No evidence that hormonal contraceptive use or circulating sex steroids predict complex emotion recognition

Talia N. Shirazi[a], **Kevin A. Rosenfield**[a], **Rodrigo A. Cárdenas**[b], **S. Marc Breedlove**[c], **David A. Puts**[a]

[a] Department of Anthropology, Pennsylvania State University; 419 Carpenter Building, University Park, PA, 16802

[b] Department of Psychology, Pennsylvania State University; Moore Building, University Park, PA, 16802

[c] Department of Neuroscience, Michigan State University; 240 Giltner Hall, East Lansing, MI 48824

Abstract

Relatively little is known about the effects of endogenous and exogenous steroid hormones on ecologically relevant behavioral and cognitive phenotypes in women, such as emotion recognition, despite the widespread use of steroid hormone-altering hormonal contraceptives (HCs). Though some previous studies have examined the effect of HC use, estradiol, progesterone, and testosterone on emotion recognition in women, they have been limited by cross-sectional designs, small sample sizes (total $n < 100$), and compromised statistical power to detect significant effects. Using data from two test sessions in a large sample of naturally cycling women (NC; $n = 192$) and women on HCs ($n = 203$), we found no group differences in emotion recognition; further, the lack of group differences in emotion recognition were not modulated by item difficulty or emotional valence. Among NC women who provided saliva samples across two sessions that were assayed for estradiol and progesterone concentrations, we found no compelling evidence across models that between-subject differences and within-subject fluctuations in these ovarian hormones predicted emotion recognition accuracy, with the exception that between-subjects estradiol predicted emotion recognition for emotions of neutral valence ($p = 0.042$). Among HC women who provided saliva samples across two sessions that were assayed for testosterone, we found no compelling evidence that between-subjects differences and within-subject fluctuations in testosterone predicted emotion recognition accuracy. Overall, our analyses provide little support for the idea that circulating endogenous or exogenous ovarian hormones influence emotion recognition in women.

Correspondence: Dr. David Puts, 218 Carpenter Building, University Park, PA 16802, dap27@psu.edu

1. Introduction

Though initial investigations of the side effects associated with hormonal contraceptive (HC) use focused largely on physical symptoms such as weight gain, nausea, and intermenstrual spotting, research over the last several decades has increasingly focused on psychobehavioral effects (Montoya & Bos, 2017; Pletzer & Kerschbaum, 2014). HC use has been associated with increased depression (Skovlund, Mørch, Kessing, & Lidegaard, 2016) and mood swings (Gingnell et al., 2013), decreased life satisfaction (Zethraeus et al., 2017) and libido (Sanders, Graham, Bass, & Bancroft, 2001; but see Pastor, Holla, & Chmel, 2013), shifts in mate preferences (Cobey, Little, & Roberts, 2015; but see Marcinkowska, Hahn, Little, Debruine, & Jones, 2019), and altered performance on cognitive tasks (e.g., verbal fluency; Griksiene & Ruksenas, 2011).

Recent evidence also suggests that HC use impairs emotion recognition (Hamstra, Kloet, Hemert, Rijk, & van der Does, 2015; Hamstra, Rover, Rijk, & Van der Does, 2014; reviewed in Osório, Cassis, & Sousa, 2018). Impairments in emotion recognition may be of particular importance in daily life, as the ability to accurately detect and interpret the facial emotions of others is vital to success in social contexts (Osório et al., 2018). Differences in emotion recognition between naturally-cycling (NC) women and women using HCs may result from differences in brain activation patterns during emotion recognition (Gingnell et al., 2013) in regions implicated in face and emotion processing, such as the amygdala and insula, as well as in gross neural structure (Petersen, Touroutoglou, Andreano, & Cahill, 2015; see Pletzer & Kerschbaum, 2014 for review). These differences may reflect HC-induced changes in progestogen, estrogen, and testosterone concentrations (Fleischman, Navarrete, & Fessler, 2010; Pletzer & Kerschbaum, 2014; Zimmerman, Eijkemans, Coelingh, Blankstein, & Fauser, 2014). In NC women, both emotion recognition and its associated neural activity may be modulated by cycle phase and by circulating progesterone and estradiol concentrations (Osório et al., 2018), but likely not by testosterone (van Honk & Schutter, 2007), supporting the hypothesis that differences between HC and NC women in emotion recognition are driven by differential production and activity of estradiol and progesterone.

A recent systematic review on the relationship between ovarian hormones and emotion recognition (Osório et al., 2018) found that emotion recognition accuracy generally increased during the follicular phase, characterized by high estradiol and low progesterone (Derntl, Hack, Kryspin-exner, & Habel, 2013; Derntl, Kryspin-Exner, Fernbach, Moser, & Habel, 2008; Derntl, Windischberger, et al., 2008; Rubin et al., 2012; but see Gingnell, Morell, Bannbers, Wikström, & Sundström, 2012; Rubinow, Smith, Schenkel, Schmidt, & Dancer, 2007; Zhang, Zhou, & Ye, 2013); measured estradiol concentrations correlated with accuracy for the recognition of certain emotions in some studies (Hamstra, Kloet, Quataert, Jansen, & Does, 2017; Kamboj, Krol, & Curran, 2015; Pearson & Lewis, 2005) but not in others (Derntl, Kryspin-Exner, et al., 2008; Derntl, Windischberger, et al., 2008; Rubin

et al., 2012); and similarly, measured progesterone concentrations correlated with emotion recognition in some studies (Derntl, Kryspin-Exner, et al., 2008) but not in others (Derntl, Windischberger, et al., 2008; Maner & Miller, 2014; Rubin et al., 2012; van Wingen et al., 2008). Several reports have found hormonally-modulated neural activity during emotion recognition tasks, but with associations often in opposing directions (Derntl et al., 2008; van Wingen et al., 2008).

Variability across studies in whether cycle phase and hormone concentrations predict emotion recognition in NC women, and whether HC and NC women differ in emotion recognition (positive results: Hamstra et al., 2015, 2017; Maner & Miller, 2014; Pahnke et al., 2019; null results: Gingnell et al., 2013; Radke & Derntl, 2016) may be due to methodological limitations. The majority of previous studies have employed cross-sectional rather than longitudinal designs, which may be underpowered to detect differences in behaviors that are hypothesized to shift with ovarian hormone production across the ovulatory cycle (Gangestad et al., 2016).

Cycle phase has often been determined through self-report and counting methods, and reliance on such methods may lead to high error rates in classification (Blake, Dixson, O'Dean, & Denson, 2016; Gangestad et al., 2016; Gonzales & Ferrer, 2016), which reduce the power to detect differences and changes across cycle phases. Finally, in the studies reviewed by Osório et al. (2018), the average sample size for studies assessing differences across cycle phases was 33, while the average sample size for studies assessing differences between NC and HC women was 66. Recommended sample sizes for detecting medium-sized effects are significantly larger for ovulatory shift research (Gangestad et al., 2016; Gonzales & Ferrer, 2016); for example, with moderately valid measures of cycle phase or conception risk, 456 women are required for 80% power to detect a medium-sized effect in cross-sectional studies, while at least 45 women are required for longitudinal studies. These recommended sample sizes suggest that previous studies may have been too small to provide stable, reliable model estimates.

