Psychoneuroendocrinology 74 (2016) 77–91

Contents lists available at ScienceDirect

Psychoneuroendocrinology

journal homepage: www.elsevier.com/locate/psyneuen

# Risky decision-making in adolescent girls: The role of pubertal hormones and reward circuitry

Zdeňa A. Op de Macks[a,*], Silvia A. Bunge[a,b], Orly N. Bell[a], Linda Wilbrecht[a,b], Lance J. Kriegsfeld[a], Andrew S. Kayser[c,d], Ronald E. Dahl[e,f]

[a] Department of Psychology, University of California, Berkeley, 2310 Tolman Hall, Berkeley, CA 94720-1650, USA
[b] Helen Wills Neuroscience Institute, University of California, Berkeley, 175 Li Ka Shing Center, Berkeley, CA 94720, USA
[c] Department of Neurology, University of California, San Francisco, 675 Nelson Rising Lane, San Francisco, CA 94143, USA
[d] Department of Neurology, VA Northern California Health Care System, 150 Muir Road, Martinez, CA 94553, USA
[e] School of Public Health, University of California, Berkeley, 50 University Hall, Berkeley, CA 94720-7360, USA
[f] Institute of Human Development, University of California, Berkeley, 1121 Tolman Hall, Berkeley, CA 94720-1690, USA

ARTICLE INFO

*Article history:*
Received 7 April 2016
Received in revised form 19 July 2016
Accepted 15 August 2016

*Keywords:*
fMRI
Puberty
Testosterone
Estradiol
Nucleus accumbens
Orbitofrontal cortex

ABSTRACT

Adolescence is a developmental period characterized by a greater tendency to take risks. While the adult literature has shown that sex steroids influence reward-related brain functioning and risk taking, research on the role of these hormones during puberty is limited. In this study, we examined the relation between pubertal hormones and adolescent risk taking using a probabilistic decision-making task. In this task, participants could choose on each trial to play or pass based on explicit information about the risk level and stakes involved in their decision. We administered this task to 58 11-to-13-year-old girls while functional MRI images were obtained to examine reward-related brain processes associated with their risky choices. Results showed that higher testosterone levels were associated with increased risk taking, which was mediated by increased medial orbitofrontal cortex activation. Furthermore, higher estradiol levels were associated with increased nucleus accumbens activation, which in turn related to decreased risk taking. These findings offer potential neuroendocrine mechanisms that can explain why some adolescent girls might engage in more risk taking compared to others.



## 1. Introduction

### 1.1. Pubertal hormones and adolescent risk taking

Adolescence, the developmental period between childhood and adulthood, is a dynamic time of transition characterized by dramatic biological, cognitive, social, emotional, and behavioral changes (Dahl, 2004). According to most developmental researchers, the onset of adolescence is marked by puberty (Dorn et al., 2006), a biological process that involves a substantial rise in sex steroids, such as testosterone and estradiol (Shirtcliff et al., 2009; Biro et al., 2014). While the rise in sex steroids during puberty is associated with some of the physical and physiological changes necessary for sexual reproduction, such as the development of secondary sex characteristics, it has also been suggested that pubertal hormones influence the developing adolescent brain (Schulz et al., 2009). As such, pubertal hormones are thought to play an important role in activating the behaviors that characterize adolescents, such as risk taking (Forbes and Dahl, 2010; Peper and Dahl, 2013).

Indeed, adolescent risk taking has been associated with higher levels of testosterone and estradiol, independent of age (De Water et al., 2013; Vermeersch et al., 2008a, 2008b). A neurobiological model proposed to explain increases in risk taking across adolescence emphasizes the role of pubertal hormones in altering subcortical brain functioning, such as reward-related brain activation (Crone and Dahl, 2012). While enhanced reward processing during adolescence, as measured by increased ventral striatum/nucleus accumbens activation compared to children and/or adults, is thought to underlie the tendency to take risks (Galvan, 2010; Somerville et al., 2010; Steinberg, 2010), the influence of pubertal hormones on reward processing often remains untested. Thus, relatively few studies have concomitantly captured data related to hormone levels, brain activity, and risk-taking behavior.

* Corresponding author at: Prevention Science Institute, University of Oregon, 1600 Millrace Drive, Suite 106, Eugene, OR, 97403-6217, USA.
*E-mail address:* zopdemac@uoregon.edu (Z.A. Op de Macks).

http://dx.doi.org/10.1016/j.psyneuen.2016.08.013

Importantly, in addition to developmental influences, individual differences in risk taking tendencies resulting from genetic (Harden and Mann, 2015) and/or personality differences (Braams et al., 2015; Van Duijvenvoorde et al., 2014) likely contribute to differences in adolescent risky behavior (and associated brain activation). Moreover, these individual and developmental differences are likely to interact across adolescent development, meaning that developmental changes might result in different behavioral and/or neural phenotypes depending on genetics or personality characteristics. These findings suggest that individual differences in adolescent risk taking (and associated brain processes) reflect both developmental and individual differences (Chick, 2015), which might both be captured by individual differences in pubertal hormone levels.

### 1.2. Pubertal hormones and reward-related neural processes

To date, few longitudinal neuroimaging studies have shown that structural changes in subcortical brain regions involved in reward processing, such as the nucleus accumbens, are related to pubertal stage (Goddings et al., 2014) and pubertal hormone levels (Peper et al., 2011; Herting et al., 2014), and these findings have not been consistent in all studies (Koolschijn et al., 2014). In addition to structural changes, functional changes in reward-related brain regions have also been reported, including enhancing effects of testosterone and estradiol administration on reward processes associated with risk taking in adults (Hermans et al., 2010; Thomas et al., 2014). However, evidence for the role of hormones in subcortical brain functioning during puberty is both limited and conflicting (Forbes et al., 2010; Op de Macks et al., 2011; Braams et al., 2015).

In a prior cross-sectional study, we began to address this gap by using a newly designed probabilistic decision-making task called the Jackpot task. In this task, participants chose to play or pass based on presented information about the chance of winning 10 Eurocents (Op de Macks et al., 2011). By administering this task to a sample of 33 girls and 17 boys (10–16 yrs), we demonstrated that higher levels of testosterone corresponded with enhanced ventral striatum activation when a risky decision was rewarded, whereas higher levels of estradiol corresponded with more frontal activation, although the estradiol findings were weaker (Op de Macks et al., 2011). A similar finding resulted from a longitudinal study on reward processing among 8–27 year-olds, whereby changes in nucleus accumbens activation in response to reward receipt correlated with changes in testosterone level (Braams et al., 2015). In contrast, a cross-sectional study of 39 girls aged 11–12 years who played a card-guessing game found a *reduced* striatal response to rewards in girls with higher levels of testosterone (Forbes et al., 2010). These latter studies did not report on the relation between estradiol level and reward-related brain activation. More importantly, these studies focused on the relation between testosterone and brain processes involved in decision-making, but did not test the three-way relation between testosterone level, reward-related brain processes, and associated risk-taking behavior.

Additionally, relatively less is known about hormonal influences on other components of the reward system, including areas within the medial frontal cortex that are linked to valuation (O'Doherty, 2007). Specifically, the ventromedial prefrontal cortex, or medial orbitofrontal cortex, is considered part of the "reward circuit" given its anatomical connections with other reward-related brain regions (e.g., ventral striatum, including nucleus accumbens) and its responsiveness to both primary and secondary rewards (Haber and Knutson, 2010). More importantly, anatomical differences in the medial orbitofrontal cortex have been shown to mediate the relation between testosterone level and risk taking on the balloon analogue risk-taking (BART) task in adolescent boys and girls (Peper et al., 2013). However, the implications of structural differences for reward-related brain functioning remain unknown. Furthermore, orbitofrontal cortex activation during reward anticipation has been shown to differ across the menstrual cycle in adult women (Dreher et al., 2007) and ventromedial prefrontal cortex activation has been shown to increase upon estradiol administration in postmenopausal women (Thomas et al., 2014). Together, these findings suggest that the rise in sex steroids during puberty might enhance reward processing in medial frontal regions, in addition to subcortical regions.

### 1.3. The present study

To directly test the three-way relation between pubertal hormone levels, reward-related brain functioning, and risky behavior, we designed a study to examine the relation between levels of testosterone as well as estradiol and risk taking, and to test whether this relation was mediated by increased reward-related (cortical and subcortical) brain activation. Furthermore, to optimize our ability to test these relations independent of age within a cross-sectional design, we focused on a narrow age range around the onset of puberty to capture the developmental window during which individual differences in pubertal hormone levels are the largest – given the large individual variation in the onset and speed of pubertal changes – while keeping age relatively constant (Dorn et al., 2006; Peper and Dahl, 2013). We used a modified version of the Jackpot task based on Op de Macks et al. (2011) in which, in addition to manipulating the *probability* of reward, we also manipulated the *magnitude* of the potential reward (i.e., stakes). The study that used the previous version of the Jackpot task, as well as other studies that used a decision paradigm with the same probabilities (33% vs. 67%), have shown that adolescents more often made risky decisions when the chance to win was higher and engaged in riskier decisions when stakes were higher (Op de Macks et al., 2011; Van Leijenhorst et al., 2008, 2010). Therefore, we aimed to increase possibilities for risk taking (e.g., when the chance to win was relatively low but the potential reward was relatively large) and thereby optimize our ability to investigate reward-related brain processes associated with risky decisions. Besides the types of decisions adolescents made, we also examined decision speed, which has been shown to vary by stakes (but not by probability), such that decision speed decreased with increasing stakes (Van Leijenhorst et al., 2010). Based on these findings, we hypothesized that adolescents in this study would show greater risk taking in the low-risk condition and when a larger reward was at stake. Furthermore, we hypothesized that adolescents with higher testosterone and estradiol levels would show increased risk taking (Vermeersch et al., 2008a, 2008b), particularly in the high-risk condition, as adolescents showed greater individual variation in the propensity to take risks within this condition (Op de Macks et al., 2011).

We focused on individual differences in activation of nucleus accumbens (Haber and Knutson, 2010), a region known to be involved in reward anticipation and/or outcome processing and often reported to show increased activation in adolescents compared to children and adults in the context of risky decision-making (reviewed in Galvan, 2010). Concomitantly, to ensure that behaviorally-relevant activation within other regions of the reward circuit would be captured, we also conducted whole-brain analyses that included the medial frontal cortex. We hypothesized that nucleus accumbens activation would be elevated during risky decisions (Op de Macks et al., 2011). Furthermore, we predicted that the relation between pubertal hormones and risk taking would be mediated by increased reward-related brain activation (according to the model proposed by Crone and Dahl, 2012).

## 2. Materials and methods

### 2.1. Participants

A total of 78 healthy, adolescent girls were recruited through advertising via an email network used by over 33,000 parents in the San Francisco Bay Area (the Berkeley Parents Network), word of mouth, and by re-contacting families that participated in prior studies in the lab. We focused on girls only to optimize our power to investigate the relation between individual differences in pubertal measures and differences in risk taking and reward processing, given that the nature as well as the timeline of puberty differs dramatically between boys and girls (Dorn et al., 2006; Shirtcliff et al., 2009). For example, while both boys and girls show an increase in testosterone level during puberty, the rise in testosterone level is less dramatic in girls (Braams et al., 2015). Similarly, while girls and boys show increases in sensation seeking during adolescence, girls show lower levels of sensation seeking (and higher levels of impulse control) than boys (Shulman et al., 2015), indicating a potential differential role of testosterone in boys and girls.

Participants were screened in a phone interview with their parent or legal guardian; they were included in the study if they were (1) right-handed, (2) native English speakers, (3) enrolled in elementary or middle school, (4) medically healthy (i.e., no history of neurological or psychiatric disorders and/or past or present use of psychoactive medications), and (5) free from contraindications to MRI. Before entering the study, written informed consent was obtained from the parent or legal guardian of the participant, and assent was obtained from the participant. Sixty-eight participants completed the entire study, including an MRI scan on the second lab visit (see Section 2.2). These participants received $130 in gift cards at the end of the study, which included compensation for their travel time, the time spent in the lab, and additional task winnings. The University of California Berkeley Institutional Review Board approved all procedures.

Ten participants were excluded from analysis for the following reasons: (1) task-related imaging data were invalid due to technical problems[1] ($n = 6$) or movement ($n = 3$), and (2) response rate on the task was low (i.e., no response was recorded on 25% of the trials; $n = 1$). Thus, the results presented here are based on 58 participants: 23 11-year-olds, 19 12-year-olds, and 16 13-year-olds ($M$ age = 12.4 ± 0.92). Among the included participants 46.6% were Caucasian, 10.3% Asian, 5.2% Hispanic/Latin, 3.4% African-American, 24.1% were multi-racial, and 10.4% did not provide information about their race or ethnicity. Results of the Child Behavior Checklist (CBCL; Achenbach, 1991), a questionnaire for the assessment of behavioral problems that was completed by the parent or legal guardian, showed that all participants scored within the 'normal' range based on their total age-corrected score (i.e., the sum score across all subscales ≤ 65). Furthermore, there were no age-related differences in cognitive functioning, as measured by their performance on the matrix-reasoning subtest of the Wechsler Abbreviated Scale of Intelligence (WASI-MR; Wechsler, 1991). See Supplementary Table S1 for the means, standard deviations, and ranges for each age group.

