Supplemental Material for

Lasting effects of oral contraceptives on fear responses to a safe context.

Lisa-Marie Davignon[ab], Alexandra Brouillard[ab], & Marie-France Marin[ab]

[a]Research Center of the Institut Universitaire en Santé Mentale de Montréal, 7331 Hochelaga Street, Montreal, Canada, H1N 3J4;

[b]Department of Psychology, Université du Québec à Montréal, 100 Sherbrooke Street W, Montreal, Canada, H2X 2P3.

*Marie-France Marin.
Email: marin.marie-france@uqam.ca
Address: 100, Sherbrooke W Street
Montreal, H2X 3P2

Running title: Oral contraceptives and fear responses to safe contexts

Keywords: Sex hormones, Estradiol, Oral contraceptives, Fear extinction, Contextual signaling, Skin conductance, functional MRI

2. Materials and Methods

2.1 Participants

This project reports data from the second and third sessions of a 3-session larger project. From the 218 participants in the larger project, 39 did not take part in this phase of the study for the following reasons: 15 were familiar with the fear protocol, 14 discontinued after the first session of the larger project (attrition), 1 was claustrophobic, 4 no longer met inclusion criteria (e.g., medication onset after the first session of the larger project), 3 had MRI contraindications (e.g., metal), and 2 did not report accurate cue-shock contingencies after conditioning.

2.2 Fear conditioning and extinction protocol

A validated fear conditioning and extinction protocol was employed [Figure 1; 1–5]. During habituation, participants viewed six cue-context combinations (cues: red, blue, yellow; contexts: office, library), without electrical stimulations. Prior to each subsequent phase, participants were informed of potential shocks. Fear conditioning occurred in the office context (CX+), where a lamp (off for 3 s) illuminated in red, blue, or yellow for 6 s. Two lamp colors (CS+1, CS+2) co-terminated with a 0.5 s shock on 5 of 8 trials (partial reinforcement), while the third (CS−) was presented 16 times without reinforcement. Conditioning was divided into two consecutive blocks: CS+1/CS− and CS+2/CS−. Extinction learning followed immediately in a different context (library; CX−), with one CS+ and the CS− presented 16 times each without reinforcement, resulting in extinction of the selected CS+. Contingency awareness was assessed after conditioning and extinction. Two participants were excluded due to inaccurate responses (see section 3.1).

Twenty-four hours later, extinction recall, and fear renewal were tested in the extinction (CX−) and conditioning (CX+) contexts, respectively. The extinguished CS+, non-extinguished CS+, and CS− were presented 8-, 8-, and 16-times during extinction recall, and 4-, 4-, and 8-times during fear renewal, without reinforcement. Both phases included two blocks: extinguished CS+/CS− and non-extinguished CS+/CS−. Color–stimulus assignments, block order, and context presentation were counterbalanced across eight conditions. Inter-trial intervals ranged from 12–18 s. Stimulus presentation was controlled using SuperLab (Cedrus).

2.3 Sex hormone collection and data analysis

Saliva samples (2 mL) were collected via passive drooling and stored at -20ºC for up to 2 months, before transferring to a -80ºC freezer for long-term storage. Samples were shipped on dry ice to ZRT Laboratory (Beaverton, Oregon, USA) for analysis using liquid chromatography-tandem mass spectrometry (LC-MS/MS) with an AB Sciex Triple Quad 5500 system. Samples were combined with internal standards, and hormones were separated using CI 8 column chromatography, dried under nitrogen, derivatized, diluted, and injected [6]. Analytical separation utilized an Agilent Poroshell 120 EC-C8 column, with ionization via atmospheric pressure chemical ionization.

The lower limits of quantification (LLOQ) were 0.30 pg/mL for E2, 5.00 pg/mL for progesterone, 3.00 pg/mL for testosterone, and 0.40 pg/mL for EE. For OC users in the active phase of their regimen, a detection limit of 0.11 pg/mL was applied for EE values below the LLOQ. A common issue with LC-MS/MS quantification of salivary sex hormones in healthy adults is the significant proportion of data below the LLOQ (i.e., undetected). We have previously shown that undetected hormone levels followed the expected pattern, with higher proportions in low-hormone groups [7–9]. Thus, undetected data were treated as indicative of low hormone concentrations. To address non-detection, multiple imputation from a fitted log-normal distribution addressed this issue using the *fitdistrplus* and *EnvStats* R packages [10,11]. Ten datasets were generated, with undetected data randomly imputed between 0 and the LLOQ. Imputed data were log-transformed due to right-skewed distributions. For between-group analyses, mean *p*-values (p_M) are reported, with effect sizes and confidence intervals not reported due to averaging across ten datasets, posing interpretability issues.

2.4 Questionnaires

To control for potential pre-existing psychological differences across groups, participants completed a series self-report questionnaires targeting constructs relevant to fear regulation and/or OC use.

2.4.1 State and Trait Anxiety Inventory – State/Trait (STAI-S, STAI-T). Anxiety was evaluated using the French adaptation of the STAI [12–14], which consists of 40 items divided equally between assessments of state (STAI-S) and trait (STAI-T) anxiety. Participants rated items on a 4-point Likert scale, with responses ranging from ‘Not at all’

to ‘A lot’ (state scale), and from ‘Hardly ever’ to ‘Almost always’ (trait scale). Total scores were calculated by summing item responses, with elevated scores reflecting higher anxiety. The French version demonstrates high internal consistency, coefficient values of .90 (state) and .91 (trait) [12].

