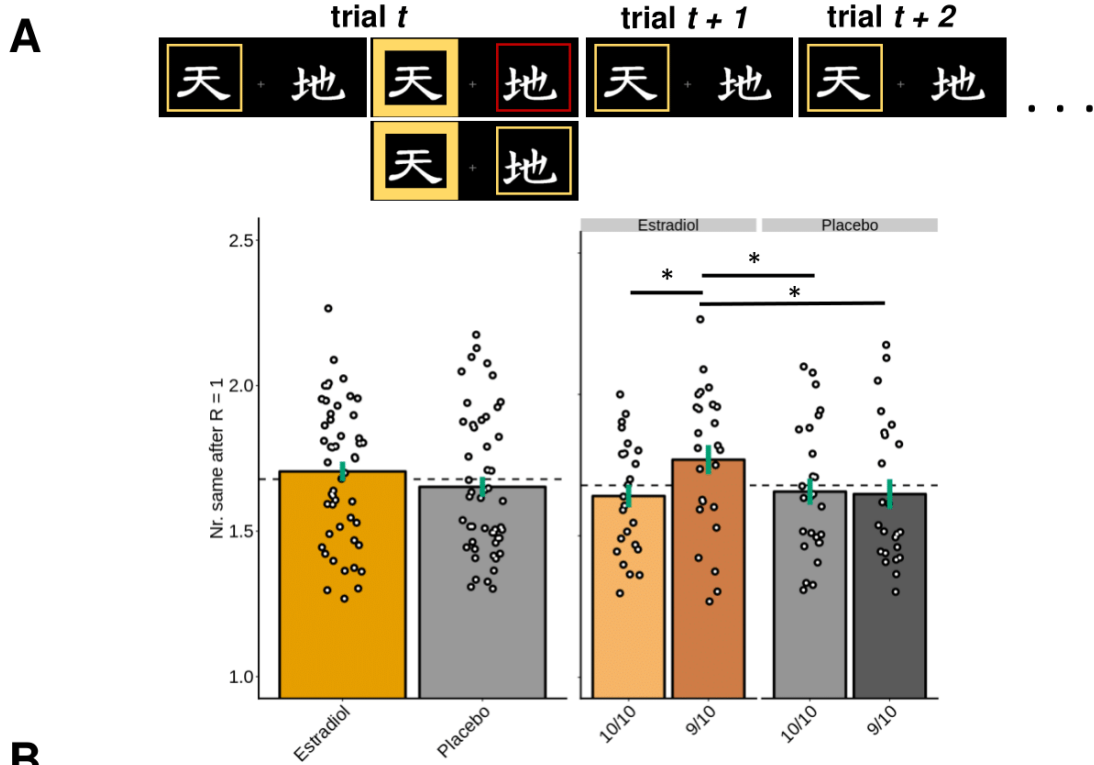
389 **Fig. 5 A)** Staying behaviour: the average number of trials the same option was selected when that
390 choice was previously rewarded. In both plots, each dot represents a subject, the green error bar
391 represents standard error of the mean. **BC)** Autocorrelation analysis showing the impact of previous
392 choices from 1 (n-1) to 8 (n-8) trials ago on the current choice. The left panel shows the averaged effect
393 for both options irrespective of whether they were previously rewarded for that choice. The right panel
394 shows the averaged effect for both options when they were previously rewarded for that choice. Both
395 line plots are the mean and SEM of z-scored regressor weights. **B)** shows the split according to estradiol
396 and placebo while **C)** shows a further split according to the 9/10 and 10/10 DAT1 polymorphism. ** $p <$
397 .01, * $p <$ .05.

398

Discussion

399

400 We examined the causal role of estradiol in reward processing in men. We show that estradiol

401 affects reward processing and learning in men by increasing reward sensitivity, affecting

402 choice autocorrelation, and increasing explicit awareness of changing reward probabilities.

403 Furthermore, we show that estradiol's effect on accuracy is moderated by subjects' striatal

404 dopaminergic functioning (DAT1 genotype) and reward responsiveness. Finally, we show that

405 from a computational perspective, its effect can be characterised as an increase in learning

406 rates.

407 First, we found that estradiol increased reward sensitivity during reinforcement

408 learning. We observed that both groups systematically choose different options across trials,

409 and on a subset of trials the options each group chose was statistically significantly different.

410 Furthermore, estradiol increased choice autocorrelation for recent choices compared to

411 placebo, while decreasing it for rewarded choice made several trials ago. These findings jointly

412 show that by increasing reward sensitivity through upregulated striatal reward prediction

413 errors[8], estradiol administration caused recent choices and outcomes to be more salient

414 compared to ones occurring further in the past. This interpretation is further supported by the

415 estradiol group having higher learning rates for both positive and negative prediction errors,

416 and subjects explicitly reporting to have observed a higher degree of reward probability

417 changing. Furthermore, it is supported by the effect of estradiol on accuracy being moderated

418 by reward responsiveness. The predicted increase in reward prediction errors caused more

419 reward responsive subjects to switch more, in turn reducing their accuracy, while this

420   relationship did not exist with placebo. The key contribution of the results above is showing,

421   for the first time, the behavioural and computational effect of estradiol administration on

422   reinforcement learning in men.

423   Next, we found that the effect of estradiol on accuracy was moderated by the DAT1

424   genotype, replicating previous correlational studies in women[6,75]. This result provides

425   evidence for the hypothesis that estradiol may act by amplifying dopamine D1 receptor

426   signalling in humans[2], as such an effect would also have been predicted by dopamine

427   precursor administration [27] [26,36]. To determine the robustness of this effect, we computed

428   several behavioural metrics and Bayes factors to show converging evidence for a moderating

429   role of DAT1. We found that estradiol administration significantly increased accuracy in

430   subjects with the 9/10 (i.e. high striatal dopamine) compared to ones with the 10/10 genotype

431   (i.e. low striatal dopamine) and that this increase in accuracy was likely driven by increased

432   staying behaviour. That is, subjects with high striatal dopamine chose the same option on

433   more trials, on average, if they were previously rewarded for that choice, compared to the

434   other subgroups. Furthermore, even when they were not rewarded, they showed stronger

435   choice autocorrelation compared to subjects with high striatal dopamine who received a

436   placebo.

437   The key contribution of the DAT1-related findings is in reconciling discrepancies of

438   previous correlational work that have been attributed to differences in baseline dopamine

439   levels [7,51–53]. That is, we show for the first time that baseline differences in dopamine, indeed,

440   play an important role in how estradiol influences reward processing and that these need to

441   be considered in future work aiming at better understanding their relationship. Crucially, the

442   effects we report for accuracy and staying behaviour were not explained by other mechanistic

443   explanations such as those related to androgen receptor functioning, androgen to estrogen

444   conversion, or estrogen receptor functioning (see Supplementary Materials).

445   Our results highlight a further feature of how estradiol influences reinforcement

446   learning. We found that estradiol improved accuracy specifically in the high striatal dopamine

19

447   group which is in contrast to previous research with the dopamine precursor L-dopa where

448   decreased accuracy in such subjects was reported but increased accuracy in subjects with

449   low striatal dopamine was found [26]. However, it is known that dopamine precursors impact

450   behaviour in a dose-dependent manner [54,55]. This would imply that our estradiol dose acted

451   akin to a "low dosage" of a dopamine precursor which is supported by two estradiol

452   administration studies in women. One showing 12 mg of estradiol (i.e. 6 times our dose)

453   decreasing working memory performance[34], which was interpreted as an overstimulation of

454   dopaminergic transmission. With the other showing a dopamine-like dose-dependent effect

455   of estradiol on hippocampal activity with doses between 2 and 12mg.[35].

456   In addition to the effects reported for DAT1, we also provide preliminary evidence that

457   the effect of estradiol is moderated by COMT (see Supplementary Materials). This is in support

458   of the hypothesis that COMT activity is inhibited through estradiol metabolites which in turn

459   increases dopamine availability [25 2,56 13,57].

460   The current study encountered limitations. We recommend future work employing

461   pharmacogenetics to increase their sample size. While our sample size was approximately

462   twice as large compared to most previous work [7,25,34,36,52,53,58,59] and in line with suggestions

463   for the field [2], we suspect that an increase in sample size, as in work on other hormones [60,61],

464   would have enabled us to make more precise claims at the level of polymorphism subgroup.

465   As no previous estradiol administration study investigated either DAT1 or COMT, we had no

466   basis for the minimal viable sample size, except the general recommendation in [2]. Our results

467   therefore await replication in future administration work.

468

469

470

471

472 **Conclusion**

473 In conclusion, we have shown that estradiol influences choice behaviour by increasing reward

474 sensitivity in healthy young men with distinct behavioural and computational signatures and

475 that these effects are moderated by striatal dopamine-related genes (DAT1) and personality

476 traits related to reward sensitivity. The approach and findings of this study show that

477 understanding the role of estradiol in reward processing has important implications for a better

478 understanding of the biology and neuroscience of human cognition that is moderated by genes

479 in both health and disorder.

480

481 # Acknowledgements

486

487 # Conflict of interests

488 The authors declare no conflict of interest with respect to the research, authorship, and/or

489 publication of this article. RL received travel grants and/or conference speaker honoraria within

490 the last three years from Shire, Heel, Bruker, and support from Siemens Healthcare regarding

491 clinical research using PET/MR. He is a shareholder of BM Health GmbH since 2019.

492

493

494

References

495

496 1. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward.

497   *Science (80-.).* **275**, 1593–1599 (1997).

498 2. Diekhof, E. K. Estradiol and the reward system in humans. *Curr. Opin. Behav. Sci.* **23**,

499   58–64 (2018).

500 3. Saldanha, C. J., Remage-Healey, L. & Schlinger, B. A. Synaptocrine signaling:

501   Steroid synthesis and action at the synapse. *Endocr. Rev.* **32**, 532–549 (2011).

502 4. Luine, V. N. Estradiol and cognitive function: Past, present and future. *Horm. Behav.*

503   **66**, 602–618 (2014).

504 5. Thomas, J., Météreau, E., Déchaud, H., Pugeat, M. & Dreher, J. Hormonal treatment

505   increases the response of the reward system at the menopause transition : A

506   counterbalanced randomized placebo-controlled fMRI study.

