Supporting Information

Sex differences in economic decision-making:
exogenous estradiol has opposing effects on fairness framing in women and men

Marie Coenjaerts, Frederike Pape, Virginia Santoso, Franziska Grau, Birgit Stoffel-Wagner, Alexandra Philipsen, Johannes Schultz, René Hurlemann, and Dirk Scheele*

* *Corresponding authors:*
Marie Coenjaerts
Division of Medical Psychology
University of Bonn
Venusberg-Campus 1
53105 Bonn, Germany
Phone: +49 (0)228 287 19704
Fax: +49 (0)228 287 19125
E-mail: Marie.coenjaerts@ukbonn.de

Dirk Scheele, Ph.D.
Department of Psychiatry
University of Oldenburg
Hermann-Ehlers-Str. 7
26160 Bad Zwischenahn, Germany
+49 (0)441 9615 1504
E-mail: Dirk-Scheele@gmx.de

This PDF file includes:

Supplementary Results

Demographic and psychometric baseline characteristics

Baseline demographics and psychometric assessments of the participants are displayed in Table S1. There were no significant differences between treatment groups across and within sexes (all *p*s > 0.05). Across treatments, there was a trend-to-significant effect that the male participants were older than the female subjects ($t_{(209)}$ = -1.97, p = 0.05, d = 0.27). Women reported significantly higher social anxiety (Liebowitz scale, $t_{(210)}$ = 3.77, p < 0.01, d = 0.52) and increased trait anxiety (STAI Trait, $t_{(206.05)}$ = 1.97, p = 0.05, d = 0.27). Childhood maltreatment, alexithymia, depressive symptoms and autistic-like traits were not significantly different between the sexes.

Stereotypical beliefs

A sample of 133 healthy subjects (85 females; mean age ± SD = 25.41 ± 5.44 years), who participated in another study with estradiol treatment, were asked about their specific knowledge of estradiol. In this sample, 22.6% of the participants stated concrete ideas. Half of them addressed estradiol's function as a female sex hormone and its importance for the menstrual and reproductive cycle. The other half mainly linked estradiol to emotionality and empathy.

In addition, to assess stereotypical beliefs, participants were asked to rate how strong they associate a list of terms with estradiol and testosterone (cf. Table S3). Terms were taken from the Positive and Negative Affect Schedule (PANAS) questionnaire (Watson et al., 1988) and complemented by seven adjectives (weak, sympathetic, loving, empathetic, friendly, emotional and disciplined. Compared to testosterone, estradiol was rated as weak, but also emotional,

empathetic, loving and friendly (all *p*s < 0.001). In contrast, testosterone was significantly more associated with the terms active, strong, hostile and determined (all *p*s < 0.001). Thus, the mere belief of receiving estradiol might have guided participants to accept more unfair offers due to the fact that they believe estradiol induces an increased forgiving and tolerant behavior.

Furthermore, post-hoc t-tests for the acceptance rates in the unframed UG showed that in the placebo group, participants who believed that they had received estradiol accepted significantly more 2€ offers ($t_{(39.05)}$ = 2.75, p_{cor} < 0.05, *d* = 0.691) than those who believed that they had received a placebo. By contrast, the believed treatment was not significantly related to the acceptance rate of different offer sizes in the estradiol group (all *p*s > 0.05).

Moreover, we computed post-hoc t-tests to disentangle the significant three-way-interaction of sex, offer size and believed treatment ($F_{(2.86, 568.38)}$ = 2.76, *p* < 0.05, η_p^2 = 0.014) for the acceptance rate in the computer UG. Men who believed that they had received estradiol treatment showed increased acceptance rates for the 0€ ($t_{(40.09)}$ = 2.66, p_{cor} < 0.05, *d* = 0.69) compared to men who believed that they had received a placebo treatment. By contrast, the believed treatment did not alter the acceptance rates of the different offer sizes in women (all *p*s > 0.05).

