



# No Robust Association between Static Markers of Testosterone and Facets of Socio-Economic Decision Making

Laura Kaltwasser<sup>1\*</sup>, Una Mikac<sup>2</sup>, Vesna Buško<sup>2</sup> and Andrea Hildebrandt<sup>3</sup>

<sup>1</sup> Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin, Germany, <sup>2</sup> Faculty of Humanities and Social Sciences, University of Zagreb, Zagreb, Croatia, <sup>3</sup> Department of Psychology, Ernst-Moritz-Arndt-Universität Greifswald, Greifswald, Germany

Digit ratio (2D:4D) and facial width-to-height ratio (WHR) are supposedly static indicators of testosterone exposition during prenatal and pubertal lifetime, respectively. Both measures have been linked to aggressive and assertive behavior in laboratory economic games, as well as in real world scenarios. Most of the research—often limited to male subjects—considers the associations between these behaviors, traits, and hormonal markers separately for 2D:4D and WHR. Reported associations are weak and volatile. In the present study we had independent raters assess 2D:4D and WHR in a sample of  $N = 175$  participants who played the ultimatum game (UG). Respondent behavior in UG captures the tendency to reject unfair offers (negative reciprocity). If unfair UG offers are seen as provocations, then individuals with stronger testosterone exposition may be more prone to reject such offers. Economists argue that negative reciprocity reflects altruistic punishment, since the rejecting individual is sacrificing own resources. However, recent studies suggest that self-interest, in terms of status defense plays a substantial role in decisions to reject unfair offers. We also assessed social preferences by social value orientation and assertiveness via self-report. By applying structural equation modeling we estimated the latent level association of 2D:4D and WHR with negative reciprocity, assertiveness and prosociality in both sexes. Results revealed no robust association between any of the trait measures and hormonal markers. The measures of 2D:4D and WHR were not related with each other. Multigroup models based on sex suggested invariance of factor loadings allowing to compare hormone-behavior relationships of females and males. Only when collapsing across sex greater WHR was weakly associated with assertiveness, suggesting that individuals with wider faces tend to express greater status defense. Only the right hand 2D:4D was weakly associated with prosocial behavior, indicating that individuals with lower prenatal testosterone exposure are more cooperative. Rejection behavior in UG was not related with 2D:4D nor WHR in any of the models. There were also no curvilinear associations between 2D:4D and prosociality as theorized in the literature. Our results suggest that previous studies over-estimated the role of static markers of testosterone in accounting for aggression and competition behavior in males.

**Keywords:** testosterone, 2D:4D, facial width-to-height ratio, economic decision making, social preferences, assertiveness

## OPEN ACCESS

### Edited by:

Pablo Brañas-Garza,  
Middlesex University, United Kingdom

### Reviewed by:

Giovanni Benedetto Ponti,  
University of Alicante, Spain  
Teresa García-Muñoz,  
University of Granada, Spain

### \*Correspondence:

Laura Kaltwasser  
laura.kaltwasser@hu-berlin.de

**Received:** 31 August 2017

**Accepted:** 11 December 2017

**Published:** 20 December 2017

### Citation:

Kaltwasser L, Mikac U, Buško V and  
Hildebrandt A (2017) No Robust  
Association between Static Markers of  
Testosterone and Facets of  
Socio-Economic Decision Making.  
*Front. Behav. Neurosci.* 11:250.  
doi: 10.3389/fnbeh.2017.00250

## INTRODUCTION

### The Impact of Testosterone

The steroid hormone testosterone, produced in the male testes, and to a lesser extent in female ovaries, circulates the human brain throughout life and it is assumed to impact behavior and its development. Relationships between hormonal activity and behavior are complex, consisting of both endocrine effects on behavior and, vice versa, behavioral effects onto endocrine function. On the one hand, endocrines have been shown to affect attachment and sex (Carter, 1998; Insel and Young, 2001), aggression (Koolhaas et al., 1990; Dabbs et al., 1995) and social status (Mazur and Booth, 1998; Josephs et al., 2003). On the other hand, sexual behavior, competition for status or fighting can alter endocrine levels (Mazur and Lamb, 1980; Elias, 1981; Carmichael et al., 1994).

Previous research in primates and humans suggests that high levels of testosterone promote behaviors intended to enhance one's status over other individuals and to climb up the social hierarchy. According to the biosocial model of status (Mazur, 1985), status defense can overtake a form of dominance or aggression. An individual is *dominant* if its intent is to gain or defend high status over another member of its species. An *aggressive* individual will have the intent to inflict physical and psychological injury on a conspecific. Sometimes dominant behavior takes the form of aggressive or antisocial behavior such as violence or law breaking. However, the distinction between dominance and aggression is particularly important in humans, where dominance is often asserted without any intent to cause injury. For instance, Ehrenkranz et al. (1974) showed that both, aggressive prisoners and dominant, but non-aggressive prisoners had a significantly higher level of plasma testosterone as compared with non-aggressive and low dominance prisoners.

### Measuring Antisocial Behavior in the Lab

In the laboratory, socio-economic games are widely used to study non-aggressive anti-social behavior. Socio-economic games are social decision-making trials simulating real-world strategic interactions (Camerer, 2003). Involved individuals make monetary choices based on an interdependent pay-off matrix. The two bargaining partners are given a set of rules and they face limited information since they are confronted with uncertainty about the other's intentions (see below for details). Importantly, the individuals' choices alter not only their own outcome, but also the outcome of the other, allowing the researcher to study game-theoretical constructs such as fairness, reputation building and status defense. While prosocial behavior or altruism are often the target dependent variables of investigation, recent attempts have been made to use socio-economic games for measuring anti-social or assertive behavior as in the tendencies to punish and retaliate (Falk et al., 2005; Nikiforakis, 2008; Yamagishi et al., 2012). The public goods game is a stylized model of situations that require cooperation to obtain socially beneficial outcomes in the presence of incentives for free riders. By using this game, Herrmann et al. (2008) showed that antisocial punishment exists in different participant pools around the world. The punishment of unfair behavior such as free riding

may arise from negative emotions that are evoked through feeling exploited. Accordingly, emotions such as anger or moral disgust make individuals disregard the immediate consequences of their behavior, allowing them to preserve a reputation over time as someone who is reliably committed to this behavior (Yamagishi et al., 2009).

The ultimatum game (UG) allows to study the tendency to punish unfair behavior (negative reciprocity) in the responder. The UG (Güth et al., 1982) is a two stage socio-economic game in which a proposer is given a monetary endowment, which he can split and share with a responder. Only if the responder accepts, both players receive their share according to the proposer's split. Thus, the proposer has the power to postulate an ultimatum to the responder. Economists argue that negative reciprocity reflects altruistic punishment (Fehr and Gächter, 2002), since the rejecting individual is sacrificing own resources. However, recent studies suggest that self-interest, in terms of status defense, plays a substantial role in decisions to reject unfair offers (Yamagishi et al., 2012; Kaltwasser et al., 2016). According to the above mentioned biosocial model of status (Mazur, 1985), individuals with high levels of testosterone should be more likely to retaliate, e.g., have a greater desire to harm those who committed unfair acts. While most studies focused on the responder behavior in UG in order to quantify negative reciprocity as the tendency to reject unfair offer, for each participant, we obtained data in both roles of the UG—as proposer and responder. This “dual” version of the UG is valuable not only in order to obtain preferences for fear of punishment (strategic behavior) in the proposer data, but also in order to study whether the assigned role affects cooperation behavior in general. For example Brañas-Garza et al. (2006) investigated behavior in a dual UG with illiterate gypsies in Vallecas, Madrid, acting as both proposer and responder. In this set-up, the responder's acceptance of a zero offer was not a rare case, but the modal value, and 97% of the subjects proposed an equal split in the role of the proposer.

### Ratio of Second-Finger-Length to Fourth-Finger-Length (2D:4D)

The scientific study of the impact of sex steroids on brain and behavior has been separated into activational and organizational effects. Activational effects are temporal and occur throughout life depending on current hormone levels. Organizational effects are permanent and mainly occur in two phases: early in development when most neural structures are formed and during adolescence (Phoenix et al., 1959). However, empirical evidence speaks against a rigid dichotomy between both classes of effects (Arnold and Breedlove, 1985). Studies provided by the animal model suggest that organizational hormones may prime the brain by changing its responsivity to hormones that are present later in life (Clark and Galef, 1998).

There is some evidence for prenatal organizational effects of sex steroids (for a review see Auyeung et al., 2013). For example, twin studies have been conducted following the assumption that females from pairs of opposite-sex twins are exposed to higher levels of prenatal testosterone compared to same-sex twins. While free circulating testosterone levels were not

yet systematically related to different personality traits, a sex difference in aggression proneness has been observed. Opposite-sex girls of the twin dyad studied show a more masculine pattern of aggression proneness than same-sex girls (Cohen-Bendahan et al., 2005a).

Furthermore, females with Congenital Adrenal Hyperplasia (CAH), a genetic disorder which causes excessive androgen levels during early development, show a masculinization of their behaviors, for example in playing (Hines, 2003) and spatial navigation (Hampson et al., 1998), as well as with respect to cognitive abilities (Resnick et al., 1986) and personality traits (Berenbaum and Resnick, 1997; Mathews et al., 2009). The studies with CAH participants suggest that differences between males and females are due to androgens as testosterone, but they are less informative about the role of androgens in producing typical variations (Cohen-Bendahan et al., 2005b).

Similar to persons with CAH, individuals with androgen insensitivity syndrome, who have androgen levels typical for males and XY generic structure but do not react to androgens due to dysfunction of androgen receptors, show a higher ratio of second-finger-length to fourth-finger-length (Berenbaum et al., 2009; van Hemmen et al., 2017; 2D:4D). Therefore, 2D:4D with smaller values is considered to mark stronger prenatal testosterone exposure (Manning et al., 1998) and it is taken to be a static indicator of prenatal testosterone in normally developing humans.

This interpretation is partly endorsed by similar timing of both, the prenatal digit development and the highest prenatal testosterone levels (Vaillancourt et al., 2012), and the relation of sex hormones and bone growth established in research on mammals (Kondo et al., 1997). One of the most cited papers providing evidence for the usability of 2D:4D as an indicator of organizational effects of sex steroids reported a negative correlation of right-hand 2D:4D with the ratio of testosterone and estrogen in the amniotic fluid mid gestation (Lutchmaya et al., 2004). However, this finding should be interpreted with caution. The reason is first the used methodology (Hollier et al., 2015; Yeung and Tse, 2017) and second, the fact that the relation of sex-hormone levels in amniotic fluid with levels of sex-hormone in the fetus blood are not well-established (Cohen-Bendahan et al., 2005b). When sex steroid levels were measured in umbilical cord, no systematic relation to 2D:4D could be established (Hollier et al., 2015; Mitsui et al., 2016), which might also result from differing levels of steroid hormones during prenatal development.

