Testosterone reactivity to infant crying and caregiving in women: The role of oral contraceptives and basal cortisol.

Voorthuis, A.; Bakermans-Kranenburg, M.J.; IJzendoorn, M.H. van

Testosterone reactivity to infant crying and caregiving in women: The role of oral contraceptives and basal cortisol

A. Voorthuis, M.J. Bakermans-Kranenburg, & M.H. van IJzendoorn

Several studies have shown that mothers and fathers have significant lower levels of testosterone (T) than non-mothers and non-fathers, and that caregiving is related to a decrease in T in males. To date, only a few studies have examined T in women. We examined T reactivity to a crying infant simulator in 160 women. Use of oral contraceptives (OC), basal cortisol (CORT) levels and childhood experiences of maternal love withdrawal were taken into account. T was significantly higher in women not using OC, and in women not using OC, participants with high basal CORT showed higher initial T levels and larger decreases of T during caregiving than participants with low basal CORT. No effect of CORT was found in women with OC use. Childhood experiences of maternal love withdrawal did not affect T levels. This is the first study to show support for a decrease of T in women while taking care of a crying infant, supporting the *Challenge hypothesis* and the *Steroid/Peptide Theory of Social Bonds*.

Keywords: Testosterone in women, basal Cortisol, Caregiving, Crying infant, Oral Contraceptives

1. Introduction

In many species, higher testosterone (T) levels are linked to increased mating efforts (Wingfield, Hegner, Dufty, & Ball, 1990), whereas lower T levels are associated with increased parenting effort. Indeed, research with humans shows that fathers have significantly lower T than non-fathers (Gray, Yang, & Pope, 2006; Kuzawa, Gettler, Muller, McDade, & Feranil, 2009) and that T is lower in mothers than in non-mothers (Barret et al., 2013; Kuzawa, Gettler, Huang, & McDade, 2010). Compared to males with higher baseline T, males with lower baseline T show more paternal empathy in response to infant cries (Fleming, Corter, Stallings, & Steiner, 2002), and more affectionate touch, vocalization and gaze during father-child interaction (Weisman, Zagoory-Sharon, & Feldman, 2014). Males with lower basal T also hold a baby doll longer and are more responsive to infant cues (Storey, Walsh, Quinton, & Wynne-Edwards, 2000). Moreover, males who are more responsive to infant cues have a greater decrease in T in response to holding a baby doll and listening to infant crying sounds (Storey et al., 2000). These findings are in line with the *Challenge hypothesis* which states that T levels increase in males in the context of competition and decrease in the context of caregiving (Archer, 2006).

However, evidence for the Challenge hypothesis is not unequivocal. One study found that T levels increased rather than decreased in reaction to infant crying (Fleming et al., 2002) and that administration of T enhanced neural responsivity to baby cries in women (Bos, Hermans, Montoya, Ramsey, & Van Honk, 2010). In response to these inconsistencies, Van Anders, Tolman and Volling (2012) suggested that the *Steroid/Peptide Theory of Social Bonds* (S/P theory) may provide a better framework for the understanding of T responsivity to baby cues than the Challenge hypothesis. The S/P theory states that not all parenting behaviors are related to low T: only those behaviors that are nurturing would be linked with lower T, while parenting behaviors related to challenge or threat would be linked to higher T (Van Anders, Goldey, & Kuo, 2011). In an experimental study, they showed that infant crying was related to decreases in T when the cries were coupled with an effective nurturing response, and to increases in T when unaccompanied by nurturing behavior (Van Anders et al., 2012). However, further examination of the data showed that almost half of the participants (46%) did not show a decrease in T when the crying was coupled with an effective nurturing response or an increase in T when the crying was uncoupled with nurturing behavior (48%), suggesting that the context of effective nurturing versus no nurturing accounts for only part of the variance in T reactivity. In the present study we aim to shed more light on individual differences in T reactivity in response to caring for a crying infant simulator.

Individual differences in T reactivity may be explained by accompanying levels of cortisol (CORT) as proposed in the *Dual-hormone hypothesis*. The Dual-hormone hypothesis proposes that basal CORT and T have a joint effect on behavioral systems implicated in status-relevant behavior

and aggression. According to this hypothesis, traits generally associated with high T manifest more in individuals with low basal CORT (Mehta & Josephs, 2010). This interaction occurs because CORT can inhibit the secretion of T at all levels of the hypothalamic-pituitary-gonadal axis (HPG axis), and T can inhibit CORT secretion by acting upon the hypothalamus (Viau, 2002). Indeed, studies show that basal CORT interacts with T to predict dominant behavior in both males and women (Mehta & Josephs, 2010), social status in women (Edwards & Casto, 2013), and criminal violence (Dabbs, Jurkovic, & Frady, 1991) and aggression in adolescent males (Popma et al., 2007). The Dual-hormone hypothesis may also apply to other behavioral systems. Zilioli, Ponzi, Henry and Maestripieri (2014) showed that self-reported empathy is related to the joint activation of T and CORT. In individuals with low basal CORT, higher T relates to lower empathy, whereas individuals with higher basal CORT show a reversed association between T and empathy. Other recent studies also show an interaction of basal CORT with T, but in the opposite direction. For example, T and psychopathy were positively correlated in men with high CORT and negatively correlated in men with low CORT (Welker, Lozoya, Campbell, Neumann, & Carré, 2014). In women, T predicted reactive aggression and state dominance only in individuals with high basal CORT (Denson, Mehta, & Ho Tan, 2013). In addition, CORT also seems to be related to T reactivity. Increases in T in response to the Trier Social Stress Test were found to be larger for participants with lower basal CORT (Bedgood, Boggiano, & Turan, 2014). As CORT is also implicated in caregiving behavior (e.g., Fleming, Steiner, & Corter, 1997; Swain, Kim, & Ho, 2011), and caregiving challenges such as a crying baby might lead to changes in T, further examination of the interplay of basal CORT with T in the context of caregiving is needed.

Animal research shows that the Hypothalamic-Pituitary-Adrenal (HPA) axis, the HPG axis and their interaction can be affected by experiences with maternal care early in life (Toufexis, Rivarola, Lara, & Viau, 2014). The role of experienced maternal love withdrawal in explaining the variance in effects of oxytocin administration on behavior has been documented (for a review, see Bakermans-Kranenburg & Van IJzendoorn, 2013). Love withdrawal is a parental strategy that involves withholding love and affection to correct or punish the child. Studies indicate that the positive effects of oxytocin may be most prominent in individuals with low levels of experienced maternal love withdrawal (e.g., Van IJzendoorn, Huffmeijer, Alink, Bakermans-Kranenburg & Tops, 2011; Riem et al., 2013). Experiences of maternal love withdrawal may also influence T reactivity in the context of caregiving.