Addressing this last limitation specifically, a recent prospective study by Pahnke and colleagues (2019) investigated emotion recognition in a sample of NC ($n = 53$) and HC ($n = 42$) women and found impaired complex emotion recognition among HC relative to NC women. Further, this effect was exaggerated for expressions that were more difficult to recognize, but did not depend on the expression's valence. No differences were found between women on HCs with androgenic versus anti-androgenic properties. In a set of exploratory cross-sectional analyses, NC women were estimated to be in either the follicular and luteal phase based on forward-counting from the date of last menstrual onset, and no differences were detected in emotion recognition as a function of estimated cycle phase. Nonetheless, the authors tentatively concluded that differences between NC and HC women in emotion recognition are most likely driven by lower estradiol and progesterone in HC women (Fleischman et al., 2010). However, hormone levels were not analyzed in NC or HC women, and although Pahnke and colleagues (2019) assigned hormone values for NC women based on estimated cycle day, the high degree of variability in hormones and cycle phases across women (Cole, Ladner, & Byrn, 2009; Fehring, Schneider, & Raviele, 2006) casts doubt upon the precision of these estimations.

To address these limitations and shed further light on any associations with HC use or ovarian hormones, we examined emotion recognition in the largest sample of NC ($n = 192$) and HC ($n = 203$) women of which we are aware. We conceptually replicated the analyses of Pahnke et al. (2019) using the same test of emotion recognition, testing for both main effects of contraceptive use and interactions between contraceptive use and expression difficulty and valence. Further, we obtained and analyzed estradiol and progesterone concentrations for NC women across two test sessions and tested whether between-subjects differences and within-subjects changes in these steroid hormones modulated emotion recognition. If lowered estradiol and progesterone are indeed the proximate mechanisms underlying impaired emotion recognition in HC women (as suggested by Pahnke et al., 2019 and others), between-subject differences and within-subject changes in these hormones in NC women should predict performance on tasks of emotion recognition. We also analyzed whether between-subjects differences and within-subjects changes in testosterone modulated emotion recognition in HC women.

2.   Materials and Methods

2.1   Participants

Participants were drawn from a larger IRB-approved study conducted at a large Midwest US university, broadly designed to investigate questions related to behavioral endocrinology, psychology, and biological anthropology. Participants included in the present study are women who completed two study sessions, provided data on hormonal contraceptive use, and completed the "Reading the Mind in the Eyes Test" (RMET; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; see below). Participants were 1 sibling trio, 125 sibling pairs, and 142 singletons. NC women ($n = 192$) were scheduled for two laboratory sessions using self-reported menstrual cycle length and beginning day of last menses. One session was scheduled within one day of the anticipated peak in estradiol production during the periovulatory phase, and a second session was scheduled within two days of the anticipated peak in progesterone production during the luteal phase. (See Puts, 2006 for a description of methods used to estimate dates of peak estradiol and progesterone production.) Session order was counterbalanced across NC women. Although this approach promotes menstrual cycle-related variation across sessions in ovarian hormone levels, no presumption is made about the precision with which sessions were scheduled (see Gangestad et al., 2016 regarding the imprecision of counting-based methods); hence, hormone levels reported by targeted cycle phase in Table 1 are presented for informational rather than analytical purposes. Because our aim was to elucidate the proximate mechanisms driving putative cycle shifts, estradiol and progesterone concentrations measured objectively through salivary immunoassays are utilized as predictors (see below). HC women were scheduled for two sessions, one week apart. All sessions were scheduled between 1300h and 1600h. Basic demographic characteristics for NC and HC women can be found in Table 1.

2.2   Procedure

Reading the Mind in the Eyes Test—The RMET was administered as a measure of complex emotion recognition. In the RMET, participants view 36 black-and-white pictures of the eye region of faces, each of which is presented with four labels that

may describe the emotion expressed in the picture. Participants choose the label that best describes the emotion expressed in each picture. As original supporting documentation for the RMET (e.g., Baron-Cohen et al., 2001) does not suggest a time limit, none was imposed for this task. The total number of correct responses is recorded. Scores were also calculated separately for positive, neutral, and negative emotion subscales (Harkness, Sabbagh, Jacobson, Chowdrey, & Chen, 1999; Lischke, Lemke, Neubert, Hamm, & Lotze, 2017), as well as subscales for easy and difficult items (Domes, Heinrichs, Michel, Berger, & Herpertz, 2007).

Hormone collection—Approximately 9 mL of saliva was collected in a polystyrene test tube during each session. To minimize contamination, participants were instructed to refrain from eating, drinking, smoking, chewing gum, or brushing their teeth for one hour before each session. Participants chewed sugar-free gum, which has been found to be inert in estradiol and progesterone assays (Moffat & Hampson, 1996; however, see van Anders, 2010) to stimulate saliva flow. Samples were stored at −20° C until analysis.

Hormone assays were conducted by the Neuroendocrinology Assay Laboratory at the University of Western Ontario. Progesterone and testosterone were assayed using Coat-A-Count assay kits (Diagnostic Products Corporation, Los Angeles, CA), and estradiol was assayed using 125I Ultra-Sensitive E2 RIA DSL-4800 kits (Diagnostic Systems Laboratories, Webster, TX), all of which are commercially available. All samples were assayed in duplicate and averaged for analyses. Assay sensitivities were 0.65 pg/mL for estradiol, 5 pg/mL for progesterone, and 5–10 pg/mL for testosterone. Intra-assay coefficients of variation were 5.1% for estradiol, 10.7% for progesterone, and 6.3% for testosterone. Raw hormone values were first visually inspected for values likely attributable to assay contamination and measurement error. Values were then log-transformed to reduce skew, and standardized. Here we present analyses both including and excluding outliers (operationalized as values > 3 standard deviations from the mean).