2D:4D shows a moderate but stable sex difference (Hönekopp and Watson, 2010) that develops early during fetal development and individual scores remain stable across development. Sex differences in 2D:4D are noticeable already at the end of the first trimester of prenatal development (Malas et al., 2006), but become relatively stable after 5 years of age and do not change during puberty. There are three stages during development in boys when testosterone reaches levels similar to those in adult men: (a) during 10th to 18th week of prenatal development, (b) 1–2 weeks after birth, and (c) from 8 weeks until 4–6 months of age (McIntyre, 2006). Thus, based on these findings, 2D:4D might be considered an indicator of perinatal organizational effects.

Interestingly, circulating steroid levels are unrelated to 2D:4D, suggesting that relationships between 2D:4D and target variables reflect effects of prenatal testosterone exposition (Hönekopp et al., 2007). Notwithstanding, evidence regarding the association between 2D:4D and trait variables, such as personality or facets of socio-economic decision-making is mixed.

A meta-analysis comprising 64 samples with  $N = 6,617$  females and males (Hönekopp and Watson, 2011) found no evidence for 2D:4D predicting aggression at different levels of behavior, ranging from physical and verbal aggression to anonymous contacts. The study only revealed a small negative association ( $r = -0.06$ ) between 2D:4D and aggression in males, which was absent in females. No evidence was found that either hand would predict aggression better than the other—a finding that is corroborated with other target variables such as athletic prowess (Hönekopp and Schuster, 2010). Apicella et al. (2008) showed in a sample of  $N = 98$  men that risk-taking in an investment game correlates positively with salivary testosterone levels ( $r = 0.29$ ) and facial masculinity ( $r = 0.27$ ), with the latter being a proxy for pubertal hormone exposure (see section on WHR below). 2D:4D on the other hand did not correlate with risk preferences.

Another personality trait that has been studied in conjunction with testosterone is assertiveness, the quality of being self-assured and confident. Depending on the scale used to measure assertiveness, this trait is correlated with aggression or status-imposing behavior (Buss and Perry, 1992; Yamagishi et al., 2012). While Hampson et al. (2007) found lower 2D:4D ratios to be associated with increased aggressiveness and sensation seeking, no such relationship was present for assertiveness. The absence of a relationship between 2D:4D (for both sexes and hands) and assertiveness was further confirmed by a study with a larger sample of 491 men and 627 women (Voracek, 2009).

Studies relating 2D:4D to socio-economic bargaining suggest that the broader picture of the relationship between static markers of the “status-hormone” with prosocial vs. antisocial or status-enhancing behavior is complex. Recent evidence suggests a non-monotonic, i.e., u-shaped, impact of prenatal testosterone exposure on altruism in the sense that individuals with both, high and low digit ratios give less than individuals with intermediate digit ratios (Brañas-Garza et al., 2013; Galizzi and Nieboer, 2015). Moreover, a study administering testosterone to women showed a substantial increase in *fair* bargaining behavior in the UG (Eisenegger et al., 2010). Interestingly, participants who believed that they received testosterone (regardless of whether they actually received it) showed more unfair behavior than those who were treated with placebo—providing evidence for the power of folk wisdom on participant’s expectations about testosterone as a status or even aggression inducing hormone. A later publication commenting the latter study suggests that static marker of prenatal testosterone may interact with administered testosterone, in that social cooperation increases after testosterone administration but only in participants with low levels of prenatal testosterone measured by right hand’s 2D:4D (van Honk et al., 2012).

## Facial Width-to-Height Ratio (WHR)

Another characteristic that has been related to testosterone is the WHR, that is, the face width divided by upper-face height. Research on this topic stemmed mostly from the observation that WHR is a sexually dimorphic face characteristic (Weston et al., 2007; Carré and McCormick, 2008), although a meta-analysis lead to equivocal conclusions regarding the existence of this dimorphism (Geniole et al., 2015; Kramer, 2017). Taking the finding into account that WHR dimorphism develops during adolescence (Weston et al., 2007), and because boys' craniofacial growth has shown to be enhanced by testosterone administration (Verdonck et al., 1999), WHR was suggested as a proxy for organizational hormonal effects in adolescence (Carré and McCormick, 2008). However, research on how changes in testosterone levels during adolescence are related with WHR gave equivocal results (Hodges-Simeon et al., 2016; Welker et al., 2016). Similar to 2D:4D, WHR showed no relationship to circulating testosterone levels in adulthood (Bird et al., 2016). As expected based on the idea that WHR is an organizational hormonal effects' proxy specifically of adolescence, adult WHR showed no relation to umbilical testosterone levels (Whitehouse et al., 2015). In the same study, WHR also showed no relationship with 2D:4D (ranging between  $r(N = 75) = -0.22$ , n.s., for female left hand, to  $r(N = 82) = 0.11$ , n.s., for male right hand). To our knowledge, this is the only research inspecting the relationship of 2D:4D and WHR.

Two meta-analyses were recently published on the relation of WHR to aggression (Haselhuhn et al., 2015) and threatening and dominant behaviors (Geniole et al., 2015). The first study included only men and a narrower range of behavior and published papers. These studies concluded a weak, albeit significant relation of WHR and status-enhancing behavior in men, with the effect size ranging between  $r = 0.11$  and  $0.16$ ,  $p \leq 0.01$ . For women, the effect was significant only in case of dominant behavior. Different related psychological constructs have been proposed as mediators between WHR and aggressive behavior, such as fearless dominance (Geniole et al., 2014; Anderl et al., 2016) and psychological sense of power (Haselhuhn and Wong, 2011).

The socio-economic choices mostly fit into this pattern, with men having higher WHR exploiting others' trust more in a trust game (Stirrat and Perrett, 2010) and cheating more in a lottery (Haselhuhn and Wong, 2011; Geniole et al., 2014). However, Stirrat and Perrett (2012) demonstrated that WHR is not necessarily related with antisocial behavior. In their experiment, WHR predicted higher cooperation, leading to the player's individual loss, when it benefited their group at the expense of an out-group. This might be a strategy to enhance one's status in the in-group, and is in accordance with the postulated relation of testosterone and status. Moreover, it reflects behavior in line with the male warrior hypothesis which suggests that men have a stronger tendency to treat in-group members benevolently and out-group members malevolently compared to women (Van Vugt et al., 2007).

## Current Study

In the light of the above reviewed studies, evidence on the relationship of testosterone with facets of socio-economic decision-making such as status defense are provided by two sources: First, there is research on acute effects of testosterone and human social decisions. That research includes studies administering testosterone and investigating its consequences on social decisions by using laboratory paradigms (for a review see Bos et al., 2012). Since testosterone not only affects behavior but it also responds to it, it can also serve as the dependent variable in experimental procedures where social interaction parameters, such as status, are modulated and testosterone is measured and an outcome (Carney et al., 2010). Second, stable trait-like dispositions with regard to testosterone can be the matter of study—including static markers of testosterone, which are consequences of developmental differences in testosterone exposition. In the current study we investigate the association of such static markers of testosterone with facets of socio-economic decision making in a typically developing population of young adults. As far as we know, this is the first study to relate WHR and 2D:4D to facets of socio-economic decision making within one statistical model, therefore allowing to estimate the shared variance of different markers of exposure to testosterone during early stages of development.

## Hypotheses

Based on the reviewed literature on testosterone and facets of socio-economic decision making, we expected participants with lower 2D:4D to show increased assertive and less prosocial behavior. If unfair UG offers are seen as provocations, then individuals with stronger prenatal testosterone exposition may be more prone to reject such offers.

Regarding WHR we hypothesized that individuals with wider faces show more masculinized behavior—reflected in more assertive and less prosocial behavior.

Since the evidence for gender effects in the associations between both static markers of testosterone and the target variables is rather inconsistent, we modeled the relationships separately and together for both genders.

More recent literature discussed above suggest an inverted U-shaped relationship between prenatal testosterone exposition and prosocial behavior. We thus tested whether individuals with lower vs. higher 2D:4D show less prosocial behavior as compared with persons with intermediate 2D:4D. Following the same argumentation, a U-shaped relationship may be predicted for rejections in the Ultimatum Game indicating negative reciprocity due to provocative behavior.

## METHODS

### Participants

The reported data stems from a sample of 84 females and 91 males ( $N = 175$ ) who took part in a larger study investigating socio-emotional processes and abilities. Participants gave consent of their pictures being used for further investigations (Kaltwasser et al., 2016). The mean age of this sample was 27.62 ( $SD = 5.4$ ).



Participants were recruited through the university's participant pool and public announcement in newspapers as well as on local websites. The study conformed to the guidelines of the ethics committee of the Department of Psychology, Humboldt-Universität zu Berlin. All experiments were in accordance with the Declaration of Helsinki. The protocol was approved under the approval number 2013-17. All participants provided written consent before starting the experimental procedures. They received a compensation of 8 € per hour and were informed that they could win further money during the UG, depending on their choices. Each participant received an additional amount of 5 € as payout from UG. Seventy-seven percent of the participants had completed German high school of which 35% had a university degree. Forty-six percent of the sample were still studying while the rest was working full-time or unemployed (16%).

## Procedure

The experiment consisted of two sessions. During the behavioral session that lasted 2 h, participants completed computerized self-report measures of personality and fairness preferences, as well as several ability measures of face and object cognition, which are not analyzed for the scope of this paper. All questionnaires were programmed in Inquisit software (Inquisit 4.0.0.1, 2012; Millisecond Software, Seattle, WA), and responses were given via computer mouse. In the laboratory session, taking place 1–2 weeks after the behavioral session, participants were photographed and 2D:4D measurements were acquired by means of a photocopy machine. Additionally, they played the UG as proposer and responder. During data acquisition of the UG, in the responder condition we also measured the participants' EEG. Electrophysiological measures are however not the scope of this paper.

## Assertiveness

We applied the assertiveness scale of the German Inventory of Personality Styles and Disorders (Persönlichkeits-Stil-und-Störungs-Inventar) (Kuhl and Kazén, 2009). The scale consists of 10 items ( $\alpha = 0.82$ ) measuring the tendency to impose oneself onto others and the tendency to defend ones' status. This tendency may extend to ruthless and antisocial behavior. A sample item is "If others want something which I need, I normally prevail." Responses are given on four-point Likert scales (disagree strongly, disagree somewhat, agree somewhat and agree strongly). For the analyses, we formed three parcels of three to four items each based on the underlying motivation for assertiveness as reflected in the content of the item (aggressive, egoistic, or assertive behavior).