507   *Psychoneuroendocrinology* **50**, 167–180 (2014).

508 6. Dreher, J. *et al.* Menstrual cycle phase modulates reward-related neural function in

509   women. *PNAS* **104**, 2465–2470 (2007).

510 7. Diekhof, E. K. & Ratnayake, M. Menstrual cycle phase modulates reward sensitivity

511   and performance monitoring in young women: Preliminary fMRI evidence.

512   *Neuropsychologia* **84**, 70–80 (2016).

513 8. Bayer, J., Rusch, T., Zhang, L., Gläscher, J. & Sommer, T. Dose-dependent effects of

514   estrogen on prediction error related neural activity in the nucleus accumbens of

515   healthy young women. *Psychopharmacology (Berl).* (2019) doi:10.1007/s00213-019-

516   05409-7.

517 9. Lévesque, D. & Di Paolo, T. Rapid conversion of high into low striatal D2-dopamine

518   receptor agonist binding states after an acute physiological dose of 17β-estradiol.

519   *Neurosci. Lett.* **88**, 113–118 (1988).

520   10.   Becker, J. B. Gender Differences in Dopaminergic Function in Striatum and Nucleus

521   Accumbens. *Pharmacol. Biochem. Behav.* **64**, 803–812 (1999).

522   11.   Becker, J. B. Direct effect of 17β-estradiol on striatum: Sex differences in dopamine

523   release. *Synapse* **5**, 157–164 (1990).

524   12.   Pasqualini, C., Olivier, V., Guibert, B., Frain, O. & Leviel, V. Acute Stimulatory Effect

525   of Estradiol on Striatal Dopamine Synthesis. *J. Neurochem.* **65**, 1651–1657 (1995).

526   13.   Ball, P., Knuppen, R., Haupt, M. & Breuer, H. Interactions Between Estrogens and

527   Cateehol Amines III. Studies on the Methylation of Catechol Estrogens, Catechol

528   Amines and other Catechols by the Catechol-O-Methyltransferase of Human Liver. *J.*

529   *Clin. Endocrinol. Metab.* **34**, 736–746 (1972).

530   14.   Yoest, K. E., Quigley, J. A. & Becker, J. B. Rapid effects of ovarian hormones in

531   dorsal striatum and nucleus accumbens. *Horm. Behav.* **104**, 119–129 (2018).

532   15.   Schultz, W., Stauffer, R. W. & Lak, A. The phasic dopamine signal maturing: from

533   reward via behavioural activation to formal economic utility. *Curr. Opin. Neurobiol.* **43**,

534   139–148 (2017).

535   16.   Glimcher, P. W. Understanding dopamine and reinforcement learning: the dopamine

536   reward prediction error hypothesis. *PNAS* **108**, 15647–54 (2011).

537   17.   Frank, M. J., Seeberger, L. C. & O'Reilly, R. C. By carrot or by stick: Cognitive

538   reinforcement learning in Parkinsonism. *Science (80-.).* **306**, 1940–1943 (2004).

539   18.   Cools, R. *et al.* Striatal Dopamine Predicts Outcome-Specific Reversal Learning and

540   Its Sensitivity to Dopaminergic Drug Administration. *J. Neurosci.* **29**, 1538–1543

541   (2009).

542   19.   den Ouden, H. E. M. *et al.* Dissociable Effects of Dopamine and Serotonin on

543   Reversal Learning. *Neuron* **80**, 1090–1100 (2013).

544   20.   Jocham, G., Klein, T. A. & Ullsperger, M. Dopamine-Mediated Reinforcement

545   Learning Signals in the Striatum and Ventromedial Prefrontal Cortex Underlie Value-

546   Based Choices. *J. Neurosci.* **31**, 1606–1613 (2011).

547   21.   Jocham, G., Klein, T. a & Ullsperger, M. Differential Modulation of Reinforcement

548   Learning by D2 Dopamine and NMDA Glutamate Receptor Antagonism. *J. Neurosci.*

549   **34**, 13151–13162 (2014).

550   22.   Eisenegger, C. *et al.* Role of dopamine D2 receptors in human reinforcement learning.

551   *Neuropsychopharmacology* **39**, 2366–2375 (2014).

552   23.   Swart, J. C. *et al.* Catecholaminergic challenge uncovers distinct Pavlovian and

553   instrumental mechanisms of motivated (in)action. *Elife* **6**, 1–36 (2017).

554   24.   Frank, M. J. *et al.* Genetic triple dissociation reveals multiple roles for dopamine in

555   reinforcement learning. *PNAS* **104**, 16311–16316 (2007).

556   25.   Jacobs, E. & D'Esposito, M. Estrogen Shapes Dopamine-Dependent Cognitive

557   Processes: Implications for Women's Health. *J. Neurosci.* **31**, 5286–5293 (2011).

558   26.   Eisenegger, C. *et al.* DAT1 Polymorphism Determines L-DOPA Effects on Learning

559   about Others ' Prosociality. *PLoS One* **8**, e67820 (2013).

560   27.   Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J. & Frith, C. D. Dopamine-

561   dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*

562   **442**, 1042–1045 (2006).

563   28.   Doll, B. B., Bath, K. G., Daw, N. D. & Frank, M. J. Variability in dopamine genes

564   dissociates model-based and model-free reinforcement learning. *J. Neurosci.* **36**,

565   1211–1222 (2016).

566   29.   Nemmi, F. *et al.* Interaction between striatal volume and DAT1 polymorphism predicts

567   working memory development during adolescence. *Dev. Cogn. Neurosci.* **30**, 191–
568   199 (2018).

569   30.   Boksem, M. A. S., Tops, M., Kostermans, E. & De Cremer, D. Sensitivity to
570   punishment and reward omission: Evidence from error-related ERP components. *Biol.*
571   *Psychol.* **79**, 185–192 (2008).

572   31.   Boksem, M. A. S., Tops, M., Wester, A. E., Meijman, T. F. & Lorist, M. M. Error-
573   related ERP components and individual differences in punishment and reward
574   sensitivity. *Brain Res.* **1101**, 92–101 (2006).

575   32.   Lee, S. H., Ham, B. J., Cho, Y. H., Lee, S. M. & Shim, S. H. Association study of
576   dopamine receptor D2TaqI A polymorphism and reward-related personality traits in
577   healthy Korean young females. *Neuropsychobiology* **56**, 146–151 (2008).

578   33.   Reuter, M., Schmitz, A., Corr, P. & Hennig, J. Molecular genetics support Gray's
579   personality theory: The interaction of COMT and DRD2 polymorphisms predicts the
580   behavioural approach system. *Int. J. Neuropsychopharmacol.* **9**, 155–166 (2006).

581   34.   Sommer, T. *et al.* Effects of the experimental administration of oral estrogen on
582   prefrontal functions in healthy young women. *Psychopharmacology (Berl).* **235**, 3465–
583   3477 (2018).

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

587   36.   Jakob, K., Ehrentreich, H., Holtfrerich, S. K. C., Reimers, L. & Diekhof, E. K. DAT1-
588   genotype and menstrual cycle, but not hormonal contraception, modulate
589   reinforcement learning: Preliminary evidence. *Front. Endocrinol. (Lausanne).* **9**, 60
590   (2018).

591   37.   Stricker, R. *et al.* Establishment of detailed reference values for luteinizing hormone ,

592   follicle stimulating hormone , estradiol , and progesterone during different phases of

593   the menstrual cycle on the Abbott ARCHITECT □ analyzer. *Clin. Chem. Lab. Med.* **44**,

594   883–887 (2006).

595   38.   Eisenegger, C., von Eckardstein, A., Fehr, E. & von Eckardstein, S. Pharmacokinetics

596   of testosterone and estradiol gel preparations in healthy young men.

597   *Psychoneuroendocrinology* **38**, 171–178 (2013).

598   39.   Steyer, R., Schwenkmezger, P., Notz, P. & Eid, M. Testtheoretische Analysen des

599   Mehrdimensionalen Befindlichkeitsfragebogen. *Diagnostica* **40**, 320–328 (1994).

600   40.   Patton, J. H., Stanford, M. S. & Barratt, E. S. Factor structure of the barratt

601   impulsiveness scale. *J. Clin. Psychol.* **51**, 768–774 (1995).

602   41.   Carver, Charles, S. & White, Teri, L. Behavioral Inhibition, Behavioral Activation, and

603   Affective Responses to Impending Reward and Punishment: The BIS/BAS Scales. *J.*

604   *Pers. Soc. Psychol.* **67**, 319–333 (1994).

605   42.   Kim, S. H., Yoon, H. S., Kim, H. & Hamann, S. Individual differences in sensitivity to

606   reward and punishment and neural activity during reward and avoidance learning.

607   *Soc. Cogn. Affect. Neurosci.* **10**, 1219–1227 (2014).

608   43.   Sali, A. W., Anderson, B. A. & Yantis, S. Reinforcement learning modulates the

609   stability of cognitive control settings for object selection. *Front. Integr. Neurosci.* **7**, 95

610   (2013).

611   44.   Unger, K., Heintz, S. & Kray, J. Punishment sensitivity modulates the processing of

612   negative feedback but not error-induced learning. *Front. Hum. Neurosci.* **6**, 186

613   (2012).

614   45.   Lighthall, N. R., Gorlick, M. A., Schoeke, A., Frank, M. J. & Mather, M. Stress

615   Modulates Reinforcement Learning in Younger and Older Adults. *Psychol Aging* **28**,

616   35–46 (2013).

617   46.   Cools, R. & D'Esposito, M. Inverted-U-shaped dopamine actions on human working

618   memory and cognitive control. *Biol. Psychiatry* **69**, 113–125 (2011).

619   47.   Heinz, A. *et al.* Genotype influences in vivo dopamine transporter availability in human

620   striatum. *Neuropsychopharmacology* **22**, 133–139 (2000).

621   48.   Ahn, W.-Y., Haines, N. & Zhang, L. Revealing Neurocomputational Mechanisms of

622   Reinforcement Learning and Decision-Making With the hBayesDM Package. *Comput.*

623   *Psychiatry* **1**, 24–57 (2017).