2.3 Statistical analyses

Analyses were run using R statistical software (R Core Team, 2014), and α was set *a priori* at 0.05. We used multilevel models, nesting observations within participants, and participants within sibling pairs using the packages *lme4* (Bates, Mächler, Bolker, & Walker, 2014) and *lmertest*. For analyses comparing emotion recognition among NC and HC women, we ran two separate models. The first model's predictors included session to control for putative learning effects across sessions, age, group (NC or HC), difficulty (easy or difficult), and the interaction between group and difficulty. Similarly, the second model's predictors included session, age, group, valence (positive, negative, or neutral), and a group × valence interaction. Difficulty and valence variables were deviation coded which facilitates making ANOVA-style inferences, with one contrast term for item difficulty and two contrast terms for item valence (Barr, 2019)

For analyses testing whether hormones predict emotion recognition in NC women, estradiol and progesterone were subject-mean centered, creating variables indexing both between-subject differences (i.e., individual subject averages across sessions) and within-subject

changes in estradiol and progesterone. This model's predictors included session, age (per prior work linking age and emotion recognition; e.g., (Mill, Allik, Realo, & Valk, 2009), terms for between-subject differences in estradiol and progesterone as well as their interaction, and terms for within-subject changes in estradiol and progesterone as well as their interaction. The interaction term for within-subject changes of estradiol × within-subject changes in progesterone has been used in several previous studies aiming to use hormone concentrations to study cyclic shifts in women's socio-cognitive processes (e.g., Roney & Simmons, 2013; Shirazi et al., 2019). Such interaction terms are interpreted as follows: the magnitude and/or direction of the effect of changes in one hormone on the phenotype of interest depend on the magnitude and/or direction of changes in a second hormone. The inclusion of both between-subjects and within-subjects hormone terms allowed us to address two questions: First, do individual differences in estradiol and progesterone predict differences in emotion recognition when controlling for within-subjects fluctuations? Second, do cyclic fluctuations in estradiol and progesterone predict intraindividual changes in emotion recognition when controlling for between-subjects differences? The simultaneous inclusion of between-subjects and within-subjects model terms thus allows us to partition, and hence more fully understand, the different sources of hormonal variation that may modulate women's emotion recognition.

Similar analyses were run to test whether testosterone predicts emotion recognition in HC women. Estradiol values were assigned to women based on their self-reported hormonal contraceptive type (USDHHS, 2014; see ESM Table 1) and were treated as a between-subjects term in models. Because different generations of progestins differ in their progestational activity (Goldstuck, 2011; Sitruk-Ware, 2004), it was not possible to assign progestin values to women (Beltz, Hampson, & Berenbaum, 2015). Thus, these models addressed the following two questions: First, do individual differences in estimated estradiol and measured testosterone predict differences in emotion recognition? Second, do fluctuations in testosterone predict intraindividual changes in emotion recognition?

All data and code files have been uploaded as electronic supplementary material.

3.   Results

3.1   Difficulty-dependent group differences in emotion recognition

This primary model investigated the main effects of group and emotion difficulty, and the group × difficulty interaction, on emotion recognition while controlling for session and age. The effect of session was not significant (estimate = 0.001, $t = 1.76$, $p = 0.079$). There was a significant main effect of difficulty (estimate = 0.08, $t = 10.13$, $p < 0.001$; Figure 1), with participants scoring higher on the easy subscale than on the difficult subscale. The main effect of group (estimate = −0.003, $t = −0.30$, $p = 0.767$) and the group × difficulty interaction (estimate = 0.01, $t = 1.41$, $p = 0.160$) were not statistically significant (Fig. 1).

3.2   Valence-dependent group differences in emotion recognition

This primary model investigated the main effects of group and emotion valence, and the group × valence interaction, on emotion recognition while controlling for session and age.

The effect of session was significant (estimate = 0.01, t = 2.46, p = 0.014). Contrast terms suggested differences in emotion recognition by valence (contrast term 1 [positive versus negative items] estimate = −0.05, t = −5.51, p < 0.001; contrast term 2 [positive versus neutral items] estimate = −0.01, t = −1.68, p = 0.023). Post-hoc tests further revealed that emotion recognition accuracy was higher in neutral relative to negative emotions (estimate = 0.04, t = 6.39, p < 0.001), positive relative to neutral emotions (estimate = 0.03, t = 4.27, p < 0.001), and positive relative to negative emotions (estimate = 0.06, t = 9.46, p < 0.001; Figure 1). There was no significant effect of group in the main model (estimate < −0.001, t = −0.02, p = 0.988; Figure 1), and no significant group × valence interaction for either contrast term (group × contrast term 1 estimate = −0.02, t = −1.93, p = 0.054; group × contrast term 2 estimate = −0.02, t = −1.77, p = 0.077). Though these group × valence interactions were not significant at α = 0.05, we ran three exploratory post-hoc tests comparing emotion recognition accuracy in HC and NC women for positive, negative, and neutral emotions separately. There were no significant effects of group for positive (estimate = 0.02, t = 1.12, p = 0.263), negative (estimate = −0.01, t = −76, p = 0.447), or neutral (estimate = −0.01, t = −0.65, p = 0.516) emotions.

3.3 Hormones and emotion recognition in naturally cycling women

These analyses investigated the main effects of within-subject fluctuations in salivary progesterone, estradiol, and their interaction (with these within-subject fluctuations denoted with the ' ' symbol' in the text and in tables), and the main effects of between-subject differences (i.e., mean values across sessions) in progesterone, estradiol, and the progesterone × estradiol interaction on emotion recognition accuracy in NC women. Analyses included the aforementioned within-subjects and between-subjects terms, and their interactions with either difficulty (section 3.3.1) or emotion valence (section 3.3.2).

3.3.1 Hormones and difficulty-dependent emotion recognition in naturally cycling women—Results of this model excluding outliers are displayed in Table 2. No effects were statistically significant; in a model including these outlier values (an additional 6 observations from 3 women), there was a statistically significant three-way interaction between changes in estradiol, changes in progesterone, and difficulty (p = 0.037; p = 0.054 in model excluding outliers). To elucidate this interaction, separate models were then run for easy and difficult composites. The estimate for E × P did not significantly predict scores on the easy subscale (estimate = 0.01, t = 0.52, p = 0.603 for full sample; estimate = 0.01, t = 0.38, p = 0.705 when excluding outliers); the estimate for E × P was similarly not significant (estimate = −0.03, t = −1.36, p = 0.176 for both full sample and when excluding outliers) in predicting scores on the difficult subscale.

3.3.2 Hormones and valence-dependent emotion recognition in naturally cycling women—Results of this model excluding outliers (n = 9 observations from 3 women) are displayed in Table 3 and Figure 2. The only statistically significant effects were those for the between-subjects estradiol × contrast 2 (positive versus neutral items) interaction (estimate = −0.04, t = −2.38, p = 0.018; estimate = −0.03, t = −2.92, p = 0.022 in full sample) and between-subjects progesterone × contrast 2 interaction (estimate = 0.03, t = −2.25, p = 0.02; estimate = 0.03, t = 2.19, p = 0.029 in full sample). We ran

separate models for each valence category to elucidate these interactions. No hormone terms significantly predicted scores on the positive valence subscale. Between-subjects estradiol predicted scores on the negative valence subscale at $p = 0.057$ (estimate = -0.02, $t = -1.92$; estimate = -0.02, $t = -1.89$, $p = 0.062$ in full sample), and scores on the neutral valence subscale (estimate = -0.03, $t = -2.05$, $p = 0.042$; estimate = -0.03, $t = -2.10$, $p = 0.037$ in full sample). Between-subjects progesterone did not significantly predict scores on any individual subscale.