## Social Value Orientation (SVO)

The magnitude of concern people have for others can be measured by a six-item questionnaire ( $\alpha = 0.89$ ), where participants indicate how they would share resources with an anonymous stranger (Murphy et al., 2011). Each item is a resource allocation over a continuum of joint payoffs. For example, the participant has to choose a value  $x_{\text{self}}$  between 50 and 100, knowing that the anonymous partner will get  $x_{\text{other}} = 150 - x_{\text{self}}$ . According to the pay-off structure, the participant

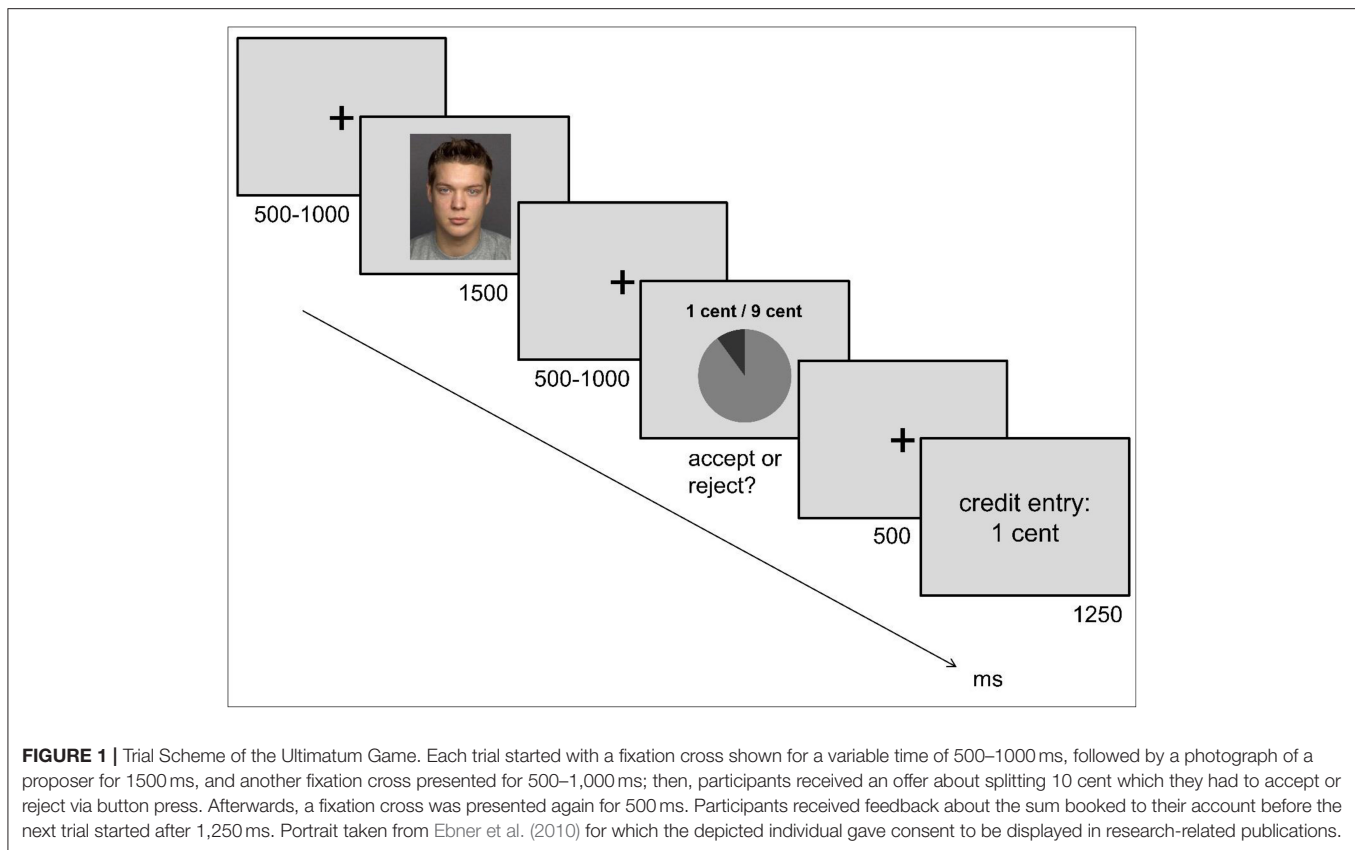
is assigned a continuous value of social orientation, which can be categorized to competitive, individualistic, prosocial and altruistic. Previous research indicates that SVO is a valid predictor of the cooperative tendency in social dilemmas (Bogaert et al., 2008; Balliet et al., 2009). In the analyses, we formed three parcels out of two SVO items each to serve as indicators for the latent factor of prosociality next to the indicator of total offers in the responder part of the Ultimatum Game (see next section).

## Ultimatum Game (UG)

Upon arrival to the laboratory, participants were introduced to the rules of the UG, informing them that they would play with other participants, which would require having their picture taken. Moreover, participants were asked to play the proposer in the UG, making 12 offers on a query sheet. In each offer, the participant could divide 10 cents into two shares: one for her/him and one for the other player. There were three predefined proposals: 9/1 (nine for the proposer, one for the responder), 7/3 and 5/5. Participants were informed that these offers would later be presented to other players together with their picture. They were instructed that the other player could then decide whether to accept or reject each offer. Participants were told that they would receive the corresponding amount of money if the offer was accepted by the responder. After providing their offers on a sheet, participants played the computerized version of the UG in the role of the responder while EEG was recorded (288 trials). They were explained that they would receive monetary offers made by six previous participants, but the actual offers came from six pseudo-proposers (50% females). Due to the EEG methodology whose data is published elsewhere (Kaltwasser et al., 2016) we required an experimental protocol of the UG which allows for a specific offer distribution and high signal-to-noise ratio, e.g., many trial repetitions. Hence, it was necessary to deceive the participants in the origin of the proposals they saw. These proposers were represented by portraits taken from a standardized stimulus set, the FACES database (Ebner et al., 2010). We included portraits of the proposer prior to the offers in order to create a social bargaining situation, since previous work suggests that social cues affect cooperation behavior (Haley and Fessler, 2005). The responder version of the UG comprised trials with fair (5/5), slightly unfair (7/3), or highly unfair (9/1) offers which were paired with the same proposer identities, so that the participant could learn over the course of the experiment, that two proposers always made fair offers, two always made unfair offers, and two made mixed offers. The rejection rates of unfair offers for each of the three experimental blocks served as indicators for the latent factor of negative reciprocity. A typical trial of the responder version of the UG with an unfair offer is depicted in **Figure 1**.

## Facial Photographs

Full frontal facial photographs were taken of all participants without glasses or head wear with a Panasonic HDC-SD707 on a tripod in front of a gray background. The distance between the camera and the subject was kept consistent with 1.5 m. The portraits were preprocessed and cut into rectangular facial images



of the same size (e.g., removing the presence of the neck and the remaining space above head) using Photoshop. Pictures of the participants who gave consent of their pictures being used in further studies were used for the analyses reported below. Eighty-six percent of the sample of Kaltwasser et al. (2016) agreed and their data is reported here.

### WHR Measurement

Two raters independently measured facial width and height on the full frontal photographs using ImageJ 1.48 software (Schneider et al., 2012). Width was defined as the distance between the points on the picture where ears and face meet. Height was the distance from the point where the brow touches the root of the nose to the highest point of the lips (Weston et al., 2007; Carré and McCormick, 2008).

### 2D:4D Measurement

The ratio of second-finger-length to fourth-finger-length was acquired for the left and right hand independently. A see-through foil with a printed standard ruler was placed on the scanner for each participant (in accordance with Kemper and Schwerdtfeger, 2009). Before scanning, the proximal crease was marked with a water-soluble marker as to ease the determination of ventral proximal crease (in accordance with Voracek et al., 2007). Participants were instructed to press lightly with both hands at the same time. The experimenter verified that participants followed the instruction and checked that their hand position was

in accordance to the guidelines provided by Mayhew et al. (2007). As suggested by Hiraishi et al. (2012), white cloth was put on the hands by the experimenter in order to achieve more contrast and an easy determination of points on the scanned pictures. Scans were made using HP Scanjet 7650 and the resolution was kept standard. Two raters with previous experience with 2D:4D measurement independently measured digit lengths using specialized open source software AutoMetric (DeBruine, 2004).

### Data Analysis

Latent factors of 2D:4D, WHR, prosociality, negative reciprocity and assertiveness, along with their mutual relationships were estimated in measurement and structural models using structural equation modeling conducted with the lavaan package (Rosseel, 2012) in the R software for statistical computing (R Core Team, 2017). For testing specific relationships due to sex between those latent variables, multi-group structural equation models (e.g., Little et al., 2007) were fitted using the same software. Structural equation models (SEM) can be used to test theories on linear relationships between multiple psychological entities by explicitly accounting for measurement error and the specificity of the measurement method (Bollen, 1989). SEMs estimate latent variables based on their measured, observable indicators. The basic idea behind latent variables is that all psychological measurements are error prone and contain measurement method specificity. For example, the measured values of 2D:4D from hand image scans by two different raters will not completely

overlap. Using the multiple rating values provided by different raters as indicators of a latent variable to be estimated on the basis the indicators' covariances allows taking rater specific measurement error into account. Thus, latent variables are quantifying the true score variance of 2D:4D, WHR and the traits to be studied in the present work. The quality of SEMs can be assessed by multiple formal statistical tests and fit indices: Chi-square statistics, the root mean square error of approximation (RMSEA, should be lower than 0.08), standardized root mean square residual (should be SRMR < 0.08) and the Comparative Fit Index (should be CFI > 0.95; see e.g., Bollen, 1989 for details).

For testing the non-linear relationship between 2D:4D with prosociality and with negative reciprocity, we used an exploratory method called Local Structural Equation Modeling (LSEM; Hildebrandt et al., 2016). This method allows estimating an SEM along the values of a moderator. Because we are interested to explore curvilinear relationships between 2D:4D and prosociality, we aim to estimate the measurement models of prosociality and negative reciprocity along continuously sampled values of 2D:4D within its possible range of measured values. The LSEM modeling approach allows to investigate whether the mean of the latent prosociality and negative reciprocity factors are different across varying values of 2D:4D. Based on LSEM estimates, the latent factor means of prosociality and negative reciprocity can be plotted along the values of 2D:4D. Thus, for the present research the range of the 2D:4D left vs. right hand variables was taken as a continuous scale along which the latent factor mean of prosociality and negative reciprocity may vary, following an inverted U-shaped or U-shaped curve, respectively (see hypotheses on non-linear relations above). We thus provide parameter plots estimated by LSEM to illustrate how average prosociality and negative reciprocity varies along the measured values of 2D:4D. In summary, these gradients visualize curvilinear relations between prosociality and negative reciprocity, respectively, with the 2D:4D measurements (see Hildebrandt et al., 2016 for details on LSEM). LSEM was conducted with the *sirt* package in R (Robitzsch, 2015).