624   49.   Ahn, W. *et al.* Decision-making in stimulant and opiate addicts in protracted

625   abstinence : evidence from computational modeling with pure users. *Front. Psychol.* **5**,

626   1–15 (2014).

627   50.   Daw, N. D., Doherty, J. P. O., Dayan, P., Seymour, B. & Dolan, R. J. Cortical

628   substrates for exploratory decisions in humans. *Nature* **441**, 876–879 (2006).

629   51.   Smith, C. T., Sierra, Y., Oppler, S. H. & Boettiger, C. A. Ovarian Cycle Effects on

630   Immediate Reward Selection Bias in Humans: A Role for Estradiol. *J. Neurosci.* **34**,

631   5468–5476 (2014).

632   52.   Diekhof, E. K. Be quick about it. Endogenous estradiol level, menstrual cycle phase

633   and trait impulsiveness predict impulsive choice in the context of reward acquisition.

634   *Horm. Behav.* **74**, 186–193 (2015).

635   53.   Reimers, L., Büchel, C. & Diekhof, E. K. How to be patient. The ability to wait for a

636   reward depends on menstrual cycle phase and feedback-related activity. *Front.*

637   *Neurosci.* **8**, 401 (2014).

638   54.   Chowdhury, R. *et al.* Dopamine restores reward prediction errors in old age. *Nat.*

639   *Neurosci.* **16**, 648–653 (2013).

640   55.   Chowdhury, R., Guitart-Masip, M., Bunzeck, N., Dolan, R. J. & Düzel, E. Dopamine

641   modulates episodic memory persistence in old age. *J. Neurosci.* **32**, 14193–14204

642   (2012).

643   56.   Männistö, P. T. & Kaakkola, S. Catechol-O-methyltransferase (COMT): Biochemistry,

644   molecular biology, pharmacology, and clinical efficacy of the new selective COMT

645   inhibitors. *Pharmacol. Rev.* **51**, 593–628 (1999).

646   57.   Colzato, L. S. & Hommel, B. Effects of estrogen on higher-order cognitive functions in

647   unstressed human females may depend on individual variation in dopamine baseline

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

649   58.   Colzato, L. S., Hertsig, G. & Wildenberg, van den Hommel, B. Estrogen modulates

650   inhibitory control in healthy human females: evidence from the stop-signal paradigm.

651   *Neuroscience* **167**, 709–715 (2010).

652   59.   Colzato, L. S., Pratt, J. & Hommel, B. Estrogen modulates inhibition of return in

653   healthy human females. *Neuropsychologia* **50**, 98–103 (2012).

654   60.   Geniole, S. N. *et al.* Using a Psychopharmacogenetic Approach To Identify the

655   Pathways Through Which—and the People for Whom—Testosterone Promotes

656   Aggression. *Psychol. Sci.* **30**, 481–494 (2019).

657   61.   Losecaat Vermeer, A. B. *et al.* Exogenous testosterone increases status-seeking

658   motivation in men with unstable low social status. *Psychoneuroendocrinology* **113**,

659   (2020).

660   62.   Sheehan, D. V. *et al.* The validity of the Mini International Neuropsychiatric Interview

661   (MINI) according to the SCID-P and its reliability. *Eur. Psychiatry* **12**, 232–241 (1997).

662   63.   Fishman, J., Boyar, R. M. & Hellman, L. Influence of body weight on estradiol

663   metabolism in young women. *J. Clin. Endocrinol. Metab.* **41**, 989–991 (1975).

664   64.   Schneider, J. *et al.* Effects of Obesity on Estradiol Metabolism: Decreased Formation

665   of Nonuterotropic Metabolites. *J. Clin. Endocrinol. Metab.* **56**, 973–978 (1983).

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

668   66.   Eisenegger, C., Naef, M., Snozzi, R., Heinrichs, M. & Fehr, E. Prejudice and truth
669   about the effect of testosterone on human bargaining behaviour. *Nature* **463**, 356–359
670   (2010).

671   67.   Aeberli, I., Molinari, L. & Zimmermann, M. B. A composite score combining waist
672   circumference and body mass index more accurately predicts body fat percentage in
673   6- to 13-year-old children. *Eur. J. Nutr.* **52**, 247–253 (2013).

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

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

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

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

681   72.   Rutledge, R. B. *et al.* Dopaminergic Drugs Modulate Learning Rates and
682   Perseveration in Parkinson ' s Patients in a Dynamic Foraging Task. *J. Neurosci.* **29**,
683   15104–15114 (2009).

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

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

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

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

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

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

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

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

702   81.   Gillies, G. E. & McArthur, S. Estrogen actions in the brain and the basis for differential
703   action in men and women: A case for sex-specific medicines. *Pharmacol. Rev.* **62**,
704   155–198 (2010).

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

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

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

714   85.   Ma, S. L. *et al.* Polymorphisms of the estrogen receptor (ESR1) gene and the risk of

715   Alzheimer's disease in a southern Chinese community. *Int. Psychogeriatrics* **21**, 977–

716   986 (2009).

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736 **Supplementary Materials:**

737 **A causal role for estradiol in human reinforcement learning**

738 Sebastijan Veselic[1,2,3]*, Gerhard Jocham[4], Christian Gausterer[5], Bernhard Wagner[6], Miriam

739 Ernhoefer-Reßler[6], Rupert Lanzenberger[7], Clauss Lamm[1,8], Christoph Eisenegger[1] & Annabel

740 Losecaat Vermeer[1]*

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

Methods

Subjects

The short version of e-MINI [62] was used to screen and exclude those who had a non-diagnosed, disclosed, or a diagnosed psychiatric disorder. 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.

Measurement Instruments

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.

Reinforcement Learning: In addition to the main 500 trials, subjects were trained on 10 practice trials with two initial options to learn how the task works. The initial two options were then changed (i.e. the sensory cues) before the main trials started. We did this to avoid carry-over effects from practice to the main task. As shown in Fig. 1, each trial included three stages: (1) a cue onset stage (5 sec) where subjects had to decide between the two options and press the corresponding key. If they did not respond within that time frame, they would see a warning message indicating they should respond and try to be faster next time; (2) a choice feedback stage (1 sec) where subjects received information about both the chosen (thick frame) and unchosen (thin frame) option (yellow - correct, red - wrong); and (3) an inter-trial interval ($M =$ 1.5 sec, jittered between 0.9 to 2.1 sec).

Working memory capacity: As described in the manuscript, subjects were presented with a sequence of letters one-by-one. For each letter, they had to decide if the current letter was the

33

782   same as the one presented *N* trials ago by pressing "R", in case it was not the same they had to

783   press "O". For example, in the 3-back condition, the letter sequence "***A B D A A***" would require

784   subjects to press "R" only to the second occurrence of A, as this was the same letter as the one

785   3 trials ago. The last A in this example sequence is defined as a lure trial, while the other letters

786   were nontarget trials. Lure trials were present only in the 2-BACK and 3-BACK conditions as

787   in [25], and while lure trials were added to keep the task consistent with their implementation,

788   we did not further analyse them separately as they were not relevant for our question. In total,

789   there were four blocks per condition. Each block was announced by an instruction lasting for

790   2 sec (Fig. 1A), a fixation cross (1 sec) and a sequence of 20 trials. Each trial was presented

791   for 1 sec with a 1 sec feedback phase and a 1 sec inter-stimulus interval. After every 20 trials,

792   subjects had a 3 sec resting period, before the next block was announced. A lack of response

793   to any cue was considered a miss.

Procedure

795   During the first session, we assessed height, weight, abdominal, and visceral fat because

796   these variables could impact estradiol metabolisation [63,64]. On the second session subjects

797   applied a topical transparent gel on their chest and shoulders that either contained 2 mg of

798   estradiol (Divigel, Orion Pharma AG, Zug Switzerland) or a placebo. A male experimenter was

799   present to ensure that the subjects applied the gel correctly.

Statistical analysis

Behavioural analysis

802   To quantify statistical significance for the cumulative choice difference reported in the

803   manuscript, we employed permutation testing (2000 iterations) by shuffling the responses of

804   each subject and thereby decoupling the label from the responses, thus building a null

805   distribution. The null distribution shows the difference that would be expected from random

806   allocation to the group. To determine significance, we computed z-scores as measures of

807   standardized effect size (e.g. [65]) by subtracting from the quantity of interest the mean of the

34

808   null distribution and dividing it by the standard deviation of the null distribution. From this, we

809   were able to use the Fisher-z-transformation to obtain p-values.

810   We used two-sample proportion z-tests to determines the percentage of trials on which

811   the number of subjects who chose option A in one group was statistically significantly different

812   from the other group on a trial-by-trial basis.

813   For all general linear models reported in the main results, we regressed out several

814   nuisance regressors known to impact estradiol metabolism or affect reinforcement learning

815   behaviour. These included weight, BMI, abdominal and visceral fat [63,64], post-administration

816   cortisol levels [45], and beliefs related to having received the drug. This was done because of

817   our previous work showing the impact of beliefs about a hormone on subsequent behaviour,

818   irrespective of whether subjects underwent treatment or received placebo [66].

819   All linear models were compared with BIC and AIC. Unless stated otherwise in the

820   main text, for all reported results the winning model regressed out cortisol levels following

821   administration, beliefs about having received the drug, the certainty of that belief and whether

822   they had observed any changes in themselves, a composite score of weight and BMI that

823   were summed together [67] because of their high intrinsic correlation ($r = 0.89$), visceral, and

824   abdominal fat. For general linear models involving accuracy, we also regressed out reaction

825   times to control for accuracy-speed trade-offs. All nuisance regressors were z-scored.

826   We compared both treatment groups for age and other bodily characteristics (i.e. BMI,

827   height, weight, visceral, and abdominal fat) and potential differences in self-reported mood

828   (MDBF), impulsiveness (BIS-11) and reward responsiveness (BIS/BAS) (see Questionnaires,

829   Table S4 and S5). We used two-tailed independent samples Welch t-tests, or Wilcoxon

830   signed-rank test if assumptions of normality were not met, to test whether the groups matched

831   on all variables. To test for mood differences after administration between the treatment

832   groups, we performed an ANCOVA for each of the three subscales of the MDBF questionnaire

833   where we controlled for baseline mood scores. Two-way ANOVAs were further performed on

834   the individual subscales of the BIS-11 questionnaire to investigate whether there was an

835   interaction between the group (estradiol, placebo) and session (pre, post) on impulsiveness.