### 2.2. Study procedure

Each participant visited the lab on two separate occasions, which were spaced an average of 20 ± 21 days apart (range: 0–125 days[2]). Across the two lab visits, participants completed interviews, computer tasks, pen-and-pencil questionnaires, and an MRI scan. Saliva samples for hormone assessment were collected at home, during the time in between the two lab visits. The participants were instructed on how to conduct saliva donation by passive drool during the first lab visit (further described below), and they brought the samples that were collected at home to the lab on their second visit. Collection of the first saliva sample occurred, on average, 10 days (±14.1 days) before the second (MRI) lab visit.[3]

During the second lab visit, participants underwent MRI scanning. Before they entered the scanner, participants were instructed on how to play the fMRI task and completed 12 practice trials. Each participant then completed five scans: a structural scan, a resting state scan, two task-related scans, and another resting-state scan, in that order (results from the resting state scans are reported in Kayser et al., 2016). Upon completion of the first structural scan, we visually inspected it for signs of excessive movement; if present, we collected an additional structural image at the end of the scanning procedure. In total, participants spent up to one hour in the scanner.

### 2.3. Pubertal hormones

Testosterone and estradiol levels were measured based on two saliva samples provided by each participant. Salivary testosterone levels are strongly correlated with serum-based testosterone levels, thus providing an accurate indication of testosterone production, even in pre-pubertal boys and girls (Ostatníková et al., 2002). The relation between saliva-based and serum-based estradiol levels among children and adolescents has yet to be tested, although a strong association has been found in postmenopausal women on hormone treatment (Tivis et al., 2005). We used the passive drool method for saliva collection to minimize discomfort and maximize compliance (Shirtcliff et al., 2001). Participants were instructed to collect the two saliva samples at home on separate – preferably consecutive – mornings between 7:00 a.m. and 9:00 a.m. To aid saliva collection, participants were provided with 2 mL tubes, which they were instructed to fill up at least halfway, and straws. To prevent (blood) contamination of the saliva samples, participants were asked to avoid (1) brushing their teeth or eating a major meal for at least 1 h prior to collection, (2) eating anything acidic or high-sugar within 20 min before collection, and (3) consuming something that stimulated the production of saliva (e.g., chewing gum). They were also asked to rinse their mouth with water about 10 min prior to collection. Finally, participants were instructed to store the samples in their freezer at home immediately upon collection and to return the samples to the lab during the second (MRI) visit using an ice pack.

Saliva samples brought into the lab were immediately stored in a freezer at −20 °C. Testosterone (T) assays were conducted in the Kriegsfeld laboratory at UC Berkeley, using Salimetrics salivary testosterone enzyme immunoassay kits with rabbit antibodies and a sensitivity of 1 pg/mL (www.salimetrics.com). These assays were run in duplicate and were repeated for samples with an *intra*-assay coefficient of variability (CV) above 7%. Of these repeats, the assay results with the lowest intra-assay CV (i.e., highest reliability) were included for analysis ($M$ intra-assay CV = 2.2%, SD = 1.9%, range: 0–9.4%). All samples were analyzed across a total of 6 separate assays with an *inter*-assay CV of 21.3%. Estradiol (E) assays were conducted at the University of New Orleans, Louisiana, under supervision of Dr. E.A. Shirtcliff, using Salimetrics salivary estradiol enzyme immunoassay kits with rabbit antibodies and a sensitivity

[1] Only one run (instead of two) of task-related imaging data was collected ($n = 2$), the task did not work ($n = 3$), and the data were not saved ($n = 1$).

[2] The participant who completed the second lab visit 125 days later than the first lab visit was an outlier; the second-longest inter-session interval was 67 days.

[3] Information about the dates and times of saliva collection was missing for 13 participants included in this study.

of 0.1 pg/mL (M intra-assay CV = 6.3%, SD = 6.0%, range: 0.08–28.8%; inter-assay CV = 12.8%).

One participant (12 yrs) was excluded because her sample could not be located after initial storage in the lab. There were no significant differences between the two samples collected from each participant (T: $t(55) = 0.21$, $p = 0.83$; E: $t(53) = 1.7$, $p = 0.10$). Thus, hormone levels were calculated as the average across the two samples collected by each participant, unless one of the samples was excluded due to any of the following reasons: (1) too dirty to be analyzed ($n = 1$), (2) contained insufficient quantity of saliva ($n = 1$; estradiol only), (3) had an intra-assay CV > 30% ($n = 2$; estradiol only), or (4) had a concentration > 3 SDs above the mean ($n = 1$; estradiol only); in these cases the value of the valid sample was used (instead of the average). This led to the exclusion of one additional participant (13 yrs) for estradiol only. Hence, results for testosterone are based on 57 participants, whereas those for estradiol are based on 56 participants. See Table 1 for the means, standard deviations, and ranges for each age group. Note that while there were no differences in testosterone level between age groups, there was an age-related increase in estradiol level (Table 1); post-hoc Tukey tests revealed that 12- and 13-year-olds had higher estradiol levels than 11-year-olds ($p = 0.040$ and $p = 0.002$, respectively), but did not differ from each other ($p = 0.51$). Finally, because testosterone levels were positively skewed ($z = 2.8$, $p = 0.005$) and therefore not normally distributed (Shapiro–Wilk test statistic = 0.94, $p = 0.006$) but did not contain zero values, analyses were conducted using the square-root transformed values for testosterone level (Shapiro–Wilk test statistic = 0.97, $p = 0.14$).

Importantly, the saliva samples that were obtained in this study showed substantial inter-individual variability in measured estradiol levels (range: 0.43–2.66; $M = 1.51 \pm 0.53$). Substantial inter-subject variability derives in large part from the fact that, based on our study design, some of our participants had not yet started menstruating ($n = 26$ out of 56) and others who had more recently reached menarche were less likely to have developed regular menstrual cycles. Thus, obtaining estradiol levels that are locked to a particular phase of the menstrual cycle could not be done across our participants. However, by obtaining morning estradiol levels without regard to cycle, as we did, we might also be likely to encounter greater within-subject variability in estradiol measurements in regularly cycling girls.

Despite the issue of not being able to phase-lock the estradiol measure for our sample of girls, it is likely that variability across developmental stages – the factor in which we were interested – would be greater than variability within individuals. Previous work in a very similar age group, for example, has demonstrated that average estradiol levels in a group of 9-year-old twins followed longitudinally through age 12 showed a 3-fold increase between those ages in the girls ($n = 87$), a time during which the average Tanner score increased from 1 to 3–4 (Koenis et al., 2013). Variations about this higher average value during menstrual cycles would therefore still be expected to lead to higher estradiol levels in post-menarche than pre-menarche girls. To test this idea, we correlated estradiol levels with our other pubertal measures: testosterone level (which varies minimally with menstrual cycle: Liening et al., 2010) and PDS score. If within-subject variance in estradiol levels outweighed between-subject variance, we would anticipate a failure to identify significant relationships between estradiol and these additional measures. However, estradiol level was significantly correlated with both (square-root transformed) testosterone level ($r = 0.38$, $p = 0.004$) and PDS score ($r = 0.52$, $p < 0.001$), arguing that inter-individual variability in estradiol levels could be validly assessed in our sample of girls, despite potential differences in the timing of salivary sample acquisition relative to menstrual cycle phase. Of note, while estradiol level corresponded with all other pubertal measures, including testosterone level and PDS score (see Section 2.4), testosterone level only corresponded with estradiol level. Furthermore, both testosterone and estradiol level showed a (marginally) significant association with age ($r = 0.26$, $p = 0.056$ and $r = 0.44$, $p = 0.001$, respectively).

### 2.4. Self-reported pubertal stage

During the first lab visit, participants completed the Pubertal Development Scale (PDS; Petersen et al., 1988), a self-report measure of pubertal stage. We used this measure based on evidence showing that this self-report measure can be compared to the scores derived from physical examination done by a nurse practitioner (Shirtcliff et al., 2009), and because it has high reliability or internal consistency, as indicated by a high Cronbach's alpha: $\alpha = 0.73$ for the current sample, compared to $\alpha = 0.81$ for the girls reported in Shirtcliff et al. (2009). The PDS consists of five questions about the physical changes associated with puberty (i.e., changes in height, body hair, and skin, as well as breast development and the presence of menarche) that were scored from "no physical changes" (1) to "development seems complete" (4), except for the item about menarche, which was scored as either "no" (1) or "yes" (4). The average of all five items (i.e., the total score) was calculated to provide an index of pubertal stage (see Table 1 for the means, standard deviations, and ranges for each age group). Of note, if there was a lag of more than 45 days between the two lab visits, the PDS was re-administered during the second visit to control for any pubertal changes during the time in between visits. For the participants who filled out the PDS twice ($n = 4$), we used the total score based on the second-time completion.[4] Differences in the PDS scores between first and second-time completion ranged from 0.0–0.6, indicating no change or increases in PDS score.

As shown in Table 1, PDS (total) scores differed between the age groups; post-hoc Tukey tests revealed that both 12- and 13-year-olds scored higher than 11-year-olds ($p = 0.026$ and $p = 0.001$, respectively), but did not differ from each other ($p = 0.37$). While PDS score was positively correlated with age ($r = 0.47$, $p < 0.001$) and with estradiol level ($r = 0.52$, $p < 0.001$), there was no correlation with testosterone level ($r = 0.16$, $p = 0.25$), which is consistent with a previous study in 10-to-16-year-olds that showed a positive correlation between testosterone level and PDS score in boys, but no significant correlation in girls (Herting et al., 2011). Furthermore, the lack of a significant relation between self-reported pubertal stage and testosterone level was consistent across PDS subscores (adrenal: $r = 0.18$, $p = 0.18$; gonadal: $r = 0.14$, $p = 0.32$), as calculated according to the scoring method used in Shirtcliff et al. (2009). In fact, none of the individual item scores correlated with testosterone level ($-0.03 < r < 0.17$, $0.20 < p < 0.83$), except for the item about skin changes, for which scores showed a marginally significant positive relation with testosterone level ($r = 0.25$, $p = 0.061$), indicating that girls with higher testosterone levels tended to report more skin changes. In contrast, all item scores correlated positively with estradiol level ($0.26 < r < 0.51$, $p \leq 0.05$). Together, these findings support the idea that, in girls, individual variation in levels of estradiol, rather than testosterone, relates more strongly to differences in pubertal stage (Shirtcliff et al., 2009). Finally, the mean age to start menstruating was 11 years and 10 months (±8 months); among the 11-year-olds 30% reported having started to menstruate, while for the 12- and 13-year-olds 58% and 81% had started menstruating, respectively.

[4] Two 11-year-olds and two 12-year-olds completed the PDS a second time (interval range: 47–125 days). One 11-year-old returned after an interval of 53 days, but did not complete the PDS again. For this participant, we used the first-visit PDS score. Exclusion of this participant did not change the results.

**Table 1**
Descriptives of the hormone and self-report measures: Mean ± standard deviation (and range).

| | 11 years ($n$ = 23) | 12 years ($n$ = 19) | 13 years ($n$ = 16) | Group differences |
|---|---|---|---|---|
| | ($n$ = 23) | ($n$ = 18) | ($n$ = 16) | |
| Testosterone level (in pg/mL) | 46.3 ± 14.1 (17.9–75.6) | 53.4 ± 17.4 (32.2–91.3) | 56.6 ± 20.1 (33.1–102.6) | $F(2, 54) = 1.9$, $p = 0.15$ |
| | ($n$ = 23) | ($n$ = 18) | ($n$ = 15) | |
| Estradiol level (in pg/mL) | 1.2 ± .51[a, b] (0.43–2.6) | 1.6 ± 0.51[a] (0.91–2.4) | 1.8 ± 0.41[b] (1.1–2.7) | $F(2, 53) = 6.9$, $p = 0.002$ |
| PDS score (Range: 1–4) | 2.2 ± 0.56[c, d] (1.2–3.2) | 2.7 ± 0.68[c] (1.6–3.8) | 3.0 ± 0.55[d] (2.2–3.8) | $F(2, 55) = 8.5$, $p = 0.001$ |
| RQ score (Range: 0–24) | 0.91 ± 1.4 (0–6) | 1.7 ± 1.5 (0–6) | 2.1 ± 2.6 (0–9) | *$F(2, 55) = 2.3$, $p = 0.11$* |

[a, b, c, d] Age groups that differed significantly from one another based on a post-hoc Tukey test ($p < 0.05$). PDS = Pubertal Development Scale, RQ = Risk Questionnaire.