3.4   Hormones and emotion recognition in women using hormonal contraceptives

These analyses investigated the main effects of within-subject fluctuations in salivary testosterone (with these within-subject fluctuations denoted with the ' ' symbol' in the text and in tables), and the main effects of between-subject differences (i.e., mean values across sessions) in estimated estradiol on emotion recognition accuracy in HC women. Analyses included the aforementioned within-subjects and between-subjects terms, and their interactions with either difficulty (section 3.4.1) or emotion valence (section 3.4.2). There were no testosterone outliers.

3.4.1   Hormones and difficulty-dependent emotion recognition in women using hormonal contraceptives—The effects of between-subjects estimated estradiol (estimate = -1.97, $t = -1.51$, $p = 0.136$), between-subjects testosterone (estimate = 0.02, $t = 1.94$, $p = 0.056$), and within-subjects testosterone (estimate = -0.01, $t = -0.92$, $p = 0.357$) were not statistically significant, nor were any interactions between these hormone terms and difficulty (all $p > 0.189$).

3.4.2   Hormones and valence-dependent emotion recognition in in women using hormonal contraceptives—The effects of between-subjects estimated estradiol (estimate = -1.92, $t = -1.46$, $p = 0.147$), between-subjects testosterone (estimate = 0.02, $t = 1.87$, $p = 0.065$), and within-subjects testosterone (estimate = -0.01, $t = -1.17$, $p = 0.243$) were not statistically significant, nor were any interactions between these hormone terms and valence (all $p > 0.123$).

4.   Discussion

The present analyses sought to address several limitations of previous work on the effects of contraceptive use and the reproductive hormones estradiol and progesterone on emotion recognition in women. Previous studies have been underpowered to detect effects small-to-medium in magnitude, have relied on error-prone self-reports to estimate cycle phase, and have then relied on these error-prone estimates to infer hormone concentrations. Here, we present analyses of the largest sample of NC and HC women included in an investigation of contraceptive use and emotion recognition of which we are aware, and we utilized multiple hormone measurements taken from women.

Our results accord with previous studies suggesting no link between contraceptive use and emotion recognition (Gingnell et al., 2013; Radke & Derntl, 2016), and contradict others finding a link (Hamstra et al., 2015, 2017; Maner & Miller, 2014; Pahnke et al., 2019). Pahnke et al. (2019) hypothesize that previous studies finding no effect of contraceptive

use on emotion recognition used tasks that were too difficult, and that differences in task difficulty contributed to discrepancies in the literature. However, the present study utilized the same task as Pahnke et al. (2019), and yet unlike the study of Pahnke et al., we did not detect an effect of contraceptive use on emotion recognition. Further, measures of central tendency and spread are comparable across both datasets, making it unlikely that ceiling effects, floor effects, or issues of restricted range contribute to our discrepant findings. However, we do not consider the question of whether hormonal contraceptives influence emotion cognition answered, as it is only with the accumulation of single study estimates that meta-analyses can be performed to calculate stable population estimates of a hypothesized effect.

Between-subjects differences in estradiol and progesterone did not statistically predict emotion recognition among NC women. Though there was a statistically significant effect of between-subjects differences in estradiol on neutral emotion recognition, this effect would not survive Bonferroni correction for multiple comparisons (i.e., $\alpha/3 = 0.017$). That differences in average estradiol and progesterone levels among NC women did not predict emotion recognition makes it unlikely that previous findings of impaired emotion recognition among HC women (such as those in Pahnke et al., 2019) are explained by HC-induced decreases in these hormones. We also did not find evidence that within-subject changes in estradiol and progesterone statistically predict within-subject changes in emotion recognition. Within previous ovulatory shift research, the lack of significant effects of changes estradiol and progesterone on cognition has been interpreted as a lack of cycle shift (e.g., Roney & Simmons, 2013; Shirazi et al., 2019). Interpreted within this framework, our results suggest that emotion recognition abilities do not shift across the menstrual cycle, which is supported by some studies (Gingnell, Morell, Bannbers, Wikström, & Sundström, 2012; Rubinow, Smith, Schenkel, Schmidt, & Dancer, 2007; Zhang, Zhou, & Ye, 2013) but not others (Derntl, Hack, Kryspin-Exner, & Habel, 2013; Derntl, Kryspin-Exner, Fernbach, Moser, & Habel, 2008; Derntl et al., 2008; Rubin et al., 2012). It nevertheless remains possible that emotion recognition abilities in fact shift across the ovulatory cycle, but that these shifts are independent of fluctuating estradiol and progesterone.

Analyses of HC women also provide little support for a role of circulating steroid hormones in affecting emotion recognition. Between-subjects differences in estimated estradiol, between-subjects differences in measured testosterone, and within-subject changes in testosterone did not significantly predict emotion recognition in HC women. It is possible that the decreased emotion recognition among HC relative to NC women observed by Pahnke et al. (2019) was a result of HC-related decreases in androgens (Zimmerman et al., 2014). Although we found trends toward positive between-subjects relationships between T and our measures of emotion recognition, these relationships were not statistically significant, and we did not observe the differences in emotion recognition between HC and NC women that such relaionships would predict.

Taken together, our results provide little evidence for a role of circulating estradiol, progesterone, or testosterone in emotion recognition in adult women. However, steroid hormones are nevertheless implicated in the development of emotion recognition abilities. For example, there is a sex difference in emotion recognition with females outperforming

males (Thayer & Johnsen, 2001), and this difference can be detected during infancy and childhood (see McClure, 2000 for meta-analysis and review). Sex differences in prenatal hormonal millieu can exert large, permanent differences in cognition (Hines, 2010); similarly, within-sex perturbations in prenatal hormonal action, typically studied within the context of disorders of sexual development, also appear to exert permanent effects on cognition (e.g., Hines, Fane, Pasterski, & Mathews, 2003; Puts, McDaniel, Jordan, & Breedlove, 2008; Resnick, Berenbaum, Gottesman, & Bouchard, 1986) and on brain activation patterns related to emotion detection and processing (Ernst et al., 2007). Such cases have shown that neural and cognitive sexual differentiation is largely androgen-driven (see Puts & Motta-Mena, 2018 for review). In addition to continuing to investigate the effects of circulating hormones on emotion recognition in different samples and using different experimental paradigms, future studies may also examine patients with disorders of sexual development to elucidate the effect of hormones during certain developmental critical periods, such as during the prenatal and peripubertal windows (Berenbaum, Beltz, & Corley, 2015; Schulz, Molenda-Figueira, & Sisk, 2009; Schulz & Sisk, 2016), wherein hormones (namely, androgens) are capable of exerting permanent effects on sexualy differentiated cognitive phenotypes such as emotion cognition.