2.4.2 Beck Depression Inventory-II (BDI-II). Depressive symptoms were measured using the French version of the 21-item BDI-II [15,16]. Each item is scored from 0 to 3, reflecting symptom severity. Total scores were calculated by summing item responses, with higher scores reflecting more severe depressive symptoms. Internal consistency has been reported at .92 in a French-speaking university student sample [17].

2.4.3 Santa Clara Strength of Religious Faith Questionnaire (SCSRFQ). Religious faith strength was assessed using the French version of the 10-item SCSRFQ [18]. Participants responded on a 4-point Likert scale ranging from ‘Strongly disagree’ to ‘Strongly agree’. Total scores were calculated by summing item responses, with higher scores reflecting stronger religious faith. The original English version exhibits excellent reliability, with coefficients ranging from .94 to .97 [18]. The French version also showed excellent internal consistency (.98) in the current sample (n = 146).

2.4.4 Emotion Regulation Questionnaire (ERQ). Emotion regulation strategies (cognitive reappraisal, expressive suppression) were assessed using the French version of the ERQ [19,20]. The questionnaire includes 6 items for cognitive reappraisal and 4 items for expressive suppression. Responses were collected on a 7-point Likert scale ranging from ‘Strongly disagree’ to ‘Strongly agree’. Subscale scores were computed by summing respective item responses, with higher totals indicating more frequent use of the strategy. Internal consistency values reported for the French version are .76 for cognitive reappraisal and .72 for expressive suppression [19].

2.4.5 NEO Personality Inventory 3 (NEO-PI-3). Core personality traits were assessed using the French version of the NEO-Five Factor Inventory comprising 60 items, with 12 measuring each of the five traits: neuroticism, extraversion, conscientiousness, openness, and agreeableness [21]. Items were rated on a 5-point scale from ‘Strongly disagree’ to ‘Strongly agree’. Total scores for each trait were determined by summing the relevant items. Reported internal consistency coefficients for the French version range from .83 to .90 across the five traits [22].

2.4 fMRI analyses: Image acquisition, preprocessing, regions of interest definition

Scanning was performed on a Siemens MAGNETOM Prisma 3T scanner at the Montreal General Hospital, using a 64-channel head coil and an interleaved multiband sequence (repetition time [*TR*] = 867 ms, echo time [*TE*] = 20 ms, Flip angle = 58º, slice number = 51, voxel size = 3 mm^3). To prevent signal loss in regions of interest (e.g., 11), field maps were rotated 30° clockwise relative to the AC-PC axis. Analyses were conducted in Matlab R2018b (The Mathworks Inc., Natick, Massachusetts, USA) and Statistical Parametric Mapping (SPM12; Wellcome Trust Centre for Neuroimaging, www.fil.ion.ucl.ac.uk).

Pre-processing involved slice-timing correction, realignment, co-registration with structural scans, normalization to MNI space, and smoothing using an 8 mm full-width half-maximum Gaussian kernel. High-pass filtering (128 s) and autoregressive (AR-1) modeling addressed signal drift and physiological noise. Artifacts were identified using ArtRepair, and motion parameters were included as regressors in first-level analyses.

Second-level analyses were applied small volume correction using the following regions of interest (ROIs): hippocampus, amygdala, vmPFC, and ACC. The vmPFC was defined as in De La Vega et al. [23]. The hippocampus, amygdala, and ACC were defined using the Anatomical Automatic Labeling atlas [24]. These masks were selected due to their liberal delimitations. ROIs were combined into a single mask to correct for multiple comparisons [25]. Whole-brain analyses were also conducted.

3. Results

3.1 Participant exclusions

Skin conductance responses were qualitatively and quantitatively inspected (see Methods). Nineteen participants (1 man, 6 never users [2 E2-, 4 E2+], 4 past users [2 E2-, 1 E2+], 8 OC users [5 low EE doses, 3 high EE doses]) were excluded due to non-responsiveness. Exclusions were balanced between participant groups (p = .245 for E2-based groups; p = .263 for OC-history-based groups; p = .939 for EE-based groups). Given the potential bias introduced by excluding non-responders [26,27], sensitivity analyses including all participants were performed. Results remained consistent, although some significant post-hoc effects shifted to trend level ($ps \leq .08$). Nine participants were

excluded due to technical issues on day 1 (n = 7) or on day 2 (n = 2), and 4 withdrew on day 2. The final sample for SCR analyses consisted of 147 participants, consistent with Davignon et al. [28]. Four additional participants were excluded from fMRI analyses due to excessive motion (> 4 mm) but retained for SCR analyses.

3.2 Between-group differences in salivary hormone levels

Hormone means and standard deviations are reported in **Table 1**. In E2-based groups, both E2- and E2+ women showed higher testosterone levels than OC users (p_M = .006; p_M < .001). In OC-history-based groups, never and past users exhibited higher testosterone levels than current users (p_M < .001, p_M = .003), and past users had elevated E2 levels compared to current users (p_M = .006). Across both types of groupings, men had significantly higher testosterone levels than all female groups (p_Ms < .001). Among OC users, EE+ women showed higher EE concentrations than EE- women. As expected, progesterone levels did not differ across groups, irrespective of the analysis set.