### 2.5. Self-reported risky behaviors

During the first lab visit, participants also completed the Risk Questionnaire (RQ; modified from Steinberg et al., 2009), a self-report measure for risky behaviors. This questionnaire consisted of eight risky behaviors (ride in a car with a drunk driver, drink alcoholic beverages, smoke cigarettes, engage in vandalism, shoplift, go into a dangerous neighborhood, get into a heated argument, get into a physical fight) and participants were asked to report how many times they did those things in the last six months by choosing from the options "None" (scored as 0), "Once or twice" (1), "Two to five times" (2), or "More than 5 times" (3). With the prospect of conducting a 1–2 year follow-up, we selected this measure to capture real-life high-risk behaviors that are thought to become more prevalent across adolescence. RQ scores were calculated as the sum of all items (see Table 1 for the means, standard deviations, and ranges for each age group). Note that there were no age-related differences in RQ scores (Table 1). Furthermore, although RQ scores contained zero values and were highly positively skewed ($z = 6.3$, $p < 0.001$) and therefore not normally distributed (Shapiro–Wilk test statistic = 0.76, df = 58, $p < 0.001$), transformation of the RQ scores was ineffective (i.e., log-transformed values of RQ remained positively skewed). Thus, we used a non-parametric test (Spearman's rho) to assess the relations between RQ score and our developmental measures; RQ scores were positively associated with testosterone level ($\rho = 0.31$, $p = 0.021$), estradiol level ($\rho = 0.47$, $p < 0.001$), and PDS score ($\rho = 0.31$, $p = 0.017$), but not with age ($\rho = 0.21$, $p = 0.12$).

### 2.6. fMRI paradigm

All participants played a modified version of the Jackpot task (based on Op de Macks et al., 2011) while in the MRI scanner. In this two-choice probabilistic decision task, the chance to win on a given trial is presented visually in a way that is intuitive to children. On each trial, a slot machine appeared with two out of three slots showing plums. The three possible outcomes for the third slot were shown in a yellow frame above the slot machine. To win, all three slots need to show plums. In the low-risk condition, the presented chance to win was 67% (2/3); in the high-risk condition, the presented chance to win was 33% (1/3). Additionally, information about the reward at stake (1 or 3 points) was presented in a green frame above the slot machine. Thus, there were four types of trials (Fig. 1a–d): low-risk/high stakes (LR-3pts), low-risk/low stakes (LR-1pt), high-risk/high stakes (HR-3pts), and high-risk/low stakes (HR-1pt), which were presented in random order across the entire task.

Based on the presented information about risk level and stakes, the participant could choose to play (i.e., take the risk to win or lose 1 or 3 points), or to pass (i.e., skip the trial), as indicated by a button press with the right index or middle finger, respectively. The option to pass was added because prior studies demonstrated that outcome monitoring (or reward processing) is more salient when the outcome is the result of an *active* choice (Leotti and Delgado, 2014). Upon the button press (or after 2 s, in the absence of a response), the outcome was presented. When the participant chose to play, the outcome could be positive (gain) or negative (loss). During gain, participants saw three plums in a row and the words "you won" (Fig. 1e). During loss, participants saw a different fruit (orange or cherries) in the third slot accompanied by the words "try again" (Fig. 1f). Upon the choice to pass, the outcome was neutral (no gain or loss), and the third slot showed an "X" with the words "passed" (Fig. 1g). If participants failed to respond, the third slot showed an orange frame with the words "too slow" and they lost 1 point (Fig. 1h). This additional loss upon failure to respond was done to encourage task engagement.

The task was administered across two runs of scans, separated by a self-paced break (Fig. 2). Participants completed four blocks of 24 trials (i.e., 96 trials in total). Each trial started with a 500 ms fixation cross, which was jittered for an additional 0–8 s at 2 s increments. The stimulus was then presented until the subject responded (by button press), up to a maximum of 2 s. The button press was immediately followed by a 750 ms anticipation phase. During this phase, the third slot would spin (upon 'play'), or – to equate the visual experience of the anticipation phase across trial types – an "X" (for 'pass') or orange frame (for no response) would flicker in the third slot. The anticipation phase was followed by the outcome, which was presented for 2 s. To ensure that each trial had the same duration despite the RT-dependent choice phase, we subsequently extended the time that the fixation cross was presented during the inter-trial interval by [2 s – RT]. After every six trials, participants received feedback about their cumulative task performance; these *feedback phases* each lasted 4 s and were followed by 1 s of fixation. There were two types of feedback (monetary and social rank), which were each presented throughout two of the four task blocks (see Supplementary material for a detailed description of the feedback). For this paper, results were collapsed across feedback type, as there were no effects of feedback type on risk taking (see Section 3.1.3). Results for the effects of feedback type – collapsed across risk level and stakes – on brain and behavior, as well as individual differences in neural and behavioral influences of feedback type and their associations with pubertal differences, are reported elsewhere (Op de Macks et al., in press).

Participants were instructed that they had $5 in play money and that they could increase this amount up to $30 if they chose to play. All participants were told that they would be paid according to their final score – in points – which was translated into a monetary amount at the end of the experiment. In actuality, all participants won $10 because the choice to play resulted in positive feedback in 50% of the trials, regardless of the presented risk. In keeping with our previous work (Op de Macks et al., 2011), this choice was made to facilitate comparisons between gains and losses in each of the four task conditions – i.e., to have a similar number of observations for both trial outcomes, thereby better enabling direct comparison of the brain response associated with gain and loss – while limiting scanner time in order to reduce subject fatigue. Importantly, the discrepancy between the *presented* probability of winning (i.e., 33% or 67%) and the *experienced* probability of winning (i.e., 50%) did not affect choice behavior at the group level for a number of reasons. First, because each condition included 24 trials, the absolute devi-

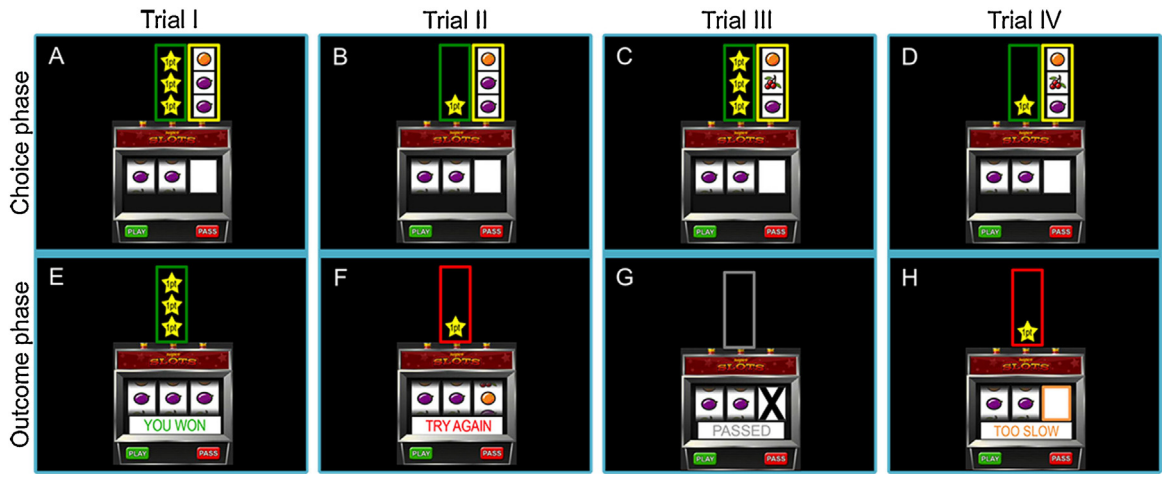


**Fig. 1.** Examples of trials in the Jackpot task (modified from Op de Macks et al., 2011). During the choice phase (a–d), four different stimuli conditions were presented in random order. During the outcome phase (e–h), four possible outcomes (depending on participants' choices) were presented. Trial I: the chance to win 3 points was 67%; the participant chose to play and won. Trial II: the chance to win 1pt was 67%; the participant chose to play and lost. Trial III: the chance to win 3pts was 33%; the participant chose to pass and nothing happened. Trial IV: the chance to win 1 point was 33%; the participant did not respond and lost 1pt. Note that while all possible stimuli and outcomes are depicted in this figure, in actuality other combinations of choices and outcomes were possible, depending on the participant's decisions.

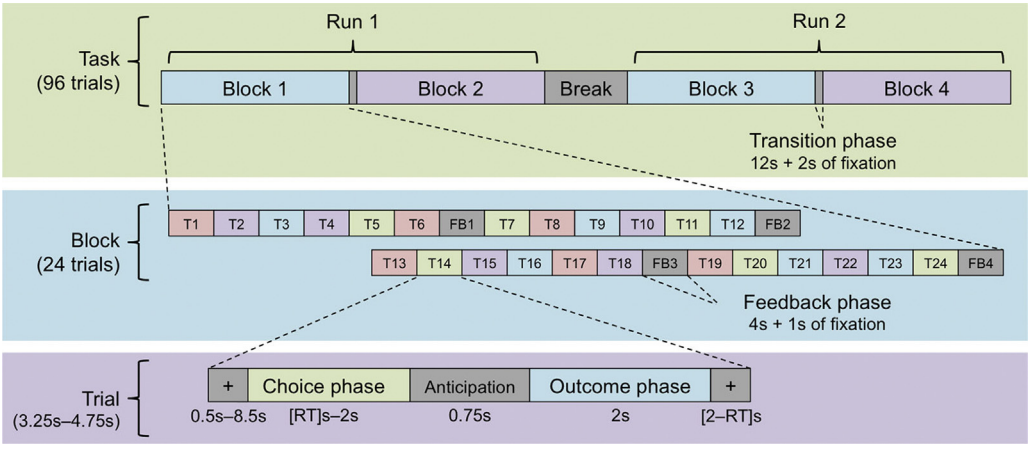


**Fig. 2.** *Upper panel:* The task was administered across two runs of scans with a self-paced break in between. Before each run, participants were told which feedback type would be presented first; in between blocks (within the same run) they were visually prompted about the transition in feedback type (i.e., transition phase). A fixation cross was presented at the start of every run for 2 s and after each feedback phase (1 s) and transition phase (2 s). Throughout each block, feedback type (Social rank or Money) was held constant and the order was counterbalanced between participants (MSMS or SMSM), within each age group. *Middle panel:* Each block consisted of 24 trials – 6 trials of each condition – presented in random order. Feedback phases occurred after every 6 trials – i.e., 4 times in each block. *Bottom panel:* Trials consisted of a choice phase, in which participants chose to play or pass based on information about risk level (33% vs. 67%) and stakes (1pt vs. 3pts), and an outcome phase, during which participants were shown whether they won or lost (upon the choice to play), or that nothing changed (upon the choice to pass). Before each trial, a fixation cross was presented for the duration of 500 ms + an additional 0–8 s that was jittered at 2 s increments. After each trial, a fixation cross was presented for 2 s – [RT] (i.e., the remaining duration of the choice phase).

ation between the expected and actual number of gain trials was limited to 4 of 24 trials in any one condition (e.g., 16 versus 12 gains in the low risk condition); and when summed across all conditions, the expected total number of gain trials remained unchanged. Second, subjects did not recognize these small discrepancies, even after experience with the task. Specifically, while the percentage of play choices differed significantly between the four task conditions, $F(3, 55) = 83.0$, $p < 0.001$ (see Supplementary Fig. S1a for a distribution of the percentage of play choices in each of the four task conditions), there were no differences across the four task blocks of 24 trials each, $F(3, 55) = 0.99$, $p = 0.41$. Furthermore, there was no interaction between conditions and blocks, $F(9, 49) = 0.74$, $p = 0.67$. These results indicate that participants adjusted their choices based on the information provided about risk level and stakes, but did not change their choice behavior over time (i.e., based on their task experience). The absence of a learning effect was similar across task conditions (see Supplementary Fig. S1b).

### 2.7. MRI image acquisition

MRI scanning was conducted on a Siemens MAGNETOM Trio 3T MR Scanner using a 12-channel head coil at the Henry H. Wheeler, Jr., Brain Imaging Center at the University of California, Berkeley. During the structural scan, we collected an anatomical image of 160 slices acquired using a T1-weighted MP-RAGE protocol (TR = 2300 ms; TE = 2.98 ms; FOV = 240 × 256 mm; matrix size = 240 × 256 mm; voxel size = 1 $mm^3$; 160 volumes/run; 1 run). During the task-related scans, we collected two runs of functional images (285 volumes/run of 6.5 minutes each), which consisted of 24 axial slices acquired with an ascending interleaved gradient echoplanar imaging protocol (TR = 1370 ms; TE = 27 ms; FOV = 225 × 225 mm; matrix size = 96 × 96 mm; voxel size = 2.3 × 2.3 × 3.5 mm; inter-slice gap ~0.3 mm). The fMRI task was programmed and presented using Visual Basic 6.0 software (microsoft-visual-basic.en.softonic.com) and projected onto

a screen behind the head coil within the scanner bore. A mirror was placed on top of the head coil to allow the participant to see the display. Participants made their responses on an MRI-safe fiber optic response pad (Inline Model HH-1x4-L; www.crsltd.com). Head motion was restricted by foam inserts that surrounded the head.