Only a few studies examined the relation of T and caretaking in women. In women, T is secreted by different organs and through different physiological systems than in males. Males have significant higher T levels than women (e.g., Endendijk et al., 2016), more variability in T and a more pronounced and more stable diurnal rhythm (Granger et al., 2003). Using presentations of facial stimuli as proxies for social interactions (e.g., Van Honk, et al., 2000), differences between males and women in the relation between T and social behavior have been found. While women show an

increase in T when viewing angry faces compared to happy faces, this increase is not present in males (Zilioli, Caldbrick, & Watson, 2014). In women T was found to increase the reward value of infant faces (Hahn, DeBruine, Fisher, & Jones, 2015) and exogenous T enhanced the neural responsivity to baby cries (Bos et al., 2010). Moreover, the T system might act differently in relation to parenting behavior in males and women (Endendijk et al., 2016).

Oral contraceptive (OC) use is known to lower T levels because OCs suppress ovarian steroid production, and increases sex hormone binding globulin (SHBG) resulting in a reduction in unbound or free T (e.g. Boyd, Zegarac, Posvar, & Flack, 2001; Jung-Hoffman & Kuhl, 1987; Thorneycroft et al., 1999; Van der Vange, Blankenstein, Kloosterboer, Haspels, & Thijssen, 1990). OC use is associated with lower basal T levels in many studies (e.g., Crewter, Hamilton, Casto, Kilduff, & Cook, 2015; Edwards & O'Neal, 2009; Liening, Stanton, Saini, & Schultheiss, 2010; Vibarel-Rebot, Rieth, Lasne, Jaffré, & Collomp, 2015), but the relation of OC use with T responsivity is less clear. For example, one study reported that female athletes with OCs show virtually the same increase in T over the course of the competition as women without OCs (Edwards & O'Neal, 2009), whereas another study showed reduced T responses to training and competition activities in women athletes with OCs compared to women athletes without OCs (Crewter et al., 2015).

In the present study we tested whether and how basal CORT, basal T, OC use and childhood experiences of maternal love withdrawal influence T reactivity in response to infant crying. First, we examined T reactivity in response to infant crying and caretaking, and we explored the role of OC use. Second, we examined the role of basal CORT in relation to T reactivity, and we explored the influence of childhood experiences on T reactivity taking basal CORT into account.

2. Method

2.1. Participants

A total of 353 undergraduate students from the departments of education and child studies and psychology at Leiden University participated in the first phase of the study. In this phase, the participants completed online questionnaires on caregiving experiences and some demographic details. One participant was excluded due to random responses. Ten male participants and five women with children of their own were excluded to avoid variance in gender and parenting experience. The remaining 337 participants were then invited for the second phase of the study in which they had to take care of a 'life-like' infant simulator for two evenings at home, and during a 30-min lab session. One hundred and eighty-six participants (55.2%) were willing to participate (see also Voorthuis et al., 2013); they did not differ from the non-participants on demographic characteristics (all $p > .05$). Twenty participants were excluded due to medication use (antidepressants, stimulants, antibiotics and steroids) and 6 participants were excluded due to smoking (more than 2 cigarettes less than 2 hours before the session [see Kuldielka, Hellhammer, & Wust, 2009]), resulting in a total

of 160 healthy nulliparous female participants, aged between 17 and 28 years (*M* = 19.83, *SD* =1.59) in the current study. More than 95% of the participants were born in the Netherlands. Sixty-six percent (n = 105) used OCs .

2.2 Procedure and materials

2.2.1. Procedure

Participants were asked to take care of an infant simulator for two evenings at their home (5 p.m. - 10 p.m.) and during a lab session as if it were their own baby. During the two evenings, the simulator was programmed with various cry episodes so that participants could get familiar with the infant simulator (Bhandari et al., 2013). Following the two evenings, the participants took care of the infant simulator in a 30-minute lab session with increasing competing demands (10 min free play, 10 min mildly competing demand, 10 min moderately competing demand, for details see Appendix and Voorthuis et al., 2013). The lab room was decorated as a living room with a dresser with a changing mat, a baby cot, an infant carrier, a desk with a chair, an easy chair, magazines, children's books, and toys. The participants were told that they were allowed to use all materials in the room while taking care of the simulator as if it was a real infant. Participants were instructed not to use alcohol, caffeine, or cigarettes on the day of the lab session. The session started either at 11.00 (morning group) or 15.00 hours (afternoon group). In the hour before the collection of the first saliva sample, participants were instructed about the lab session, they practiced saliva collection, and had a resting period (reading neutral magazines), during which baseline measures of cardiovascular activity were obtained. Saliva samples were collected four times during the session: one hour after arrival, before taking care of the infant simulator (T1; baseline), immediately after care-taking (T2), 15 minutes after care-taking (T3), and 30 minutes after care-taking (T4). During the caretaking in the mildly competing demand episode, participants completed a questionnaire on potential physiological confounds such as menstrual cycle, OC use, hearing quality, psychiatric disorders, alcohol use and smoking, and use of medication. The procedure was approved by the Departmental Ethics Committee and informed consent was obtained from all participants.

2.2.2. Infant simulator

The infant simulator (RealCare Baby II-Plus; Realityworks, Eau Claire, WI, USA) is a doll resembling a real infant of about 6 weeks in physical appearance, size and weight (2.95 kg). The physical features of the infant simulator are modeled in a way that requires taking care like one would do with a real infant. For example, it has a lifelike neck that falls back if not supported. The doll is programmed as an infant with realistic cry sounds (beginning with mild fussing eventually increasing to full-blown crying) as well as breathing, burping, giggling and suckling sounds. Moreover, cry episodes are programmed for which specific caregiving activity is required, such as feeding,

burping, diaper change, and rocking. The schedules (cry frequencies, gap between cry episodes, etc.) are based on diaries kept by parents of real infants. Starting with fussing, the cry sounds progressively become more intense in case of no care or inappropriate care. The infant simulator responds with a short giggle or by being calm again if the appropriate care is provided. Special inbuilt sensors in combination with external sensors (in diapers and feeding bottles) register whether correct caregiving was provided or not. For the home sessions, a "difficult baby" schedule was used in order to have sufficient cry episodes. During the lab session, the sensor was not used and the infant simulator could not register care. As a result the crying started with fussing and increased to full-blown crying because caregiving attempts were ineffective. The infant simulator would cry for approximately five minutes (M = 4.72, SD = 0.35) of each of the three 10 min observations (Free play, Mildly competing demand, Moderately competing demand). After about four minutes, the infant started with 30 seconds of whining, followed by three minutes of intense crying, five seconds quiet, one minute crying, eight seconds quiet, one minute crying, and finally one minute silent. The doll used in our study was representing a Caucasian female infant (Voorthuis et al., 2013).