3.3 Preliminary analyses: Correlations between psychological and demographic variables and outcome measures

Beyond the between-group comparisons presented in **Table 1**, we explored whether any control variables were significantly correlated with outcome measures across all participants. Neuroticism was marginally negatively correlated with fear return to the CX- (r(145) = –.16, p = .060), suggesting that higher neuroticism was associated with lower fear responses. This is contrary to expectations, as neuroticism typically predicts heightened fear responding [29–31] or shows no reliable association with fear responses in experimental paradigms [32,33].

A significant association also emerged for relationship status (t(99.93) = -3.71, p < .001), with partnered individuals showing greater fear return to the CX- than single participants. Because current OC users were more likely to be partnered than never users, relationship status may represent a partial confound in OC-history-based analyses (but not in EE-based analyses where EE- women are nominally more likely to be partnered than EE+ women). However, to our knowledge, prior literature does not support a direct influence of relationship status on conditioned fear responses in laboratory protocols. While attachment dynamics may shape fear expression in close relationships [34–37], attachment styles or relationship quality were not assessed. We also considered whether

relationship status might reflect a broader psychosocial profile associated with OC use that could include other demographic, psychological, or personality characteristics. Among our control variables, only conscientiousness was marginally associated with relationship status (p = .059), with partnered participants scoring slightly higher. However, as noted above, conscientiousness itself was not significantly related to fear return. Therefore, although residual confounding by relationship status cannot be entirely excluded, OC use remains the more plausible explanatory factor of fear returns [38,39], with relationship status likely representing a correlated but independent variable.

It should be noted that these analyses derive from a large exploratory correlation matrix that was not corrected for multiple comparisons. Given that these tests were not driven by prior hypotheses nor strongly supported by the literature, the results should be interpreted with caution, as correction procedures would further diminish statistical significance and reduce confidence in these associations.

3.4 Sex differences across the protocol

Multivariate analyses of SCRs revealed no Group (men vs. women) X Trial interactions or Group main effect during fear conditioning (F(3,4557) = 0.31, p = .821; F(1,147) = 0.34, p = .561), extinction learning (F(3,4557) = 1.97, p = .116; F(1,147) = 0.30, p = .588), extinction recall (F(3,4557) = 0.70, p = .553; F(1,147) = 0.00, p = .978), or fear renewal (F(3, 2205) = 1.11, p = .343; F(1,147) = 0.37, p = .543). Between-group comparisons of fear returns likewise revealed no significant effects of Group for the CX- (F(1,145)=1.51, p=.221, η^2_p=0.01) and the CX+ (F(1,145)=0.08, p=.783, η^2_p=0.00). In addition, no significant between-group difference emerged on BOLD contrasts (CX- and CX+).

3.5 Blood oxygen level-dependent (BOLD) responses (fMRI analyses)

We performed one sample t-tests across all participants for the contrasts of interest (CX- and CX+). Significant effects at both the ROI and whole-brain levels are reported in **Table S2**. Subsequently, between-group analyses were conducted. For the CX+ contrast (early fear renewal vs. late fear conditioning), both OC-history-based groups and EE-based groupings revealed a significant cluster in the left dorsal ACC (OC-history: -10, 40, -4, k = 40, F(3, 139) = 7.72, pFWE = .025; EE: -8, 40, -4, k = 33, F(2, 76) = 11.32, pFWE = .033). Post-hoc comparisons of extracted beta estimates indicated significantly greater ACC

activation in men, past and current OC users compared to never users (all *p*s ≤ .009). Similarly, in the EE-based grouping, both EE− and EE+ women exhibited greater ACC activation relative to never users ($p < .001$; $p = .007$).

To better characterize these differences, we further examined between-group effects separately for each phase (fear conditioning and fear renewal), using the same ROI mask. OC-history-based analyses revealed no significant clusters. However, the EE-based analysis identified a significant cluster in the left dorsal ACC during the fear conditioning phase, overlapping with the coordinates previously observed in the CX+ contrast (−8, 40, −4, $k = 81$, $F(2, 76) = 15.81$, *p*FWE = .001). This suggests that the observed CX+ effect is primarily driven by heightened dorsal ACC activation in never users, relative to both EE-current users ($p < .001$) and EE+ current users (in trend; $p = .060$) during fear conditioning.

Although traditional clusterwise analyses revealed no significant between-group differences for the CX- contrast (see above), mean hippocampal and vmPFC activation across the entire ROIs significantly correlated with fear return to the CX- (reported in the main text). To explore whether neural activation patterns mirrored physiological group differences, we conducted between-group ANOVAs specifically on these extracted values. In the OC-history-based grouping, a significant main effect of Group emerged for hippocampal activation ($F(3,139) = 3.50$, $p = .017$; **Figure S3A**). Post hoc comparisons indicated that never users showed significantly greater hippocampal activation than men ($p = .010$) and marginally greater activation than current OC users ($p = .079$). In the EE-dose-based grouping, we observed a marginal main effect of Group ($F(2,76) = 2.85$, $p = .064$; **Figure S3B**), with never users exhibiting slightly higher hippocampal activation than EE+ users ($p = .064$). No between-group differences were observed for vmPFC activation (*p*s ≥ .105).

3.6 Exploratory analyses – Objective C: EE-based groups

To tentatively isolate the relationship between EE and fear returns to the CX-, we compared never users with low-progestin (≤250 μg, n=34) and high-progestin users (1000-3000 μg, n=18; **Figure 1**). Groupings based on EE and progestin doses were independent ($\mathcal{X}^2$=0.00, p=.964). However, the progestin-based analysis also revealed a significant Group effect ($F(2,79)$=5.10, p=.008, η^2_p=0.11), with high-progestin users showing higher fear returns than never users (p=.006, 95%CI=[80.91, 522.72], d=1.10). When both

predictors were included, EE dose explained unique variance (p=.042), whereas progestin dose showed a trend (p=.078). However, model fit indices were comparable (EE: AIC=1181, BIC=1192; Progestin: AIC=1182, BIC=1191), limiting strong conclusions regarding the superior predictive value of either hormone.