Further hormonal assessments

Post-hoc t-tests showed that the estradiol level significantly increased after the estradiol treatment in both women ($t_{(53)}$ = -13.82, p_{cor} < 0.001, *d* = -2.633) and men ($t_{(52)}$ = -11.00,

$p_{cor} < 0.001$, $d = -2.109$), while there was no significant change in the placebo group (all *p*s > 0.05).

We also tested whether estradiol, testosterone and progesterone levels moderated treatment effects in the unframed Ultimatum Game (UG), computer UG and the delayed discounting task. Separately including these baseline concentrations as covariates in the analyses of variance (ANOVAs) did not yield significant main or interaction effects of treatment. Furthermore, we assessed the influence of the endogenous estradiol, testosterone and progesterone baseline levels on the acceptance rates of fair-framed and unfair-framed offers. Testosterone and progesterone did not correlate with either acceptance rates (all *p*s > 0.05). However, there was a negative correlation between the endogenous estradiol levels and the acceptance rate of unfair-framed offers under placebo (women: $r_{(50)} = -0.31$, $p = 0.03$; men: $r_{(47)} = -0.29$, $p = 0.045$). Correlation coefficients are displayed in Table S4 and Table S5.

Sex of the proposer

An ANOVA with the sex of the proposer and the framing as within-subject factors, the participants' sex and treatment as between-subject variables, and the average acceptance rate as dependent variable yielded a significant interaction between the sex of the proposer and the sex of the participants ($F_{(1,206)} = 4.02$, $p = 0.046$, $\eta_p^2 = 0.02$). Male participants accepted more offers than female subjects and this effect was more pronounced for offers made by a male proposer. However, there were no significant interaction effects between the sex of the proposer and the framing or treatment (all *p*s > 0.05).

Reaction times in the unframed UG, computer UG and delayed discounting

There were no significant treatment effects on the reaction times in the unframed UG, the computer UG or the delayed discounting task. However, a mixed-design ANOVA with the

response times in the unframed UG as dependent variable, treatment and sex as between-subject factors and offer size as within-subject variable yielded significant main effects of offer size ($F_{(4.19,866.87)}$ = 22.12, p < 0.01, η_p^2 = 0.10) and sex ($F_{(1,207)}$ = 4.78, p = 0.03, η_p^2 = 0.02). Response times were higher for medium size offers than for very low or very high offers and women showed significantly faster reaction times (mean ± SD = 1.29 ± 0.34 sec) compared to men (1.41 ± 0.49 sec). Furthermore, a mixed-design ANOVA with the response times in the computer UG as dependent variable, treatment and sex as between-subject factors and offer size as within-subject variable yielded a significant main effect of offer size ($F_{(4.46,919.38)}$ = 16.34, p < 0.01, η_p^2 = 0.07) and a significant interaction between the sex of the participants and the offer sizes ($F_{(4.46,\ 919.38)}$ = 2.86, p = 0.02, η_p^2 = 0.014). Women demonstrated faster reaction times than men if the offer was low. However, with increasing offer sizes the male subsample showed faster reaction times than the female subsample.

Minimum acceptable offer and fairness ratings

After completion of all tasks, participants were asked to state their minimum acceptable offer and rate the fairness of different offers. An ANOVA with the participants' sex and treatment as between-subject variables, and the minimum acceptable offer as dependent variable yielded a trend-to-significant interaction between the treatment and the sex of the participants ($F_{(1,206)}$ = 3.21, p = 0.075, η_p^2 = 0.02). In line with the results observed for the framed UG, female participants in the estradiol group expected significantly more money to be offered from a proposer than women in the placebo group ($t_{(106)}$ = -2.28, p = 0.025, d = -0.44). This effect was not found in the male subsample ($t_{(100)}$ = 0.45, p = 0.65, d = 0.09).

The perception of fairness increased with the offer sizes ($F_{(5.41,1114.39)}$ = 1210.78, p < 0.001, η_p^2 = 0.86). Furthermore, there was a trend-to-significant interaction of the treatment and sex

($F_{(1,206)}$ = 3.10, p = 0.08, η_p^2 = 0.02). In the female subsample, the fairness ratings were modulated by the treatment ($F_{(1,106)}$ = 3.45, p = 0.07, η_p^2 = 0.03), such that lower fairness ratings were evident in the estradiol subsample than in the placebo subsample. This trend was not evident in the male subsamples ($F_{(100)}$ = 0.39, p = 0.53, η_p^2 = 0.004).