2.3 Hormone assays

Participants rinsed their mouths with water before providing the first saliva sample. They provided up to 2 ml of saliva via unstimulated passive drool into a sterile polypropylene vial. Vials were sealed immediately after collection and placed in a frozen storage. After the session was completed, all four samples were stored in an -80 C freezer before being send to the Research Center for Psychobiology at the University of Trier. In Trier, saliva samples were stored at -20 C until analysis. After thawing, saliva samples were centrifuged at 2000 *g* for 10 minutes, which resulted in a clear supernatant of low viscosity. For CORT, 100ul of saliva was used for duplicate analysis. CORT levels were determined with an in-house method employing a competitive solid phase time-resolved fluorescence immunoassay with flourometric end point detection (DELFIA, PerkinElmer) using a rabbit anti-cortisol antibody. Sensitivity of DELFIA is <0.5 pg/ml (for more details on DELFIA, see: Dressendörfer, Kirschbaum, Rohde, Stahl, & Strasburger, 1992). The intra-assay coefficient of variation of this immunoassay was between 4.0% and 6.7% and the corresponding inter-assay coefficients of variation were between 7.1%-9.0%. T was assessed by solid phase enzyme-linked immunosorbent assays (ELISA), employing an anti-testosterone antibody (rabbit polyclonal antibody), which is based on the principle of competitive binding using the Demeditec Testosterone free in Saliva ELISA Kit with a sensitivity of 2.2 pg/ml (for more details see: http://www.immunoassay.co.uk/download/DES6622.pdf). The intra-assay coefficient of variation was between 5.6% and 9.7%, and the corresponding inter-assay coefficients of variation were between 7.0%-8.0%. All samples were assayed in duplicate and the average of the two samples was used in all analyses.

Three participants who provided insufficient saliva for processing CORT at baseline were excluded from the analysis. In addition, for some samples the amount of collected saliva was only sufficient for the analysis of one marker. In such cases we prioritized CORT assessments. For T, one participant was excluded from all analyses due to missing data on all time-points. Two participants had missing data at T1, T2 and T3, one participant had missing data at T3 and T4, five participants had missing data at T1 and five participants had missing data at T4. For the main analyses, all 17 participants with missing data on either CORT or T were excluded (10,6%), resulting in a total N = 143.

2.4 Caregiving experiences

Maternal use of love withdrawal was assessed using an 11-item questionnaire (Huffmeijer, Tops, Alink, Bakermans-Kranenburg, & Van IJzendoorn, 2011) based on the Child's Report of Parental Behavior Inventory (CRPBI; Beyers & Goossens, 2003; Schludermann & Schludermann, 1970) and the Parental Discipline Questionnaire (PDQ; Hoffman & Saltzste, 1967; Patrick & Gibbs, 2007). Each item (e.g. *My mother was a person who is less friendly with me, if I do not see things her way*) was rated on a 5-point scale ranging from *not applicable* to *fully applicable*. Due to a technical error one of the items was missing in some questionnaires. The score for this item was substituted with the mean score of the other items. Internal consistency of the scale was high (Cronbach's α = .92). The distribution of maternal love withdrawal scores was positively skewed and values were log transformed to approximate a normal distribution.

2.5 Statistical analyses

Simple t-tests were used to check for associations of OCs, menstrual cycle and morning or afternoon sessions with hormonal levels. To examine the influence of age, quality of hearing and cry duration on the hormonal levels, Pearson's *r* correlations were used. Stability of hormonal levels over the time-points and relations between hormones were examined using Pearson's correlation coefficients. To examine T reactivity to the caretaking session, we first performed repeated measures ANOVAs with OC use as between-subject variable to examine changes in T. Further data analyses were performed separately for participants with and without reported use of OCs because of the large differences in T levels between these groups (which is common, see, e.g., Virabel-Rebot et al., 2015; Stanton, Mullette-Gillman & Huettel, 2011). Repeated measures ANOVAs with basal CORT and maternal love withdrawal as covariates were used to examine the relation of basal CORT and maternal love withdrawal with T reactivity. In the first step, basal CORT was entered to examine the association between basal CORT and T reactivity. In the second step, maternal love withdrawal was entered to examine the association between maternal love withdrawal and T reactivity controlling for the effects of basal CORT.

3. Results

3.1. Preliminary analyses

T and CORT distributions were positively skewed and values were log transformed to approximate normal distributions (e.g. Bedgood et al., 2014; Welker et al., 2014). Univariate outliers (> 3SD) were found for T at time 1 (n = 1), time 2 (n = 1) and time 4 (n = 1) and these values were winsorized to 3 *SDs* below the mean (Wilcox, 2001). Mean CORT and T values are shown in Table 1.

There was no relation between the hormonal levels and age or hearing quality. Menstrual cycle (follicular compared to luteal phase) was unrelated to CORT or T (p >.40). CORT was significantly higher in the morning group compared to the afternoon group, but T did not differ between these two groups (Table 2). Basal CORT did not differ between participants with and without OC use, but, as expected, T was significantly higher in the group without OC at all time-points (Table 2).

Participants with and without OC use showed high correlations between T levels over time (Table 3). Correlations between basal CORT and T were only significantly positive in the group without OC use (Table 3).

3.2. Testosterone reactivity to caretaking

A repeated measure ANOVA to test for T reactivity showed a significant main effect of time (Wilks' Lambda = .754, $F(3, 139) = 15.15$, $p < .001$ (partial η^2=.25)) and a main effect of OC use $F(1, 141)=48.80$, p <.001 (partial η^2=.26)) (see Figure 1). No interaction between OC use and time was found (Wilks' Lambda = .992, $F(3, 139) = .355$, $p = .79$ (partial η^2=.01)). Pairwise comparisons with Bonferroni correction showed a significant decrease between all time-points except between T2 and T3. All further analyses were performed separately for the participants with and without reported use of OCs.