836   To compare working memory capacity assessed by the N-BACK task, we analysed

837   target accuracy, reaction times, and d-prime. We analyzed this with an ANOVA containing the

838   between-subject variable group (estradiol, placebo) and within-subject variable for condition

839   together with an interaction term for group and condition.

840   In the supplementary results we used generalized linear mixed effects models using

841   the lme4 package in R [68,69] to investigate whether the interaction between drug, DAT1 or

842   COMT and trial would be predictive of the subjects' chosen option.

Computational modelling

844   To test whether estradiol would increase reward sensitivity and thereby learning, we

845   formalized behaviour within a reinforcement learning framework and fitted several Q-learning

846   models [70] with softmax choice rules:

847   Q-learning model (equation 1):

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

849   Softmax choice rule (equation 2):

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

851   Where, t is time, A is option A, Q is subjective value, α is the learning rate, R is the obtained

852   reward, and τ is the temperature parameter. Equations 1 and 2 represent our first model

853   (model 1). In Q-learning, the basic idea is that agents learn subjective values of actions they

854   perform in their environment. Subjective values are learned and updated through a value

855   function (Equation 1) following feedback after each action. A teaching signal known as the

856   learning rate-weighted prediction error dictates how strongly the subjective value will be

857   updated on each action. The prediction error corresponds to the difference between the

858   obtained and expected reward (i.e. the subjective value prior to making the new choice). Within

36

859 this process, the learning rate dictates how heavily new information will be weighted in

860 proportion to previous information about the option, and therefore how strongly the subjective

861 value will change from its current estimate. The softmax equation then yields the probability

862 of selecting an action given the learning rate and the temperature parameter, which reflects

863 stochasticity of choice behaviour.

864   By employing computational modelling of this sort, we were able to obtain parameter

865 estimates that quantify the difference in subjects' behaviour, captured by a difference in

866 learning rates. To obtain a more precise account of the effect of estradiol on reward

867 processing, we extended the basic Q-learning model in several ways, as described below.

868   The first extension (equation 3a and 3b) allowed for separate learning rates for α_{Pos}

869 and α_{Neg} that would differentiate between learning from positive and negative prediction

870 errors.

$$Q_{t+1}^A = Q_t^A + \alpha_{Pos}(R_t^A - Q_t^A) \tag{3a}$$

$$Q_{t+1}^A = Q_t^A + \alpha_{Neg}(R_t^A - Q_t^A) \tag{3b}$$

874 The updating with a positive learning rate occurs when the prediction error term $(R_t^A - Q_t^A)$

875 evaluates to positive while updating with the negative learning rate occurs when the prediction

876 error term evaluates to negative. Furthermore, due to reward stochasticity of our task reward

877 probability distribution (obtained by a Gaussian random walk – Fig. 1B), we added an

878 additional parameter ξ, representing a lapse or irreducible noise parameter [71] in our choice

879 rule (equation 4):

$$p = \frac{1}{1 + e^{\tau(Q_t^A - Q_t^B)}}(1 - \xi) + \frac{\xi}{2} \tag{4}$$

882 Finally, we added a perseverance parameter λ [72] (equation 5):

$$883 \qquad p \ = \ \frac{1}{1 \ + \ e^{\tau(Q_t^A - Q_t^B + \lambda C)}} (1 - \xi) + \frac{\xi}{2} \qquad\qquad (5)$$

884

885   Where C = 1, if the same option was chosen on trial *n* and trial *n+1*, and C = -1 if the opposite

886   was true. In summary, our full model had separate learning rates for positive and negative

887   prediction errors, a choice stochasticity, irreducible noise, and perseverance parameter. All

888   other models were reduced cases of this model and all possible combinations of the described

889   free parameters therefore yielded eight models in total for which we estimated parameters:

890   **Table S0. Models and parameters**

Model 1	Learning rate, temperature
Model 2	Learning rate, temperature, lapse
Model 3	Learning rate, temperature, perseverence
Model 4	Learning rate, temperature, lapse, perseverence
Model 5	Positive and negative learning rate, temperature
Model 6	Positive and negative learning rate, temperature, lapse
Model 7	Positive and negative learning rate, temperature, perseverence
Model 8	Positive and negative learning rate, temperature, lapse, perseverence

891

892   The model fitting was performed using JAGS and the rjags (v 4.9) package in R (v 3.6.0).

893   Each model was run with 5000 samples each with 1000 burn-in samples on three chains.

894   Priors over parameters and hyperparameters were set to default as described in [48]. We

895   computed the leave one out information criterion using the loo package [73] and used this metric

896   to compare the models. Furthermore, we performed Bayesian model comparison by

897   computing the (protected) exceedance probability [74] using the VBA toolbox [75] to determine the

898   best model and compare its congruency with the LOOIC measure. Finally, we extracted the

899   posterior predictive density for each subject as a measure of predictive power of the best

900   model. This was then compared to the actual behaviour as a measure of static (accuracy

901   collapsed across time) and dynamic (accuracy at each trial across subjects) predictive

902   accuracy. From the obtained maximum a-posteriori estimates we then generated synthetic

903   datasets for each subject and refit the model using synthetic data to assess parameter

904   recovery. We did this by correlating our original and recovered parameters to determine

905   whether the learning rate parameters of the winning model were correlated.

Genotyping

DNA extraction and quantification

908   Buccal swabs were collected using sterile cotton swabs (Sarstedt AG, Germany). DNA was

909   extracted from swabs using the QIAamp DNA Mini kit (Qiagen, Hilden, Germany) and eluted

910   in a final volume of 50 μL of QIAamp buffer AE (Qiagen). Human nuclear DNA was quantified

911   using the Applied Biosystems (AB) 7500 real-time PCR instrument (Thermo Fisher Scientific,

912   Waltham, MA) and the Quantifiler Human Plus quantification Kit (AB) following manufacturer's

913   recommendations.

Typing of repeat length polymorphisms

915   Genomic DNA fragments that contain polymorphic repeat sequences were amplified in two

916   separate reactions: i.e. a multiplex PCR (simultaneously targeting AR(CAG)n, DAT1 VNTR,

917   Erα(TA)n and Erβ(CA)n) and a singleplex PCR (targeting solely AR(GGN)n), respectively.

918   The multiplex PCR was performed using 5 ng template DNA in a reaction mix (total

919   volume of 25 μL) consisting of 1 × GeneAmp PCR buffer (AB), 0.25 mM each dNTP, 2.5 units

920   AmpliTaq Gold polymerase (AB) and target specific primers (AR(CAG), DAT1, ERα and Erβ;

921   including 5'-fluorescent-dye-labeled forward primers; details provided in Table 1). The

922   following protocol was applied using the Veriti 96-well thermal cycler (AB): 35 cycles at 95 °C

923   for 30 seconds, 55 °C for 1 minute, and 72 °C for 1 minute. Before the first cycle, an initial

924   denaturation (95 °C for 5 minutes) was included, and the last cycle was followed by a final

925   extension step at 72 °C for 45 minutes.

926   The singleplex PCR was conducted using 5 ng template DNA in a reaction mix (total

927   volume of 20 µL) containing target specific primers (AR(GGN)n, details provided in Table S1)),

928   0.5 µL Phire Hot Start II DNA polymerase (Thermo Fisher) in 1 × Phire reaction buffer (Thermo

929   Fisher). Amplification was carried out on the Veriti thermal cycler (AB) and included an initial

930   denaturation step at 98 °C for 30 seconds, followed by 33 cycles of 10 seconds at 98 °C, 30

931   seconds at 60 °C and 30 seconds at 72 °C. The last cycle was followed by a final extension

932   at 72 °C for 10 minutes.

933   Aliquots of PCR products were diluted with Hi-Di formamide (AB), mixed with internal lane

934   standard LIZ 600 v.2 (AB) and separated on the ABI 3500 Genetic Analyzer applying standard

935   conditions. The number of repeats predicted by the GeneMapper ID-X software (AB) was in

936   full agreement to the actual repeats determined by direct sequencing of PCR products using

937   the BigDye Terminator Sequencing Kit v3.1 (AB) in selected DNA samples.

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

940   [a]

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	

941   [a] AR(GGN)n primers only used in singleplex PCR; all other primers combined in a multiplex PCR
942   [b] Chromosome number and genomic location of targeted sequence (orientation provided in brackets) according to
943   UCSC version hg38 (http://genome.ucsc.edu/)
944   [c] The non-specific primer tail is underlined in Italics
945   [d] The final primer concentrations in the reaction mix
946
947
948

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 S2) 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.

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	

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

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

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

[a] Chromosome number and genomic location of primer binding site (orientation provided in brackets) according to UCSC version hg38 (http://genome.ucsc.edu/). The targeted SNP rs4680 is located at position chr22:19963748.

41

972   [b] The non-specific primer tail is underlined in Italics
973

974   ExoSAP-IT treatment was again applied for the clean-up of the minisequencing reaction. 5 µl

975   of purified minisequencing reaction product was then mixed with 9.3 µL Hi-Di formamide (AB)

976   and 0.2 µL of GeneScan-LIZ 120 internal size standard (AB). After a denaturing step for 5 min

977   at 98 °C followed by cooling to 4 °C the fragments were separated on an ABI PRISM 310

978   Genetic Analyzer (AB) with POP4 polymer and analysed with GeneMapper v3.2 software.

979   Calling of SNP variants based on minisequencing was in full agreement to results from direct

980   sequencing of PCR products in selected DNA samples.

Hormone concentrations

982   Quantification of estrone and estradiol in saliva samples was performed with derivatization

983   using pentafluorobenzoyl chloride (PFBCl) and the addition of the isotopically labeled internal

984   standards estrone-d_4 and estradiol-d_5. Organic saliva was reacted with 1.0 mL 1% PFBCl and

985   0.1 mL pyridine at 60°C for 30 min. The derivatization agents were evaporated, the sample

986   was reconstituted with 0.5 mL NaHCO3 and extracted with 1 mL n-hexane. The organic phase

987   was substituted with 0.2 mL dodecane and subjected to optimized GC-MS/MS analysis using

988   an Agilent 7890 GC with Agilent DB-17ht 15 m x 0.25 mm x 0.15 µm capillary column

989   connected to an Agilent 7010 tandem mass spectrometer operated in MRM mode using

990   negative chemical ionization at 150°C with methane as a reaction gas (40%, 2 mL/min).