### 2.8. fMRI preprocessing

Functional images were converted from DICOM to 4D NIfTI format using MRIcron (www.mccauslandcenter.sc.edu). Preprocessing was performed using SPM8 (Wellcome Trust Center for Neuroimaging; www.fil.ion.ucl.ac.uk). Motion correction was performed using a two-pass procedure in which the images were first registered to the first image, after which they were registered to the resulting mean image. Images were corrected for slice-timing offsets using the first slice as a reference. Coregistration was performed with the mean image as a reference image, and the anatomical image as source image. Coregistered images were segmented using tissue probability maps, and were warped using the International Consortium for Brain Mapping (ICBM) space template. Next, the images were smoothed with a 6 mm full-width half maximum (FWHM) Gaussian kernel. Finally, we used ArtRepair, a toolbox for SPM (Mazaika et al., 2009) to identify and remove volumes that showed scan-to-scan movement greater than 0.5 mm, which were subsequently replaced by values resulting from linear interpolation from the neighboring unrepaired scans. Participants were excluded from analysis if more than 20% of their volumes had to be removed ($n = 3$; see Section 2.1). The total number of repaired volumes across both runs of scans ($M = 9.5 \pm 18.1$, range: 0–69 volumes) did not correlate with any of the developmental measures (age: $\rho = -0.02$, $p = 0.89$; testosterone level: $\rho = 0.07$, $p = 0.63$; estradiol level: $\rho = 0.13$, $p = 0.35$; PDS score: $\rho = -0.11$, $p = 0.43$). Note that non-parametric Spearman's rho tests were used, rather than transforming the total number of repaired volumes, because the total number of repaired volumes remained positively skewed even after transformation.

### 2.9. fMRI analyses

Statistical analyses were performed on individual subjects' data using the general linear model (GLM) in SPM8. To capture neural responses during the choice phase, trials were modeled as zero-duration events starting at the onset of stimulus presentation. Note that while each trial consisted of a stimulus, anticipation, and outcome phase, these phases were not modeled separately due to the absence of jittered periods in between the phases within each trial. Feedback phases were also modeled as zero-duration events starting at the onset of feedback presentation. Transition phases were modeled as 12-s events starting at the onset of the transition screen presentation. Here, we report the results of analyses collapsed across feedback type.

Three separate subject-specific design matrices were created to test for brain responses associated with risk taking (choice model), reward processing (outcome model), and task condition (conditions model). The choice model included three regressors of interest that modeled Play, Pass, and Miss trials. The outcome model included four regressors of interest: Gain, Loss, Pass, and Miss trials. Note that the only difference between the choice and outcome models is the further categorization of play trials into (1) play trials that resulted in gains, and (2) play trials that resulted in losses, which allowed for the comparison of gain and loss outcomes following the choice to play. The conditions model included four regressors of interest, one for each task condition: LR-1pt, LR-3pts, HR-1pt, and HR-3pts. For this model, we collapsed across choices to maximize our power despite a limited number of trials, used to reduce subject fatigue or disengagement given the pediatric sample. For each of these first-level statistical models, regressors of no interest were added for the (1) feedback phases (collapsed across feedback type), (2) transition phases, and (3–8) the movement parameters (roll, pitch, yaw and displacement in superior, left and posterior directions).

To examine *risk taking*-related brain activation across the entire group, second-level statistical analyses were conducted to test the contrast of Play > Pass trials. To examine *reward*-related brain activation across the group, we tested the contrast of Gain > Loss trials. To examine *risk*- and *stakes*-related brain activation across the group, we computed the following contrasts: high-risk (HR) > low-risk (LR) trials and high-stakes (3pts) > low-stakes (1pt) trials, respectively. Task-related responses were considered significant if they consisted of at least 10 contiguous voxels that exceeded a family-wise error (FWE) corrected threshold of $p < 0.05$ in SPM.

Region-of-interest (ROI) analyses were conducted to examine individual differences in brain activation using the MarsBar toolbox for SPM8 (Brett et al., 2002). Parameter estimates were extracted from two ROIs: bilateral nucleus accumbens (NAc), which was created by drawing a 4 mm-radius sphere around its MNI coordinates ($x = \pm 10$, $y = 12$, $z = -3$; based on Haber and Knutson, 2010) and medial orbitofrontal cortex (mOFC), a functional ROI that resulted from an exploratory whole-brain analysis with risk taking as a covariate (see Section 3.2.3).

Regression analyses were conducted to examine the relations between individual differences in activation of our ROIs (NAc and mOFC), the pubertal measures (testosterone level, estradiol level, and pubertal stage), and (task-related and self-reported) risk taking. Given the (marginally) significant associations between the pubertal measures and age (T: $r = 0.26$, $p = 0.056$; E: $r = 0.44$, $p = 0.001$; PDS: $r = 0.47$, $p < 0.001$), we added age as a covariate in both the behavioral and imaging regression analyses. Furthermore, to test whether the relations between the pubertal measures and risk taking were mediated by reward-related (NAc and mOFC) activation associated with risky decisions during the task, we conducted mediation analyses using a bootstrapping method with 1000 iterations provided by the PROCESS module for SPSS (Hayes, 2015; www.processmacro.org).

## 3. Results

### 3.1. Behavioral results

#### 3.1.1. Effects of risk level and stakes on task-related risk taking

Risk taking was measured as the percentage of trials on which the participant chose to play. Results of a repeated-measures ANOVA that included task-related risk taking as the dependent variable and feedback type (social rank or monetary), presented risk level (33% or 67%), and stakes (1 or 3 points) as predictors showed significant main effects for both risk level, $F(1, 57) = 208.8$, $p < 0.001$, and stakes, $F(1, 57) = 5.0$, $p = 0.030$. Follow-up analyses revealed that the girls chose to play more often in the low-risk condition ($M = 90.9 \pm 0.10\%$) compared to the high-risk condition ($M = 45.5 \pm 0.22\%$), $t(57) = 14.5$, $p < 0.001$. Furthermore, they played more often when a larger reward (3pts) was at stake ($M = 70.6 \pm 0.15\%$) compared to when a small reward (1pt) was at stake ($M = 44.5 \pm 0.11\%$), $t(57) = 13.1$, $p < 0.001$ (Fig. 3a). There was no significant interaction between risk level and stakes, $F(1, 57) = 0.56$, $p = 0.46$, indicating that the effect of stakes on task-related risk taking was similar across risk levels.

Results of a regression analysis with task-related risk taking as the dependent variable and (square-root transformed) testosterone level as predictor, while controlling for age, showed that girls with higher testosterone levels tended to engage in more

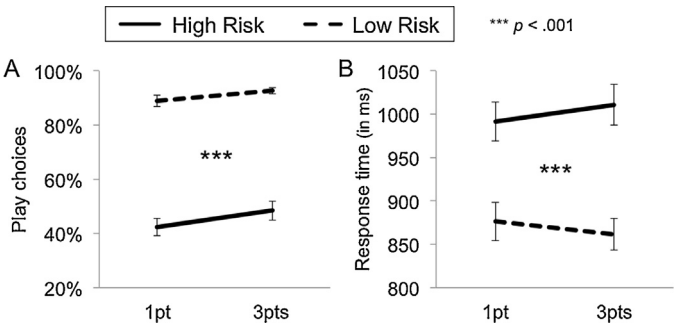


**Fig. 3.** Task behavior across all participants ($n = 58$). Group averages of risk taking (a) and response times (b) plotted separately for the low-risk and high-risk conditions. Besides the depicted main effects of risk level on risk taking and response times, there was also a significant main effect of stakes on risk taking ($p < 0.001$) and an interaction between risk level and stakes on response times ($p = 0.043$). Error bars represent the standard errors.

task-related risk taking ($\beta = 0.32$, $p = 0.020$), regardless of their age ($\beta = -0.13$, $p = 0.34$); however, this model only explained 10% of the variance in task-related risk taking, $F(2, 54) = 3.0$, $p = 0.06$. To further examine whether the relation between testosterone level and task-related risk taking depended on the task condition, we conducted a repeated-measures ANCOVA with testosterone level and age as covariates and risk taking in each condition as the dependent variable. Results showed that, besides a main effect of testosterone level on task-related risk taking, $F(1, 54) = 5.8$, $p = 0.019$, there was a third-order interaction between testosterone level and condition, $F(1, 54) = 5.0$, $p = 0.03$. Further analyses showed that girls with higher testosterone levels – independent of age – tended to play more often when 1pt was at stake ($\beta = 0.36$, $p = 0.008$; $F(2, 54) = 3.8$, $p = 0.028$), but not when 3pts were at stake, $F(2, 54) = 0.94$, $p = 0.40$. While this finding suggests that the relation between testosterone level and risk taking depended on stakes, the interaction between stakes and testosterone level was not significant, $F(1, 54) = 2.7$, $p = 0.11$. There was no interaction between risk level and testosterone level, $F(1, 54) = 0.92$, $p = 0.34$.

No associations were found between task-related risk taking and estradiol level, $F(2, 53) = 0.10$, $p = 0.91$, or PDS score, $F(2, 55) = 0.06$, $p = 0.95$, while controlling for age. These findings suggest that individual differences in task-related risk taking were associated with differences in the testosterone level, but not with differences in estradiol level or self-reported pubertal stage. Of note, there was no association between task-related and self-reported risk taking ($r = -0.13$, $p = 0.32$); this finding was true across task conditions (LR-1pt: $r = 0.10$, $p = 0.44$; LR-3pts: $r = -0.20$, $p = 0.13$; HR-1pt: $r = -0.08$, $p = 0.53$; HR-3pts: $r = -0.16$, $p = 0.22$).

#### 3.1.2. Effects of risk level and stakes on task-related decision speed

Decision speed was measured as the average time in milliseconds between stimulus onset and the button press, excluding trials on which the participant failed to respond. Results of a repeated-measures ANOVA with decision speed as the dependent variable and feedback type, risk level, and stakes as predictors showed a main effect of risk level, $F(1, 57) = 92.7$, $p < 0.001$, and an interaction between risk level and stakes, $F(1, 57) = 4.3$, $p = 0.043$. Follow-up analyses revealed that the girls took longer to make their decisions in the high-risk condition ($M = 1001 \pm 144$ ms) compared to the low-risk condition ($M = 869 \pm 167$ ms), $t(57) = 9.6$, $p < 0.001$. Furthermore, in the low-risk condition the girls were *faster* to decide when 3pts were at stake ($M = 862 \pm 139$ ms) than when 1pt was at stake ($M = 876 \pm 168$ ms), whereas in the high-risk condition the girls were *slower* to decide when 3pts were at stake ($M = 1011 \pm 179$ ms) than when 1pt was at stake ($M = 992 \pm 171$ ms) (Fig. 3b). However, while the interaction effect was significant, direct comparison of the means using paired-sample *t*-tests revealed non-significant differences for both contrasts (LR-1pt vs. LR-3pts: $t(57) = 1.04$, $p = 0.30$; HR-1pt vs. HR-3pts: $t(57) = 1.43$, $p = 0.16$). These results suggest that the girls' decision speed was mostly influenced by risk level.

Results of a regression analysis with task-related decision speed as the dependent variable showed no associations with testosterone level, $F(2, 54) = 1.8$, $p = 0.18$, although results of an ANCOVA with both testosterone level and age as covariates showed a marginally significant interaction between risk level and testosterone level, $F(1, 54) = 3.2$, $p = 0.079$, and a significant three-way interaction between risk level, stakes, and age, $F(1, 54) = 5.1$, $p = 0.028$, indicating that the effects of risk level and stakes depended on age (after controlling for testosterone level). Further analyses showed that while there was a main effect of risk level on decision speed across the three age groups (all $p$'s $< 0.001$), there was a significant interaction between risk level and stakes in the 12-year-olds only, $F(1, 18) = 9.1$, $p = 0.007$. No association was found between decision speed and estradiol level, $F(2, 53) = 1.9$, $p = 0.17$ or self-reported pubertal stage, $F(2, 55) = 0.32$, $p = 0.73$, while controlling for age.

Taken together, these findings suggest that the type of risky decisions, rather than decision speed, is associated with testosterone level, and neither the type nor speed of decision-making during the Jackpot task was related to estradiol level or self-reported pubertal stage.