3.3. Effects of basal cortisol levels and maternal love withdrawal on testosterone reactivity

Repeated measures ANOVAs with session (morning vs afternoon) as between-subjects factor and basal CORT as covariate were performed to examine the influence of basal CORT on T reactivity in participants with and without OC use. In the second step, maternal love withdrawal was entered to examine the influence of maternal love withdrawal on T reactivity.

When basal CORT was entered as covariate, no significant main or interaction effects were found for participants with oral contraceptive use. In the group of participants without OC use no significant main effect of time (Wilks' Lambda = 0.92, $F(3, 42) = 1.20$, $p = .32$ (partial η^2=.08)) or session $F(1, 44)=.007$, $p = .93$ (partial η^2=.00), was found, but a significant main effect of basal CORT emerged, $F(1, 44)=16.23$, $p < .001$ (partial η^2=.27), and a marginally significant interaction between

time and basal CORT, Wilks' Lambda = 0.83, $F(3, 42) = 2.83$, $p = .050$ (partial η^2=.17). Figure 2 shows the different patterns in T reactivity in participants with low and high basal CORT: participants with high basal CORT showed higher basal T and a larger decrease of T compared to individuals with low basal CORT. There were no main or interaction effects of session (morning vs afternoon).

In the second step, maternal love withdrawal was added as covariate. No significant main or interaction effects were found for participants with OC use. In the group of participants without OC use a significant main effect of basal CORT was found, $F(1, 42)=4,95$, $p = .03$ (partial η^2=.11), no other main or interaction effects were found.

4. Discussion

Young women showed a decrease in T levels in response to caring for a crying infant simulator. This result supports the theoretical framework of the Steroid/Peptide Theory of Social Bonds (Van Anders et al., 2012) in that nurturing parental behavior is related to decreasing T levels. Others also reported that caregiving for the crying infant simulator decreased T in male respondents (Van Anders, et al., 2011). While these results seem to complement each other, an important difference between the studies should be made explicit: the nurturance in the Van Anders et al. (2011) study with male respondents was *effective nurturing* whereas the nurturance in the current study was ineffective. Van Anders et al (2011) made a distinction between 3 groups; (1) only listening to crying sounds, (2) effective nurturing (i.e. infant simulator responds to care) and (3) ineffective nurturing (i.e. infant simulator does not respond to care). They found that only the *effective nurturing* was coupled with decreases in T. The ineffective nurturing in our study may however be comparable with the effective nurturing condition in the earlier study, as our participants experienced effective nurturing with the infant simulator for two nights. They may therefore have associated the infant simulator with their past experiences of being an effective caregiver.

In line with the findings of Liening et al (2010), menstrual cycle did not affect T or CORT levels, and time of day had an effect on CORT but not on T. Compared to participants without OC use, participants with OC use had significantly lower levels of T at all time points, whereas no such difference was found for basal CORT. These findings are comparable to previous studies in which women using OC showed lower T and no difference in CORT compared to women not using OC (e.g., Crewther et al., 2015; Liening et al., 2010). The impact of OC use was large, not only for average levels of T but also for correlational patterns, and it therefore seems crucial to control for contraceptive use in T studies or, better still, to study T only in women without OC use.

When basal CORT was taken into account, a marginally significant interaction between basal CORT and T reactivity was found in participants who did not use OC; participants with high basal CORT tended to show higher initial T levels and a larger decrease of T compared to individuals with low basal CORT. No main or interaction effects were found for participants with oral contraception

use. The finding that participants with higher basal CORT showed stronger T reactivity than those with low basal CORT seems to be in contrast with Dual-hormone hypothesis and the study of Bedgood et al (2014) showing a stronger reactivity in T in males with low basal CORT. There are two important aspects that may explain why these findings diverge. First, Bedgood et al (2014) used a social stress paradigm while we used caregiving stimuli. Earlier studies showed different T responses to different stimuli and the interaction between basal CORT and T reactivity may also differ depending on the stimuli involved. Second, T and CORT may interact in a different way in males and women. T has been found positively related to aggression in males when CORT is low (Popma et al., 2007) and to aggression in women when CORT is high (Denson et al., 2013), indicating that some traits or behaviors associated with T may be more pronounced in males with *low* CORT but in women with *high* CORT. To our knowledge this is the first study on T reactivity in relation to basal CORT in women and more research is clearly needed.

Experiences of love withdrawal did not predict or moderate T reactivity to caregiving. The relative small sample size of the participants without OC use may have limited power to find such effects, the more so because of underrepresentation of more severe cases of maternal love withdrawal. In larger or clinical samples untoward childhood experiences may have larger effects on hormonal responses to caregiving and infant crying. The current study has two other limitations to consider. First, our sample consisted of women without children of their own and hormonal changes related to pregnancy and childbirth that would normally play a role in mothers of 6 week-old infants, could not be taken into account. However, in our more homogeneous sample we were better able to examine the pure effects of caregiving behavior without individual variation in hormones related to pregnancy and childbirth. Second, future studies should include an experimental condition with caregiving for a non-crying infant to be able to disentangle the influence of caring for the crying infant and caregiving behavior not related to crying.

Most studies examining T have for obvious practical reasons included males and not women, but findings might not easily generalize across gender and OC use might play a crucial role. Previous studies reported differences between males and women in basal T (e.g., Zilioli, Caldbick, et al., 2014; Endendijk et al., 2016), and in the interaction of CORT and T (e.g., Denson et al., 2013; Popma et al., 2007). Some studies examining women have only included women using OC (e.g., Bos et al., 2010), women not using OC (e.g., Hahn et al., 2015; Zilioli, Calbrick, et al., 2014), or mixed samples (Costa, Correia, & Oliveira, 2015; Kuzawa et al., 2010). Our results show that not only the levels of T differ between women with and without OC use; the interaction between T and CORT is also significantly different. Thus, when examining T in women, OC use should always be taken into account as a potential moderator.

The nature of the stressor is an important design feature and standardized stress procedures are of great importance. Examining hormonal reactivity during caregiving in response to standardized

infant behavior has the advantage that results are not influenced by random and non-random variation in child behavior. Although the caregiving behavior in this standardized procedure with controlled infant stimuli is of course limited by the simulator's abilities to respond to the caregiver, we have shown that the participants thought that the experience was realistic (Voorthuis et al., 2013), and that sensitive caregiving behavior could be assessed in a meaningful way (Bakermans-Kranenburg, Alink, Biro, & Van IJzendoorn, 2015; Voorthuis et al., 2013). The infant simulator has great potential to become a standardized stressful procedure to study hormonal correlates of parenting.