## RESULTS

To test our hypotheses, we run a series of measurement and structural models including latent variables representing organizational effects of hormones measured by estimations of (1) 2D:4D (left and right hand) provided by two different raters and (2) of WHR estimated by two raters as well. Furthermore, (3) prosociality, (4) negative reciprocity, and (5) assertiveness was modeled based on multiple measured behavioral indicators. Thus, in a first step we estimated a measurement model of 2D:4D and WHR, including three latent factors because 2D:4D has been measured on the right as well as on the left hand side by two different raters. Consequently, there are two indicators (provided by two different raters) for each of the three latent variables representing prenatal and pubertal organizational effects of hormones. In a second step we aimed to establish a measurement model for the behavioral indicators of prosociality, negative reciprocity and assertiveness. We estimated the latent factor of

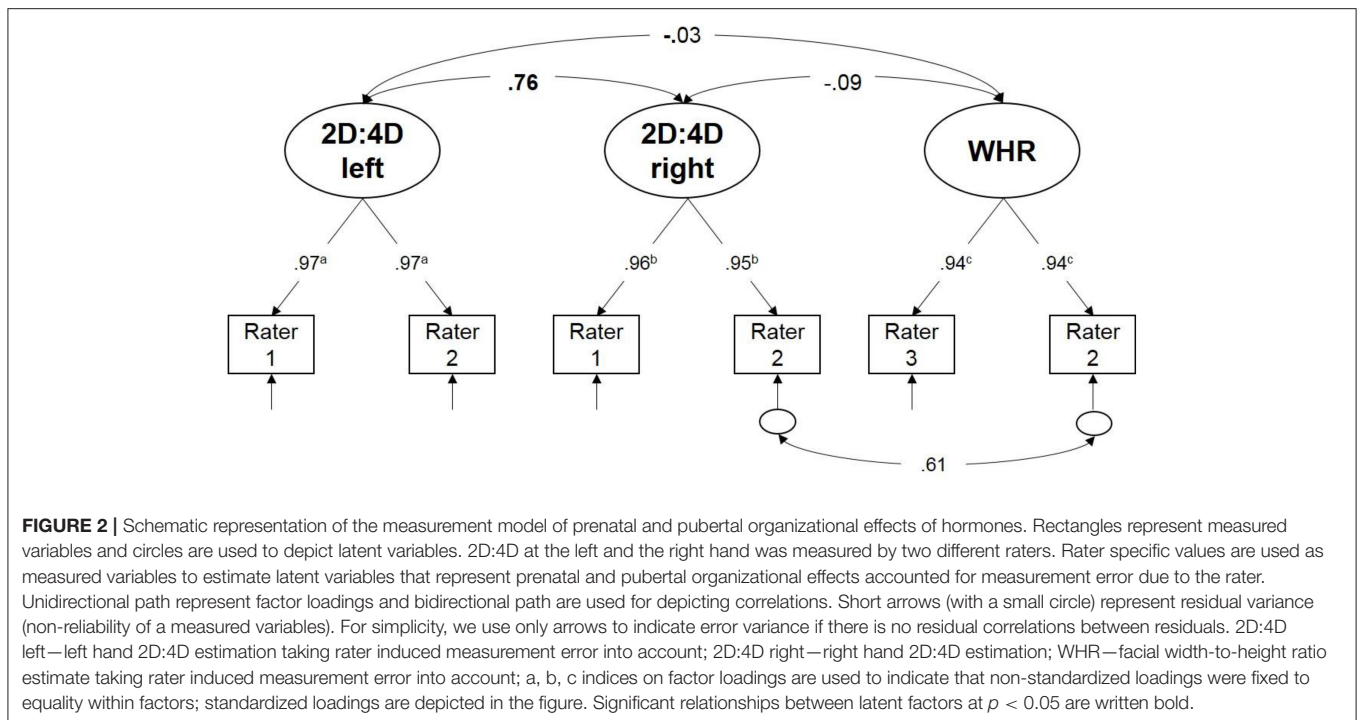
prosociality by means of three parcels of SVO responses (see above) and a further indicator of total offers in the proposer part of the Ultimatum Game. Negative reciprocity as a latent variable is measured by rejection rates of unfair offers in three independent experimental blocks and assertiveness is reflected by three indicators of different underlying motivations for assertiveness (aggressive, egoistic, or assertive behavior; see also task descriptions in the method section above). Third, the two measurement models were related to each other in a structural equation model of hormone-behavior relations. Fourth, the structural model was simultaneously estimated for males and females using the well-established technique of multiple group modeling. As customary in multiple group analyses (see Little et al., 2007), the sex specificity of hormone-behavior relations was tested after establishing measurement invariance across sex. This is to ensure that the factors can be interpreted as isomorphic (equivalent) for males as compared with females. If indicators are measuring the latent variables with the same precision—thus, factor loadings would be equal for males and females—we could conclude that the association between hormones and behavior are statistically and substantially comparable across sex because the meaning of the factors are equivalent. Last, we tested a curvilinear association between hormones and prosocial behavior vs. negative reciprocity in the whole sample to investigate whether their relationship is rather inverted U-shaped vs. U-shaped in case of negative reciprocity, and not linear (see discussion above and the data analyses section for details on the LSEM procedure).

### Measurement Model of 2D:4D and WHR

2D:4D at the left and right hand and WHR were estimated by three different raters (see **Figure 2**). These ratings for each person included in the final sample were used as indicators for measuring three latent factors—2D:4D left, 2D:4D right and WHR—to be established in the measurement model of organizational effects of hormones. The model depicted in **Figure 2** fitted the data very well:  $\chi^2(8) = 3.79$ ,  $p = 0.88$ , CFI = 1, RMSEA = 0.00, SRMR = 0.02. Because only two indicators were available for each factor, their non-standardized loadings were fixed to equality within each factor (note that standardized loadings are depicted in **Figure 2**). The model fit was excellent in spite of equality constraints on the factor loadings. High standardized factor loadings depicted in **Figure 2** suggested that 2D:4D and WHR measurements were highly consistent across raters based on the above described measurement procedure. Latent factor correlations revealed that 2D:4D is not related with WHR, whereas left and right hand 2D:4D are substantially ( $r = 0.76$ ), but not perfectly correlated. Having the same rater across different indicators led to a correlated error between Rater 2 of right 2D:4D and WHR (see **Figure 2**) which needs to be included in order to achieve good model fit.

### Measurement Model of Prosociality, Negative Reciprocity and Assertiveness

In the second measurement model displayed in **Figure 3**, behavioral indicators described in the method section were



used to estimate three latent factors—prosociality, negative reciprocity and assertiveness. The measurement model, including one theoretically expected residual covariance between indicators of SVO due to similar pay-off structures, had a very good fit to the data:  $\chi^2(31) = 34.63$ ,  $p = 0.30$ , CFI = 0.99, RMSEA = 0.03, SRMR = 0.04. Standardized factor loadings (see **Figure 3**) were all significantly different from zero and were substantial in their magnitude. Prosociality showed a small negative association with negative reciprocity and assertiveness, whereas the relation between assertiveness and negative reciprocity did not reach statistical significance.

### Structural Model of Organizational Hormonal Effects and Behavior

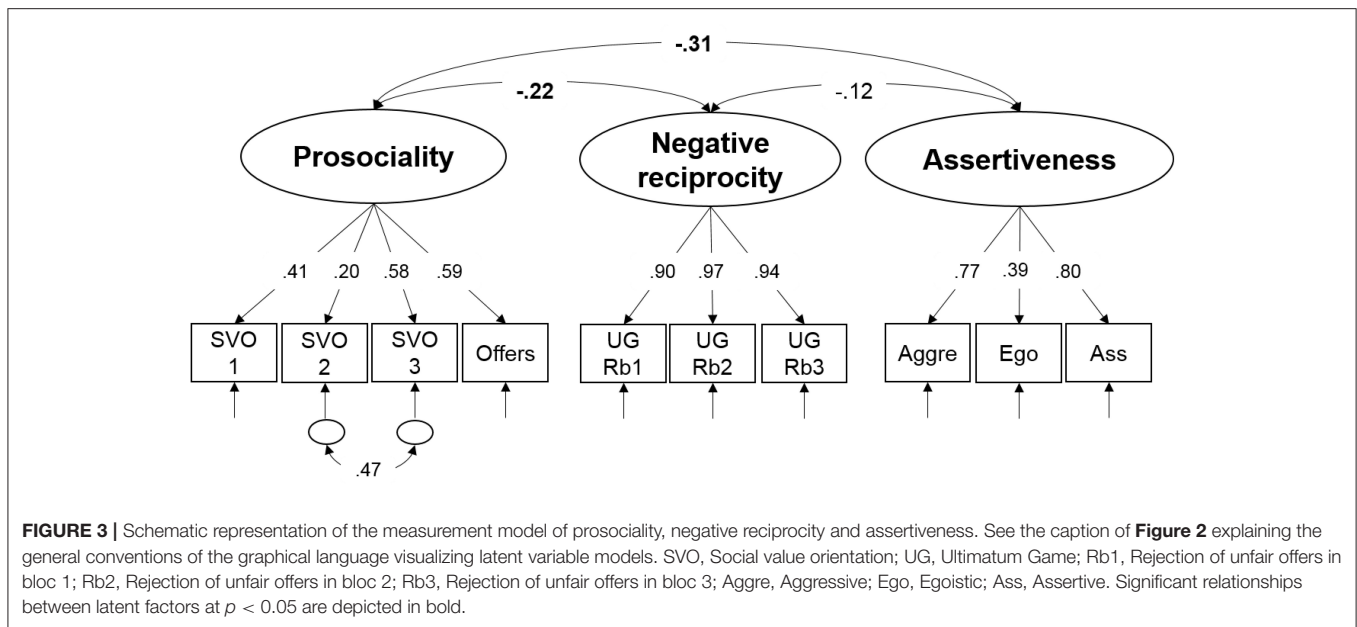
To estimate the relationship between prenatal and pubertal organizational effects of hormones and prosociality, negative reciprocity and assertiveness, the two measurement models established above were related to each other in a full structural equation model. The measurement models were completely equivalent to those described above. All bivariate relationships between latent factors were estimated. The structural model also had an excellent fit to the data:  $\chi^2(90) = 90.23$ ,  $p = 0.47$ , CFI = 1, RMSEA = 0.00, SRMR = 0.04. The correlations between 2D:4D, WHR and trait factors are provided in **Figure 4**. There was no association between organizational effects of prenatal and pubertal hormones and traits, except for a small positive association between WHR and assertiveness, suggesting that persons with higher facial width-to-height ratio are more assertive. A further positive association prevailed between right hand 2D:4D and prosociality, suggesting that persons with higher 2D:4D are somewhat more prone to prosocial decisions.