#### 3.1.3. Effects of feedback type on task-related risk taking and decision speed

No main effects of feedback type were found for either task-related risk taking, $F(1, 57) = 0.05$, $p = 0.82$, or for decision speed, $F(1, 57) = 0.01$, $p = 0.91$. These findings indicate that, regardless of whether they were receiving monetary or social rank feedback, the girls, in the aggregate, made similar choices across the four task conditions. Thus, we collapsed across feedback type for all remaining analyses. Individual differences in the effect of feedback type on task behavior as well as neural differences associated with feedback type are reported elsewhere (Op de Macks et al., in press).

### 3.2. Imaging results

#### 3.2.1. Whole-brain results: activation of reward circuitry

To define the neural correlates of task-related risk taking, we evaluated a series of voxel-wise fMRI statistical analyses for each subject. Whole-brain results for Play > Pass trials across all participants revealed clusters of activation in bilateral striatum (caudate, putamen, globus pallidus, and nucleus accumbens), midbrain, and bilateral anterior insula (Fig. 4a). Furthermore, whole-brain results for Gain > Loss trials revealed clusters of activation in bilateral ventral striatum and medial prefrontal cortex (Fig. 4b). Finally, whole-brain results for Low-risk (LR) > High-risk (HR) trials also revealed clusters of activation in bilateral ventral striatum (Fig. 4c). (See Supplementary Table S2 for the MNI coordinates of these regions.) The opposite contrasts (Pass > Play, Loss > Gain, and HR > LR trials), as well as the contrasts High-stakes (3pts) > Low-stakes (1pt) and vice versa, did not reveal any clusters of activation after correction for multiple comparisons. (See Supplementary Fig. S2 for the clusters that appeared when we lowered the threshold to $p < 0.001$ uncorrected, $k \geq 10$ voxels, and see Supplementary Table S2 for their MNI coordinates.) Together, these results suggest the involvement of reward- and/or approach-related brain regions during risk taking, winning, and reward anticipation, respectively.

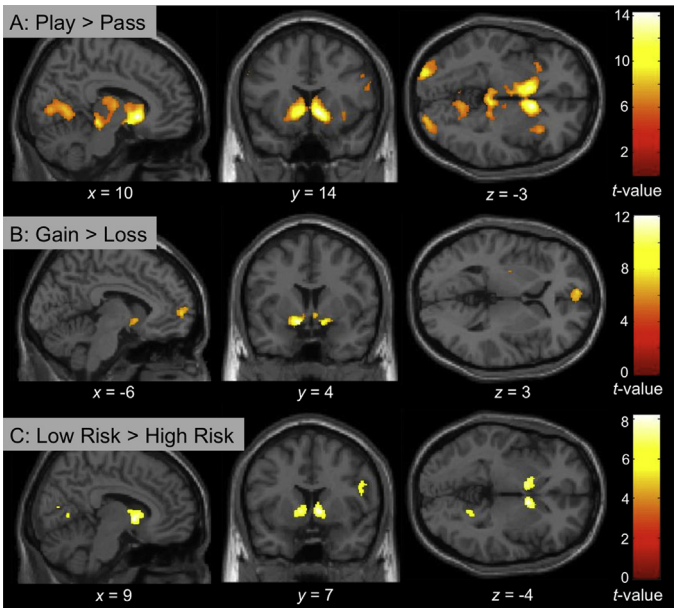


**Fig. 4.** Imaging results across all participants (n = 58). Whole-brain results for (a) Play > Pass trials, (b) Gain > Loss trials (upon the choice to play), and (c) Low-risk > High-risk trials (collapsed across choices and outcomes), corrected for multiple comparisons (FWE) at $p < 0.05$, $k \geq 10$ voxels.

### 3.2.2. ROI results: nucleus accumbens

*3.2.2.1. Nucleus accumbens activation across the group.* Our bilateral nucleus accumbens (NAc) ROI overlapped with the regions of activation that resulted from the contrasts of Play > Pass and Gain > Loss, suggesting that across the group NAc was significantly more active during task-related risk taking and winning. This possibility was confirmed by the time-series of this region, which showed that across the group NAc activation increased with the decision to play, but not with the decision to pass (i.e., Play – fixation > Pass – fixation). Moreover, NAc activation during Play trials remained elevated when play choices resulted in gains, whereas NAc activation returned to baseline more rapidly for play choices that resulted in losses (i.e., Gain – fixation > Loss – fixation; Fig. 5a).

Of note, our NAc ROI overlapped with the regions of activation that resulted from the contrast of LR > HR trials, suggesting that across the group NAc was significantly more active when the chance to win was greater. More specifically, results of a repeated-measures ANOVA with NAc activation as the dependent variable and risk level as well as stakes as predictors showed significant main effects of both risk level, $F(1, 57) = 43.4$, $p < 0.001$, and stakes, $F(1, 57) = 13.5$, $p = 0.001$, and a significant interaction between risk level and stakes, $F(1, 57) = 6.3$, $p = 0.015$. Follow-up analyses revealed that NAc activation was highest for the LR-3pts condition, intermediate for the LR-1pt condition, and lowest for the high-risk conditions; there were no significant activation differences between the HR-1pt and HR-3pts conditions, $t(57) = 1.2$, $p = 0.22$ (Fig. 5b). These results indicate that, regardless of the choices participants made during the task, activity within the NAc depended on a combination of the presented risk level and stakes.

*3.2.2.2. Individual differences in NAc activation.* Individual differences in NAc activation for Play > Pass trials, collapsed across all task conditions, were negatively associated with differences in task-related risk taking ($r = -0.28$, $p = 0.033$), indicating that girls who chose to play more often (across the task) showed *less* differential NAc activation between play and pass trials. This relation was most pronounced for risk taking in the HR-1pt condition ($r = -0.33$, $p = 0.011$; Supplementary Fig. S3); no associations between NAc (Play > Pass) activation and risk taking were found for the HR-3pts condition ($r = -0.12$, $p = 0.39$), or the LR conditions (LR-1pt: $r = -0.17$, $p = 0.22$; LR-3pts: $r = -0.06$, $p = 0.65$). These findings suggest that the extent to which the girls' NAc distinguished between Play and Pass trials across the task was most strongly associated with the extent to which they engaged in risk taking on HR-1pt trials.

Furthermore, when play- and pass-related NAc activations were evaluated separately (compared to fixation), results showed that girls who chose to play more often showed less NAc activation during Play trials ($r = -0.37$, $p = 0.004$), but similar NAc activation during Pass trials ($r = -0.14$, $p = 0.29$). The relation between risk taking and NAc activation was present for risk taking in the HR conditions (HR-1pt: $r = -0.31$, $p = 0.018$; HR-3pts: $r = -0.34$, $p = 0.009$), but not for risk taking in the LR conditions (LR-1pt: $r = -0.10$, $p = 0.45$; LR-3pts: $r = -0.01$, $p = 0.96$). These findings suggest that individual differences in risk taking-related NAc activation were more strongly associated with risk taking when the presented chance to win was relatively low (i.e., 33%). No association was found between individual differences in NAc activation for Play > Pass trials and differences in RT ($r = 0.21$, $p = 0.11$), although girls who showed more differential NAc activation between play and pass trials tended to make decisions more slowly on LR-1pt trials ($r = 0.31$, $p = 0.018$); no such relation existed for the other task conditions ($0.09 > r > 0.21$, $0.12 < p < 0.49$).

Next, we determined whether activity within the NAc might be associated with pubertal differences. To control for the possibility that individual differences in NAc activation for Play > Pass trials, collapsed across task conditions, were driven by behavioral differences on the task, we included risk taking in the HR-1pt condition, in addition to age, as a covariate in our regression analyses. Results of these regression analyses showed that girls with higher estradiol levels ($\beta = 0.29$, $p = 0.047$), in addition to taking fewer risks on HR-1pt trials ($\beta = -0.28$, $p = 0.035$), showed increased NAc activation for Play relative to Pass trials, regardless of their

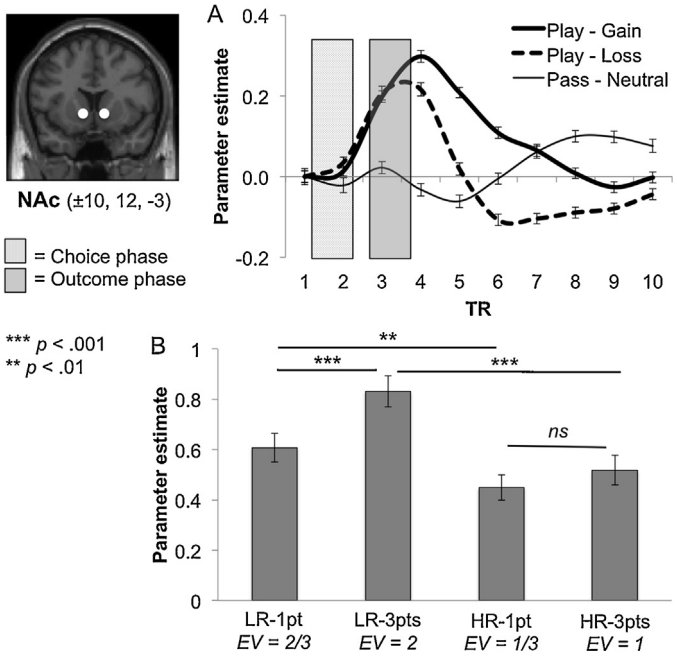


**Fig. 5.** (a) Time-series showing nucleus accumbens activation plotted separately for trials on which participants chose to play (separately for gain and loss outcomes) and to pass. (b) Average NAc activation across the group for each task condition, collapsed across play and pass choices. Error bars represent standard errors. NAc = nucleus accumbens, TR = repetition time, LR = low risk, HR = high risk, EV = expected value (i.e., risk level * stakes), ns = not significant.

age ($\beta = -0.13$, $p = 0.37$); this model explained 16% of the variance in NAc activation, $F(3, 52) = 3.2$, $p = 0.030$. No associations were found between individual differences in NAc activation and differences in testosterone level ($\beta = -0.10$, $p = 0.48$), or pubertal stage ($\beta = 0.22$, $p = 0.13$); for both models, differences in risk taking (in the HR-1pt condition) accounted for the variance in NAc activation: $\beta = -0.30$, $p = 0.034$, $F(3, 53) = 2.3$, $p = 0.085$ and $\beta = -0.33$, $p = 0.011$, $F(3, 54) = 3.1$, $p = 0.033$, respectively.

Furthermore, we conducted regression analyses with individual differences in NAc activation for the other contrasts (Gain > Loss and LR > HR) as dependent variables. Individual differences in risk taking (on HR-1pt trials) were associated with differences in NAc activation for LR > HR trials, collapsed across play and pass decisions ($r = -0.55$, $p < 0.001$), but not with NAc activation for Gain > Loss trials, within play trials collapsed across task conditions ($r = -0.04$, $p = 0.79$). Thus, we included risk taking on HR-1pt trials as a predictor in the regression analyses conducted for the LR > HR contrast only. No associations were found between individual differences in NAc activation and differences in testosterone level, estradiol level, or PDS score, for either contrast. Instead, individual differences in NAc activation for LR > HR trials were accounted for by differences in task-related risk taking (Supplementary Table S3), such that girls who took more risks in the HR-1pt condition showed *less* differential NAc activation for LR > HR trials. Together, these findings suggest that while both differences in task behavior and in estradiol levels contributed to individual differences in NAc activation for Play > Pass trials, only differences in task behavior contributed to individual differences in NAc activation for LR > HR trials (neither differences in task behavior or in any of the pubertal measures contributed to individual differences in NAc activation for Gain > Loss trials).

### 3.2.3. Exploratory whole-brain regression: medial orbitofrontal cortex

To examine whether any regions besides NAc correlated with task-related risk taking, we performed an exploratory whole-brain analysis for Play > Pass trials with the percentage of play choices as a covariate of interest. No regions of activation survived correction for multiple comparisons. However, when we added the percentage of play choices in the HR-1pt condition only – the same task condition in which NAc (Play > Pass) activation showed the strongest relation with behavior – we found a cluster of activation in right medial orbitofrontal cortex (mOFC) with a peak at $x = 9$, $y = 45$, $z = -14$ (peak-level FWE-corrected $p = 0.039$). Activation in this mOFC region showed a positive association with risk taking, indicating that girls who more often chose to play on HR-1pt trials tended to show *more* differential mOFC activation between Play and Pass trials across the task. More specifically, girls who took more risks in the HR-1pt condition activated this mOFC region *less* during trials on which they chose to pass compared to fixation ($r = -0.27$, $p = 0.041$), but showed no differences in mOFC activation during trials on which they chose to play compared to fixation ($r = 0.11$, $p = 0.40$). (See Supplementary Fig. S4a for the time-series of mOFC activation during Play and Pass trials across all participants.) No brain regions showed a negative association between differences in activation and risk taking on the Jackpot task.