In sum, T was significantly higher in women without OC use compared to women with OC use. In women without OC use, participants with higher basal CORT showed higher initial T levels and a larger decrease of T compared to individuals with lower basal CORT whereas no effect of CORT was found in women with OC use. The interplay of CORT and T in parenting should therefore be studied without the confounding influence of contraceptive hormones. Controlling for contraceptive use, we found some support for a decrease of T in women who had to take care of a crying infant. We found no support for the Dual-hormone hypothesis as we found stronger T reactivity in individuals with high CORT and not in those with low CORT, as would be expected based on the Dual-hormone hypothesis. Our results do support the Challenge hypothesis and more specific the *Steroid/Peptide Theory of Social Bonds*, as T decreased in the context of nurture and caregiving.

Acknowledgements

Funding: Marinus van IJzendoorn and Marian Bakermans- Kranenburg are supported by the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (Gravitation program, SPINOZA, VICI). Marian Bakermans- Kranenburg is also supported by the European Research Council (ERC AdG 669249)

References

Archer, J. (2006). Testosterone and human aggression: an evaluation of the Challenge hypothesis. *Neuroscience and Biobehavioral Reviews, 30,* 319-345. doi: 10.1016/j.neurbiorev.2004.12.007.

Barret, E.S., Tran, V., Thurston, S., Jasienska, G., Furberg, A., Ellison, P.T., & Thune, I. (2013). Marriage and motherhood are associated with lower testosterone concentrations in women. *Hormones and Behavior, 63*, 72-79. doi: 10.1016/j.yhbeh.2012.10.012.

Bakermans-Kranenburg, M.J., Alink, L.R.A., Biro, S., Voorthuis, A., & Van IJzendoorn, M.H. (2015). The Leiden Infant Simulator Sensitivity Assessment (LISSA): Parenting an infant simulator as your own baby. *Infant and Child Development, 24*, 220-227. doi: 10.1002/icd.1950.

Bakermans-Kranenburg, M.J., & Van IJzendoorn, M.H., (2013). Sniffing around oxytocin: review and meta-analyses of trials in healthy and clinical groups with implications for pharmacotherapy. *Translational Psychiatry, 3*, e258. doi: 10.1038/tp.2013.34.

Bedgood, D., Boggiano, M.M., & Turan, B. (2014). Testosterone and social evaluative stress: The moderating role of basal cortisol. *Psychoneuroendocrinology, 47*, 107-115. doi: 10.1016/j.osyneuen.2014.05.007.

Beyers, W., & Goossens, L. (2003). Psychological separation and adjustment to university: Moderating effects of gender, age and perceived parenting style. *Journal of Adolescent Research, 18,* 363-382. doi: 10.1177/0743558403018004003

Bhandari, R., Voorthuis, A., Van der Veen, R., Out, D., Bakermans-Kranenburg, M.J., & Van IJzendoorn, M.H. (2013). Temperamental sensitivity moderates the effects of maternal love-withdrawal on perception of infant crying. *Family Science, 3*, 246-254. doi: 10.1080/19424620.2013.789977

Bos, P.A., Hermans, E.J., Montoya, E.R., Ramsey, N.F., & Van Honk, J. (2010). Testosterone administration modulates neural responses to crying infants in young women. *Psychoneuroendocrinology, 35*, 114-121. doi: 10.1016/j.psyneuen.2009.09.013.

Boyd, R.A., Zegarac, E.A., Posvar, E.L., & Flack, M.R. (2001). Minimal androgenic activity of a new oral contraceptive containing norethindrone acetate and graduated doses of ethinyl estradiol. *Contraception*, 63, 71-76. doi: 10.1016/S0010-7824(01)00179-2

Costa, R.M., Correia, M., & Oliveira, R.F. (2015). Does personality moderate the link between women's testosterone and relationship status? The role of extraversion and sensation seeking. *Personality and Individual Differences, 76*, 141-146. doi: 10.1016/j.paid.2014.12.003.

Crawford, JR & Henry, JD (2004) The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 43*, 245–265

Crewther, B.T., Hamilton, D., Casto, K., Kilduff, L.P., & Cook, C.J. (2015). Effects of oral contraceptive

use on salivary testosterone and cortisol responses to training sessions and competitions in elite women athletes. *Physiology & Behavior, 147*, 84-90. doi: 10.1016/j.physbeh.2015.04.017.

Dabbs, J.M. Jr., Jurkovic, G.J., & Frady, R.L. (1991). Salivary testosterone and cortisol among late adolescent male offenders. *Journal of Abnormal Child Psychology, 19(4),* 469-478. Doi: 10.1007/BF00919089

Denson, T.F., Mehta, P.H., & Ho Tan, D. (2013). Endogeneous testosterone and cortisol jointly influence reactive aggression in women. *Psychoneuroendocrinology, 38,* 416-424. doi: 10.1016/j.psyneuen.2012.07.003.

Dressendorfer, RA, Kirschbaum, C, Rohde, W, Stahl, F, Strasburger, CJ (1992). Synthesis of a cortisol biotin conjugate and evaluation as a tracer in an immunoassay for salivary cortisol measurement. *Journal of Steroid Biochemistry and Molecular Biology, 43,* 683-692. DOI: 10.1016/0960-0760(92)90294-S

Edwards, D.A., & Casto, K.V. (2013). Women's intercollegiate athletic competition: cortisol, testosterone, and the Dual-hormone hypothesis as it relates to status among teammates. *Hormones and Behavior, 64,* 153-160. doi: 10.1016/j.physbeh.2015.01.037.

Edwards, D.A., & O'Neal, J.L. (2009). Oral contraceptives decrease saliva testosterone but do not affect the rise in testosterone associated with athletic competition. *Hormones and Behavior, 56,* 195-198. doi: 10.1016/j.yhbeh.2009.01.008.

Endendijk, J.J., Hallers-Haalboom, E.T., Groeneveld, M.G., Van Berkel, S.R., Van der Pol, L.D., Bakermans-Kranenburg, M.J., & Mesman, J. (2016). Diurnal testosterone variability is differentially associated with parenting quality in mothers and fathers. *Hormones and Behavior, 80*, 68-75. doi: 10.1016/j.yhbeh.2016.01.016

Fleming, A.S., Corter, C., Stallings, J., & Steiner, M. (2002). Testosterone and prolactin are associated with emotional responses to infant cries in new fathers. *Hormones and Behavior, 42*, 399-413. doi: 10.1006/hbeh.2002.1840.