991   Method validation was performed using ion transition m/z 464 -> 400 as a quantifier for estrone

992   and m/z 660 -> 596 for estradiol, whereas a LLOQ of 1.92 fg o.c. and 1.94 fg was obtained,

993   respectively.

994   Quantification of hydrocortisone and testosterone in saliva samples was performed

995   using liquid chromatography tandem mass spectrometry (LCMS/MS), with an Agilent 6460

996   with electrospray ionization in positive mode coupled to a 1290 UHPLC system. Collision

997   energy was optimized for specific MRM transitions of Hydrocortisone (363.2/121.1 m/z;

998   363.2/91.1 m/z), Testosterone (289.2/109.1; 289.2/97.1 m/z), 2,3,4-13C3-Hydrocortisone

999   (366.2/124 m/z) and 2,3,4-13C3-Testosterone (292.2/100 m/z). Agilent Poroshell 120 EC-C18

1000 was used for chromatographic separation under reversed phase conditions. The internal

1001 standard preparation and internal standard mixture was prepared containing 2,3,4-13C3-

1002 Hydrocortisone; 2,3,4-13C3-Testosterone, 2,4,16,16,17-d5-17b-Estradiol and concentration

1003 of 5ng/mL each. Samples were prepared by adding 100 µl internal standards (5 ng/mL) to

1004 500µl plasma or saliva and the steroids were extracted using 4 mL MTBE. After 10 min.

1005 overhead shacking, the samples were centrifuged for 5 min. at 3000 rpm and the top MTBE

1006 layer was transferred to a test tube. MTBE was evaporated using a centrivap concentrator at

1007 40°C (Labconco). The residual sample was then re-dissolved in methanol and analyzed by

1008 LC-MS/MS.

1009 **Questionnaires**

1010 *Mood*: To control for a potential confound of mood, tiredness, or alertness from the treatment

1011 affecting subjects' performance [22], we assessed subjects' self-reported mood before and after

1012 administration of the treatment, using the German Multidimensial Mood State Questionnaire

1013 ("Der Mehrdimensionale Befindlichskeitfragebogen - MDBF) [39] *Both versions of this*

1014 *questionnaire (A and B) contain 12 items with a 5-level Likert scale and three subscales that*

1015 *test for different continuums of mood (Good-Bad [α_{pre} = .81, α_{post} = .77], Awake-Tired [α_{pre} =*

1016 *.84, α_{post} = .87], Calm-Nervous [α_{pre} = .73, α_{post} = .75]).*

1017 *Impulsiveness*: We used the Barratt Impulsiveness Scale (BIS-11; [40] to measure subjects'

1018 impulsiveness as [52] observed that variations in estradiol levels differentially affected women

1019 with low trait as opposed to high trait impulsiveness. BIS-11 is a widely used measure for

1020 impulsiveness with 30 items describing common behaviour and preferences related to

1021 (non)impulsiveness which individuals have to rate on a 4-point scale (1 - rarely/never, almost

1022 always/always - 4). The General Impulsiveness (α_{pre} = .71, α_{post} = .75) factor together with its

1023 three second-order factors (Motor Impulsiveness (α_{pre} = .47, α_{post} = .54) Nonplanning

1024 Impulsiveness (α_{pre} = .6, α_{post} = .63), Attentional Impulsiveness (α_{pre} = .49, α_{post} = .52) are

1025 reported.

1026 ***Behavioural inhibition and activation:*** we measured the trait behavioural activation and

1027 inhibition with the Behavioural inhibiton/Behavioural Activation Scales (BIS/BAS; [41]. The BAS

1028 scale is a 24-item questionnaire answered on a four-level scale (1- very true for me, 4 - very

1029 false for me). It is subdivided into Drive (α= .74), Fun Seeking (α= .67), and Reward

1030 Responsiveness (α= .6) while the BIS scale (α= .77) is unidimensional. Drive is thought to

1031 measure the persistent pursuit of goals (e.g. "I go out of my way to get the things I want"), Fun

1032 Seeking: the desire for new rewards and willingness to approach events that would be

1033 potentially rewarding (e.g. "I crave excitement and new sensations"), while Reward

1034 Responsiveness focuses on positive responses that would occur if a reward is anticipated

1035 (e.g. "When I am doing well at something I love to keep doing it"). Finally, the BIS scale

1036 measures sensitivity to negative events (e.g. "Criticism or scolding hurts me quite a bit").

Belief probes

1038 In addition, we probed subjects' beliefs and confidence about receiving estradiol (e.g. whether

1039 they believed they received estradiol or a placebo, how certain they were of this answer, and

1040 whether they noticed any changes). This was done to later regress out the potential

1041 contribution of beliefs arising, for example, from subjects researching potential side effects of

1042 the hormone prior the experiment. Namely, subjects' beliefs about having received a hormone

1043 and beliefs about the effects of a hormone on their performance have previously shown to

1044 modulate behaviour independent of whether subjects had actually received it[66].

1045

1046

1047

1048

1049

1050

1051

1052

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. 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 [38]. 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 checked whether 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 [26,76]. 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: $F_{(1, 96)} = 0.30$, $p = 0.58$; Alertness: $F_{(1, 96)} = 1.35$, $p = .25$; Calmness: $F_{(1, 96)} = 1.34$, $p = .25$). Similarly, we observed no interaction between group membership and post-administration score (Mood: $F_{(1, 96)} = 0.06$, $p = .81$; Alertness: $F_{(1, 96)} = 1.88$, $p = .17$; Calmness: $F_{(1, 96)} = 1.55$, $p = .22$).

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

| | Group | | | | |
	Estradiol (n=50)	Placebo (n=50)	*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

46

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

1099 Note: Values in cells denote M, parentheses denote SD. The superscript 1 denotes the Mann-Whitney-Wilcoxon
1100 W value. For the remaining group comparisons, two-tailed independent samples Welch t-tests were employed. In
1101 cases where n is not equal to $N = 100$, data was not recorded for that particular variable. p-values are Bonferroni
1102 corrected.
1103
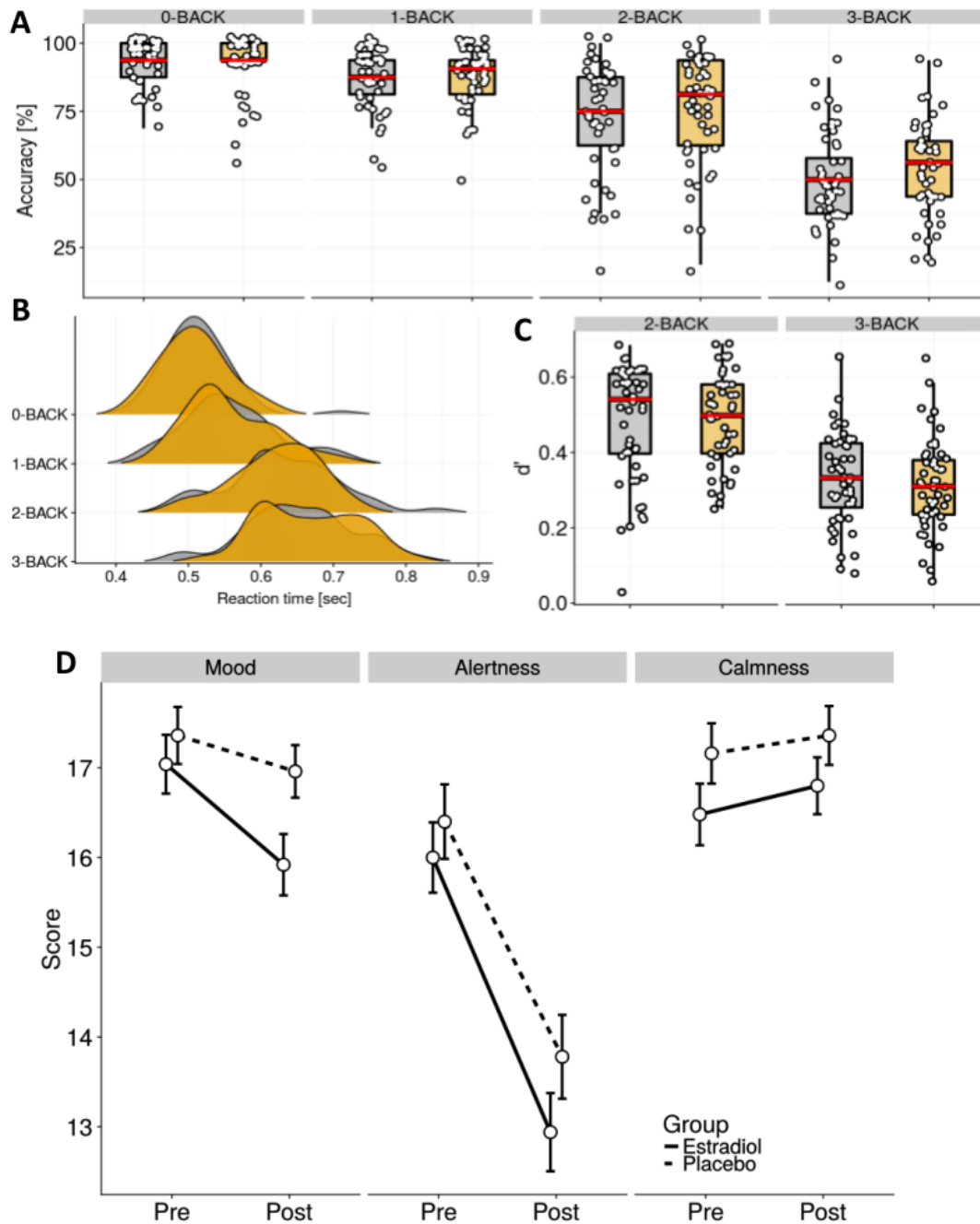
Figure S1. A) Accuracy for individual conditions. The red bar represents the median, the box plot represents the 75% middle most data points, with the whiskers representing 1.95*IQR. Orange depicts the estradiol and gray the placebo group. That is, they represent the division of subjects according to 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.