Moreover, a prior analysis in this sample identified an association between EE dose and progestin androgenicity, whereby low-EE formulations were more frequently paired with highly androgenic progestins, whereas high-dose EE formulations tended to contain low androgenic progestins [7]. As in our previous publications [7,8], OCs were further categorized as anti-androgenic (n=17), low androgenic (n=15), or high androgenic (n=20), and compared to never users. We first examined whether participants using OCs with different progestin androgenicity profiles differed in psychological or demographic traits (similar to Table 1). Androgenicity was significantly associated with age ($p < .001$), education ($p = .028$), agreeableness ($p = .022$), and conscientiousness ($p = .024$): compared to low-androgenic users, users of anti-androgenic formulations were marginally younger ($p = .066$), less educated ($p = .040$), and scored lower on agreeableness ($p = .025$). Anti-androgenic users also scored higher on conscientiousness relative to high-androgenic users ($p = .024$). As reported in section 3.3, none of these variables were significantly associated with fear returns.

Groups were then compared on their fear returns. A significant main effect of Group was observed (F(3,78)=3.78, p=.014, η^2_p=0.13), with users of anti-androgenic progestins showing higher fear returns to the CX- than never users (p=.008, 95%CI=[74.50, 574.00], d=1.31). In a model including both EE dose and progestin androgenicity, EE dose accounted for unique variance (p=.038), whereas progestin androgenicity did not (p=.109). Again, model fit indices were comparable across models (EE: AIC=1181, BIC=1191; androgenicity: AIC=1183, BIC=1195).

3.7 Sensitivity analyses

To assess the robustness of the results, we repeated group comparisons on fear return percentages after reintroducing participants who had been excluded due to poor conditioned responses. In E2-based groups, the main effect of Group remained non-significant for both the CX- ($F(3,160) = 0.97$, $p = .408$) and CX+ contexts ($F(3,160) = 0.44$, $p = .727$). In OC-history-based groups, the main effect of Group reached trend level

for the CX- ($F(3,160) = 2.66$, $p = .050$), with higher fear return in both current ($p = .046$) and past OC users ($p = .050$) compared to never users, replicating the pattern observed in the main analyses. No significant effect was found for the CX+ ($F(3,160) = 0.65$, $p = .583$). In EE-based groups, the main effect of Group was significant for the CX- ($F(2,91) = 5.32$, $p = .007$), with higher fear return in high-EE users relative to never users ($p = .026$), again consistent with the main results. No significant effect was found for the CX+ ($F(2,91) = 1.60$, $p = .208$).

We then removed the 5 OC users in their placebo pill phase. In E2-based groups, the main effect of Group remained non-significant for both the CX- ($F(3,138) = 1.55$, $p = .203$) and CX+ contexts ($F(3,138) = 0.53$, $p = .666$). In OC-history-based groups, the main effect of Group was significant for the CX- ($F(3,138) = 3.87$, $p = .011$), with higher fear return in both current ($p = .014$) and past OC users ($p = .014$) compared to never users. No significant effect was found for the CX+ ($F(3,138) = 0.78$, $p = .508$). In EE-based groups, the main effect of Group was significant for the CX- ($F(2,74) = 4.50$, $p = .014$), with higher fear return in high-EE users relative to never users ($p = .041$). The effect of Group was also significant for the CX+ ($F(2,74) = 3.23$, $p = .045$), with no significant post-hoc contrasts ($ps \geq .067$).

Overall, the results remained consistent across all sensitivity analyses, supporting the robustness of the main findings.

References

1. Linnman C, Zeidan MA, Pitman RK, Milad MR. Resting cerebral metabolism correlates with skin conductance and functional brain activation during fear conditioning. Biol Psychol. 2012;89:450–459.
2. Marin M-F, Song H, VanElzakker MB, Staples-Bradley LK, Linnman C, Pace-Schott EF, et al. Association of resting metabolism in the fear neural network with extinction recall activations and clinical measures in trauma-exposed individuals. Am J Psychiatry. 2016;173:930–938.
3. Milad MR, Pitman RK, Ellis CB, Gold AL, Shin LM, Lasko NB, et al. Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. Biol Psychiatry. 2009;66:1075–1082.
4. Milad MR, Quirk GJ, Pitman RK, Orr SP, Fischl B, Rauch SL. A role for the human dorsal anterior cingulate cortex in fear expression. Biol Psychiatry. 2007;62:1191–1194.
5. Milad MR, Wright CI, Orr SP, Pitman RK, Quirk GJ, Rauch SL. Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. Biol Psychiatry. 2007;62:446–454.
6. Nave G, Nadler A, Dubois D, Zava D, Camerer C, Plassmann H. Single-dose testosterone administration increases men's preference for status goods. Nat Commun. 2018;9:2433.
7. Brouillard A, Davignon L-M, Turcotte A-M, Marin M-F. Morphologic alterations of the fear circuitry: the role of sex hormones and oral contraceptives. Front Endocrinol. 2023;14:1228504.
8. Davignon L-M, Brouillard A, Juster R-P, Marin M-F. The role of sex hormones, oral contraceptive use, and its parameters on visuospatial abilities, verbal fluency, and verbal memory. Horm Behav. 2024;157:105454.
9. Brouillard A, Davignon L-M, Cernik R, Giguère C-É, Findlay H, Juster R-P, et al. Comparing immunoassay and mass spectrometry techniques for salivary sex hormone analysis. Psychoneuroendocrinology. 2025;174:107379.