To determine whether activity within the mOFC might be associated with pubertal differences, we next conducted separate regression analyses with mOFC activation for Play > Pass trials across the task as the dependent variable and (1) testosterone level, (2) estradiol level, and (3) PDS score as predictors, while controlling for age. Results showed that girls with higher testosterone levels ($\beta = 0.32$, $p = 0.020$), engaged mOFC more for Play than Pass trials, regardless of their age; however, this model only explained 10% of the variance in mOFC activation, $F(2, 54) = 2.9$, $p = 0.062$. No associations with mOFC activation were found for estradiol level, $F(2, 53) = 0.05$, $p = 0.95$, or PDS score, $F(2, 55) = 1.4$, $p = 0.26$. When additionally controlling for risk taking on HR-1pt trials, differences in testosterone level ($\beta = 0.18$, $p = 0.17$) no longer contributed to individual differences in mOFC activation; instead, differences in risk taking ($\beta = 0.44$, $p = 0.001$) explained 27% of the variance in mOFC activation, $F(3, 53) = 6.7$, $p = 0.001$. Similar findings resulted from the model including estradiol level; differences in risk taking ($\beta = 0.47$, $p < 0.001$), not estradiol ($\beta = 0.05$, $p = 0.70$) explained the variance in mOFC activation, $F(3, 52) = 4.9$, $p = 0.005$. However, differences in PDS score marginally contributed to individual differences in mOFC activation ($\beta = 0.25$, $p = 0.063$), in addition to differences in risk taking ($\beta = 0.49$, $p < 0.001$), $F(3, 54) = 7.2$, $p < 0.001$. These findings suggest that, despite marginal contributions of pubertal measures (testosterone level and PDS score), differences in task behavior contributed most to differences in mOFC (Play > Pass) activation.

To evaluate the specificity of these findings, we also conducted regression analyses with individual differences in mOFC activation for the other contrasts (Gain > Loss and LR > HR) as dependent variables. Because individual differences in risk taking on HR-1pt trials were not associated with mOFC activation for either Gain > Loss trials ($r = 0.03$, $p = 0.84$), or LR > HR trials ($r = 0.14$, $p = 0.30$), we did not include risk taking on HR-1pt trials as a predictor in the regression analyses conducted for these contrasts. No associations were found between individual differences in mOFC activation and differences in testosterone level, estradiol level, or pubertal stage, for either contrast. Instead, individual differences in mOFC activation for LR > HR trials were positively associated with age (Supplementary Table S3), such that older girls showed more differential mOFC activation for LR > HR trials. (See Supplementary Fig. S4b for the average mOFC activation during each task condition across all participants.)

Finally, to confirm whether activation differences in both NAc and mOFC contributed to the individual differences in risk taking, we conducted a regression analysis including (Play > Pass) activation for both regions as predictors of the percentage of play choices in the HR-1pt condition. Results showed that increased risk taking was associated with both *decreased* differential activation in NAc ($\beta = -0.36$, $p = 0.002$) and *increased* differential activation in mOFC ($\beta = 0.51$, $p < 0.001$). Together, these brain activation differences accounted for 36% of the variance in risk taking in the HR-1pt condition, $F(2, 55) = 15.7$, $p < 0.001$.

## 3.3. Linking hormones, brain, and behavior

To test the three-way relation between hormones, task-related risk taking, and associated brain activation, we conducted mediation analyses between these variables with age as a covariate. Results showed that, after controlling for age, the relation between testosterone level and task-related risk taking on HR-1pt trials was mediated by *increased* mOFC activation (Fig. 6a), such that girls with higher testosterone levels engaged mOFC *more* during trials on which they chose to play compared to pass, which in turn was associated with *more* risk taking. Interestingly, despite the absence of a direct relation between estradiol level and task-related risk taking, there was an indirect relation between estradiol level and risk taking on HR-1pt trials through *decreased* NAc activation (Fig. 6b), such that girls with higher estradiol levels engaged NAc *more* during trials on which they chose to play compared to pass, which in turn was associated with *less* risk taking. Together, these findings suggest that both hormones (testosterone and estradiol) might contribute (indirectly) to differences in task behavior through increased mOFC and decreased NAc activation during risk taking on the Jackpot task.

Additionally, given that individual differences in self-reported risky behaviors, as indicated by RQ scores, were associated with both testosterone and estradiol level (see Section 2.5), we

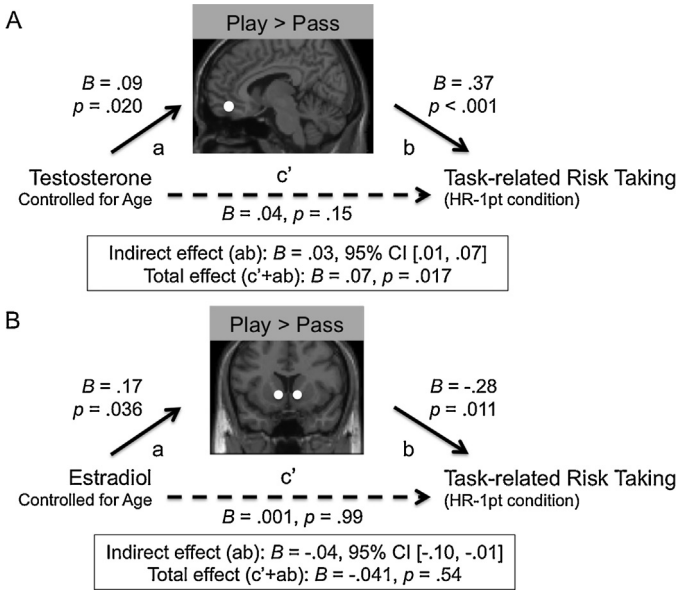


**Fig. 6.** Results of the mediation analyses. (a) Girls with higher testosterone levels, independent of age, showed more differential medial orbitofrontal cortex activation for Play > Pass trials across the task, which in turn was associated with *more* risk taking on HR-1pt trials of the Jackpot task. (b) Girls with higher estradiol levels, independent of age, showed more differential nucleus accumbens activation for Play > Pass trials across the task, which in turn was associated with *less* risk taking on HR-1pt trials of the Jackpot task.

conducted mediation analyses to test whether mOFC and NAc activation during *task-related* risk taking mediated the relations of testosterone and estradiol level, respectively, with *self-reported* risky behaviors. Results showed that girls with higher testosterone levels, independent of their age, showed increased mOFC activation during Play > Pass trials, which in turn predicted higher RQ scores (indirect effect: $B = 0.010$, 95% CI [0.001, 0.262]). However, NAc activation did not mediate the relation between estradiol level and RQ score (indirect effect: $B = 0.04$, 95% CI [−0.16, 0.36]). These findings suggest that the relation between testosterone level and self-reported risky behaviors – outside the laboratory setting – could be, at least partially, explained by heightened mOFC activation during risk taking on the Jackpot task, even in the absence of a direct association between RQ scores and mOFC activation ($\rho = 0.16$, $p = 0.24$).

## 4. Discussion

The goal of this study was to understand the three-way relation between hormone levels, risk taking, and associated reward-related brain processes during puberty. We tested this relation in a sample of 11-to-13-year-old girls and predicted that girls with higher levels of testosterone and estradiol, independent of age, would show increased risk taking (Vermeersch et al., 2008a, 2008b), which would be mediated by increased reward-related brain activation (Crone and Dahl, 2012). Indeed, self-reported risk taking was associated with all three pubertal measures (testosterone level, estradiol level, and PDS score), but not with age. For task-related risk taking, there was a significant positive relation with testosterone level, but no significant association with estradiol level (or PDS score), after controlling for age. Interestingly, the relation between testosterone level and task-related risk taking was mediated by mOFC activation, and there was an indirect relation between estradiol level and task-related risk taking through NAc activation. Specifically, girls with higher levels of testosterone tended to show increased mOFC activation during Play > Pass trials across the task, which in turn corresponded with *increased* risk taking on the Jackpot task. In contrast, girls with higher levels of estradiol tended to show increased NAc activation during Play > Pass trials across the task, which in turn corresponded with *decreased* risk taking on the Jackpot task. These three-way relations were most pronounced for risk taking in the HR-1pt condition. These findings are discussed below in terms of how each of these regions is thought to contribute to risk-taking behavior.

### 4.1. The role of mOFC in risk taking

Girls who engaged in more risk taking, particularly in the HR-1pt condition, showed increased mOFC activation during trials on which they chose to play as opposed to trials on which they chose to pass. The OFC is known to play a key role in determining the subjective value of a potential reward by integrating sensory, affective, and motivational signals from sensory and subcortical brain regions (Wallis, 2007; Levy and Glimcher, 2012). Particularly when faced with complex decisions in which reward value is not straightforward, the OFC is thought to calculate reward value by performing a cost-benefit-analysis based on different aspects of a decision, such as how much energy is required, whether it fulfills the individual's need, and what alternatives are present (Wallis, 2007). Moreover, the OFC is thought to be specifically involved in representing the reward value on a common scale in order to enable direct comparison of different alternatives, thus enabling complex decision-making (Levy and Glimcher, 2012). Furthermore, while the lateral OFC has connections with the sensory cortices and has been implicated in primary reward/punishment processing, medial OFC has connections with the limbic regions (including NAc) and is thought to be involved in determining subjective hedonic value for more abstract rewards, such as money (Peters and Büchel, 2010), especially the more anterior part of mOFC (Sescousse et al., 2013). Together, these findings suggest that girls who engaged in more risk taking when the expected value was the lowest (i.e., in the HR-1pt condition) experienced the choice to play as more rewarding overall compared to the choice to pass.

Importantly, individual differences in mOFC activation for Play > Pass trials across the task mediated the relation between testosterone level (independent of age) and both task-related and self-reported risk taking. These findings suggest that the mOFC may be a potential target for testosterone (see also Van Wingen et al., 2011). Furthermore, these findings are consistent with results from a cross-sectional study conducted in a large sample of 8- to 25-year-olds (males and females) that showed that individual differences in mOFC structure mediated the relation between individual differences in testosterone level and risk taking on the balloon analogue risk-taking (BART) task (Peper et al., 2013). Specifically, this study found that females with higher levels of testosterone tended to have a smaller surface area in mOFC, which in turn was associated with increased risk taking. These results were interpreted as mOFC volume acting as a suppressor of risk taking by regulating behavior (Peper et al., 2013). Since the relation between brain structure and function remains unknown, it is unclear how our finding of increased mOFC activation mediating risk taking can be reconciled with these structural findings. Furthermore, the structural analyses were conducted across a much larger region of mOFC compared to where we found activation differences. Given the heterogeneity of the mOFC (Peters and Büchel, 2010; Sescousse et al., 2013), it remains to be tested in future studies whether functional and structural differences associated with increased risk taking occur in overlapping regions. Future studies might also consider using other measures of real-world risk taking, given that RQ scores were not associated with risk taking on the Jackpot task.

### 4.2. The role of NAc in risk taking

As expected, our study showed that risk taking (Play trials) as opposed to opting out of a trial (Pass trials) was associated with increased activation in various reward-related regions, including the striatum (Delgado, 2007). Specifically, both dorsal and ventral striatum were activated during the choice to play, but only ventral striatum was activated during gains, which corresponds with the idea that, during decision-making, dorsal striatum is involved in action selection (Balleine et al., 2007), whereas ventral striatum is involved in reward evaluation (Haber and Knutson, 2010). We focused on the nucleus accumbens (NAc), a brain region located within the ventral striatum, because this region has been shown to be active during the anticipation and receipt of rewards during decision-making (Haber and Knutson, 2010). Indeed, during the Jackpot task, NAc activation was more active during trials on which girls decided to play compared to pass, and remained elevated for risky (i.e., play) decisions that resulted in gains, but not losses. Furthermore, NAc was more active during low-risk compared to high-risk trials, regardless of the girls' decisions. These findings correspond with the notion that, in contrast to mOFC, the NAc consistently responds to reward magnitude and sometimes probability, but does not take into account features like delay in reward receipt and effort involved in obtaining the reward (Haber and Knutson, 2010).

Interestingly, girls who engaged in more risk taking on the task showed *less* differential NAc activation for Play > Pass trials (and for LR > trials), suggesting that reward-related processes represented by the NAc were more similar across decisions (and risk levels) in girls who engaged in more risk taking on the task. More specifically, girls who took more risks showed *less* NAc activation on Play trials. This finding seems to contradict a common account for the increase in risk taking among adolescents, which is their increased sensitivity to rewards, as evidenced by their elevated NAc/ventral striatum response to rewards relative to children and/or adults (reviewed in Galvan, 2010). However, there is also evidence for a reduced NAc response in adolescents relative to adults, particularly during reward anticipation (Bjork et al., 2004, 2010). This relatively reduced NAc response has been interpreted as a potential driving force for adolescents to engage in risk taking, namely in order to increase activation of an otherwise blunted NAc. While our finding that girls who showed reduced NAc activation during Play trials actually engaged in *more* risk taking seems to be in line with the latter interpretation, we must be cautious with this interpretation for the following reasons.