Fleming, A.S., Steiner, M., & Corter, C. (1997). Cortisol, hedonics, and maternal responsiveness in human mothers. *Hormones and Behavior, 32,* 85–98. doi: 10.1006/hbeh.1997.1407

Granger, D.A., Shirtcliff, A., Zahn-Waxler, C., Usher, B., Klimes-Dougan, B., & Hastings, P. (2003). Salivary testosterone diurnal variation and psychopathology in adolescent males and women: Individual differences and developmental effects. *Development and Psychopathology, 15*, 431-449. doi: 10.1017.S0954579403000233.

Gray, P.B., Yang, C.J., & Pope, H.G. Jr. (2006). Fathers have lower salivary testosterone levels than unmarried men and married non-fathers in Beijing, China. *Proceedings of the Royal Society B: Biological Sciences, 273,* 333-339. doi: 10.1098/rspb.2005.3311.

Hahn, A.C., DeBruine, L.M., Fisher, C.I., & Jones, B.C. (2015). The reward value of infant facial cuteness tracks within-subject changes in women's salivary testosterone. *Hormones and Behavior, 67,* 54-59. doi: 10.1016/j.yhbeh.2014.11.010

Hoffman, M.L., & Saltzste, H.D., 1967. Parent discipline and the child's moral development. *Journal of Personality and Social Psychology*, 5, 45-47. doi: 10.1037/h0024189

Huffmeijer, R., Tops, M., Alink, L.R.A., Bakermans-Kranenburg, M.J., & Van IJzendoorn, M.H. (2011). Love withdrawal is related to heightened processing of faces with emotional expressions and incongruent emotional feedback: evidence from ERPs. *Biological Psychology, 86,* 307-313. doi: 10.1016/j.biopsycho.2011.01.003.

Jung-Hoffman, H., & Kuhl, H. (1987). Divergent effects of two lose-dose oral contraceptives on sex hormone-binding globulin and free testosterone. *American Journal of Obstetrics and Gynecology, 156*, 199-203. doi: 10.1016/0002-9378(87)90238-9

Kuldielka, B.M., Hellhammer, D.H., & Wust, S. (2009). Why do we respond so differently? Reviewing determants of human salivary cortisol responses to challenge. *Psychoneuroendocrinology, 34*, 2-18. doi: 10.1016/j.psyneuen.2008.10.004.

Kuzawa, C.W., Gettler, L.T., Muller, M.N., McDade, T., & Feranil, A.B. (2009). Fatherhood, pairbonding and testosterone in the Philippines. *Hormones and Behavior, 56,* 429-435. doi: 10.1016/j.yhbeh.2009.07.010.

Kuzawa, C.W. Gettler, L.T., Huang, Y., & McDade, T.W. (2010). Mothers have lower testosterone than non-mothers: Evidence from the Philippines. *Hormones and Behavior, 57,* 441-447. doi: 10.1016/j.yhbeh.2010.01.014.

Lennarsson, A.K., Kushnir, M.M., Bergquist, J., Billig, H., & Jonsdottir, I.H. (2012). Sex steroid levels temporarily increase in response to acute psychosocial stress in healthy men and women. *International Journal of Psychophysiology, 84*, 246-253. doi: 10.1016/j.ijpsycho.2012.03.001.

Liening, S.H., Stanton, S.J., Saini, E.K., & Schultheiss, O.C. (2010). Salivary testosterone, cortisol, and progesterone: Two-week stability, interhormone correlations, and effects of time of day, menstrual cycle, and oral contraceptives use on steroid hormone levels. *Physiology and behavior, 99,* 8-16. doi: 10.1016/j.physbeh.2009.10.001.

Mehta, P.H., & Josephs, R.A. (2010). Testosterone and cortisol jointly regulate dominance: Evidence for a Dual-hormone hypothesis. *Hormones and Behavior, 58,* 898-906. doi: 10.1016/j.yhbeh.2010.08.020.

Patrick, R.B., & Gibbs, J.C. (2007). Parental expression of disappointment: should it be a factor in Hoffman's model of parental discipline? *The Journal of Genetic Psychology, 168*, 131-145. doi: 10.3200/GNTP.168.2.131-146.

Popma, A., Vermeiren, R., Geluk, C. A. M.L., Rinne, T., van den Brink, W., Knol, D.L., . . . Doreleijers, T.A.H. (2007). Cortisol moderates the relationship between testosterone and aggression in

delinquent male adolescents. *Biological Psychiatry, 61*, 405-411. doi: 10.1016/j.biopsych.2006.06.006.

Riem, M.M.E., Van IJzendoorn, M.H., Tops, M., Boksem, M.A.S., Rombouts, S.A.R.B., & Bakermans-Kranenburg, M.J. (2013). Oxytocin effects on complex brain networks are moderated by experiences of maternal love withdrawal. *European Neuropsychopharmacology, 23*, 1288-1295. doi: 10.1016/j.euroneuro.2013.01.011

Salvador, A. (2005). Coping with competitive situations in humans. *Neuroscience & Biobehavioral Reviews, 29*, 195-205. doi:10.1016/j.neubiorev.2004.07.004.

Salvador, A., Suay, F., González-Bono, E., & Serrano, M.A. (2003). Anticipatory cortisol, testosterone and psychological responses to judo competition in young men*. Psychoneuroendocrinology, 28*, 364-75. doi: 10.1016/S0306-4530(02)00028-8.

Schludermann, E., & Schludermann, S. (1970). Replicability of factors in Children's Report of Parent Behavior (CRPBI). *The Journal of Psychology Interdisciplinary and Applied, 76*, 239-249. doi: 10.1080/00223980.1970.9916845

Schoofs, D., & Wolf, O.T. (2011). Are salivary gonadal steroid concentrations influenced by acute psychosocial stress? A study using the Trier Social Stress Test (TSST). *International Journal of Psychophysiology, 80*, 36-43. doi: 10.1016/j.ijpsycho.2011.01.008.

Stanton, S.J., Mullette-Gillman, O, A., & Huettel, S.A. (2011). Seasonal variation of salivary testosterone in men, normally cycling women, and women using hormonal contraceptives. *Physiology & Behavior, 104*, 804-808. doi: 10.1016/j.physbeh.2011.07.009.