10. Canales RA, Wilson AM, Pearce-Walker JI, Verhougstraete MP, Reynolds KA. Methods for handling left-censored data in quantitative microbial risk assessment. Appl Environ Microbiol. 2018;84:e01203-18.
11. Herbers J, Miller R, Walther A, Schindler L, Schmidt K, Gao W, et al. How to deal with non-detectable and outlying values in biomarker research: Best practices and recommendations for univariate imputation approaches. Compr Psychoneuroendocrinology. 2021;7:100052.
12. Gauthier J, Bouchard S. Adaptation canadienne-française de la forme révisée du State–Trait Anxiety Inventory de Spielberger. Can J Behav Sci Rev Can Sci Comport. 1993;25:559–578.
13. Spielberger CD, Gorsuch RW, Lushene RE. Manual for the state trait anxiety inventory. Palo Alto Consult Psychol Press. 1970. 1970.
14. Spielberger CD, Gorsuch RW, Lushene RE, Vagg PR, Jacobs GA. Manual for the state trait anxiety inventory (Form Y). Palo Alto Consult Psychol Press. 1983. 1983.
15. Beck AT, Steer RA, Brown G. Manual for the Beck Depression Inventory-II. Psychological Corporation. San Antonio, TX; 1996.
16. Centre de psychologie appliquée. Manuel du BDI-II. Editions du centre de psychologie appliquée. Paris; 1996.
17. Bourque P, Beaudette D. Étude psychometrique du questionnaire de dépression de Beck auprès d'un échantillon d'étudiants universitaires francophones. Can J Behav Sci Rev Can Sci Comport. 1982;14:211–218.
18. Plante TG. The Santa Clara Strength of Religious Faith Questionnaire: Assessing faith engagement in a brief and mondenominational manner. Religions. 2010;1:3–8.
19. Christophe V, Antoine P, Leroy T, Delelis G. Évaluation de deux stratégies de régulation émotionnelle : La suppression expressive et la réévaluation cognitive. Rev Eur Psychol Appliquée. 2009;59:59–67.
20. Gross JJ, John OP. Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. J Pers Soc Psychol. 2003;85:348–362.
21. McCrae RR, Costa PT. Brief versions of the NEO-PI-3. J Individ Differ. 2007;28:116–128.

22. Rolland JP, Parker WD, Stumpf H. A psychometric examination of the french translations of NEO-PI-R and NEO-FFI. J Pers Assess. 1998;71:269–291.
23. De La Vega A, Chang LJ, Banich MT, Wager TD, Yarkoni T. Large-scale meta-analysis of human medial frontal cortex reveals tripartite functional organization. J Neurosci. 2016;36:6553–6562.
24. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated Anatomical Labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage. 2002;15:273–289.
25. Roiser JP, Linden DE, Gorno-Tempini ML, Moran RJ, Dickerson BC, Grafton ST. Minimum statistical standards for submissions to Neuroimage: Clinical. NeuroImage Clin. 2016;12:1045–1047.
26. Lonsdorf TB, Menz MM, Andreatta M, Fullana MA, Golkar A, Haaker J, et al. Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. Neurosci Biobehav Rev. 2017;77:247–285.
27. Marin M-F, Barbey F, Rosenbaum BL, Hammoud MZ, Orr SP, Milad MR. Absence of conditioned responding in humans: A bad measure or individual differences? Psychophysiology. 2020;57.
28. Davignon L-M, Brouillard A, Marin M-F. Current and lasting effects of oral contraceptives on fear regulation. under review.
29. Drabant EM, Kuo JR, Ramel W, Blechert J, Edge MD, Cooper JR, et al. Experiential, autonomic, and neural responses during threat anticipation vary as a function of threat intensity and neuroticism. NeuroImage. 2011;55:401–410.
30. Lommen MJJ, Engelhard IM, Van Den Hout MA. Neuroticism and avoidance of ambiguous stimuli: Better safe than sorry? Personal Individ Differ. 2010;49:1001–1006.
31. Norris CJ, Larsen JT, Cacioppo JT. Neuroticism is associated with larger and more prolonged electrodermal responses to emotionally evocative pictures. Psychophysiology. 2007;44:823–826.