48

1115   Furthermore, our working memory (N-BACK) task revealed a comparable picture for accuracy

1116   (Figure S2A), reaction times (Figure S2B), and d-prime (Figure S2C). That is, there was no

1117   statistically significant difference between the estradiol and placebo group in accuracy,

1118   average reaction times, and d-prime. We did observe an expected drop in performance in

1119   terms of decreased accuracy (0-BACK: 92.94 ± 9.34, 1-BACK: 88.06 ± 10.78, 2-BACK: 74.25

1120   ± 19.38, 3-BACK: 51.56 ± 17.37), and d-prime (2-BACK: 0.48 ± 0.14, 3-BACK: 0.32 ± 0.12),

1121   and increased reaction times (0-BACK: 0.51 ± 0.05, 1-BACK: 0.56 ± 0.06, 2-BACK: 0.63 ±

1122   0.07, 3-BACK: 0.66 ± 0.07) as the condition became more difficult (i.e. went from 0-BACK to

1123   3-BACK). Separate linear models were used to compute to check for main effects of drug ($F_{(1,}$

1124 $_{196)}$ = 2.01, p = .16) and an interactive effect of drug and condition on d-prime ($F_{(1, 196)}$ = 0.82,

1125 p = .37). As mentioned above, we also did this for accuracy (main effect of drug: $F_{(1, 392)}$ = 1.07,

1126 p = .30; drug*condition interaction: $F_{(3, 392)}$ = 2.30, p = .08), and reaction times (main effect:

1127 $F_{(1, 347)}$ = 1.31, p = .25; drug*condition interaction: $F_{(1, 347)}$ = 0.99, p = .39).

1128

1129   **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)

1130   Note: values in cells denote *M*, values in parentheses denote *SD*.

1131

1132   We observed no correlation between the certainty in the subjects' belief as to whether

1133   they had received the drug or placebo (r = 0.02, p = .82), or between the reported observed

1134   changes and actually receiving estradiol (r = -0.08, p = .42). This shows that our double-blind

1135   procedure was successful, and that our placebo gel preparation was indistinguishable from

1136   the actual drug.

1137 **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

1138 Note: the split according to both COMT and DAT does not sum to 100 because for a few subjects it was not possible
1139 to determine their polymorphism.
1140

1141

1142 **Reinforcement learning task**

1143 **The role of CYP 19A1, ERα, ERβ, CAG, and GGN**

1144 Because the results for accuracy, staying, and switching behaviour that we report in the

1145 manuscript could also be moderated through other candidate mechanisms, we further

1146 analysed these candidate mechanisms together by providing theoretical motivation for the

1147 analyses. Here, we first briefly outline their importance and then summarize the observed

1148 results.

1149 It is known that androgens are converted to estrogen [77]. This means that the increase

1150 in estrogen levels arises from this conversion process and the administration more directly.

1151 Furthermore, variation in the length of two functional polymorphisms (CAG – polyglutamine,

1152 and GGN – polyglycine) are known to modulate the functioning of the androgen receptor gene

1153 [78]. This is important for two reasons. First, our procedure has previously shown to increase

1154 circulating testosterone levels which could have raised estradiol levels whilst being moderated

1155 by subjects' androgen receptor characteristics [38]. Following from this, previous work has

1156 shown that brain regions important for memory and learning contain androgen receptors [79].

50

1157 Therefore, it could be possible that interindividual differences in both functional polymorphisms

1158 could have moderated our observed results due to interindividual variability. For example,

1159 greater CAG repeat length has previously been associated with lower scores in different

1160 cognitive tests in older men [78], and GGN repeat length has been associated with immediate

1161 and delayed logical memory recall in women [80]. The described results show a correlation

1162 between individual variability in androgen receptor functioning and cognitive performance,

1163 giving rise to the possibility that CAG and GGN polymorphisms being potential candidate

1164 mechanisms moderating the observed effect of estradiol.

1165 Throughout the conversion process from androgens to estrogens, the CYP19A1 gene

1166 encodes instructions for aromatase – the enzyme converting androgens to estrogens [81]. The

1167 single nucleotide polymorphisms (SNPs) associated with the CYP19A1 gene regulate the

1168 metabolism of androgens and mediate brain estrogen activity. Two specific SNPs (rS700518,

1169 rs936306) have been previously shown to have a role in cognitive functioning in humans. For

1170 example, men with the homozygous AA allele have higher estradiol serum levels and greater

1171 bilateral posterior hippocampal gray matter volume compared to men with the homozygous

1172 GG allele [82]. While other work has shown a differential impact of homozygous CC alleles

1173 versus homozygous TT alleles on episodic memory recall in women [83]. Given that our

1174 procedure has previously shown to increase circulating testosterone levels and that

1175 polymorphisms of the CYP19A1 gene are known to have a role in cognitive functioning, we

1176 aimed to exclude the possibility that this may have driven our observed effects, and analysed

1177 both single nucleotide polymorphisms of the CYP19A1 gene.

1178 Once androgens are converted to estrogens, estrogen action is mediated through the

1179 estrogen receptors (ERα, ERβ). Both receptors are widely distributed throughout the brain,

1180 including regions of importance for cognitive functioning and reward processing [84]. So far, it

1181 has been shown that ERα is responsible for most of estrogen-related activation. For example,

1182 it has been shown that SNPs of ERα are related to Alzheimer's disease and are associated

1183 with the likelihood of developing cognitive impairment [85]. We have, therefore, focussed on two

1184 particular SNPs of ERα: rs9340799, rs2234693. In contrast, little is known of a potential impact

1185 of ERβ. As an exploratory measure, we have included repeats of this receptor in our analysis

1186 as well.

1187   Of the described candidates (CAG, GGN, CYP 19A1, ERα, ERβ), no test revealed any

1188 effect of interest. There was no interaction between drug group (i.e. estradiol or placebo) and

1189 either the SNPs of ERα: rs9340799 ($F_{(2, 84)} = 0.66$, $p = .52$), rs2234693 ($F_{(2, 84)} = 0.63$, $p = .53$)

1190 in relation to accuracy. Furthermore, the same was true for the interaction between CAG

1191 repeats and drug group ($F_{(1, 87)} = 0.45$, $p = .51$), GGN repeats and drug group ($F_{(1, 87)} = 1.31$, p

1192 $= .26$), and SNPs of the CYP19A1 gene and drug group (rs700518 $F_{(2, 84)} = 1.84$, $p = .15$,

1193 rs936306 $F_{(2, 84)} = 0.34$, $p = .72$). In a final examination, we also looked at the repeats of ERβ

1194 to determine whether this could have driven any of the observed effects. However, this was

1195 not the case for either recorded variant of ERβ (ERβ1: $F_{(1, 87)} = 0.02$, $p = .89$, ERβ2: $F_{(1, 87)} =$

1196 0.00, $p = .96$).

1197   Identical results were obtained for switching behaviour. While we observed a

1198 statistically significant interaction between estradiol administration and the COMT

1199 polymorphism (see next section), this was not true for any of the other mechanistic

1200 explanations. That is, no model showed an interaction between drug group and either of the

1201 SNPs of ERα: rs9340799 ($F_{(2, 84)} = 2.90$, $p = .06$), rs2234693 ($F_{(2, 84)} = 2.88$, $p = .06$), CAG

1202 repeats ($F_{(1, 87)} = 0.10$, $p = .76$), GGN repeats ($F_{(1, 87)} = 1.32$, $p = .25$), and SNPs of the CYP19A1

1203 gene (rs700518 $F_{(2, 84)} = 1.81$, $p = .17$, rs936306 $F_{(2, 84)} = 1.08$, $p = .35$) in relation to switching

1204 behaviour. As in the case of accuracy, we also looked at the repeats of ERβ. Again, there was

1205 no statistically significant contribution to switching behaviour from this predictor for either

1206 recorded variant of ERβ (ERβ1: $F_{(1, 87)} = 3.05$, $p = .08$; ERβ2: $F_{(1, 87)} = 0.96$, $p = .33$).

1207   We finally repeated the set of analyses for staying behaviour with no effects found of

1208 any of these variables: SNPs of ERα: rs9340799 ($F_{(2, 84)} = 1.69$, $p = .19$), rs2234693 ($F_{(2, 84)} =$

1209 1.79, $p = .17$), CAG repeats ($F_{(1, 87)} = 0.38$, $p = .54$), GGN repeats $F_{(1, 87)} = 0.30$, $p = .59$), SNPs

1210 of the CYP19A1 gene (rs700518 $F_{(2, 84)} = 1.27$, $p = .29$, rs936306 $F_{(2, 84)} = 0.59$, $p = .55$), and

1211 variant of ERβ (ERβ1: $F_{(1, 87)} = 1.35$, $p = .25$; ERβ2: $F_{(1, 87)} = 0.86$, $p = .36$).

1212   In brief, we have shown that the effects related to accuracy, switching, and staying

1213   reported in the manuscript did not depend on (1) the overall androgen receptor functioning

1214   assessed by CAG and GGN repeat polymorphisms, (2) interindividual variability in the

1215   androgen-to-estrogen conversion process via the CYP19A1 gene polymorphism, and (3)

1216   estrogen receptor polymorphisms (ERα) or repeats (ERβ).

1217

1218   The effect of estradiol administration on choice behaviour is moderated by

1219   polymorphisms of both COMT and DAT1

1220   We directly tested whether the effect of estradiol on choice behaviour is moderated by

1221   polymorphisms of dopamine-related genes (e.g. COMT, DAT1) by using generalized linear

1222   mixed models. We specifically tested whether the interaction between drug, polymorphism

1223   (COMT or DAT1), and trial are a significant predictor of the options they would choose.

1224   Based on the inverted U-shape dopamine hypothesis [46], we predicted that estradiol

1225   administration would upregulate reward sensitivity in subjects with low prefrontal dopaminergic

1226   activity (i.e. Val/Val) but would not, or would even impair it, in those with high prefrontal

1227   dopaminergic activity (i.e. Met/Met). The model predictions support this hypothesis as it

1228   predicted that subjects with a Met/Val ($\beta = 0.20 \pm 0.04$, *95% CI* [0.11, 0.28], $z = 4.56$, $p < .001$)

1229   and Val/Val genotype ($\beta = 0.37 \pm 0.06$, *95% CI* [0.26, 0.48], $z = 6.99$, $p < .001$) were more

1230   likely to select option A as trials progressed when they received estradiol (Fig. S2, S4). Option

1231   A was the more rewarding option throughout the task (percent trials rewarded: $M_{optionA} =$

1232   53.70%, $M_{optionB} = 42.91\%$).