First, individual differences in task-related risk taking were not only explained by differences in NAc activation, but also by differences in mOFC activation. While both regions have been associated with reward processing, NAc is more consistently reported during reward anticipation (Haber and Knutson, 2010), whereas mOFC is reported more for reward receipt (Diekhof et al., 2012). These findings suggest that NAc and mOFC may be contributing to different aspects of valuation and (risky) decision-making, which is supported by the finding that individual differences in mOFC activation, compared to individual differences in NAc activation, during risk taking on the Jackpot task were more strongly correlated with differences in self-reported risky behaviors. Unfortunately, the current version of the Jackpot task did not allow for the distinction between choice and outcome-related brain processes due to the absence of jittered periods in between the choice and outcome phases within each trial. Thus, it was not possible for us to reliably distinguish brain responses associated with reward anticipation from brain responses associated with reward receipt within a trial. Consequently, we were unable to determine whether the activation differences in NAc and mOFC reflected differences in valuation processes *before* the decisions, or outcome-related processes *after* the decisions (i.e., performance monitoring) that then influenced subsequent decision-making (Crone, 2014). Although other paradigms succeeded at disentangling the different reward processes associated with decision-making (Bjork et al., 2010; Forbes et al., 2010), these studies have not been able to capture their role in risk taking. Future research using pupillometry or electroencephalography (EEG), in addition to fMRI, might be a better method (given their better temporal resolution) to capture individual differences in arousal associated with risky decisions and outcome processing.

Second, it should be noted that we used a different study design compared to those used in developmental studies that found NAc hyper-activation in adolescents. The developmental studies that found adolescent NAc hyper-activation based their results on group comparisons; they contrasted NAc activation in a group of adolescents with that of a group of children and/or adults to make inferences about developmental changes (Galvan, 2010). In contrast, the present study zoomed in on a narrow developmental window to look at individual differences in NAc activation associated with pubertal differences in girls around the same age (11–13 yrs). Thus, while the girls who took more risks showed reduced NAc activation during trials on which they decided to play, this finding does not rule out that adolescent girls, as a group, show heightened NAc activation for rewards compared to children and/or adults. The best way to test this idea would be by conducting a longitudinal study (Crone and Elzinga, 2015). To date, a few longitudinal neuroimaging studies on risk taking have been published, and report conflicting results. Two studies reported no change in reward-related NAc activation from mid-adolescence to early adulthood (Lamm et al., 2014; Van Duijvenvoorde et al., 2014), whereas a third study based on a much larger sample reported a peak in NAc activation during adolescence (Braams et al., 2015). However, these results were not related to risk-taking behavior, pointing to the need for future longitudinal studies that test the relation between reward processing and risk taking across development.

Lastly, individual differences in NAc activation were not related to differences in testosterone level, but instead with differences in estradiol level, independent of age (and after controlling for task-related behavior). The absence of a relation between NAc activation and testosterone level in girls is in line with findings from a prior study that tested the relation between individual differences in testosterone level and differences in reward processing during risky decisions within a narrow age range of 11–13 years, using a similar approach (Forbes et al., 2010). Results of that study showed no relation between testosterone level and striatal activation during reward anticipation in girls; however, in boys, higher testosterone levels corresponded with increased striatum activation during reward anticipation (Forbes et al., 2010). Furthermore, other research has shown that higher testosterone levels correspond with increased risk taking in boys (Vermeersch et al., 2008b) and higher estradiol levels correspond with increased risk taking in girls (Vermeersch et al., 2008a). Given that boys as well as girls show a developmental increase in risky behavior (Shulman et al., 2015), these findings suggest that hormone levels that more closely track with pubertal development – i.e., testosterone in boys and estradiol in girls – might be more strongly associated with developmental increases in risk taking among boys and girls, respectively. Relatedly, while both testosterone and estradiol levels increase during puberty in girls (Biro et al., 2014), in a cross-sectional study, testosterone might be more reflective of individual differences among girls, which corresponds with our finding that differences in testosterone level, but not in estradiol level, corresponded with the types of choices made by the girls. Moreover, our results of an indirect relation between estradiol level, NAc activation, and task-related risk taking, in addition to the mediation of the relation between testosterone level and task-related risk taking by mOFC activation,

suggest that these hormones might target different neural mechanisms that underlie developmental and/or individual differences in risk taking. Future longitudinal studies are needed to distinguish maturational effects on risk taking and associated brain processes from individual differences in brain and behavior.

### 4.3. Differences in risk-taking strategy

Consistent with previous studies that used a similar decision paradigm (Op de Macks et al., 2011; Van Leijenhorst et al., 2008, 2010), the girls in this study showed an overall increase in risk taking when the probability to win or the stakes were larger. Contrary to a previous study that found no differences in decision speed between low- and high-risk options (Van Leijenhorst et al., 2010), the girls in the present study were slower to decide when the probability for reward was 33% compared to when the probability for reward was 67%, regardless of the stakes involved. Furthermore, individual differences in risk taking were larger for the HR condition, consistent with findings from an earlier version of the Jackpot task that was administered in a sample of 10–16-year-olds and included the same manipulation of risk level, but no manipulation of stakes (Op de Macks et al., 2011). These findings seem to suggest that overall the girls behaved rationally during the Jackpot task, despite showing individual differences.

In contrast to our expectations, developmental and neural differences in our sample of girls corresponded most with differences in risk taking when there was a chance of 33% to win 1pt. A possible explanation is that the girls might have different strategies for making these types of risky decisions; some girls may take a more deliberate or strategic approach, whereas other girls may take a more *feeling*-based or non-strategic approach. Strategic and non-strategic decision-makers may behave more similarly when the potential reward is relatively high (3pts), but not when the potential reward is low (1pt). In other words, while strategic decision-makers may choose to pass, non-strategic decision-makers may choose to play in the condition where the choice to play is not very strategic (i.e., the expected value is the lowest). According to this idea, the girls with higher testosterone levels and more discriminative mOFC responses, and girls with lower estradiol levels and less discriminative NAc responses, might be more likely to rely on a non-strategic approach, which could explain why they engaged in more risk taking when expected value was the lowest (and showed no differences when expected value was higher). Future research is needed to understand how strategy use during task-related risk taking and associated neural processes relate to real-world risky behaviors.

### 4.4. Conclusion

For this study, we administered a simple two-choice paradigm with explicit information about the risk level and stakes involved in each decision to a group of adolescent girls in the MRI scanner. We used this paradigm to test whether individual differences in risk taking and associated reward-related brain processes could be explained by differences in saliva-based testosterone and estradiol levels around the onset of puberty. Furthermore, we tested whether differences in reward-related brain processes *mediated* the relation between pubertal hormones and (task-related and self-reported) risk taking. The findings of this study provide evidence for the contribution of both testosterone and estradiol during puberty to adolescent risk taking in the context of the Jackpot task and offer a potential mechanism – i.e., through differences in subjective decision valuation, as reflected by differences in nucleus accumbens and medial orbitofrontal cortex activation – to explain why some girls engage in more risk taking compared to others. However, further research is needed to test whether these neuroendocrine mechanisms can explain individual differences in risk-taking tendencies outside the laboratory setting, and whether they play a role in the developmental increase in risk taking across adolescence in both sexes. Given that risk-taking tendencies often become overt in late adolescence or early adulthood (Shulman and Cauffman, 2014), the identification of an early 'biomarker' of adolescent risk taking might allow for earlier intervention with high-risk adolescents. Stratifying the vulnerability of such adolescents would permit the type and intensity of any treatments to be better individualized, in keeping with ongoing efforts in the evolution of precision medicine more generally.

## Conflicts of interest

None.

## Contributors

The corresponding author, Zdeňa A. Op de Macks, was a graduate student at the time of this study and contributed to designing the study, collecting and analyzing the data, and writing up the results for publication. Orly N. Bell was an undergraduate research assistant who contributed to collecting the data and writing up the results for publication. Silvia A. Bunge, Linda Wilbrecht, Lance J. Kriegsfeld, Andrew S. Kayser, and Ronald E. Dahl contributed to designing the study, analyzing the data, and writing up the results for publication.

## Role funding source

This work was supported by the Institute for Translational Neuroscience (grant W81XWH-11-2-0145 to A.S.K.); the National Center for Responsible Gaming/NCRG (A.S.K.); the Wheeler Center for the Neurobiology of Addiction; and funds from the state of California. The sponsors were not involved in the design of the study, the collection, analysis, or interpretation of the data, the writing of the manuscript, or the decision to submit the article for publication.

## Acknowledgements

The authors would like to thank Kiren Chand, Megan Johnson, Adelle Cerreta, Melissa Allen, Emmellia Dale, Christina Kirby, Erin Badduke, Angela Weinberg, Sohee Kim, and Monique Porsandeh for their help with data collection. We would also like to thank Sarah Munro, for programming the fMRI task, as well as Carter Wendelken, Michael Vendetti, Yana Fandakova, Amal Achaibou, Aroma Dabas, and Jenny Phan, for their help with data analysis.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.psyneuen.2016.08.013.

## References

Achenbach, T.M., 1991. Manual for the Child Behavior Checklist/4-18 and 1991 Profile. University of Vermont, Department of Psychiatry, Burlington (Vermont).

Balleine, B.W., Delgado, M.R., Hikosaka, O., 2007. The role of the dorsal striatum in reward and decision-making. J. Neurosci. 27, 8161–8165, http://dx.doi.org/10.1523/JNEUROSCI.1554-07.2007.

Biro, F.M., Pinney, S.M., Huang, B., Baker, E.R., Chandler, D.W., Dorn, L.D., 2014. Hormone changes in peripubertal girls. J. Clin. Endocrinol. Metabol. 99, 3829–3835, http://dx.doi.org/10.1210/jc.2013-4528.

Bjork, J.M., Knutson, B., Fong, G.W., Caggiano, D.M., Bennett, S.M., Hommer, D.W., 2004. Incentive-elicited brain activation in adolescents: similarities and