32. Lonsdorf TB, Merz CJ. More than just noise: Inter-individual differences in fear acquisition, extinction and return of fear in humans - Biological, experiential, temperamental factors, and methodological pitfalls. Neurosci Biobehav Rev. 2017;80:703–728.
33. Rattel JA, Miedl SF, Liedlgruber M, Blechert J, Seidl E, Wilhelm FH. Sensation seeking and neuroticism in fear conditioning and extinction: The role of avoidance behaviour. Behav Res Ther. 2020;135:103761.
34. Gilissen R, Koolstra CM, Van IJzendoorn MH, Bakermans-Kranenburg MJ, Van Der Veer R. Physiological reactions of preschoolers to fear-inducing film clips: Effects of temperamental fearfulness and quality of the parent–child relationship. Dev Psychobiol. 2007;49:187–195.
35. Gilissen R, Bakermans-Kranenburg MJ, Van IJzendoorn MH, Van Der Veer R. Parent–child relationship, temperament, and physiological reactions to fear-inducing film clips: Further evidence for differential susceptibility. J Exp Child Psychol. 2008;99:182–195.
36. Stupica B, Brett BE, Woodhouse SS, Cassidy J. Attachment security priming decreases children's physiological response to threat. Child Dev. 2019;90:1254–1271.
37. Toumbelekis M, Liddell BJ, Bryant RA. Thinking of attachment figures blocks differential fear conditioning. Soc Cogn Affect Neurosci. 2018;13:989–994.
38. Graham BM. The impact of hormonal contraceptives on anxiety treatments: From preclinical models to clinical settings. Front Neuroendocrinol. 2022;67:101030.
39. Graham BM, Milad MR. Blockade of estrogen by hormonal contraceptives impairs fear extinction in female rats and women. Biol Psychiatry. 2013;73:371–378.

Table S1. Summary of oral contraceptive formulations among current users.

Brand	Progestin	Progestin dose (µg)	EE dose (µg)	Regimen (active/placebo)	*n*
Alesse/Alysena/Aviane	Levonorgestrel	100	20	21/7	14
Leeloo	Levonorgestrel	100	20	21/7	3
Indayo	Levonorgestrel	150	30	84/7	1
Ludeal	Levonorgestrel	150	30	21/7	1
Min-Ovral	Levonorgestrel	150	30	21/7	1
Ovima	Levonorgestrel	150	30	21/7	1
Marvelon	Desogestrel	150	30	21/7	6
Tri Jordyna	Norgestimate	180, 215, 250	35	21/7	2
Linessa	Desogestrel	100, 125, 150	25	21/7	2
Tri Cyclen Lo	Norgestimate	180, 215, 250	25	21/7	1
Desorelle 30	Desogestrel	150	30	21/7	1
Freya	Desogestrel	150	30	21/7	1
Lolo	Norethindrone acetate	1000	10	24/2/2	1
Minestrin	Norethindrone acetate	1000	20	21/7	1
Mirvala	Desogestrel	150	30	21/7	1
Yaz	Drospirenone	3000	20	24/4	5
Diane 35	Cyproterone acetate	2000	35	21/7	4
Yasmin	Drospirenone	3000	30	21/7	5
Izéane	Drospirenone	3000	20	24/4	1

Note. The formulations listed reflect those currently used by participants classified as current users; several had used different brands in the past. Formulations for past users could not be reported, as fewer than half had consistent formulations over time, and many could not recall specific brands. EE = ethinyl estradiol.

Table S2. Significant BOLD contrasts across all participants.

Brain region	x	y	z	*k*	*t*-value	*p*-value (uncorr.)	*p*-value (FWE-corr.)
Safe context (CX-)							
First 4 CX of extinction recall vs. Last 4 CX of extinction learning							
L hippocampus (posterior)*	-14	-36	0	73	6.56	.006	.008
R anterior cingulate cortex (dorsal)*	0	10	28	163	5.81	< .001	< .001
R hippocampus (posterior)*	20	-32	-2	92	5.69	.003	.003
R fusiform gyrus	44	-60	-10	20394	12.89	< .001	< .001
R supplementary/premotor	6	10	50	4030	8.64	< .001	< .001
	42	8	28	973	7.46	< .001	< .001
L insula	-32	24	2	147	5.24	< .001	.009
L supramarginal gyrus	-44	-36	36	198	5.00	< .001	.001
Last 4 CX of extinction learning vs. First 4 CX of extinction recall							
R anterior cingulate cortex (rostral)*	16	50	14	1283	6.81	< .001	< .001
R anterior prefrontal cortex	16	54	24	27127	8.23	< .001	< .001
R cerebellum	32	-72	-36	2098	6.81	< .001	< .001
L cerebellum	-40	-74	-32	2159	6.65	< .001	< .001
R angular gyrus	60	-52	32	627	5.87	< .001	< .001
R medial temporal gyrus	66	-20	-22	320	4.91	< .001	< .001
	62	-6	-30	148	4.57	< .001	.008
L putamen	-26	-10	-4	124	4.84	.001	.021
L medial temporal gyrus	-54	-10	-28	313	4.65	< .001	< .001
Threat context (CX+)							
First 4 CX of fear renewal vs. Last 4 CX of fear conditioning							
L fusiform	-30	-46	-16	9166	13.97	< .001	< .001
R fusiform	32	-44	-18	10004	12.47	< .001	< .001
L supplementary/premotor	-46	2	46	359	6.31	< .001	< .001
	-4	8	48	470	4.54	< .001	< .001
Last 4 CX of fear conditioning vs. First 4 CX of fear renewal							
R ventromedial prefrontal cortex*	10	56	-6	1900	6.22	< .001	< .001
R dorsolateral prefrontal cortex	26	34	42	16294	7.63	< .001	< .001
R visuomotor cortex	6	-72	38	1775	5.45	< .001	< .001
R supratemporal gyrus	64	-16	-2	696	5.38	< .001	< .001
R primary sensory cortex	45	-10	17	148	4.82	.001	.018
R cerebellum	48	-58	-44	270	4.79	< .001	< .001
L cerebellum	-8	-60	-14	224	4.48	< .001	.002
	-32	-88	-32	319	4.47	< .001	< .001
	-52	-72	-34	158	4.14	< .001	.013
L posterior cingulate cortex (dorsal)	-14	-52	34	133	4.48	.001	.029

Note. *k* = cluster size, uncorr. = uncorrected, FWE-corr. = family-wise error corrected, CS+ = conditioned stimulus reinforced with an electrical stimulation, CS- = unreinforced stimulus, CS+E = extinguished CS+, CX = context. *Indicates small volume correction

Table S3. Model selection for multilevel modelling of skin conductance responses to contexts across all phases of the protocol.