As the present study is correlational in nature, we cannot make causal claims about the effects of hormones or contraceptive use on emotion recognition. Neverthelss, if hormonal contraceptive use or circulating estradiol, progesterone, or testosterone influence emotion recognition, then we would expect to see relationships in the present data that were not apparent. Experimental, within-subject studies incorporating exogenous hormone administration, or randomized treatment to NC and HC groups, are required to determine whether a causal link exists between these variables. Such studies should also be used to investigate the effects of variables such as duration of HC use and pill phase (Radke & Derntl, 2016) on emotion recognition. Though we collected data across two sessions, data from additional sessions would contribute to more precise estimates of both within-subject and between-subject variability, and concomitantly, greater statistical power (see Gangestad et al., 2016 for discussion of effects on measurement precision on statistical power in cycle shift research). Future studies could benefit from denser sampling schedules and a greater number of observations per subject. Finally, as different women may be differentially sensitive to HCs as well as to fluctuations in endogenous hormones (Pope, Oinonen, Mazmanian, & Stone, 2017), it is possible that different relationships between hormones, contraceptive use, and complex emotion recognition (and other cognitive traits, more broadly) would emerge if women were stratified by their sensitivity to hormones.

In conclusion, the present study contributes to the important, growing literature investigating the psychobehavioral effects of OH use, and of within-subject changes and between-subject differences in reproductive hormones. We find no evidence that emotion recognition differs between NC and HC women, and no evidence that between-subject differences and within-subject changes in circulating estradiol, progesterone, or testosterone predict emotion recognition. At a basic level, an understanding of how these hormones modulate various aspects of women's cognition may guide subsequent work on links between hormones and brain structure and function, and on hormones and social functioning. Clinically, knowledge

of such effects is crucial for physicians to guide women in making informed decisions about hormone-altering medications.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

7. References

Baron-Cohen S, Wheelwright S, Hill J, Raste Y, & Plumb I (2001). The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrom or high-functioning autism. Journal of Child Psychology and Psychiatry, 42, 241–251. [PubMed: 11280420]

Barr D (n.d.). Coding categorical predictor variables in factorial designs. Retrieved from http://talklab.psy.gla.ac.uk/tvw/catpred/

Bates DM, Mächler M, Bolker BM, & Walker SC (2014). lme4: Linear mixed-effects models using eigen and S4. Retrieved from http://cran.r-project.org/package=lme4

Beltz AM, Hampson E, & Berenbaum SA (2015). Oral contraceptives and cognition: A role for ethinyl estradiol. Hormones and Behavior, 74, 209–217. 10.1016/j.yhbeh.2015.06.012 [PubMed: 26122296]

Berenbaum SA, Beltz AM, & Corley R (2015). The Importance of Puberty for Adolescent Development: Conceptualization and Measurement. Advances in Child Development and Behavior (1st ed., Vol. 48). Elsevier Inc. 10.1016/bs.acdb.2014.11.002

Blake KR, Dixson BJW, O'Dean SM, & Denson TF (2016). Standardized protocols for characterizing women's fertility: A data-driven approach. Hormones and Behavior, 81, 74–83. 10.1016/j.yhbeh.2016.03.004 [PubMed: 27072982]

Cobey KD, Little AC, & Roberts SC (2015). Hormonal effects on women's facial masculinity preferences: The influence of pregnancy, post-partum, and hormonal contraceptive use. Biological Psychology, 104, 35–40. 10.1016/j.biopsycho.2014.11.002 [PubMed: 25460829]

Cole LA, Ladner DG, & Byrn FW (2009). The normal variabilities of the menstrual cycle. Fertility and Sterility, 91(2), 522–527. 10.1016/j.fertnstert.2007.11.073 [PubMed: 18433748]

Derntl B, Hack RL, Kryspin-Exner I, & Habel U (2013). Association of menstrual cycle phase with the core components of empathy. Hormones and Behavior, 63(1), 97–104. 10.1016/j.yhbeh.2012.10.009 [PubMed: 23098806]

Derntl B, Kryspin-Exner I, Fernbach E, Moser E, & Habel U (2008). Emotion recognition accuracy in healthy young females is associated with cycle phase. Hormones and Behavior, 53, 90–95. 10.1016/j.yhbeh.2007.09.006 [PubMed: 17976599]

Derntl B, Windischberger C, Robinson S, Lamplmayr E, Kryspin-exner I, & Gur RC (2008). Facial emotion recognition and amygdala activation are associated with menstrual cycle phase. Psychoneuroendocrinology, 33, 1031–1040. 10.1016/j.psyneuen.2008.04.014 [PubMed: 18675521]

Domes G, Heinrichs M, Michel A, Berger C, & Herpertz SC (2007). Oxytocin improves "mind-reading" in humans. Biological Psychiatry, 61, 731–733. [PubMed: 17137561]

Ernst M, Maheu S, Schroth E, Hardin J, Green L, Cameron J, … Merke DP (2007). Amygdala function in adolescents with congenital adrenal hyperplasia: A model for the study of early steroid abnormalities. Neuropsychologia, 45, 2104–2113. 10.1016/j.neuropsychologia.2007.01.019 [PubMed: 17336344]

Fehring RJ, Schneider M, & Raviele K (2006). Variability in the phases of the menstrual cycle. J Obstet Gynecol Neonatal Nurs, 35(3), 376–384. 10.1111/j.15526909.2006.00051.x

Fleischman DS, Navarrete CD, & Fessler DMT (2010). Oral contraceptives suppress ovarian hormone production. Psychological Science, 21(5), 750–752. 10.1177/0956797610368062 [PubMed: 20483856]

Gangestad SW, Haselton MG, Welling LLM, Gildersleeve K, Pillsworth EG, Burriss RP, … Puts DA (2016). How valid are assessments of conception probability in ovulatory cycle research? Evaluations, recommendations, and theoretical implications. Evolution and Human Behavior, 37(2), 85–96. 10.1016/j.evolhumbehav.2015.09.001

Gingnell M, Engman J, Frick A, Moby L, Fredrikson M, Sundstro I, & Wikstro J (2013). Oral contraceptive use changes brain activity and mood in women with previous negative affect on the pill —— A double-blinded, placebo-controlled randomized trial of a levonorgestrel-containing combined oral contraceptive. Psychoneuroendocrinology, 38, 1133–1144. 10.1016/j.psyneuen.2012.11.006 [PubMed: 23219471]

Gingnell M, Morell A, Bannbers E, Wikström J, & Sundström I (2012). Menstrual cycle effects on amygdala reactivity to emotional stimulation in premenstrual dysphoric disorder. Hormones and Behavior, 62(4), 400–406. 10.1016/j.yhbeh.2012.07.005 [PubMed: 22814368]

Goldstuck N (2011). Progestin potency – Assessment and relevance to choice of oral contraceptives. Middle East Fertility Society Journal, 16(4), 248–253. 10.1016/j.mefs.2011.08.006