Storey, E.A., Walsh, C.J., Quinton, R.L., & Wynne-Edwards, K.E. (2000). Hormonal correlates of paternal responsiveness in new and expectant fathers. *Evolution and Human Behavior, 21,* 79-95. doi: 10.1016/S1090-5138(99)0042-2

Swain, J.E., Kim, P., & Ho, S.S. (2011). Neuroendorinology of parental response to baby-cry. *Journal of Neuroendocrinology, 23*, 1036-1041. doi: 10.1111/j.1365-2826.2011.02212.x

Thorneycroft, I.H., Stanczyk, F.A., Bradshaw, K.D., Ballagh, S.A., Nichols, M., & Weber, M.E. (1999). Effect of low-dose oral contraceptives on androgenic markers and acne. *Contraception, 60*, 255-262. doi: 10.1016/S0010-7824(99)00093-1

Toufexis, D., Rivarola, M.A., Lara, H., & Viau, V. (2014). *Stress and the reproductive axis. Journal of Neuroendocrinology, 26,* 573-586. doi: 10.1111/jne.12179.

Van Anders, S.M., Goldey, K.L., & Kuo, P.X. (2011). The Steroid/Peptide theory of social bonds: Integrating testosterone and peptide responses for classifying social behavioral contexts. *Psychoneuroendocrinology, 36,* 1256-1257. doi: 10.1016/j.psyneuen.2011.06.001

Van Anders, S.M., Tolman, R.M., & Volling, B.L. (2012). Baby cries and nurturance affect testosterone in men. *Hormones and Behavior, 61,* 31-36. doi: 10.1016/j.yhbeh.2011.09.012.

Van der Vange, N., Blankenstein, M.A., Kloosterboer, H.J., Haspels, A.A., & Thijssen, J.H. (1990).

Effects of seven low-dose combined oral contraceptives on sex hormone binding globulin, corticosteroid binding globulin, total en free testosterone. *Contraception, 41*, 345-352. doi: 10.1016/0010-7824(90)90034-S.

Van Honk, J., Tuiten, A., Van den Hout, M., Koppeschaar, H., Thijssen, J., De Haan, E., & Verbaten, R. (2000). Conscious and preconscious selective attention to social threat: different neuroendocrine response patterns. *Psychoneuroendocrinology, 25,* 277.591. doi: 10.1016/S0306-4530(00)00011-1.

Van IJzendoorn, M.H., Huffmeijer, R., Alink, L.R.A., Bakermans-Kranenburg, M.J., & Tops, M. (2011). The impact of oxytocin administration on charitable donating is moderated by experiences of parental love-withdrawal. *Frontiers in Psychology, 2*, 258. doi: 10.3389/fpsyg.2011.00258.

Viau, V. (2002). Functional cross-talk between the hypothalamic-pituitary-gonadal and -adrenal axes. *Journal of Neuroendocrinology, 14,* 506-513. doi: 10.1046/j.1365-2826.2002.00798.x

Vibarel-Rebot, N., Rieth, N., Lasne, F., Jaffre, C., & Collomp, K. (2015). Oral contraceptive use and saliva diurnal pattern of metabolic steroid hormones in young healthy women. *Contraception, 91*, 245-247. doi: 10.1016/j.contraception.2014.11.010.

Voorthuis, A., Out, D., Van der Veen, R., Bhandari, R., Van IJzendoorn, M.H., & Bakermans-Kranenburg, M.J. (2013). One doll fits all: validation of the Leiden Infant Simulator Sensitivity Assessment (LISSA). *Attachment & Human Development, 15,* 5-6. doi: 10.1080/14616734.2013.820897

Weisman, O., Zagoory-Sharon, O., & Feldman, R. (2014). Oxytocin administration, salivary testosterone, and father-infant social behavior. *Progress in Neuro-Psychopharmacology & Biological Psychiatry, 49,* 47-52. doi: 10.1016/j.pnpbp.2013.11.006.

Welker, K.M., Lozoya, E., Campbell, J.A., Neumann, C.S., & Carré, J.M. (2014). Testosterone, cortisol, and psychopathic traits in men and women. *Physiology & Behavior, 129,* 230-236. doi: 10.1016/j.physbeh.2014.02.057

Wilcox, R.R. (2001). *Fundamentals of modern statistical methods: substantially improving power and accuracy.* New York: Springer-Verlag.

Wingfield, J.C., Hegner, R.E., Dufty, A. M. Jr., & Ball, G.F. (1990). The "Challenge' Hypothesis": theoretical implications for patterns of testosterone secretion, mating systems, and breeding strategies. *The American Naturalist, 136,* 829-846. doi: 10.1086/285134.

Zilioli, S., Caldbick, E., & Watson, N.V. (2014). Testosterone reactivity to facial display of emotions in men and women. *Hormones and Behavior, 65,* 461-468. doi: 10.1016/j.yhbeh.2014.04.006

Zilioli, S., Ponzi, D., Henry, A., & Maestripieri, D. (2014). Testosterone, Cortisol and Empathy: Evidence for the Dual-Hormone Hypothesis. *Adaptive Human Behavior and Physiology, 1(4)*, 421-433. doi: 10.1007/s40750-014-0017-x

Table 1

Means, standard deviations and range of untransformed CORT and T values (pg/ml)

Measure		*n*	*M (SD)*	*Range*
Cortisol	T1	157	5.30 (3.04)	1.09 – 15.30
Testosterone	T1	149	40.47 (16.88)[a]	8.39 – 105.51[a]
	T2	154	38.54 (16.89)[a]	8.28 – 107.12[a]
	T3	153	37.69 (15.77)	10.79 – 104.51
	T4	150	35.85 (14.57)[a]	8.29 – 102.38[a]

[a] including non-winsorized outliers

Table 2

Descriptive statistics of the Log transformed CORT and T values for the groups with and without oral contraceptives, and the morning and afternoon groups.

		Contraceptives		No contraceptives			Morning		Afternoon		
Measure		*n*	*M (SD)*	*n*	*M (SD)*	*p-value*[1]	*n*	*M (SD)*	*n*	*M (SD)*	*p-value*[2]
Cortisol	T1	102	0.68 (.22)	55	0.61 (.26)	.10	89	0.74 (.24)	68	0.55 (.20)	.00
Testosterone	T1	98	1.51 (.15)[a]	51	1.70 (.15)	.00	84	1.57 (.18)[a]	65	1.57 (.18)	.91
	T2	101	1.48 (.17)[a]	53	1.67 (.15)	.00	87	1.55 (.18)	67	1.55 (.19)[a]	.92
	T3	101	1.47 (.17)	52	1.66 (.14)	.00	86	1.54 (.18)	67	1.53 (.18)	.72
	T4	100	1.46 (.17)[a]	50	1.64 (.14)	.00	85	1.52 (.18)[a]	65	1.52 (.19)	.93

[a] Outliers were winsorized.