### Sex Differences in Organizational Effects of Prenatal and Pubertal Hormones and Behavior

As discussed above, in the light of the literature, sex differences are expected regarding hormone-behavior relationships depicted in **Figure 4**. As a prerequisite of comparing association in a structural equation models across groups, measurement invariance needs to be tested, because the test assures the meaning of the latent variables to be equivalent across groups. Model parameters at the level of latent variables are only comparable across groups if measurement invariance can be confirmed (see Little et al., 2007).

Measurement invariance implies a stepwise test of increasingly restricted models. In a first step a model with freely estimated parameters will be inferentially compared with a model in which factor loadings are fixed to equality across sex groups. The second step includes further cross-group equivalence restriction on intercepts. The results of these invariance tests are displayed in **Table 1**. Whereas, factor loadings are invariant for females and males, the intercepts seem to be biased for sex. Such an outcome is indeed comprehensible bearing in mind the existing sex differences in the variables quantifying hormonal influences and the high inter-rater consistency. We were however not interested to compare factor means in the multigroup model, but to investigate whether the hormone-behavior relationship differed for females and males. For group-comparison regarding relationships between latent variables invariance of factor loadings in a necessary and sufficient condition. Since factor loading invariance across sex was demonstrated for the present data (see **Table 1**), comparisons of hormone-behavior relations are possible and sound. However,





multiple group modeling of the structural model depicted in **Figure 4** revealed no statistically substantial hormone-behavior associations neither in the group of females, nor males. The magnitudes of the relations were comparable across females and males and somewhat lower as compared with those displayed in **Figure 4**.

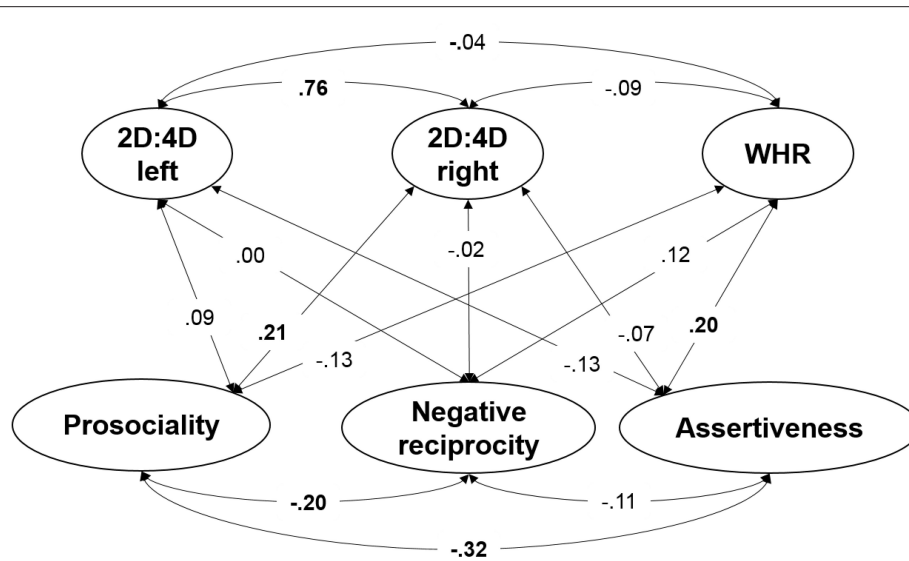
### Curvilinear Relations between Organizational Effects of Prenatal and Pubertal Hormones, Prosocial Behavior, and Related Traits

Local Structural Equation Models (LSEM, see above) were estimated for negative reciprocity and prosociality along the left vs. right hand 2D:4D measures in four separately fitted one factorial models. 2D:4D left vs. right were considered measured moderator variables for LSEM, with their values resulting by averaging the two available ratings from two different raters. LSEM models were run for the whole sample including females and males. The parameter of interest is the factor mean for negative reciprocity and prosociality as a gradient across the values of 2D:4D for left vs. right hand. Thus, latent factors were scaled by a reference indicator concerning the covariance as well as the mean structure in order to obtain estimates of latent factor means (see for example Little et al., 2007 for details regarding scaling of latent factors). **Figure 5** displays the gradients for the latent mean of the negative reciprocity factor (**Figure 5A**—left hand 2D:4D and negative reciprocity; **Figure 5B**—right hand 2D:4D and negative reciprocity) and the latent mean of the prosociality factor (**Figure 5C**—left hand 2D:4D and prosociality; **Figure 5D**—right hand 2D:4D and prosociality) along with confidence intervals. The gradients suggest an inverted U-shaped relation only for prosociality and left hand 2D:4D. Because the non-linear association is only visible at the left hand, we must treat this finding with caution.

## DISCUSSION

The aim of this study was to investigate the relationship of static markers of testosterone with facets of socio-economic decision-making. Based on the biosocial model of status (Mazur, 1985) we hypothesized static markers indicating higher levels of testosterone to be associated with status defending or assertive behavior. In order to test this hypothesis we had independent raters assess 2D:4D and WHR in a sample of  $N = 175$  participants who played the ultimatum game. Respondent behavior in UG captures the tendency to reject unfair offers (negative reciprocity). If unfair UG offers are seen as provocations, then individuals with stronger testosterone exposure may be more prone to reject such offers. Economists argue that negative reciprocity reflects altruistic punishment, since the rejecting individual is sacrificing own resources (Fehr and Gächter, 2002). However, recent studies suggest that self-interest, in terms of status defense plays a substantial role in decisions to reject unfair offers (Yamagishi et al., 2009, 2012; Kaltwasser et al., 2016). We also assessed social preferences by social value orientation (SVO) as an indicator for prosociality and assertiveness via self-report.

We estimated the latent level association of 2D:4D and WHR with negative reciprocity, assertiveness and prosociality in both sexes. To our knowledge, this is the first study combining prenatal and pubertal static indicators within one model of socio-economic decision-making. Results revealed no robust sex-specific association between any of the trait measures and hormonal markers. When collapsing across sex greater WHR was weakly associated with assertiveness ( $\beta = 0.20$ ) and the right hand 2D:4D was weakly associated with prosocial behavior ( $\beta = 0.21$ ). Furthermore, the measures of 2D:4D and WHR were not related with each other. While the study yielded mainly non-significant results, the findings are interesting and meaningful, as they seem to substantiate the inferences and conclusions offered in several recently published studies and meta-analyses.



**FIGURE 4** | Schematic representation of the structural model testing the relationship between prenatal and pubertal organizational effects of hormones and prosociality, negative reciprocity and assertiveness. Significant relationships between latent factors at  $p < 0.05$  are depicted in bold. See the caption of **Figure 2** explaining the general conventions of the graphical language visualizing latent variable models. Note that for simplicity **Figure 4** only depicts the latent variables. The measurement models of each latent variable included in this structural model was the same as shown in **Figure 2** for the latent variables representing prenatal and pubertal organizational effects and **Figure 3** for the latent variables quantifying prosociality, negative reciprocity and assertiveness.

In view of the hypothesized relationships, our results are in line with findings of various studies reporting nil correlation of 2D:4D with trait measures such as assertiveness (Hampson et al., 2007; Voracek, 2009), depression (Yeung and Tse, 2017) or indices of socio-economic behavior such as financial risk preferences (Apicella et al., 2008). As presented in **Figure 3**, only when collapsing across gender the right hand 2D:4D was significantly, albeit weakly associated with prosocial behavior, indicating that individuals with lower prenatal testosterone exposure are somewhat more cooperative. Previous research linking 2D:4D to cooperation behavior suggests that there is no linear relationship between prenatal testosterone exposure and prosociality, but that the relationship is rather U-shaped (non-monotonic) in that subjects with both high and low digit ratios give less than individuals with intermediate digit ratios. However, the existing studies supporting this claim differ in the tested sample regarding gender and the tested criterion regarding hand as well as in the applied socio-economic paradigm, so that a systematic conclusion is impossible. For example, Brañas-Garza et al. (2013) investigated the relationship between cooperation in the dictator game and 2D:4D and found an inverted U-shaped relation for left and right hands in both genders, with a more consistent relationship in men. Sanchez-Pages and Turiegano (2010) only studied the right hand in a male population and report intermediate 2D:4D as being associated with higher cooperation in a Prisoner's Dilemma. The picture gets more complicated as ethnicity also might play a role in that a robust non-monotonic association can only be replicated for Caucasian subjects in the right hand (Galizzi and Nieboer, 2015). In this respect, our study can contribute a valuable piece of evidence to the hypothesized relationship between cooperation and 2D:4D

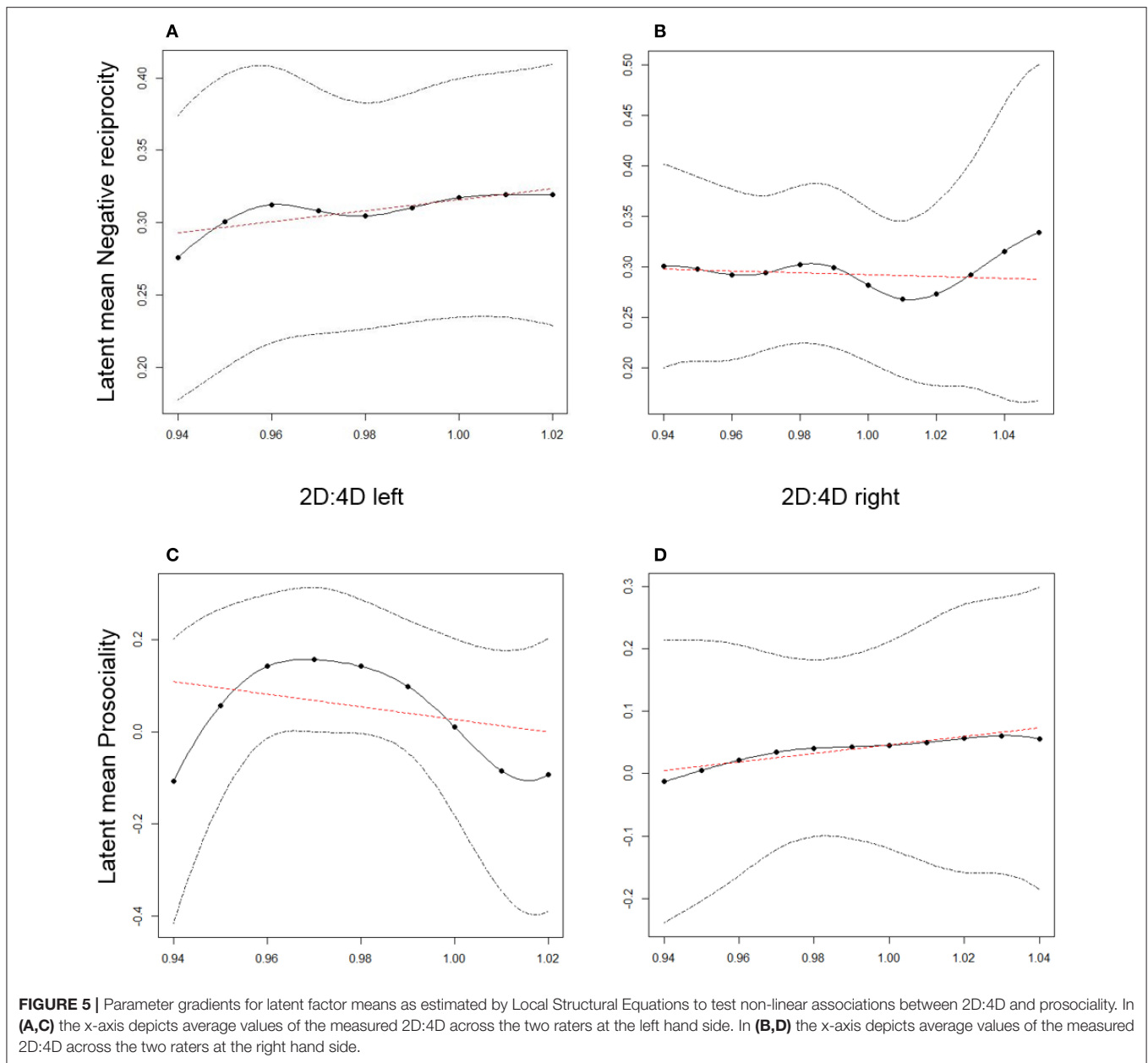
**TABLE 1** | Results of invariance testing across sex.