Gonzales JE, & Ferrer E (2016). Efficacy of methods for ovulation estimation and their effect on the statistical detection of ovulation-linked behavioral fluctuations. Behavior Research Methods, 48(3), 1125–1144. 10.3758/s13428-015-0638-4 [PubMed: 26208814]

Griksiene R, & Ruksenas O (2011). Effects of hormonal contraceptives on mental rotation and verbal fluency. Psychoneuroendocrinology, 36(8), 1239–1248. 10.1016/j.psyneuen.2011.03.001 [PubMed: 21454017]

Hamstra DA, Kloet ER De Hemert, Van AM, De Rijk, R. H., & van der Does AJW (2015). Mineralocorticoid receptor haplotype, oral contraceptives and emotional information processing. Neuroscience, 286, 412–422. 10.1016/j.neuroscience.2014.12.004 [PubMed: 25497375]

Hamstra DA, Kloet E. R. De, Quataert I, Jansen M, & Does W. Van Der. (2017). Mineralocorticoid receptor haplotype, estradiol, progesterone and emotional information processing. Psychoneuroendocrinology, 76, 162–173. 10.1016/j.psyneuen.2016.11.037 [PubMed: 27936434]

Hamstra DA, De Rover, M., De Rijk, R. H., & Van der Does, W. (2014). Oral contraceptives may alter the detection of emotions in facial expressions. European Neuropsychopharmacology, 24(11), 1855–1859. 10.1016/j.euroneuro.2014.08.015 [PubMed: 25224104]

Harkness K, Sabbagh M, Jacobson J, Chowdrey N, & Chen T (1999). Enhanced accuracy of mental state decoding in dysphoric college students. Cognition and Emotion, 19, 99–1025.

Hines M (2010). Sex-related variation in human behavior and the brain. Trends in Cognitive Sciences, 14(10), 448–456. 10.1016/j.tics.2010.07.005 [PubMed: 20724210]

Hines M, Fane BA, Pasterski VL, & Mathews GA (2003). Spatial abilities following prenatal androgen abnormality: Targeting and mental rotations performance in individuals with congenital adrenal hyperplasia. Psychoneuroendocrinology, 28, 1010–1026. 10.1016/S0306-4530(02)00121-X [PubMed: 14529705]

Kamboj SK, Krol KM, & Curran HV (2015). A specific association between facial disgust recognition and estradiol levels in naturally cycling women. PLoS ONE, 1–12. 10.1371/journal.pone.0122311

Lischke A, Lemke D, Neubert J, Hamm AO, & Lotze M (2017). Inter-individual differences in heart rate variability are associated with inter-individual differences in mindreading. Scientific Reports, 7, 1–6. 10.1038/s41598-017-11290-1 [PubMed: 28127051]

Maner JK, & Miller SL (2014). Hormones and social monitoring: Menstrual cycle shifts in progesterone underlie women's sensitivity to social information. Evolution and Human Behavior, 35, 9–16. 10.1016/j.evolhumbehav.2013.09.001

Marcinkowska UM, Hahn AC, Little AC, Debruine LM, & Jones BC (2019). No evidence that women using oral contraceptives have weaker preferences for masculine characteristics in men's faces. PLoS ONE, 1–8.

McClure EB (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. Psychological Bulletin, 126(3), 424–453. [PubMed: 10825784]

Mill A, Allik J, Realo A, & Valk R (2009). Age-related differences in emotion recognition ability: A cross-sectional study. Emotion, 9(5), 619–630. [PubMed: 19803584]

Moffat SD, & Hampson E (1996). Salivary testosterone levels in left- and right-handed adults. Neuropsychologia, 34(3), 225–233. [PubMed: 8868279]

Montoya ER, & Bos PA (2017). How oral contraceptives impact social-emotional behavior and brain function. Trends in Cognitive Sciences, 21(2), 125–136. 10.1016/j.tics.2016.11.005 [PubMed: 28089524]

Osório FL, Cassis JMDP, & Sousa J. P. M. De. (2018). Sex hormones and processing of facial expressions of emotion: A systematic literature review. Frontiers in Psychology, 9, 1–14. 10.3389/fpsyg.2018.00529 [PubMed: 29410639]

Pahnke R, Mau-moeller A, Junge M, Wendt J, Weymar M, Hamm AO, & Lischke A (2019). Oral contraceptives impair complex emotion recognition in healthy women. Frontiers in Neuroscience, 12, 1–9. 10.3389/fnins.2018.01041

Pastor Z, Holla K, & Chmel R (2013). The influence of combined oral contraceptives on female sexual desire: a systematic review. Contraception, 18(1), 27–43. 10.3109/13625187.2012.728643

Pearson R, & Lewis MB (2005). Fear recognition across the menstrual cycle. Hormones and Behavior, 47, 267–271. 10.1016/j.yhbeh.2004.11.003 [PubMed: 15708754]

Petersen N, Touroutoglou A, Andreano JM, & Cahill L (2015). Oral contraceptive pill use is associated with localized decreases in cortical thickness. Human Brain Mapping, 36(7), 2644–2654. 10.1002/hbm.22797 [PubMed: 25832993]

Pletzer BA, & Kerschbaum HH (2014). 50 years of hormonal contraception - Time to find out, what it does to our brain. Frontiers in Neuroscience, 8(8 JUL), 1–6. 10.3389/fnins.2014.00256 [PubMed: 24478622]

Pope CJ, Oinonen K, Mazmanian D, & Stone S (2017). The hormonal sensitivity hypothesis: A review and new findings. Medical Hypotheses, 102, 69–77. 10.1016/j.mehy.2017.03.012 [PubMed: 28478835]

Puts DA (2006). Cyclic variation in women's preferences for masculine traits: Potential hormonal causes. Human Nature, 17(1), 114–127. 10.1007/s12110-0061023-x [PubMed: 26181348]

Puts DA, McDaniel MA, Jordan CL, & Breedlove SM (2008). Spatial ability and prenatal androgens: Meta-analyses of congenital adrenal hyperplasia and digit ratio (2D:4D) studies. Archives of Sexual Behavior, 37, 100–111. 10.1007/s10508-007-9271-3 [PubMed: 18074217]

Puts DA, & Motta-Mena NV (2018). Is human brain masculanization estrogen receptormediated? Reply to Luoto and Rantala. Hormones and Behavior, 97, 3–4. [PubMed: 28818501]

R Core Team. (2014). R: A language and environment for statistical computing. Retrieved from http://www.r-project.org

Radke S, & Derntl B (2016). Affective responsiveness is influenced by intake of oral contraceptives. European Neuropsychopharmacology, 26(6), 1014–1019. 10.1016/j.euroneuro.2016.03.004 [PubMed: 27039036]

Resnick SM, Berenbaum SA, Gottesman II, & Bouchard TJ (1986). Early hormonal influences on cognitive functioning in congenital adrenal hyperplasia. Developmental Psychology, 22(2), 191–198. 10.1037/0012-1649.22.2.191