1233   Based on the prediction that estradiol indirectly increases striatal dopamine levels,

1234   leading to higher reward prediction errors, we expected that subjects with the 9/10 genotype

1235   (i.e. high striatal dopamine) would select the higher value option more often, while this would

1236   be less often true for subjects with the 10/10 genotype (i.e. low striatal dopamine). This

1237   prediction was supported by the model showing that subjects with the 10/10 genotype with

1238   placebo (β = -0.12 ± 0.04, *95% CI* [-0.04, -0.20], *z* = -3.03, *p* < .01) were the most likely to

1239   select the lower valued option A throughout task progression, while estradiol administration

1240   dampened this slope in subjects with the same 10/10 genotype (Fig. S4). Results from both

1241   generalized linear mixed effects models showed that once individual variation was considered,

1242   the effect of estradiol administration on choice behaviour across trials was moderated by

1243   striatal (DAT1) and prefrontal (COMT) polymorphisms.
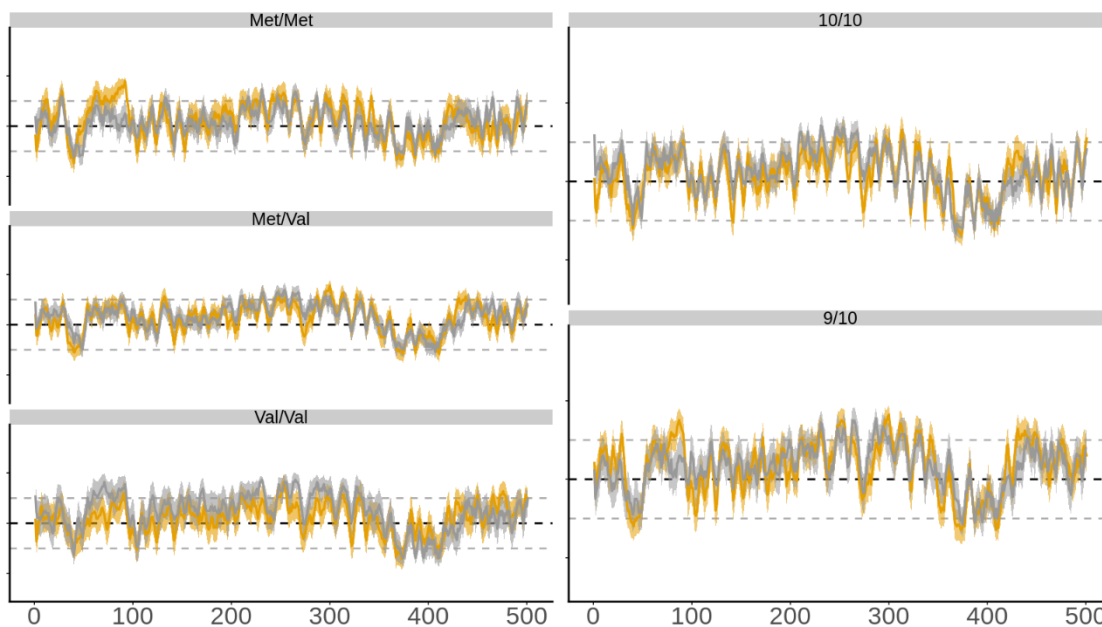
1244

1245   **Figure S2.** Relative choice probability for choosing option A (top of y-axis) vs. choosing stimulus 1 (bottom of y-
1246   axis) for the placebo (gray) and estradiol (orange) group split according to both polymorphisms assessed in the
1247   main text: COMT (left panel), DAT (right panel) across trials (1-500). Thick lines represent trial means, shaded
1248   areas denote standard error of the means. The blue line in the background denotes the empirical relative reward
1249   probability which was computed from the probability of stimulus two being rewarding (top of y-axis) - stimulus one
1250   being rewarding (bottom of y-axis). Gray dotted lines represent where subjects were on average 25% more likely
1251   to select option A (upper line) or stimulus 1 (lower line). All time-series traces are smoothed with a 5-trial moving
1252   average for visual purposes.

1253

1254   Figure S2 reveals a differential effect of estradiol administration on choice behaviour

1255   that depends on polymorphisms of both COMT and DAT. In the case of the COMT

1256   polymorphism this is most clearly visible in the lower left panel. The panel shows that placebo

1257   Val/Val subjects exhibited a clear tendency towards option A until trial ~370. After this, they

1258   did not reverse back towards choosing it more often despite option A being more rewarding

1259   from trial ~420 onwards. This is in contrast with results for subjects with other polymorphisms

1260   of COMT and results when subjects were split according to the DAT1 polymorphism. Estradiol

54

1261   Met/Met subjects exhibited choice behaviour more aligned with the reward probability

1262   distribution in the beginning at trial ~80 compared to subjects from the placebo group with the

1263   same polymorphism. When we then split subjects according DAT1 polymorphism, the

1264   estradiol 9/10 subjects can similarly be seen following the reward probability distribution more

1265   closely compared to the placebo 9/10.

1266   Because of the results above and because estradiol administration likely results in

1267   increased prefrontal dopamine levels through downregulating COMT enzyme activity [56] we

1268   predicted that the interaction between estradiol administration and COMT polymorphism

1269   would be predictive of switching behaviour [24]. As a measure of switching, we assessed the

1270   number of times the option chosen on trial *t* was different from the one chosen at trial *n + 1*

1271   (i.e. a switch), irrespective of the choice outcome on trial *t*. Estradiol administration did not

1272   significantly influence switch decisions ($M = 162.12 \pm 56.31$) compared to placebo ($M = 168.82$

1273 ± 68.13). However, we observed a significant interaction of estradiol administration by COMT

1274   genotype ($F_{(2, 80)} = 3.22$, $p = .05$, $\Omega^2 = 0.04$, Fig. S3). The interaction showed that subjects with

1275   placebo and a Val/Val genotype (i.e. low prefrontal dopamine availability) switched less often

1276   ($\beta = -84.07 \pm 33.69$, $p = .02$) compared to all other groups. As predicted by the inverted U-

1277   shaped relationship between prefrontal dopamine levels and behaviour, Val/Val placebo

1278   subjects (Val/Val: $M = 132.33 \pm 61.40$) switched less compared to Met/Met placebo subjects

1279   (i.e. associated with high prefrontal dopamine availability; Met/Met: $M = 204.27 \pm 53.52$, $t_{(15.10)}$

1280 $= 2.91$, *95% CI* [19.25, 124.54], $p = .01$, $d = 1.46$). For the estradiol group, this difference was

1281   not present (Val/Val: $M = 151.09 \pm 70.85$; Met/Met: $M = 178.5 \pm 55.34$; $t_{(18.96)} = 1.03$, *95% CI* [-

1282   28.28, 83.10], $p = .32$, $d = 0.44$). In other words, estradiol administration attenuated naturally

1283   occurring differences in switching behaviour found in subjects with the Met/Met and Val/Val

1284   genotypes that are associated with high and low prefrontal dopamine levels, respectively.
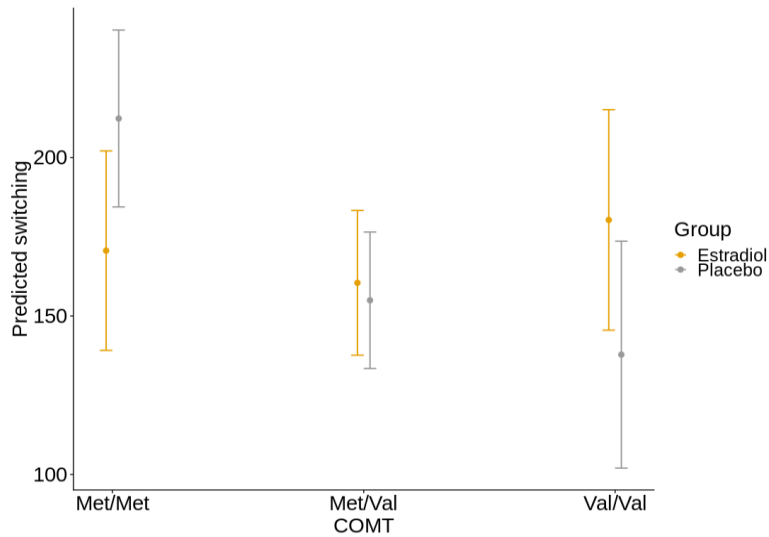
Figure S3. General linear model prediction for switching behaviour (i.e. a change in chosen stimulus on trial $t + 1$ from trial t, independent of choice outcome on trial t). Estradiol administration dampened naturally occurring differences in switching behaviour when subjects were split according to 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. Error bars represent SEM.
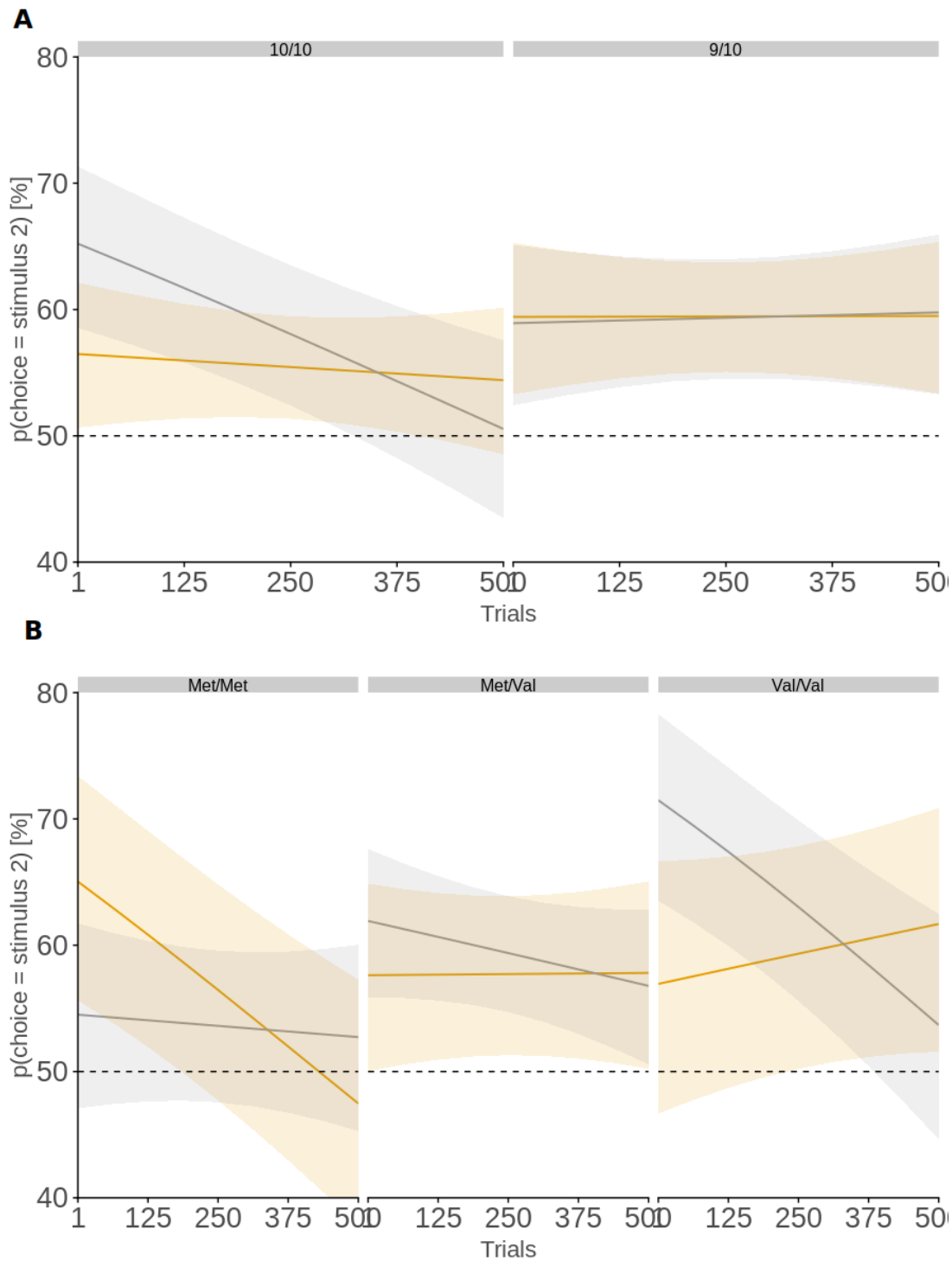
Figure S4. Predictions from winning models of the generalized linear mixed effects models for **A**) the interaction between drug, DAT, and trial on choice, and **B**) the interaction between drug, COMT, and trial on choice.