Body mass index and hormonal levels

The body mass index (BMI) of the participants was not correlated with the baseline estrogen levels ($r_{(207)}$ = 0.02, p = 0.73) irrespective of the sex. Furthermore, it affected post-treatment estrogen levels in neither the male ($r_{(51)}$ = -0.01, p = 0.93) nor in the female treatment group ($r_{(51)}$ = 0.04, p = 0.77). Thus, the estradiol treatment did not produce different peripheral levels depending on the BMI.

Delayed discounting task

An additional analysis with the discount rate k as dependent variable and sex and treatment as between-subject factors yielded no significant main or interaction effects (all ps > 0.05).

Valuation task

We applied a mixed-design ANOVA with the between-subject factors “treatment” (estradiol, placebo) and “sex” (women, men), the within-subject variables “time points of delivery” (today, in 2 weeks, in 4 weeks, in 6 weeks) and “magnitude” (low, medium, high), and the attractiveness ratings from the valuation task as dependent variable. All participants preferred sooner and larger rewards and the treatment did not affect these ratings, thus providing further evidence that estradiol did not alter the valuation of rewards options per se, but rather modulated framing sensitivity. Furthermore, we found a significant main effect of sex ($F_{(1,206)}$ =

5.47, p = 0.02, η_p^2 = 0.03) and a significant three-way interaction of the magnitude, the time points of delivery and the sex ($F_{(5.53,1138.56)}$ = 3.24, p < 0.01, η_p^2 = 0.02). Women rated the attractiveness of most options higher than men, but this effect was reversed for immediate options with high reward magnitude.

Pharmacokinetic pre-study

We conducted a pre-study involving 10 healthy participants (5 women; mean age ± SD = 24.10 ± 4.07 years), to examine the pharmacokinetics of estradiol gel (Estramon 2 mg estradiol, Hexal AG, Holzkirchen, Germany) administration. Blood samples were taken prior to estradiol administration (i.e. baseline) and in 1-hour intervals after drug application up to 5 hours post administration. An additional blood sample was taken the next day (after 18 hours). Serum estradiol levels peaked 3-4 hours after gel administration (cf. Figure S1), but a significant increase relative to baseline was already evident after 2 hours ($t_{(9)}$ = 2.44, p = 0.04, d = 1.10). Estradiol levels remained significantly elevated throughout the last measurement. A previous study tested the topical administration of a different drug (Divigel, Orion Pharma AG, Zug, Switzerland) containing 2 mg estradiol and found significantly increased estradiol serum concentrations as soon as 1 hour after administration and maximum average levels after 2 hours (Eisenegger et al., 2013). Clearly, serum estradiol levels were significantly elevated at the start of the experiments in the present study (2.5 hours after gel administration).

Power analysis

The present study was part of a larger project involving a second substance, administered after the experiments presented in this study (2x2 factorial design). For this larger project, we used G*Power 3 (Faul et al., 2007) to conduct an a-priori power analysis based on the effect size

obtained in a previous dose-response study ((Spengler et al., 2017); d = 0.78) to determine the number of participants to include. To detect an effect of this size with α = 0.05 and power = 0.75, the analysis revealed that we needed to test at least 24 participants in each of the four cells of the design (resulting from the combination of the two substance vs. placebo combinations). As sex was another factor to take into account, we aimed at including at least 100 women and 100 men, with 50 per sex to receive estradiol gel treatment and 50 placebo gel treatment.

For the purpose of the present study, we conducted an additional sensitivity power analysis excluding the second substance administration, which revealed that given our sample size (n = 212, 108 females), we have 80% power to detect an effect not smaller than Cohen's d = 0.3867 at a p-value of 0.05. This suggests that the present experiment involving estradiol is sensitive enough to capture a medium effect if present.