Roney JR, & Simmons ZL (2013). Hormonal predictors of sexual motivation in natural menstrual cycles. Hormones and Behavior, 63(4), 636–645. 10.1016/j.yhbeh.2013.02.013 [PubMed: 23601091]

Rubin LH, Carter CS, Drogos L, Jamar R, Pournajafi-Nazarloo H, Sweeney JA, & Maki PM (2012). Sex-specific associations between peripheral oxytocin and emotion perception in schizophrenia. Schizophrenia Research, 130(Mc 913), 266–270. 10.1016/j.schres.2011.06.002.Sex-specific

Rubinow DR, Smith MJ, Schenkel LA, Schmidt PJ, & Dancer K (2007). Facial emotion discrimination across the menstrual cycle in women with Premenstrual Dysphoric Disorder (PMDD) and controls. Journal of Affective Disorders, 104, 37–44. 10.1016/j.jad.2007.01.031 [PubMed: 17367867]

Sanders SA, Graham CA, Bass JL, & Bancroft J (2001). A prospective study of the effects of oral contraceptives on sexuality and well-being and their relationship to discontinuation. Contraception, 64(1), 51–58. 10.1016/S00107824(01)00218-9 [PubMed: 11535214]

Schulz KM, Molenda-Figueira HA, & Sisk CL (2009). Back to the future: The organizational-activational hypothesis adapted to puberty and adolescence. Hormones and Behavior, 55(5), 597–604. 10.1016/j.yhbeh.2009.03.010 [PubMed: 19446076]

Schulz KM, & Sisk CL (2016). The organizing actions of adolescent gonadal steroid hormones on brain and behavioral development. Neuroscience and Biobehavioral Reviews, 70, 148–158. 10.1016/j.neubiorev.2016.07.036 [PubMed: 27497718]

Shirazi TN, Self H, Dawood K, Rosen KA, Penke L, Carré JM, … Puts DA (2019). Hormonal predictors of women's sexual motivation. Evolution and Human Behavior. 10.1016/j.evolhumbehav.2019.02.002

Sitruk-Ware R (2004). Pharmacological profile of progestins. Maturitas, 47(4), 277–283. 10.1016/j.maturitas.2004.01.001 [PubMed: 15063480]

Skovlund CW, Mørch LS, Kessing LV, & Lidegaard Ø (2016). Association of hormonal contraception with depression. JAMA Psychiatry, 73(11), 1154. 10.1001/jamapsychiatry.2016.2387 [PubMed: 27680324]

Thayer J, & Johnsen BH (2001). Sex differences in judgement of facial affect: A multivariate analysis of recognition errors. Scandinavian Journal of Psychology, 41(3), 243–246.

USDHHS. (2014). Approved drug products with therapeutic equivalence evaluations. Washington, DC.

van Anders SM (2010). Chewing gum has large effects on salivary testosterone, estradiol, and secretory immunoglobulin A assays in women and men. Psychoneuroendocrinology, 35(2), 305–309. [PubMed: 19615825]

van Honk J, & Schutter D (2007). Testosterone reduces conscious detection of signals serving social correction: Implications for antisocial behavior. Psychological Science, 18(8), 663–667. [PubMed: 17680933]

van Wingen GA, van Broekhoven F, Verkes RJ, Petersson KM, Backstrom T, Buitelaar JK, & Fernandez G (2008). Progesterone selectively increases amygdala reactivity in women. Molecular Psychiatry, 13, 325–333. 10.1038/sj.mp.4002030 [PubMed: 17579609]

Zethraeus N, Dreber A, Ranehill E, Blomberg L, Labrie F, & Johannesson M (2017). A first-choice combined oral contraceptive in fluences general well-being in healthy women: A double-blind, randomized, placebo-controlled trial. Fertility and Sterility, 107(5), 1238–1245. 10.1016/j.fertnstert.2017.02.120 [PubMed: 28433366]

Zhang W, Zhou R, & Ye M (2013). Menstrual cycle modulation of the late positive potential evoked by emotional faces. Perceptual and Motor Skills, 116(3), 707–723. [PubMed: 24175446]

Zimmerman Y, Eijkemans MJ, Coelingh B, Blankstein MA, & Fauser BC (2014). The effect of combined oral contraception on testosterone levels in healthy women: a systematic review and meta-analysis. Human Reproduction Update, 20(1), 76–105. [PubMed: 24082040]

Highlights

Investigated how endogenous exogenous hormones affect emotion recognition (ER)

2 sessions from 193 naturally-cycling women and 203 on oral contraception

Oral contraceptive use did not modulate ER

Individual differences in estradiol, progesterone, and testosterone did not predict ER