57

Formal model comparison

We estimated parameters by fitting several Q-learning models. The best model (model 6, leave one out information criterion (LOOIC) = 58888, Fig. S5A) included separate learning rates for positive and negative prediction errors, 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. S5B) and performed equally well for both groups ($M_{Estradiol}$ = 66.26 % ±10.77, $M_{Placebo}$ = 64.90 % ±11.85; $t_{(97.115)}$ = 0.76, 95% CI [-0.03, 0.06], $p = .45$).
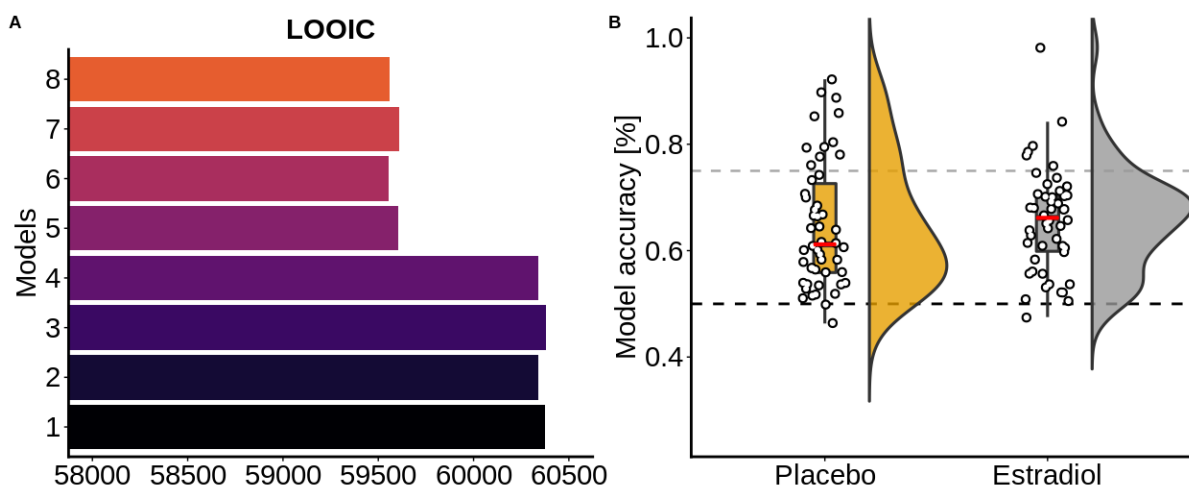
Figure S5. A) Leave one out information criterion (LOOIC) value for all employed models. Lower LOOIC indicates better model fit – model two was selected as the best model. **B)** The overall model accuracy collapsed over time obtained from the posterior predictive density shown for both drug groups separately.

In addition to computing the leave-one-out information criterion to perform model comparison [73] we similarly computed the exceedance probability of the winning model using the VBA toolbox [75]. This value showed a strong preference for the winning model P(model two) = 98%. Furthermore, we computed protected exceedance probability [74] 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

1322 rates to be present if we were to select the learning rates from models that best fit individual

1323 subjects.

Validating model

1325 We further tested the model validity and predictions by computing posterior predictive

1326 densities, i.e. what predictions does the model make on a trial-by-trial basis for subjects with

1327 the parameters such as those that were extracted from our subjects. Posterior predictive

1328 densities showed no difference in a fit between both the estradiol and placebo group and

1329 approximated the empirical reward probability distribution (Fig. S6A). To quantify this, we then

1330 compared model predictions from posterior predictive densities with actual subject behaviour

1331 to assess model accuracy collapsed across time (Fig. S6B) showing it performed above

1332 chance and equally well for both groups. We further compared accuracy on each trial across

1333 subjects to ensure that there were no unexpected drops in accuracy. This did not happen as

1334 the model (Fig. S6B) had no discernible drops in performance. We also performed parameter

1335 recovery on the winning model where we used the maximum a posteriori estimates reported

1336 in the manuscript, generated data for synthetic subjects, performed parameter estimation on

1337 the synthetic data, and correlated the newly obtained synthetic parameters with original

1338 parameters for each subject. This procedure showed that both original and recovered learning

1339 rates correlated with one another (negative: $r = 0.33$, $p < .001$, positive: r = -0.34, p < .001).

1340
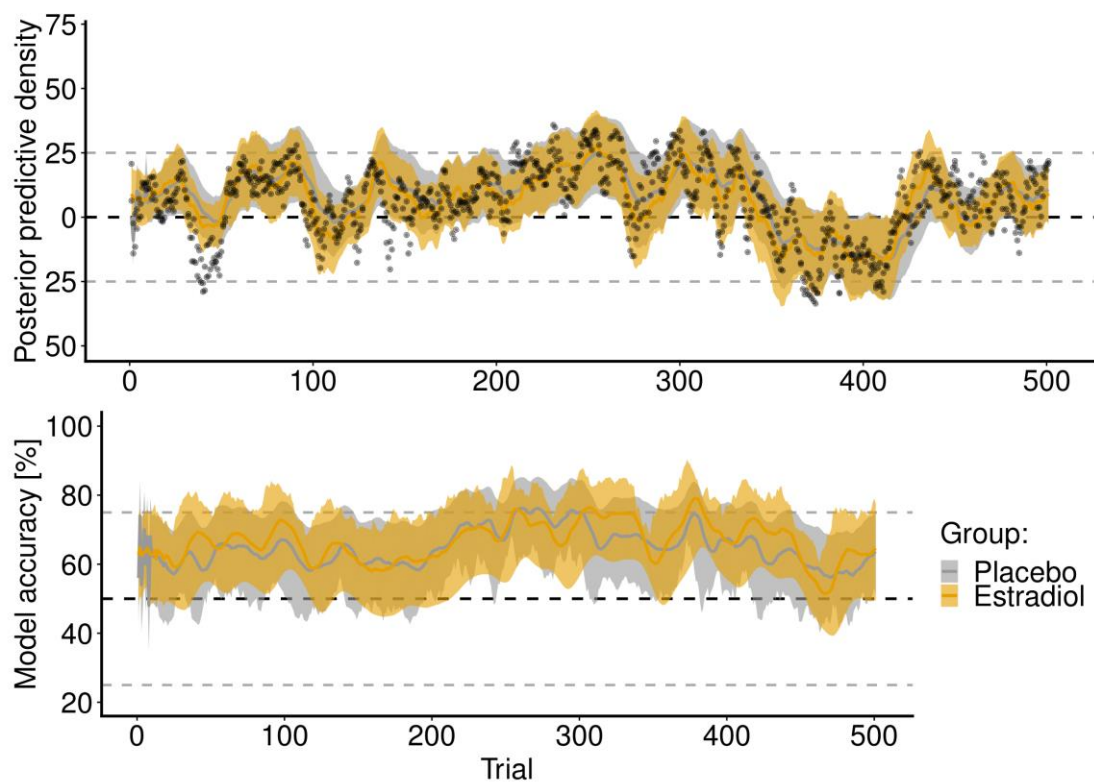
1341

Figure S6. A) Posterior predictive density computed for both drug groups with overlaid average responses for both
1343 drug groups across trials **B**) Accuracy for both drug groups obtained from the posterior predictive density for both
1344 drug groups separately.

1345

1346

1347

1348

1349

1350

1351

1352

1353

1354

1355

1356

1357

1358

1359

1360

1361

1362

1363

1364

1365

1366