Missing values

Two blood samples for measuring hormonal levels at baseline and two samples after the tasks were lost because of problems in sample assessment or analysis. Behavioral data from two participants in the framed and computer version of the ultimatum game, one participant in the unframed ultimatum game, and one participant in the delayed discounting task were not recorded due to technical issues. The believed treatment was not available from three subjects.

Supplementary Materials and Methods

Participants

The present study was part of a larger project registered in the ClinicalTrials.gov database (Identifier: NCT04330677) provided by the US National Institutes of Health. Participants were recruited from the local population by means of online advertisement and public postings. After completion of the study, participants received monetary compensation. Autistic-like traits were measured with the Autism Spectrum Quotient questionnaire (AQ) (Baron-Cohen et al., 2001), childhood trauma was measured with the Childhood Trauma Questionnaire (CTQ) (Bernstein et al., 1994), depressive and anxiety symptoms were assessed by Beck's Depression Scale (BDI, Version II) (Beck et al., 1996) and the Spielberger Trait Anxiety Inventory (STAI) (Spielberger et al., 1970), and empathy was assessed by the Saarbrücker Persönlichkeitsfragebogen, a German version of the Interpersonality Reactivity Index (IRI) (Davis, 1983; Paulus, 2009) (cf. Table S1). We screened a total of 230 subjects. 18 subjects had to be excluded because they were not fluent in German (n=6), technical malfunctions during data collection (n=5), physical illness (n=3), drug abuse (n=2), and discontinuation of study participation (n=2). After completing the UG, all subjects were asked to threshold the smallest amount of money they regarded as acceptable. Furthermore, they were instructed to rate the fairness of all offers on a scale of 1 (minimum)–7 (maximum) and to make one offer as proposer.

Valuation Task

After the delayed discounting task, participants completed the valuation task, in which the participants rated the attractiveness of 12 single options that each provided a specified monetary amount at a specified time point. The options were randomly chosen from the 36 trials of the delayed discounting task and consisted of four time points of delivery (today, in 2 weeks, in 4 weeks, in 6 weeks) crossed with three levels of reward magnitude (low: approximately EUR 30; medium: approximately EUR 45; and high: approximately EUR 60; actual values varied slightly from these approximate numbers as they were dependent on the

values presented in the delayed discounting task). The self-assessment manikin (SAM) (Lang, 2005) was presented below each option and participants rated the attractiveness on a scale of 1 (minimum) – 9 (maximum).

Statistical Analysis

The discount rate k for each participant was quantified using a standard one-parameter model of hyperbolic discounting (Mazur, 1987), captured by the following term:

$$Subjective\ value = \frac{Reward\ magnitude}{1+k*Delay},$$

where *Delay* is the time of delivery (in weeks) and *k* is the parameter that represents the participant's discount rate. Larger k values indicate steeper discounting of delayed rewards, while 0 indicates no discounting at all. We used the Matlab (Matlab R2017b, The MathWorks Inc., Natick, MA) function *fminbnd* to estimate the k value for each participant which produced the global minimum in the negative log-likelihood of individual choice probability.

Supplementary Figures

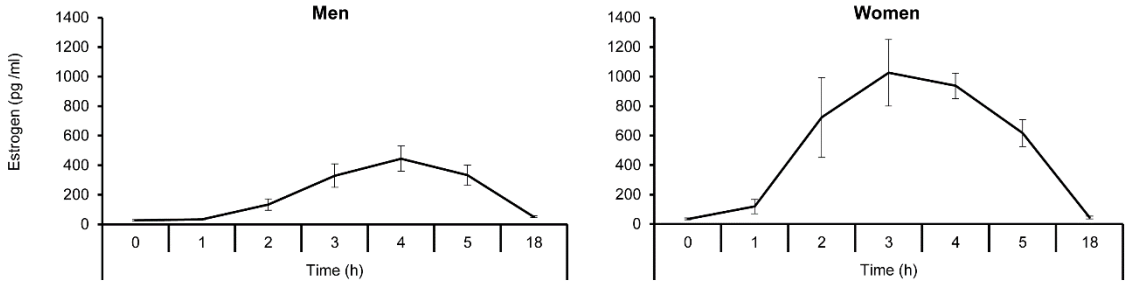

Figure S1. Estradiol serum concentrations in the pharmacokinetic pre-study following topical administration of Estramon 2 mg estradiol gel. Serum estradiol levels were significantly different from baseline after 2 hours and reached an average maximum 3-4 hours after gel administration. Error bars indicate the standard error of the mean.