Within-subject sex hormone changes across sessions did not predict ER
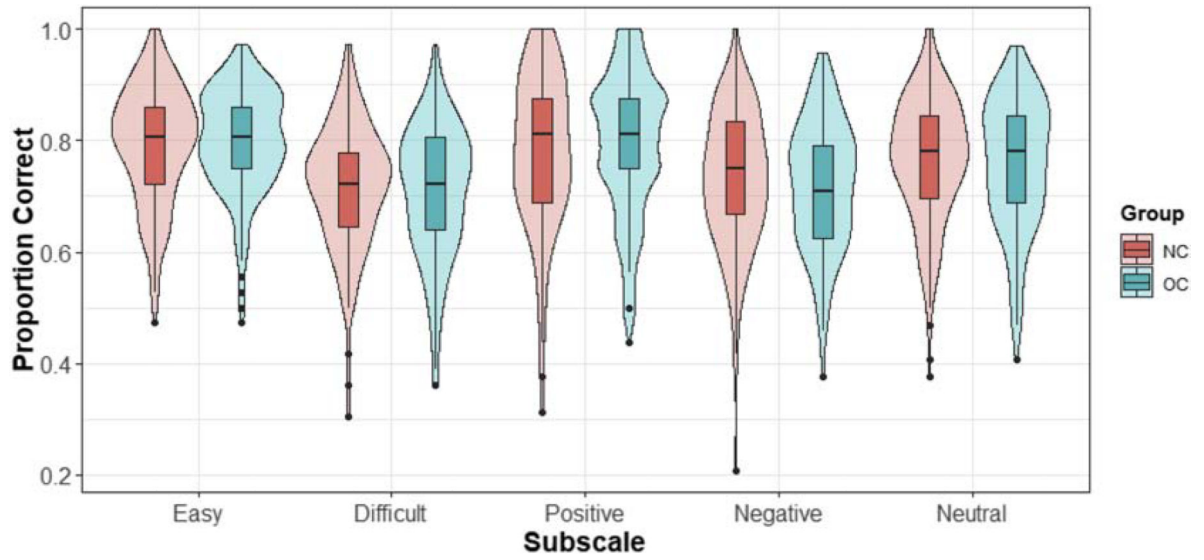
Figure 1.
Emotion recognition accuracy scores displayed by subscale and group (NC versus OC). Boxplot notches represent 25th, 50th, and 75th percentiles. Dots represent points > 1.5 and < 3 times the interquartile range below the 25th percentile. Violin plots reflect data distributions.
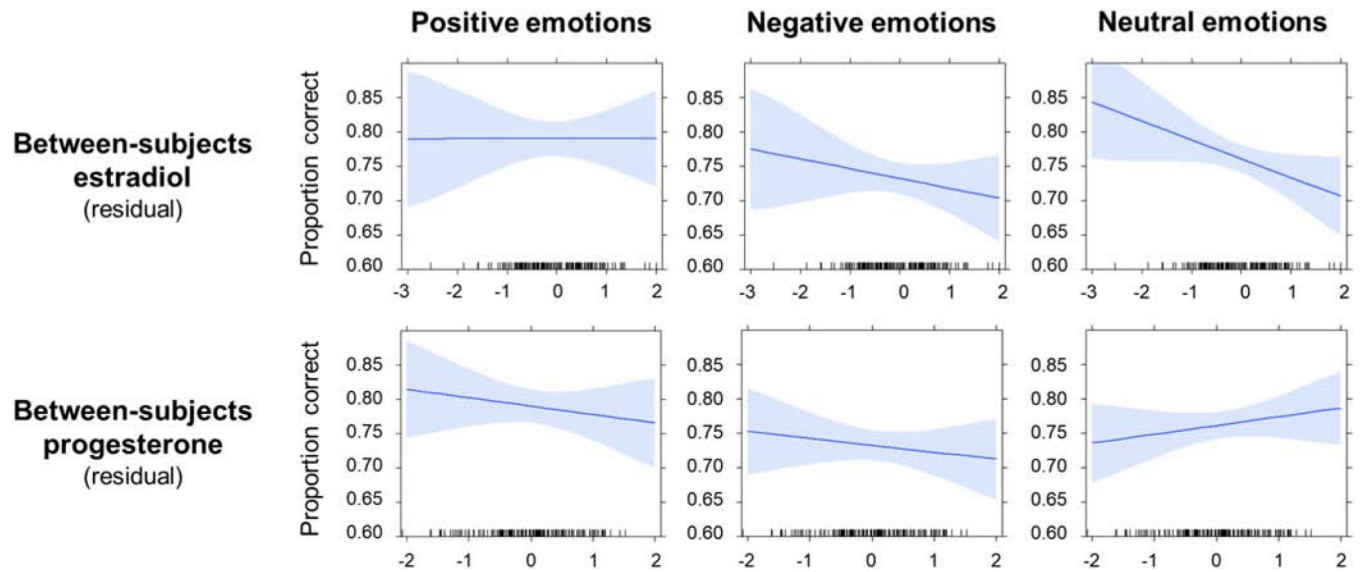
Figure 2.
Isolated effects of between-subjects estradiol and between-subjects progesterone on positive, negative, and neutral emotion subscales from models excluding outliers. Blue lines represent estimated isolated effect of hormone and shading indicates 95% confidence interval. The *x*-axis represents within-subject (residual) changes in estradiol (top row) and progesterone (bottom row).

Table 1.

Participant characteristics.

	NC women (n = 192)	HC women (n = 203)
Mean age (SE)	19.9 (0.11)	20.0 (0.12)
Race/ethnicity		
White(%)	178 (92.7%)	186 (91.6%)
Asian (%)	5 (2.6%)	8 (3.9%)
Other (%)	9 (4.7%)	9 (4.4%)
Mean estradiol (SE) pg/mL, session 1	2.09 (0.08)	n/a
Mean estradiol (SE) pg/mL, session 2	1.49 (0.06)	n/a
Mean unsigned cross-session estradiol change (SE) pg/mL	0.92 (0.07)	n/a
Mean estradiol (SE) pg/mL, targeted follicular phase	1.79 (0.08)	n/a
Mean estradiol (SE) pg/mL, targeted luteal phase	1.83 (0.07)	n/a
Mean progesterone (SE) pg/mL, session 1	81.09 (4.98)	n/a
Mean progesterone (SE) pg/mL, session 2	65.63 (4.06)	n/a
Mean unsigned cross-session progesterone change (SE) pg/mL	52.52 (5.06)	n/a
Mean progesterone (SE) pg/mL, targeted follicular phase	61.86 (3.45)	n/a
Mean progesterone (SE) pg/mL, targeted luteal phase	86.32 (5.52)	n/a
Mean testosterone (SE) pg/mL, session 1	n/a	18.07 (0.74)
Mean testosterone (SE) pg/mL, session 2	n/a	16.00 (0.90)

Table 2.

Results of models excluding hormone outliers assessing the link between within-subjects and between-subjects estradiol and progesterone terms (ΔE and ΔP and E and P, respectively), and interactions with item difficulty. Values are controlled for main effect of difficulty, session, and age.

	Estimate	t	p
Session	<0.01	0.80	0.425
ΔE	−0.01	−1.32	0.187
ΔP	0.004	0.66	0.510
ΔE × ΔP	−0.01	−0.73	0.466
E	−0.02	−1.48	0.141
P	0.001	0.12	0.904
E × P	0.01	0.67	0.502
ΔE × difficulty	−0.01	−0.71	0.477
ΔP × difficulty	−0.002	−0.13	0.899
ΔE × ΔP × difficulty	0.04	1.94	0.054
E × difficulty	0.008	0.68	0.500
P × difficulty	0.004	0.34	0.734
E × P × difficulty	0.004	0.28	0.781

Table 3.

Results of models excluding hormone outliers assessing the link between within-subjects and between-subjects estradiol and progesterone terms (ΔE and ΔP and E and P, respectively), and interactions with item valence. Values are controlled for main effect of valence, session, and age. Bold p values represent p < 0.05. Contrast 1 represents the term in linear models comparing positive and negative emotions, and contrast 2 represents the term comparing positive and neutral emotions.

	Estimate	t	p
Session	0.01	1.09	0.275
ΔE	−0.01	−1.52	0.130
ΔP	0.005	0.80	0.426
ΔE × ΔP	−0.02	−0.80	0.423
E	−0.01	−1.18	0.241
P	−0.001	−0.12	0.907
E × P	0.006	0.53	0.596
ΔE × contrast 1	0.001	0.09	0.928
ΔE × contrast 2	0.02	1.38	0.168
ΔP × contrast 1	−0.005	−0.33	0.742
ΔP × contrast 2	−0.007	−0.46	0.648
ΔE × ΔP × contrast 1	−0.009	−0.34	0.737
ΔE × ΔP × contrast 2	−0.002	−0.06	0.955
E × contrast 1	−0.02	−1.54	0.124
E × contrast 2	−0.04	−2.38	**0.018**
P × contrast 1	0.009	0.59	0.552
P × contrast 2	0.03	2.25	**0.029**
E × P × contrast 1	0.006	0.37	0.713
E × P × contrast 2	0.02	1.15	0.250