Model	$\chi^2$	df	CFI	RMSEA	$\Delta\chi^2$	$\Delta df$
Configural invariance	203.54	177	0.986	0.042	–	–
Weak (metric) invariance	230.27	193	0.981	0.048	26.72	16
Strong (scale) invariance	284.59	209	0.958	0.069	54.31*	16

\* $p < 0.01$ ; CFI, Comparative Fit Index; RMSEA, Root Mean Square Error of Approximation.

since we tested and compared both genders in both hands in a fairly large Caucasian sample. Our results suggest a small association between right-hand 2D:4D and prosocial behavior in terms of SVO and giving in UG, which neither is modulated by gender nor does it show a non-monotonic relationship for the right hand. However, there seems to be some evidence for an inverted u-shaped relationship between prosociality and left hand's 2D:4D in our sample (see **Figure 5C**).

Failure to detect significant 2D:4D effects has also been attributed to methodological weaknesses of a study, such as sample structure, its' heterogeneity or size, and also reliability issues related to 2D:4D measurement (e.g., Apicella et al., 2008). These arguments, however, cannot apply to our data having in mind the recruitment procedures and the effective degrees of freedom in this study (see Methods section) as well as the 2D:4D measurement procedure and method employed which followed the findings of previous evaluations of their reliability (Mikac et al., 2016). Moreover, as obvious from the analyses presented, all the study variables including 2D:4D measurements were defined by multiple indicators, that is, on a latent level and hence being free of measurement error.



Less clear empirical evidence is available on the role of facial WHR, with generally modest effect sizes reported where links were detected between WHR and selected target variables, typically referring to aggressive and/or dominant behavior (Geniole et al., 2015; Haselhuhn et al., 2015; Anderl et al., 2016). Comparable to the results we obtained for 2D:4D data, only after collapsing across sex greater WHR in our study appeared to be weakly associated with assertiveness, suggesting that individuals with wider faces tend to express greater status defense. Still, rejection behavior in UG was not related with 2D:4D nor WHR in any of the models. This applies to the tests of both linear and non-linear relationships between the indices of organizational effects of hormones and the behavioral measures examined. Hence,

neither hypothesized inverted U-shaped relation of digit ratio with prosociality nor U-shaped with negative reciprocity can be supported by this study.

Zero correlation found between latent 2D:4D and WHR deserves additional comment. This result is not surprising bearing in mind the upheld meaning and the rationale behind each of the two measures. While both are being considered to reflect organizational effects of exposure to sex steroids, they have been linked to different developmental stages—2D:4D being used as a proxy for pre- or perinatal testosterone exposure and WHR as a marker for pubertal hormone exposure. As no substantial correspondence is expected between perinatal and pubertal testosterone levels, the absence of a correlation between

the two indicators is plausible (although see Whitehouse et al., 2015). In a similar vein, statistical independence found between 2D:4D and several related sexually dimorphic facial metric measures (Burriss et al., 2007), as well as between each of these putative markers with circulating level of testosterone, has even been suggested as an evidence of their discriminant validity as measures of androgenization in respective time periods (Apicella et al., 2008).

Yet, there is also data advancing that sexually dimorphic features reflected in differing facial growth attributes might originate much earlier than pubertal age and that variation in facial WHR might begin as early as prenatal development (Bird et al., 2016). Whitehouse et al. (2015) showed that adult morphology happened to be more closely related to prenatal testosterone exposure than to adult concentrations, not ruling out, though, possible influences of adolescence testosterone levels. In a comprehensive 20-years follow-up study, these authors provided the direct evidence of a considerable association between prenatal testosterone exposure and human facial structure. Yet, this link was established between prenatal testosterone measured from umbilical cord blood and facial masculinity quantified by an objective algorithm based on multiple Euclidean and geodesic distances on 3D facial photography. Importantly, no relations were detected in the same study between WHR and 2D:4D indices, nor between each of the two static markers with either umbilical cord blood testosterone, adult testosterone level or the derived facial “genderness” score.

It seems that insights from this and other above mentioned studies including our own can at least partly account for the obtained overall modest and practically negligent findings on the relationships between the putative markers of testosterone exposure and behavioral trait measures. The results presented in this study support the position of a number of authors who question the status of either or both the digit ratio and facial WHR as static biomarkers for the assessment of prenatal and pubertal level of testosterone, respectively, or testosterone related traits (Hollier et al., 2015; Hodges-Simeon et al., 2016; Welker et al., 2016; Kramer, 2017; Yeung and Tse, 2017).

Our results propose that previous studies over-estimated the influence of static markers of testosterone on aggression and competition behavior in males. Moreover, when interpreting the role of testosterone in status-related behavior such as socio-economic decision making one should distinguish between static and dynamic markers of testosterone and take into account the situational dependency of the latter (Eisenegger et al., 2010; van Honk et al., 2012). Hence, we suggest that future studies should investigate the behavioral consequences of biological markers as a proxy for hormonal exposure more carefully, essentially relying on multimethod data (e.g., Brañas-Garza et al., in press) and prudently chosen methodological approaches to analyze

them, primarily depending on research design and metric quality of the data. Thus, structurally different biological markers of testosterone (static as 2D:4D and dynamic markers measured as circulating blood levels) could potentially be combined with different behavioral indicators of cooperation and analyzed preferably using latent variable modeling approach within a multi-trait-multi-method framework (MTMM; Eid and Diener, 2006).

Last but not least, we would like to emphasize that while it reflects ecologically valid real-world strategic social decision making, behavior in socio-economic games is not as uniform as it is often claimed—a matter that has been discussed recently in the literature (Wilhelm et al., 2017). For example, while economists and psychologists agree that specific socio-economic paradigms such as the dictator game and the Prisoner’s Dilemma unequivocally measure common aspects of altruistic or cooperative behavior (Levitt and List, 2007), they consent less on the question whether positive reciprocity (e.g., prosociality) and negative reciprocity (e.g., rejection of unfair offers) reflect two sides of the same coin (Yamagishi et al., 2012; Peysakhovich et al., 2014) as suggested in the theory of altruistic punishment. Furthermore, other aspects of socio-economic decision-making such as risk-taking or uncertainty avoidance should be taken into account in future studies relating facets of socio-economic decision-making to testosterone (Brañas-Garza and Rustichini, 2011).

## AUTHOR CONTRIBUTIONS

LK, Conceptual design, development of socio-economic paradigm, data acquisition, data analysis, writing of manuscript. UM, Conceptual design, data acquisition, data preprocessing, literature research. VB, Conceptual design, data analysis, writing of manuscript. AH, Conceptual design, data analysis (LSEM), writing of manuscript.

## FUNDING

This work was supported by a scholarship of Studienstiftung des deutschen Volkes to LK and a grant from the German Research Foundation to AH (grant number HI 1780/2-1). We further acknowledge support for the Article Processing Charge from the German Research Foundation and the Open Access Publication Fund of the University of Greifswald.

## ACKNOWLEDGMENTS

We thank Lena Fliedner, Alf Mante, Karsten Manske, Astrid Kiy, Friederike Ruffer, Tsvetina Dimitrova, Danyal Ansari, Katariina Mankinen, Susanne Stoll, and Nina Mader for their help in recruitment and data collection.

## REFERENCES

- Anderl, C., Hahn, T., Schmidt, A.-K., Moldenhauer, H., Notebaert, K., Clément, C. C., et al. (2016). Facial width-to-height ratio predicts psychopathic traits in males. *Pers. Individ. Dif.* 88, 99–101. doi: 10.1016/j.paid.2015.08.057
- Apicella, C. L., Dreber, A., Campbell, B., Gray, P. B., Hoffman, M., and Little, A. C. (2008). Testosterone and financial risk preferences. *Evol. Hum. Behav.* 29, 384–390. doi: 10.1016/j.evolhumbehav.2008.07.001
- Arnold, A. P., and Breedlove, S. M. (1985). Organizational and activational effects of sex steroids on brain and behavior: a