differences from young adults. J. Neurosci. 24, 1793–1802, http://dx.doi.org/10.1523/JNEUROSCI.4862-03.2004.
Bjork, J.M., Smith, A.R., Chen, G., Hommer, D.W., 2010. Adolescents, adults and rewards: comparing motivational neurocircuitry recruitment using fMRI. PLoS One 5, e11440, http://dx.doi.org/10.1371/journal.pone.0011440.
Braams, B.R., Van Duijvenvoorde, A.C.K., Peper, J.S., Crone, E.A., 2015. Longitudinal changes in adolescent risk-taking: a comprehensive study of neural responses to rewards, pubertal development and risk-taking behavior. J. Neurosci. 35, 7226–7238, http://dx.doi.org/10.1523/JNEUROSCI.4764-14.2015.
Brett, M.C., Anton, J.L., Valabregue, R., Poline, J.B., 2002. Region of interest analysis using an SPM toolbox. NeuroImage 16, 497.
Chick, C.F., 2015. Reward processing in the adolescent brain: individual differences and relation to risk taking. J. Neurosci. 35 (40), 13539–13541, http://dx.doi.org/10.1523/JNEUROSCI.2571-15.2015.
Crone, E.A., Dahl, R.E., 2012. Understanding adolescence as a period of social-affective engagement and goal flexibility. Nat. Rev. Neurosci. 13, 636–650, http://dx.doi.org/10.1038/nrn3313.
Crone, E.A., Elzinga, B.M., 2015. Changing brains: how longitudinal functional magnetic resonance imaging studies can inform us about cognitive and social-affective growth trajectories. WIREs Cogn. Sci. 6, 53–63, http://dx.doi.org/10.1002/wcs.1327.
Crone, E.A., 2014. The role of the medial frontal cortex in the development of cognitive and social-affective performance monitoring. Psychophysiology 51, 943–950, http://dx.doi.org/10.1111/psyp.12252.
Dahl, R.E., 2004. Adolescent brain development: a period of vulnerabilities and opportunities. Ann. N. Y. Acad. Sci. 1021, 1–22, http://dx.doi.org/10.1196/annals.1308.001.
De Water, E., Braams, B.R., Crone, E.A., Peper, J.S., 2013. Pubertal maturation and sex steroids are related to alcohol use in adolescents. Horm. Behav. 63, 392–397, http://dx.doi.org/10.1016/j.yhbeh.2012.11.018.
Delgado, M.R., 2007. Reward-related responses in the human striatum. Ann. N. Y. Acad. Sci. 1104, 70–88.
Diekhof, E.K., Falkai, P., Gruber, O., 2012. The orbitofrontal cortex and its role in the assignment of behavioural significance. Neuropsychologia 49 (5), 984–991, http://dx.doi.org/10.1016/j.neuropsychologia.2011.01.032.
Dorn, L.D., Dahl, R.E., Woodward, H.R., Biro, F., 2006. Defining the boundaries of early adolescence: a user's guide to assessing pubertal status and pubertal timing in research with adolescents. Appl. Dev. Sci. 10, 37–41, http://dx.doi.org/10.1207/s1532480xads1001.
Dreher, J.C., Schmidt, P.J., Kohn, P., Furman, D., Rubinow, D., Berman, K.F., 2007. Menstrual cycle phase modulates reward-related neural function in women. Proc. Natl. Acad. Sci. U. S. A. 104 (7), 2465–2470.
Forbes, E.E., Dahl, R.E., 2010. Pubertal development and behavior: hormonal activation of social and motivational tendencies. Brain Cogn. 72, 66–72.
Forbes, E.E., Ryan, N.D., Phillips, M.L., Manuck, S.B., Worthman, C.M., Moyles, D.L., Dahl, R.E., 2010. Healthy adolescents' neural response to reward: associations with puberty, positive affect, and depressive symptoms. J. Am. Acad. Child Adolesc. Psychiatry 49, 162–172.
Galvan, A., 2010. Adolescent development of the reward system. Front. Hum. Neurosci. 4, 1–9, http://dx.doi.org/10.3389/neuro.09.006.2010.
Goddings, A.L., Mills, K.L., Clasen, L.S., Giedd, J.N., Viner, R.M., Blakemore, S.J., 2014. The influence of puberty on subcortical brain development. NeuroImage 88, 242–251, http://dx.doi.org/10.1016/j.neuroimage.2013.09.073.
Haber, S.N., Knutson, B., 2010. The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology 35, 4–26, http://dx.doi.org/10.1038/npp.2009.129.
Harden, K.P., Mann, F.D., 2015. Biological risk for the development of adolescent externalizing: integrating insights from behavioral genetics and neuroscience. Child Dev. Perspect. 9, 211–216, http://dx.doi.org/10.1111/cdep.12135.
Hayes, A.F., 2015. An index and test of linear moderated mediation. Multivar. Behav. Res. 50 (1), 1–22, http://dx.doi.org/10.1080/00273171.2014.962683.
Hermans, E.J., Bos, P.A., Ossewaarde, L., Ramsey, N.F., Fernández, G., Van Honk, J., 2010. Effects of exogenous testosterone on the ventral striatal BOLD response during reward anticipation in healthy women. NeuroImage 52, 277–283.
Herting, M.M., Maxwell, E.C., Irvine, C., Nagel, B.J., 2011. The impact of sex, puberty, and hormones on white matter microstructure in adolescents. Cereb. Cortex 22 (9), 1979–1992, http://dx.doi.org/10.1093/cercor/bhr246.
Herting, M.M., Gautam, P., Spielberg, J.M., Kan, E., Dahl, R.E., Sowell, E.R., 2014. The role of testosterone and estradiol in brain volume changes across adolescence: a longitudinal structural MRI study. Hum. Brain Mapp. 35, 5633–5645, http://dx.doi.org/10.1002/hbm.22575.
Kayser, A.S., Op de Macks, Z.A., Dahl, R.E., Frank, M.J., 2016. A neural correlate of exploratory behavior in early adolescence. J. Cogn. Neurosci. 28 (2), 199–209, http://dx.doi.org/10.1162/jocn_a_00896.
Koenis, M.M.G., Brouwer, R.M., van Baal, G.C.M., van Soelen, I.L.C., Peper, J.S., van Leeuwen, M., Delemarre-van de Waal, H.A., Boomsma, D.I., Hulshoff Pol, H.E., 2013. Longitudinal study of hormonal and physical development in young twins. J. Clin. Endocrinol. Metab. 98 (3), E518–E527.
Koolschijn, P.C.M.P., Peper, J.S., Crone, E.A., 2014. The influence of sex steroids on structural brain maturation in adolescence. PLoS One 9, e83929, http://dx.doi.org/10.1371/journal.pone.0083929.
Lamm, C., Benson, B.E., Guyer, A.E., Perez-Edgar, K., Fox, N.A., Pine, D.S., Ernst, M., 2014. Longitudinal study of striatal activation to reward and loss anticipation from mid-adolescence into late adolescence/early adulthood. Brain Cogn. 89, 51–60, http://dx.doi.org/10.1016/j.bandc.2013.12.003.
Leotti, L.A., Delgado, M.R., 2014. The value of exercising control over monetary gains and losses. Psychol. Sci. 25, 596–604, http://dx.doi.org/10.1177/0956797613514589.
Levy, D.J., Glimcher, P.W., 2012. The root of all value: a neural common currency for choice. Curr. Opin. Neurobiol. 22, 1027–1038, http://dx.doi.org/10.1016/j.conb.2012.06.001.
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 contraceptive use on steroid hormone levels. Physiol. Behav. 99, 8–16, http://dx.doi.org/10.1016/j.physbeh.2009.10.001.
Mazaika, P., Hoeft, F., Glover, G.H., Reiss, A.L., 2009. Methods and software for fMRI analysis for clinical subjects. In: Poster Presentation at the Ann. Hum. Brain Mapp. Meeting, San Francisco, CA, USA.
O'Doherty, J.P., 2007. Lights, camembert, action! The role of human orbitofrontal cortex in encoding stimuli, rewards, and choices. Ann. N. Y. Acad. Sci. 1121, 254–272.
Op de Macks, Z.A., Gunther Moor, B., Overgaauw, S., Güroğlu, B., Dahl, R.E., Crone, E.A., 2011. Testosterone levels correspond with increased ventral striatum activation in response to monetary rewards in adolescents. Dev. Cogn. Neurosci. 1, 506–516, http://dx.doi.org/10.1016/j.dcn.2011.06.003.
Op de Macks, Z.A., Bunge, S.A., Bell, O.N., Kriegsfeld, L.J., Kayser, A.S., Dahl, R.E., under review. The effect of social rank feedback on risk taking and associated reward processes in adolescent girls.
Ostatníková, D., Pastor, K., Putz, Z., Dohnányiová, M., Mat'ašeje, A., Hamp, R., 2002. Salivary testosterone levels in preadolescent children. BMC Pediatrics: www.biomedcentral.com/1471-2431/2/5.
Peper, J.S., Dahl, R.E., 2013. The teenage brain: surging hormones-brain-behavior interactions during puberty. Curr. Dir. Psychol. Sci. 22, 134–139, http://dx.doi.org/10.1177/0963721412473755.
Peper, J.S., Hulshoff Pol, H.E., Crone, E.A., Van Honk, J., 2011. Sex steroids and brain structure in pubertal boys and girls: a mini-review of neuroimaging studies. Neuroscience 191, 28–37, http://dx.doi.org/10.1016/j.neuroscience.2011.02.014.
Peper, J.S., Koolschijn, P.C., Crone, E.A., 2013. Development of risk taking: contributions from adolescent testosterone and the orbito-frontal cortex. J. Cogn. Neurosci. 25, 2141–2150, http://dx.doi.org/10.1162/jocn_a_00445.
Peters, J., Büchel, C., 2010. Neural representations of subjective reward value. Behav. Brain Res. 213, 135–141, http://dx.doi.org/10.1016/j.bbr.2010.04.031.
Petersen, A.C., Crockett, L., Richards, M., Boxer, A., 1988. A self-report measure of pubertal status: reliability, validity, and initial norms. J. Youth Adolesc. 17, 117–133, http://dx.doi.org/10.1007/BF01537962.
Schulz, K.M., Molenda-Figueira, H.A., Sisk, C.L., 2009. Back to the future: the organizational-activational hypothesis adapted to puberty and adolescence. Horm. Behav. 55, 597–604, http://dx.doi.org/10.1016/j.yhbeh.2009.03.010.
Sescousse, G., Caldú, X., Segura, B., Dreher, J.C., 2013. Processing of primary and secondary rewards: a quantitative meta-analysis and review of human functional neuroimaging studies. Neurosci. Biobehav. Rev. 37, 681–696, http://dx.doi.org/10.1016/j.neubiorev.2013.02.002.
Shirtcliff, E.A., Granger, D.A., Schwartz, E., Curran, M.J., 2001. Use of salivary biomarkers in biobehavioral research: cotton-based sample collection methods can interfere with salivary immunoassay results. Psychoneuroendocrinology 26, 165–173.
Shirtcliff, E.A., Dahl, R.E., Pollak, S.D., 2009. Pubertal development: correspondence between hormonal and physical development. Child Dev. 80, 327–337, http://dx.doi.org/10.1111/j.1467-8624.2009.01263.x.
Shulman, E.P., Cauffman, E., 2014. Deciding in the dark: age differences in intuitive risk judgement. Dev. Psychol. 50 (1), 167–177, http://dx.doi.org/10.1037/a0032778.
Shulman, E.P., Harden, K.P., Chein, J.M., Steinberg, L., 2015. Sex differences in the developmental trajectories of impulse control and sensation-seeking from early adolescence to early adulthood. J. Youth Adolesc. 44, 1–17, http://dx.doi.org/10.1007/s10964-014-0116-9.
Somerville, L.H., Jones, R.M., Casey, B.J., 2010. A time of change: behavioral and neural correlates of adolescent sensitivity to appetitive and aversive environmental cues. Brain Cogn. 72, 124–133, http://dx.doi.org/10.1016/j.bandc.2009.07.003.
Steinberg, L., Graham, S., O'Brien, L., Woolard, J., Cauffman, E., Banich, M., 2009. Age differences in future orientation and delay discounting. Child Dev. 80, 28–44.
Steinberg, L., 2010. A dual systems model of adolescent risk-taking. Dev. Psychobiol. 52, 216–224, http://dx.doi.org/10.1002/dev.20445.
Thomas, J., Météreau, E., Déchaud, H., Pugeat, M., Dreher, J.C., 2014. Hormonal treatment increases the responseof the reward system at the menopausetransition: a counterbalanced randomized placebo-controlled fMRI study. Psychoneuroendocrinology 50, 167–180, http://dx.doi.org/10.1016/j.psyneuen.2014.08.012.
Tivis, L.J., Richardson, M.D., Peddi, E., Arjmandi, B., 2005. Saliva versus serum estradiol: implications for research studies using postmenopausal women. Prog. Neuropsychopharmacol. Biol. Psychiatry. 29 (5), 727–732.
Van Duijvenvoorde, A.C.K., Op de Macks, Z.A., Overgaauw, S., Gunther Moor, B., Dahl, R.E., Crone, E.A., 2014. A cross-sectional and longitudinal analysis of reward-related brain activation: effects of age, pubertal stage, and reward sensitivity. Brain Cogn. 89, 3–14, http://dx.doi.org/10.1016/j.bandc.2013.10.005.
Van Leijenhorst, L., Westenberg, P.M., Crone, E.A., 2008. A developmental study of risky decisions on the cake gambling task: age and gender analyses of

probability estimation and reward evaluation. Dev. Neuropsychol. 33 (2), 179–196, http://dx.doi.org/10.1080/87565640701884287.
Van Leijenhorst, L., Gunther Moor, B., Op de Macks, Z.A., Rombouts, S.A.R.B., Westenberg, P.M., Crone, E.A., 2010. Adolescent risky decision-making: neurocognitive development of reward and control regions. Neuroimage 51, 345–355, http://dx.doi.org/10.1016/j.neuroimage.2010.02.038.
Van Wingen, G.A., Ossewaarde, L., Bäckström, T., Hermans, E.J., Fernández, G., 2011. Gonadal hormone regulation of the emotion circuitry in humans. Neuroscience 191, 38–45, http://dx.doi.org/10.1016/j.neuroscience.2011.04.042.
Vermeersch, H., T'sjoen, G., Kaufman, J.M., Vincke, J., 2008a. Estradiol, testosterone, differential association and aggressive and non-aggressive risk-taking in adolescent girls. Psychoneuroendocrinology 33, 897–908, http://dx.doi.org/10.1016/j.psyneuen.2008.03.016.
Vermeersch, H., T'sjoen, G., Kaufman, J.M., Vincke, J., 2008b. The role of testosterone in aggressive and non-aggressive risk-taking in adolescent boys. Horm. Behav. 53, 463–471, http://dx.doi.org/10.1016/j.yhbeh.2007.11.021.
Wallis, J.D., 2007. Orbitofrontal cortex and its contribution to decision-making. Ann. Rev. Neurosci. 30, 31–56, http://dx.doi.org/10.1146/annurev.neuro.30.051606.094334.
Wechsler, D., 1991. Manual for the Wechsler Intelligence Scale for Children, 3rd ed. The Psychological Corporation, San Antonio (TX).