[1] Significance level for differences between the groups with and without contraceptives based on independent t-tests.

[2] Significance level for differences between the morning (12 pm) and afternoon groups (3 pm) based on independent t-tests.

Table 3

Correlations between hormonal levels in participants with and without oral contraceptive use

Measure	*1*	*2*	*3*	*4*	*5*
1. Cortisol Baseline	-	.63**	.58**	.49**	.44**
2. Testosterone T1	.21*	-	.87**	.85**	.84**
3. Testosterone T2	.11	.88**	-	.91**	.85**
4. Testosterone T3	.09	.85**	.91**	-	.88**
5. Testosterone T4	.12	.84**	.88**	.91**	-

Note. Upper half: without contraceptives (n = between 47 - 53), lower half: with contraceptives (n = between 96 - 101).

* p <.05, **p <.01

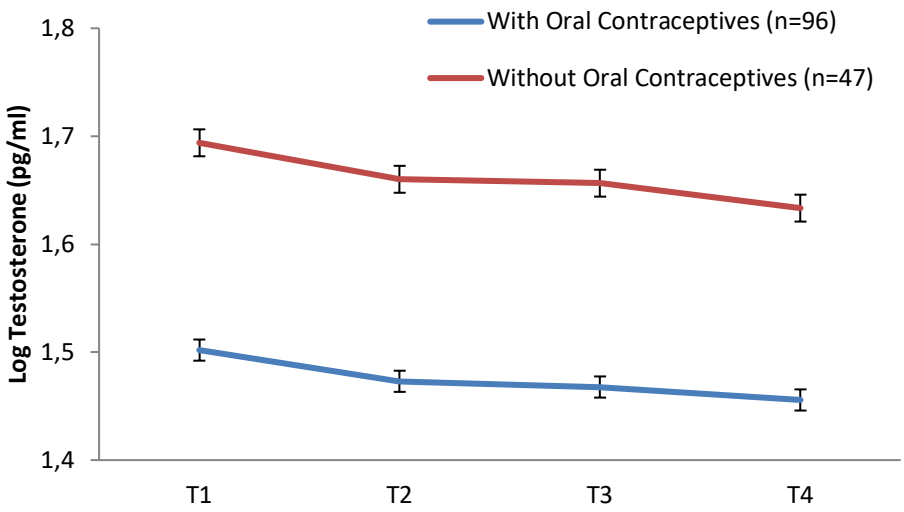

Figure 1. Testosterone levels across time in participants with and without oral contraceptive use. T1 is baseline before the caretaking; T2 is immediately after caretaking (30 min after baseline); T3 is 15 minutes after caretaking (45 min after baseline); T4 is 30 minutes after caretaking (60 min after baseline). Data reflects mean levels of testosterone ± SE.

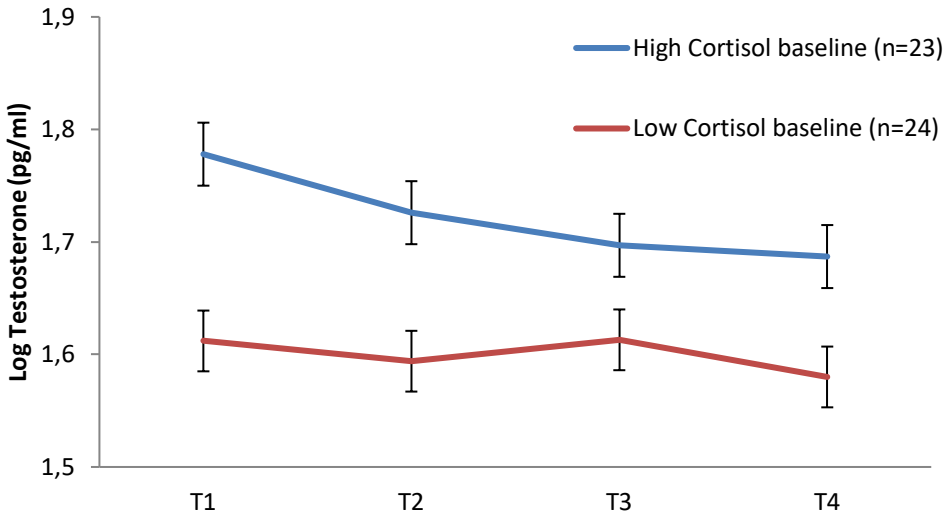

Figure 2. Testosterone levels across time as a function of cortisol at baseline (median split) in participants without oral contraceptive use. T1 is baseline before the caretaking; T2 is immediately after caretaking (30 min after baseline); T3 is 15 minutes after caretaking (45 min after baseline); T4 is 30 minutes after caretaking (60 min after baseline). Data reflects mean levels of testosterone ± SE.

Appendix

Procedure

Participants were asked to take care of an infant simulator for two evenings at their home (5 pm–10 pm) and during a lab session as if it were their own baby. During the two evenings, the simulator was programmed with various cry episodes. The participants used a hand held computer to answer questions about their perception of the crying after each cry and caretaking episode (Bhandari et al., 2013). After taking care of the infant simulator for two evenings, the participants were asked to fill in a questionnaire about their experiences with the infant simulator.

Following the two evenings, the participants took care of the infant simulator in a 30 minute lab session with increasing competing demands. The lab room was decorated as a living room with a dresser with a changing mat, a baby cot, an infant carrier, a desk with a chair, an easy chair, magazines, children's books, and toys. The participants were told that they were allowed to use all materials in the room while taking care of the simulator as if it was a real infant. In the first 10 minutes (part A), the participants were given no further instructions. In the next 10 minutes (part B), they were asked to fill in a questionnaire while taking care of the simulator (competing demand 1). In the final period of 10 minutes (part C), the participants were asked to perform a computer task (Tower of London) as fast as they could while taking care of the simulator (competing demand 2). Participants were told that those who performed the computer task fastest would receive an additional reward of five Euros, to increase the potential pressure of the competing demand. Before and after taking care of the infant simulator, the participants filled in the Positive Affect Negative Affect questionnaire (PANAS; Crawford & Henry, 2004) to indicate their current mood. Sessions were videotaped and coded for sensitivity. After caring for the infant simulator, the cry paradigm was administered. The procedure was approved by the Departmental Ethics Committee and informed consent was obtained from all participants.