- reanalysis. *Horm. Behav.* 19, 469–498. doi: 10.1016/0018-506X(85)90042-X
- Auyeung, B., Lombardo, M. V., and Baron-Cohen, S. (2013). Prenatal and postnatal hormone effects on the human brain and cognition. *Pflügers Arch.* 465, 557–571. doi: 10.1007/s00424-013-1268-2
- Balliet, D., Parks, C., and Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: a meta-analysis. *Group Process. Intergroup. Relat.* 12, 533–547. doi: 10.1177/1368430209105040
- Berenbaum, S. A., Bryk, K. K., Nowak, N., Quigley, C. A., and Moffat, S. (2009). Fingers as a marker of prenatal androgen exposure. *Endocrinology* 150, 5119–5124. doi: 10.1210/en.2009-0774
- Berenbaum, S. A., and Resnick, S. M. (1997). Early androgen effects on aggression in children and adults with congenital adrenal hyperplasia. *Psychoneuroendocrinology* 22, 505–515. doi: 10.1016/S0306-4530(97)00049-8
- Bird, B. M., Cid Jofré, V. S., Geniole, S. N., Welker, K. M., Zilioli, S., Maestripieri, D., et al. (2016). Does the facial width-to-height ratio map onto variability in men's testosterone concentrations? *Evol. Hum. Behav.* 37, 392–398. doi: 10.1016/j.evolhumbehav.2016.03.004
- Bogaert, S., Boone, C., and Declerck, C. (2008). Social value orientation and cooperation in social dilemmas: a review and conceptual model. *Br. J. Soc. Psychol.* 47, 453–480. doi: 10.1348/014466607X244970
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York, NY: Wiley.
- Bos, P. A., Panksepp, J., Bluthé, R.-M., and van Honk, J. (2012). Acute effects of steroid hormones and neuropeptides on human social-emotional behavior: a review of single administration studies. *Front. Neuroendocrinol.* 33, 17–35. doi: 10.1016/j.yfrne.2011.01.002
- Brañas-Garza, P., Cobo-Reyes, R., and Domínguez, A. (2006). “Si él lo necesita”: gypsy fairness in Vallecas. *Exp. Econ.* 9, 253–264. doi: 10.1007/s10683-006-9126-0
- Brañas-Garza, P., Galizzi, M. M., and Nieboer, J. (in press). Experimental and self-reported measures of risk taking and digit ratio (2D:4D): evidence from a large systematic study. *Int. Econ. Rev.*
- Brañas-Garza, P., Kovářík, J., and Neyse, L. (2013). Second-to-fourth digit ratio has a non-monotonic impact on altruism. *PLoS ONE* 8:e60419. doi: 10.1371/journal.pone.0060419
- Brañas-Garza, P., and Rustichini, A. (2011). Organizing effects of testosterone and economic behavior: not just risk taking. *PLoS ONE* 6:e29842. doi: 10.1371/journal.pone.0029842
- Burris, R. P., Little, A. C., and Nelson, E. C. (2007). 2D:4D and Sexually dimorphic facial characteristics. *Arch. Sex. Behav.* 36, 377–384. doi: 10.1007/s10508-006-9136-1
- Buss, A. H., and Perry, M. (1992). The aggression questionnaire. *J. Pers. Soc. Psychol.* 63, 452–459. doi: 10.1037/0022-3514.63.3.452
- Camerer, C. F. (2003). *Behavioral Game Theory – Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.
- Carmichael, M. S., Warburton, V. L., Dixen, J., and Davidson, J. M. (1994). Relationships among cardiovascular, muscular, and oxytocin responses during human sexual activity. *Arch. Sex. Behav.* 23, 59–79. doi: 10.1007/BF01541618
- Carney, D. R., Cuddy, A. J., and Yap, A. J. (2010). Power posing. *Psychol. Sci.* 21, 1363–1368. doi: 10.1177/0956797610383437
- Carré, J. M., and McCormick, C. M. (2008). In your face: facial metrics predict aggressive behaviour in the laboratory and in varsity and professional hockey players. *Proc. R. Soc. B Biol. Sci.* 275, 2651–2656. doi: 10.1098/rspb.2008.0873
- Carter, S. C. (1998). Neuroendocrine perspective on social attachment and love. *Psychoneuroendocrinology* 23, 779–818. doi: 10.1016/S0306-4530(98)00055-9
- Clark, M. M., and Galef, B. G. (1998). Effects of intrauterine position on the behavior and genital morphology of litter-bearing rodents. *Dev. Neuropsychol.* 14, 197–211. doi: 10.1080/87565649809540709
- Cohen-Bendahan, C. C., Buitelaar, J. K., van Goozen, S. H., Orlebeke, J. F., and Cohen-Kettenis, P. T. (2005a). Is there an effect of prenatal testosterone on aggression and other behavioral traits? a study comparing same-sex and opposite-sex twin girls. *Horm. Behav.* 47, 230–237. doi: 10.1016/j.yhbeh.2004.10.006
- Cohen-Bendahan, C. C., van de Beek, C., and Berenbaum, S. A. (2005b). Prenatal sex hormone effects on child and adult sex-typed behavior: methods and findings. *Neurosci. Biobehav. Rev.* 29, 353–384. doi: 10.1016/j.neubiorev.2004.11.004
- Dabbs, J. M., Carr, T. S., Frady, R. L., and Riad, J. K. (1995). Testosterone, crime, and misbehavior among 692 male prison inmates. *Pers. Individ. Dif.* 18, 627–633. doi: 10.1016/0191-8869(94)00177-T
- DeBruine, L. (2004). *AutoMetric Software for Measurement of 2D: 4D Ratios*. Verfügbar unter. Available online at: www.facelab.org/debruine/Programs/autometric.php (Accessed August 13, 2011).
- Ebner, N. C., Riediger, M., and Lindenberger, U. (2010). FACES-A database of facial expressions in young, middle-aged, and older women and men: development and validation. *Behav. Res. Methods* 42, 351–362. doi: 10.3758/BRM.42.1.351
- Ehrenkranz, J., Bliss, E., and Sheard, M. H. (1974). Plasma testosterone: correlation with aggressive behavior and social dominance in man. *Psychosom. Med.* 36, 469–475. doi: 10.1097/00006842-1974111000-00002
- Eid, M., and Diener, E. (2006). *Handbook of Multimethod measurement in Psychology*. Washington, DC: American Psychological Association.
- Eisenegger, C., Naef, M., Snozzi, R., Heinrichs, M., and Fehr, E. (2010). Prejudice and truth about the effect of testosterone on human bargaining behaviour. *Nature* 463, 356–359. doi: 10.1038/nature08711
- Elias, M. (1981). Serum cortisol, testosterone, and testosterone-binding globulin responses to competitive fighting in human males. *Aggress. Behav.* 7, 215–224. doi: 10.1002/1098-2337(1981)7:3<215::AID-AB2480070305>3.0.CO;2-M
- Falk, A., Fehr, E., and Fischbacher, U. (2005). Driving forces behind informal sanctions. *Econometrica* 73, 2017–2030. doi: 10.1111/j.1468-0262.2005.00644.x
- Fehr, E., and Gächter, S. (2002). Altruistic punishment in humans. *Nature* 415, 137–140. doi: 10.1038/415137a
- Galizzi, M. M., and Nieboer, J. (2015). Digit ratio (2D:4D) and altruism: evidence from a large, multi-ethnic sample. *Front. Behav. Neurosci.* 9:41. doi: 10.3389/fnbeh.2015.00041
- Geniole, S. N., Denson, T. F., Dixon, B. J., Carré, J. M., and McCormick, C. M. (2015). Evidence from meta-analyses of the facial width-to-height ratio as an evolved cue of threat. *PLoS ONE* 10:e0132726. doi: 10.1371/journal.pone.0132726
- Geniole, S. N., Keyes, A. E., Carré, J. M., and McCormick, C. M. (2014). Fearless dominance mediates the relationship between the facial width-to-height ratio and willingness to cheat. *Pers. Individ. Dif.* 57, 59–64. doi: 10.1016/j.paid.2013.09.023
- Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* 3, 367–388. doi: 10.1016/0167-2681(82)90011-7
- Haley, K. J., and Fessler, D. M. T. (2005). Nobody's watching?: Subtle cues affect generosity in an anonymous economic game. *Evol. Hum. Behav.* 26, 245–256. doi: 10.1016/j.evolhumbehav.2005.01.002
- Hampson, E., Ellis, C. L., and Tenk, C. M. (2007). On the relation between 2D:4D and sex-dimorphic personality traits. *Arch. Sex. Behav.* 37, 133–144. doi: 10.1007/s10508-007-9263-3
- Hampson, E., Rovet, J. F., and Altmann, D. (1998). Spatial reasoning in children with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Dev. Neuropsychol.* 14, 299–320. doi: 10.1080/87565649809540713
- Haselhuhn, M. P., Ormiston, M. E., and Wong, E. M. (2015). Men's facial width-to-height ratio predicts aggression: a meta-analysis. *PLoS ONE* 10:e0122637. doi: 10.1371/journal.pone.0122637
- Haselhuhn, M. P., and Wong, E. M. (2011). Bad to the bone: facial structure predicts unethical behaviour. *Proc. Biol. Sci.* 279, 571–576. doi: 10.1098/rspb.2011.1193
- Herrmann, B., Thöni, C., and Gächter, S. (2008). Antisocial punishment across societies. *Science* 319, 1362–1367. doi: 10.1126/science.1153808
- Hildebrandt, A., Lüdtke, O., Robitzsch, A., Sommer, C., and Wilhelm, O. (2016). Exploring factor model parameters across continuous variables with local structural equation models. *Multivariate Behav. Res.* 51, 257–258. doi: 10.1080/00273171.2016.1142856
- Hines, M. (2003). Sex steroids and human behavior: prenatal androgen exposure and sex-typical play behavior in children. *Ann. N. Y. Acad. Sci.* 1007, 272–282. doi: 10.1196/annals.1286.026
- Hiraishi, K., Sasaki, S., Shikishima, C., and Ando, J. (2012). The second to fourth digit ratio (2D:4D) in a Japanese twin sample: heritability, prenatal hormone transfer, and association with sexual orientation. *Arch. Sex. Behav.* 41, 711–724. doi: 10.1007/s10508-011-9889-z