Step	Model	Parameter	*n*	AIC	BIC	$\mathcal{X}^2$	df	*p*-value	Notes
Fear conditioning									
1	**Random intercept**		4	3265.0	3290.8				
1	*Random intercept and slope*		5	3088.5	3120.8	178.5	1	< .001	Model failed to converge
2	Linear		4	3265.0	3290.8				
2	Quadratic		5	2944.3	2976.6	322.6	1	< .001	
2	**Cubic**		6	2838.5	2877.2	107.8	1	< .001	
Extinction learning									
1	**Random intercept**		4	-612.9	-587.1				
1	*Random intercept and slope*		5	-731.0	-698.8	120.1	1	< .001	Model failed to converge
2	Linear		4	-612.9	-587.1				
2	Quadratic		5	-657.6	-625.3	46.7	1	< .001	
2	**Cubic**		6	-681.4	-642.6	25.7	1	< .001	
Extinction recall									
1	**Random intercept**		4	1024.3	1050.1				
1	*Random intercept and slope*		5	831.9	864.2	194.4	1	< .001	Model failed to converge
2	Linear		4	1024.3	1050.1				
2	Quadratic		5	889.4	921.7	136.9	1	< .001	
2	**Cubic**		6	844.6	883.3	46.8	1	< .001	
Fear renewal									
1	**Random intercept**		4	617.5	640.6				
1	Random intercept and slope		5	632.3	661.2	0	1	1.000	Model failed to converge
2	Linear		4	617.5	640.6				
2	Quadratic		5	567.9	596.7	51.7	1	< .001	
2	**Cubic**		6	548.4	582.9	21.5	1	< .001	

Notes. Selected models are bolded in the table. Model selection was based on model Bayesian information criterions (BIC), accounting for model fit and complexity. When models failed to converge, a simpler random effect structure was preferred. Italics represent models with the lower BICs that were not selected due to an overly complex random effect structure. AIC = Akaike information criterion, df = degrees of freedom.

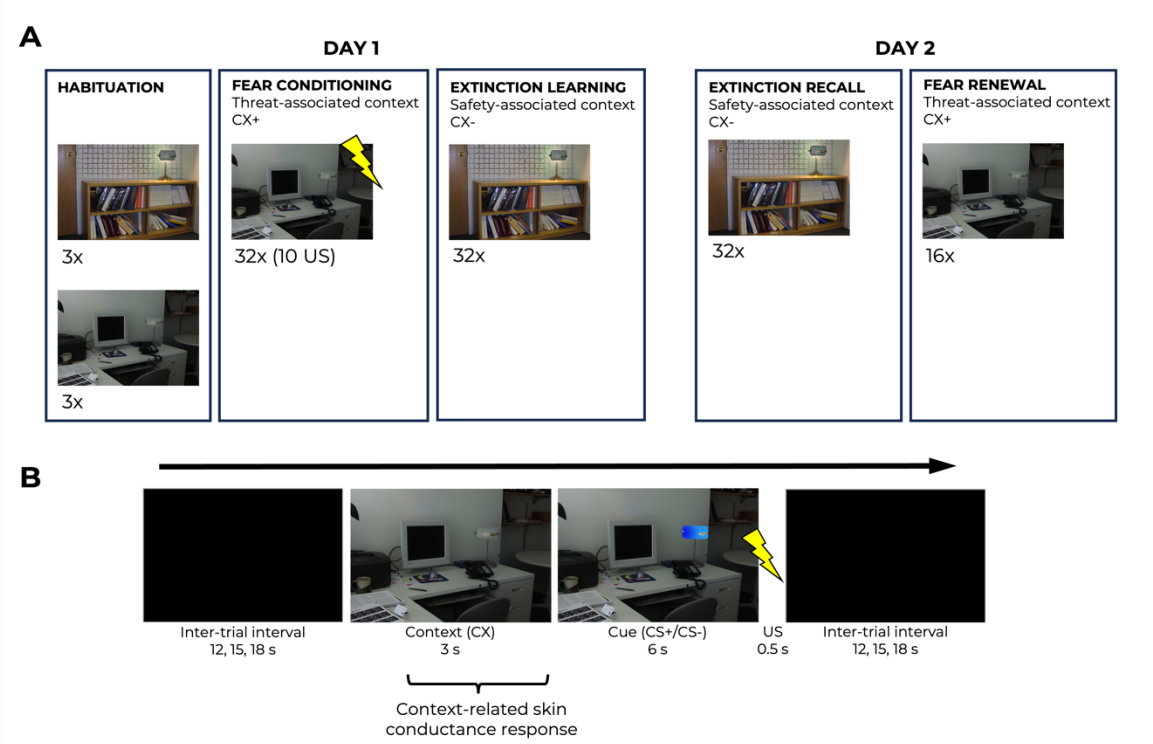

Figure S1. Schematic of the fear protocol. (A) On day 1, habituation consists of presenting the three cues (i.e., lamps) in each context. Immediately after, during fear conditioning occurring in the threatening context (CX+), two cues are paired with an aversive electric shock, while one cue is not reinforced. Then, during extinction learning in the safe context (CX-), one of the conditioned cues is presented without reinforcement, resulting in its extinction. Twenty-four hours later, extinction recall, and fear renewal are assessed by presenting the three cues in the CX- and the CX+, respectively. (B) Structure of one trial. Following an inter-trial interval (12, 15 or 18 seconds), the context (CX-/CX+) is displayed for 3 seconds. Skin conductance responses are recorded from 1-3 seconds after the context onset. The cue is then presented for 6 seconds within the context. Reinforcement (or lack thereof) is administered at the cue offset, followed by an inter-trial black screen initiating the next trial. US = Unconditioned stimulus (electrical stimulation).