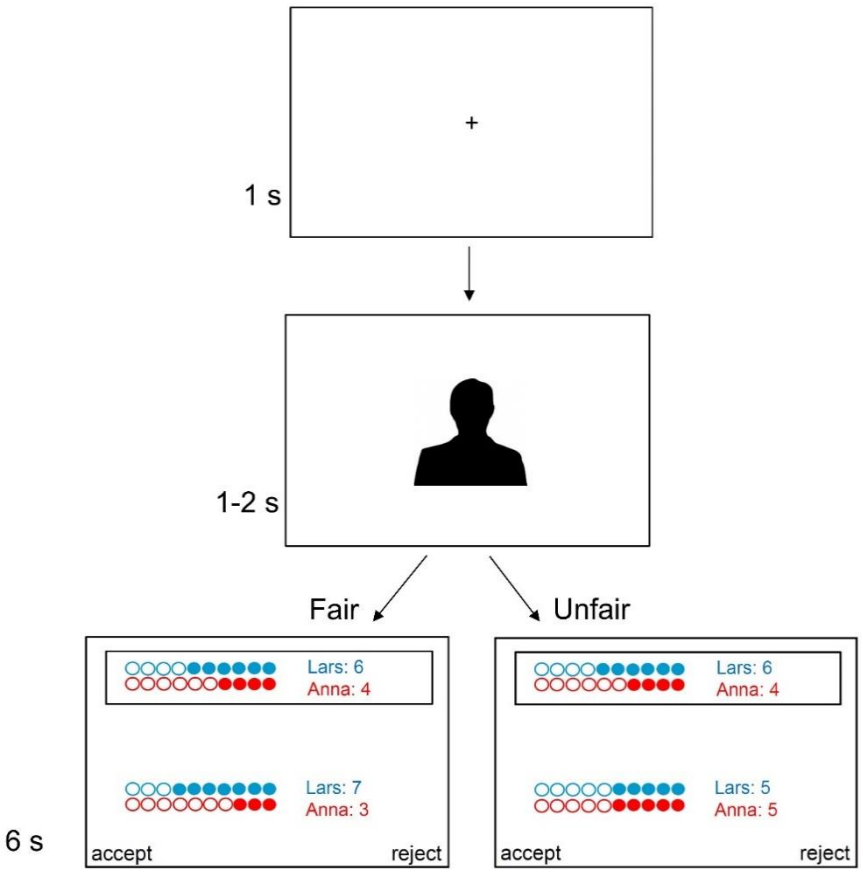

Figure S2. Design of the framed UG. First a fixation cross appeared on the screen for 1 second. Subsequently, a picture of the proposer was presented for 2 seconds, followed by the offer made to the subject. The offer was either framed as fair (left side) or as unfair (right side). The black rectangle indicates the proposer's choice which the participant can either accept or reject.

Supplementary Tables

Table S1. Demographic and psychometric baseline characteristics.

	Females (*n* = 107)		Males (*n* = 104)	
	Placebo (*n* = 54) Mean (± SD)	Estradiol (*n* = 53) Mean (± SD)	Placebo (*n* = 51) Mean (± SD)	Estradiol (*n* = 53) Mean (± SD)
Age (years)	23.00 (3.29)	23 11 (3.91)	24.65 (4.33)	23.51 (3.31)
Childhood maltreatment (CTQ[a´])	31.70 (7.76)	31.32 (5.94)	30.73 (6.53)	32.60 (7.26)
Depressive symptoms (BDI[b])	4.26 (4.94)	4.09 (4.18)	3.37 (4.15)	3.49 (3.87)
Autistic-like traits (AQ[c])	15.19 (5.85)	15.54 (6.58)	14.69 (4.34)	16.74 (6.12)
Alexithymia (TAS[d])	45.35 (9.87)	43.94 (9.73)	42.04 (9.92)	44.60 (11.62)
Trait anxiety (STAI[e])	36.41 (8.40)	37.48 (8.68)	34.26 (7.84)	35.38 (6.40)
Social anxiety (Liebowitz Total[f])	28.17 (16.53)	34.04 (17.53)	20.55 (18.20)	23.68 (16.53)