- Hodges-Simeon, C. R., Hanson Sobraske, K. N., Samore, T., Gurven, M., and Gaulin, S. J. C. (2016). Facial Width-To-Height Ratio (fWHR) is not associated with adolescent testosterone levels. *PLoS ONE* 11:e0153083. doi: 10.1371/journal.pone.0153083
- Hollier, L. P., Keelan, J. A., Jamnadass, E. S. L., Maybery, M. T., Hickey, M., and Whitehouse, A. J. O. (2015). Adult digit ratio (2D:4D) is not related to umbilical cord androgen or estrogen concentrations, their ratios or net bioactivity. *Early Hum. Dev.* 91, 111–117. doi: 10.1016/j.earlhumdev.2014.12.011
- Hönekopp, J., Bartholdt, L., Beier, L., and Liebert, A. (2007). Second to fourth digit length ratio (2D:4D) and adult sex hormone levels: new data and a meta-analytic review. *Psychoneuroendocrinology* 32, 313–321. doi: 10.1016/j.psyneuen.2007.01.007
- Hönekopp, J., and Schuster, M. (2010). A meta-analysis on 2D:4D and athletic prowess: substantial relationships but neither hand out-predicts the other. *Pers. Individ. Dif.* 48, 4–10. doi: 10.1016/j.paid.2009.08.009
- Hönekopp, J., and Watson, S. (2010). Meta-analysis of digit ratio 2D:4D shows greater sex difference in the right hand. *Am. J. Hum. Biol.* 22, 619–630. doi: 10.1002/ajhb.21054
- Hönekopp, J., and Watson, S. (2011). Meta-analysis of the relationship between digit-ratio 2D:4D and aggression. *Pers. Individ. Dif.* 51, 381–386. doi: 10.1016/j.paid.2010.05.003
- Insel, T. R., and Young, L. J. (2001). The neurobiology of attachment. *Nat. Rev. Neurosci.* 2, 129–136. doi: 10.1038/35053579
- Josephs, R. A., Newman, M. L., Brown, R. P., and Beer, J. M. (2003). Status, testosterone, and human intellectual performance. *Psychol. Sci.* 14, 158–163. doi: 10.1111/1467-9280.t01-1-01435
- Kaltwasser, L., Hildebrandt, A., Wilhelm, O., and Sommer, W. (2016). Behavioral and neuronal determinants of negative reciprocity in the ultimatum game. *Soc. Cogn. Affect. Neurosci.* 11, 1608–1617. doi: 10.1093/scan/nsw069
- Kemper, C. J., and Schwerdtfeger, A. (2009). Comparing indirect methods of digit ratio (2D:4D) measurement. *Am. J. Hum. Biol.* 21, 188–191. doi: 10.1002/ajhb.20843
- Kondo, T., Zákány, J., Innis, J. W., and Duboule, D. (1997). Of fingers, toes and penises. *Nature* 390, 29–29. doi: 10.1038/36234
- Koolhaas, J. M., Van den Brink, T. H., Roozendaal, B., and Boorsma, F. (1990). Medial amygdala and aggressive behavior: Interaction between testosterone and vasopressin. *Aggress. Behav.* 16, 223–229.
- Kramer, R. S. S. (2017). Sexual dimorphism of facial width-to-height ratio in human skulls and faces: a meta-analytical approach. *Evol. Hum. Behav.* 38, 414–420. doi: 10.1016/j.evolhumbehav.2016.12.002
- Kuhl, J., and Kazén, M. (2009). *Persönlichkeits-Stil-und-Störungs-Inventar (PSSI) [Inventory of Personality Styles and Disorders]*. Göttingen: Hogrefe.
- Levitt, S. D., and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *J. Econ. Perspect.* 21, 153–174. doi: 10.1257/jep.21.2.153
- Little, T. D., Card, N. A., Slegers, D. W., and Ledford, E. C. (2007). *Representing Contextual Effects in Multiple-Group MACS Models Modeling Contextual Effects in Longitudinal Studies*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Lutchmaya, S., Baron-Cohen, S., Raggatt, P., Knickmeyer, R., and Manning, J. T. (2004). 2nd to 4th digit ratios, fetal testosterone and estradiol. *Early Hum. Dev.* 77, 23–28. doi: 10.1016/j.earlhumdev.2003.12.002
- Malas, M. A., Dogan, S., Hilal Evcil, E. H., and Desdiciglu, K. (2006). Fetal development of the hand, digits and digit ratio (2D:4D). *Early Hum. Dev.* 82, 469–475. doi: 10.1016/j.earlhumdev.2005.12.002
- Manning, J. T., Scutt, D., Wilson, J., and Lewis-Jones, D. I. (1998). The ratio of 2nd to 4th digit length: a predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. *Hum. Reprod.* 13, 3000–3004. doi: 10.1093/humrep/13.11.3000
- Mathews, G. A., Fane, B. A., Conway, G. S., Brook, C. G., and Hines, M. (2009). Personality and congenital adrenal hyperplasia: Possible effects of prenatal androgen exposure. *Horm. Behav.* 55, 285–291. doi: 10.1016/j.yhbeh.2008.11.007
- Mayhew, T. M., Gillam, L., McDonald, R., and Ebling, F. J. (2007). Human 2D (index) and 4D (ring) digit lengths: their variation and relationships during the menstrual cycle. *J. Anat.* 211, 630–638. doi: 10.1111/j.1469-7580.2007.00801.x
- Mazur, A. (1985). A biosocial model of status in face-to-face primate groups. *Soc. Forces* 64, 377–402. doi: 10.1093/sf/64.2.377
- Mazur, A., and Booth, A. (1998). Testosterone and dominance in men. *Behav. Brain Sci.* 21, 353–363. discussion: 363–397. doi: 10.1017/S0140525X98001228
- Mazur, A., and Lamb, T. A. (1980). Testosterone, status, and mood in human males. *Horm. Behav.* 14, 236–246. doi: 10.1016/0018-506X(80)90032-X
- McIntyre, M. H. (2006). The use of digit ratios as markers for perinatal androgen action. *Reprod. Biol. Endocrinol.* 4:10. doi: 10.1186/1477-7827-4-10
- Mikac, U., Buško, V., Sommer, W., and Hildebrandt, A. (2016). Analysis of different sources of measurement error in determining second-to-fourth digit ratio, a potential indicator of perinatal sex hormones exposure. *Rev. Psychol.* 23, 39–49. doi: 10.21465/rp0023.0004
- Mitsui, T., Araki, A., Goudarzi, H., Miyashita, C., Ito, S., Sasaki, S., et al. (2016). Effects of adrenal androgens during the prenatal period on the second to fourth digit ratio in school-aged children. *Steroids* 113, 46–51. doi: 10.1016/j.steroids.2016.06.009
- Murphy, R. O., Ackermann, K. A., and Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgm. Decis. Mak.* 6, 771–781. doi: 10.2139/ssrn.1804189
- Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: can we really govern ourselves? *J. Public Econ.* 92, 91–112. doi: 10.1016/j.jpubecon.2007.04.008
- Peysakhovich, A., Nowak, M. A., and Rand, D. G. (2014). Humans display a 'cooperative phenotype' that is domain general and temporally stable. *Nat. Commun.* 5:4939. doi: 10.1038/ncomms5939
- Phoenix, C. H., Goy, R. W., Gerall, A. A., and Young, W. C. (1959). Organizing action of prenatally administered testosterone propionate on the tissue mediating mating behavior in the female guinea pig. *Endocrinology* 65, 369–382. doi: 10.1210/endo-65-3-369
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: <http://www.R-project.org>
- Resnick, S. M., Berenbaum, S. A., Gottesman, I. I., and Bouchard, T. J. (1986). Early hormonal influences on cognitive functioning in congenital adrenal hyperplasia. *Dev. Psychol.* 22, 191–198. doi: 10.1037/0012-1649.22.2.191
- Robitzsch, A. (2015). *sirt: Supplementary Item Response Theory Models*. R Package Version 0.47–36. Available Online at: <http://cran.r-project.org/web/packages/sirt/>
- Rosseel, Y. (2012). lavaan: an r package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02
- Sanchez-Pages, S., and Turiegano, E. (2010). Testosterone, facial symmetry and cooperation in the prisoners' dilemma. *Physiol. Behav.* 99, 355–361. doi: 10.1016/j.physbeh.2009.11.013
- Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Meth.* 9, 671–675. doi: 10.1038/nmeth.2089
- Stirrat, M., and Perrett, D. I. (2010). Valid facial cues to cooperation and trust. *Psychol. Sci.* 21, 349–354. doi: 10.1177/0956797610362647
- Stirrat, M., and Perrett, D. I. (2012). Face structure predicts cooperation. *Psychol. Sci.* 23, 718–722. doi: 10.1177/0956797611435133
- Vaillancourt, K. L., Dinsdale, N. L., and Hurd, P. L. (2012). Estrogen receptor 1 promoter polymorphism and digit ratio in men. *Am. J. Hum. Biol.* 24, 682–689. doi: 10.1002/ajhb.22297
- van Hemmen, J., Cohen-Kettenis, P. T., Steensma, T. D., Veltman, D. J., and Bakker, J. (2017). Do sex differences in CEOAes and 2D:4D ratios reflect androgen exposure? *A study in women with complete androgen insensitivity syndrome*. *Biol. Sex Differ.* 8:11. doi: 10.1186/s13293-017-0132-z
- van Honk, J., Montoya, E. R., Bos, P. A., van Vugt, M., and Terburg, D. (2012). New evidence on testosterone and cooperation. *Nature* 485, E4–E5. doi: 10.1038/nature11136
- Verdonck, A., Gaethofs, M., Carels, C., and de Zegher, F. (1999). Effect of low-dose testosterone treatment on craniofacial growth in boys with delayed puberty. *Eur. J. Orthod.* 21, 137–143. doi: 10.1093/ejo/21.2.137
- Voracek, M. (2009). Lack of association between digit ratio (2D:4D) and assertiveness: replication in a large sample. *Percept. Mot. Skills* 109, 757–769. doi: 10.2466/pms.109.3.757-769
- Voracek, M., Manning, J. T., and Dressler, S. G. (2007). Repeatability and interobserver error of digit ratio (2D:4D) measurements made by experts. *Am. J. Hum. Biol.* 19, 142–146. doi: 10.1002/ajhb.20581
- Van Vugt, M., De Cremer, D., and Janssen, D. P. (2007). Gender differences in cooperation and competition: the male-warrior hypothesis. *Psychol. Sci.* 18, 19–23. doi: 10.1111/j.1467-9280.2007.01842.x

- Welker, K. M., Bird, B. M., and Arnocky, S. (2016). Commentary: Facial Width-to-Height Ratio (fWHR) is not associated with adolescent testosterone levels. *Front. Psychol.* 7:1745. doi: 10.3389/fpsyg.2016.01745
- Weston, E. M., Friday, A. E., and Liò, P. (2007). Biometric evidence that sexual selection has shaped the hominin face. *PLoS ONE* 2:e710. doi: 10.1371/journal.pone.0000710
- Whitehouse, A. J., Gilani, S. Z., Shafait, F., Mian, A., Tan, D. W., Maybery, M. T., et al. (2015). Prenatal testosterone exposure is related to sexually dimorphic facial morphology in adulthood. *Proc. Biol. Sci.* 282:20151351. doi: 10.1098/rspb.2015.1351
- Wilhelm, O., Kaltwasser, L., and Hildebrandt, A. (2017). Will the real factors of prosociality please stand up? a comment on böckler, tusche, and singer (2016). *Soc. Psychol. Personal. Sci.* doi: 10.1177/1948550617707021
- Yamagishi, T., Horita, Y., Mifune, N., Hashimoto, H., Li, Y., Shinada, M., et al. (2012). Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proc. Natl. Acad. Sci. U.S.A.* 109, 20364–20368. doi: 10.1073/pnas.1212126109
- Yamagishi, T., Horita, Y., Takagishi, H., Shinada, M., Tanida, S., and Cook, K. S. (2009). The private rejection of unfair offers and emotional commitment. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11520–11523. doi: 10.1073/pnas.0900636106
- Yeung, L. A., and Tse, W. S. (2017). Why does digit ratio research fail to give any implication regarding the organizational effect of prenatal androgen? *J. Individ. Dif. Res.* 38, 36–45. doi: 10.1027/1614-0001/a000220

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Kaltwasser, Mikac, Buško and Hildebrandt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.