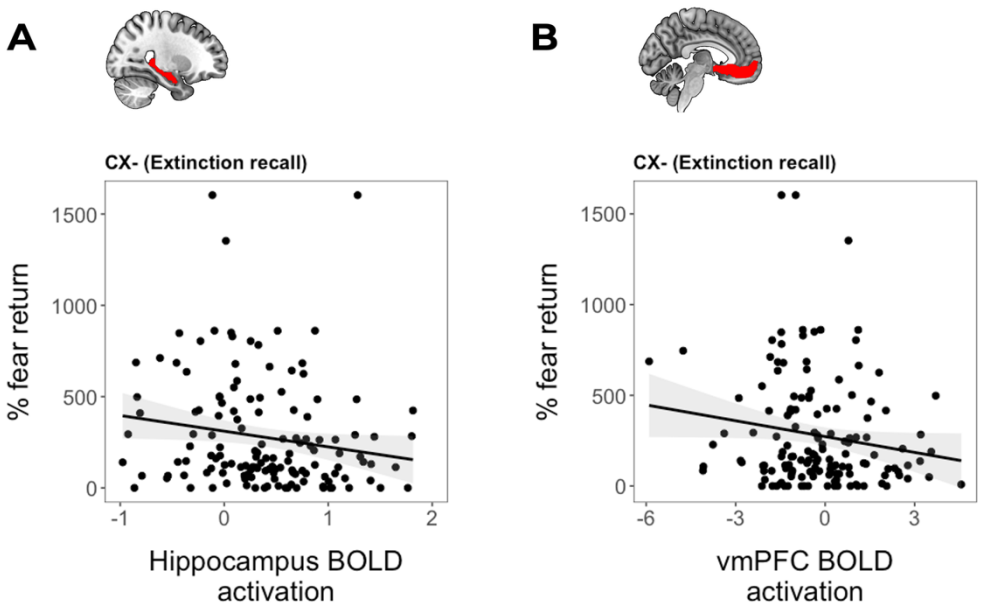

Figure S2. Correlations between blood oxygen level-dependent (BOLD) activation in bilateral regions of interest and percentages of fear return to the safe context (CX-) across all participants (n = 143). (A) Significant negative correlation between hippocampus BOLD activation and percentages of fear return to the CX-. (B) Significant negative correlation between ventromedial prefrontal cortex (vmPFC) BOLD activation and percentages of fear return to the CX-.

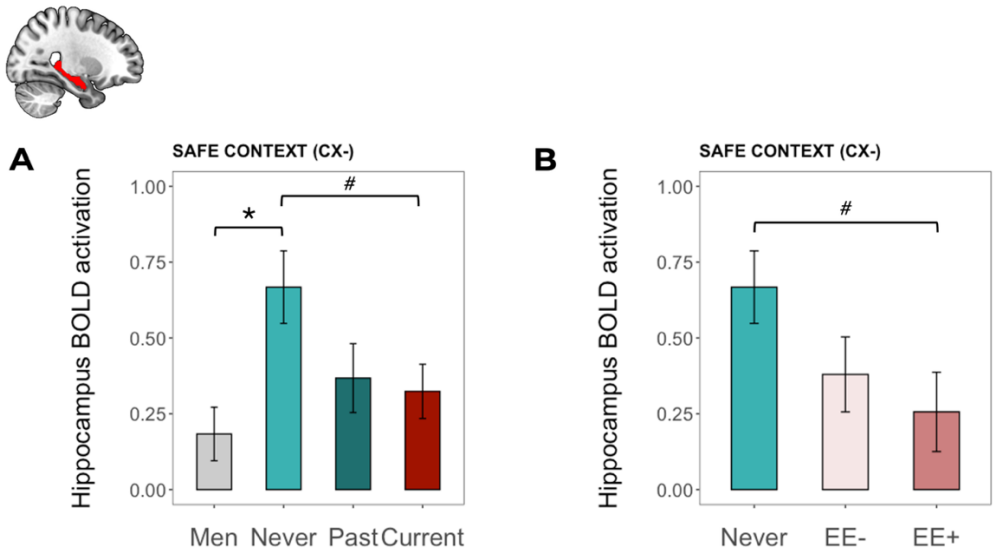

Figure S3. (A) Hippocampal BOLD activation to the safe context (CX-) contrast (early extinction recall vs. late extinction learning) as a function of oral contraceptive (OC) history. Never users showed greater activation than men and marginally greater activation than current users. (B) Hippocampal BOLD activation for the same contrast in groups based on ethinyl estradiol (EE) dosage in current OC users. Never users showed marginally greater activation than users of high EE-dose OCs. EE- = low EE dose; EE+ = high EE dose. *Indicates significant pairwise contrast ($p < .05$); #indicates statistical trend ($p < .08$).