Notes. Childhood maltreatment was assessed by the [a] CTQ (Childhood Trauma Questionnaire, Bernstein et al., 2003). Subjects rated their depressive symptoms with the [b] BDI (Becks Depression Inventory, Beck et al., 1961). Autistic-like traits were measured with the [c] AQ (Autism Spectrum Quotient, Baron-Cohen et al., 2006). Alexithymia was assessed with the [d] TAS (Toronto Alexithymia Scale, Taylor et al., 1985). The [e] STAI-Trait (State-Trait-Anxiety Inventory, Spielberger, 1983) was used to assess trait anxiety and the [f] Liebowitz questionnaire was used to measure social anxiety.

Table S2. Estradiol, progesterone and testosterone baseline and post treatment concentrations.

	Females			Males		
	PLC Mean (n, ± SD)	EST Mean (n, ± SD)	t-values	PLC Mean (n, ± SD)	EST Mean (n, ± SD)	t-values
Estradiol pre (pg/ml)	30.99 (54, 17.13)	27.89 (54, 16.00)	0.97	25.33 (49, 10.05)	24.15 (53, 11.96)	0.54
Estradiol post (pg/ml)	30.94 (53, 14.57)	909.35 (54, 473.09)	-13.69***	25.49 (50, 9.48)	303.09 (53, 186.69)	-10.81***
Progesterone pre (ng/ml)	0.16 (54, 0.14)	0.20 (54, 0.27)	-1.06	1.75 (49, 3.23)	1.26 (53, 0.44)	1.10
Progesterone post (ng/ml)	0.11 (52, 0.13)	0.11 (54, 0.17)	0.05	1.59 (50, 3.40)	1.12 (53, 0.41)	1.01
Testosterone pre (ng/ml)	0.24 (54, 0.13)	0.23 (54, 0.10)	0.41	4.04 (49, 1.58)	3.95 (53, 1.19)	0.34
Testosterone post (ng/ml)	0.20 (52, 0.12)	0.19 (54, 0.08)	0.85	3.88 (50, 1.50)	3.98 (53, 1.11)	-0.39

Notes. EST, estradiol treatment; PLC, placebo treatment. ***p < 0.001.

Table S3. Rating of associations between adjectives and estradiol and testosterone.

	Estradiol (n = 133) Mean (± SD)	Testosterone (n = 133) Mean (± SD)	t-value (df = 132)
active	2.26 (1.12)	3.8 (0.99)	-14.94***
distressed	1.88 (1.19)	1.33 (0.65)	5,44***
interested	2.37 (1.16)	2.41 (1.14)	-0.47
enthusiastic	2.41 (1.21)	3.00 (1.17)	-5.19***
upset	1.61 (1.01)	2.74 (1.17)	-9.92***
strong	2.10 (1.19)	3.99(1.10)	-15.44***
guilty	1.30 (0.65)	1.32 (0.62)	-0.36
scared	1.32 (0.68)	1.19 (0.48)	2.21*
hostile	1.27 (0.65)	2.50 (1.22)	-11.45***
inspired	2.23 (1.19)	3.32 (1.20)	-10.14***
proud	1.86 (1.06)	3.29 (1.33)	-11.99***
irritable	1.97 (1.22)	2.89 (1.23)	-7.18***
excited	2.17 (1.19)	2.57 (1.23)	-3.84***
ashamed	1.26 (0.58)	1.30 (0.64)	-0.57
alert	2.11 (1.20)	2.99 (1.30)	-7.80***
nervous	1.62 (0.92)	1.70 (0.95)	-0.79
determined	1.96 (1.18)	3.25 (1.43)	-11.16***
attentive	2.23 (1.19)	2.59 (1.31)	-3.58***
jittery	1.80 (1.06)	1.66 (0.90)	1.34
afraid	1.51 (0.95)	1.17 (0.47)	3.79***
weak	1.60 (0.86)	1.14 (0.44)	5.57***
sympathetic	2.88 (1.42)	1.56 (0.81)	10.30***
loving	2.81 (1.45)	1.65 (0.90)	9.38***
empathetic	2.80 (1.42)	1.68 (0.91)	9.07***
friendly	2.65 (1.32)	1.83 (1.02)	7.24***
emotional	3.22 (1.42)	2.38 (1.34)	5.43***
disciplined	1.95 (1.02)	2.06 (1.12)	-1.06

Notes. *p < 0.05, ***p < 0.001. Participants rated on a 5-point Likert scale (1 = very little or not at all; 5 = extremely) whether they associate adjectives with estradiol and testosterone.

Table S4. Correlations of the acceptance rates of fair-framed and unfair-framed offers with endogenous baseline levels in the female subsample.

	Estradiol (pg/ml)		Progesterone (ng/ml)		Testosterone (ng/ml)	
	Placebo	Estradiol	Placebo	Estradiol	Placebo	Estradiol
Fair-Framed Offers	-0.06	-0.25	-0.26	0.096	0.14	0.01
Unfair-Framed Offers	-0.31*	-0.24	-0.21	0.10	-0.12	0.01

Notes. (Placebo, n = 52), (Estradiol, n = 54), *p < 0.05

Table S5. Correlations of the acceptance rates of fair-framed and unfair-framed offers with endogenous baseline levels in the male subsample.

	Estradiol (pg/ml)		Progesterone (ng/ml)		Testosterone (ng/ml)	
	Placebo	Estradiol	Placebo	Estradiol	Placebo	Estradiol
Fair-Framed Offers	-0.11	0.10	-0.24	0.04	0.09	0.03
Unfair-Framed Offers	-0.29*	-0.23	-0.05	0.14	-0.19	-0.22

Notes. (Placebo, n = 49), (Estradiol, n = 53), *p < 0.05

SI References

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., Clubley, E., 2001. The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. J. Autism Dev. Disord. 31, 5-17.

Beck, A., Steer, R., Brown, G., 1996. Manual for the Beck depression inventory-II (BDI-II).

Bernstein, D.P., Fink, L., Handelsman, L., Foote, J., Lovejoy, M., Wenzel, K., Sapareto, E., Ruggiero, J., 1994. Initial reliability and validity of a new retrospective measure of child abuse and neglect. Am. J. Psychiatry 151, 1132-1136.

Davis, M.H., 1983. Measuring individual differences in empathy: Evidence for a multidimensional approach. J. Pers. Soc. Psychol. 44, 113.

Eisenegger, C., von Eckardstein, A., Fehr, E., von Eckardstein, S., 2013. Pharmacokinetics of testosterone and estradiol gel preparations in healthy young men. Psychoneuroendocrinology 38, 171-178.

Faul, F., Erdfelder, E., Lang, A.G., Buchner, A., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39, 175-191.

Lang, P.J., 2005. International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical report.

Mazur, J.E., 1987. An adjusting procedure for studying delayed reinforcement. Lawrence Erlbaum Associates, Inc. .

Paulus, C., 2009. Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index.

Spengler, F.B., Schultz, J., Scheele, D., Essel, M., Maier, W., Heinrichs, M., Hurlemann, R., 2017. Kinetics and Dose Dependency of Intranasal Oxytocin Effects on Amygdala Reactivity. Biol. Psychiatry 82, 885-894.

Spielberger, C., Gorsuch, R., Lushene, R., 1970. STAI Manual for the State-Trait Anxiety Inventory (Palo Alto, CA, Consulting Psychologists Press). Inc.

Watson, D., Clark, L.A., Tellegen, A., 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. J. Pers. Soc. Psychol. 54, 1063-1070.