

# HORMONES AND ECONOMIC BEHAVIOR

EDITED BY: Pablo Brañas-Garza, Levent Neyse, Martin Voracek,  
Ulrich Schmidt and Monica Capra  
PUBLISHED IN: Frontiers in Behavioral Neuroscience





# frontiers

## Frontiers Copyright Statement

© Copyright 2007-2019 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-735-9

DOI 10.3389/978-2-88945-735-9

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)

# HORMONES AND ECONOMIC BEHAVIOR

Topic Editors:

**Pablo Brañas-Garza**, Universidad Loyola Andalucía, Spain

**Levent Neyse**, Christian Albrechts Universität zu Kiel, Germany

**Martin Voracek**, Universität Wien, Austria

**Ulrich Schmidt**, Christian Albrechts Universität zu Kiel, Germany

**Monica Capra**, Claremont Graduate University, United States

**Citation:** Brañas-Garza, P., Neyse, L., Voracek, M., Schmidt, U., Capra, M., eds. (2019).  
Hormones and Economic Behavior. Lausanne: Frontiers Media.  
doi: 10.3389/978-2-88945-735-9

# Table of Contents

- 05    *Testosterone, Cortisol and Financial Risk-Taking***  
Joe Herbert
- 22    *Endogenous Oxytocin Release Eliminates In-Group Bias in Monetary Transfers With Perspective-Taking***  
Elizabeth T. Terris, Laura E. Beavin, Jorge A. Barraza, Jeff Schloss and Paul J. Zak
- 31    *No Evidence for a Relationship Between Hair Testosterone Concentrations and 2D:4D Ratio or Risk Taking***  
Richard Ronay, Leander van der Meij, Janneke K. Oostrom and Thomas V. Pollet
- 42    *The Dopamine Receptor D4 Gene (DRD4) and Financial Risk-Taking: Stimulating and Instrumental Risk-Taking Propensity and Motivation to Engage in Investment Activity***  
Rafat Muda, Mariusz Kicia, Małgorzata Michalak-Wojnowska, Michał Ginszt, Agata Filip, Piotr Gawda and Piotr Majcher
- 52    *Facts and Misconceptions About 2D:4D, Social and Risk Preferences***  
Judit Alonso, Roberto Di Paolo, Giovanni Ponti and Marcello Sartarelli
- 63    *Risk Preferences and Predictions About Others: No Association With 2D:4D Ratio***  
Katharina Lima de Miranda, Levent Neyse and Ulrich Schmidt
- 75    *Self-confidence, Overconfidence and Prenatal Testosterone Exposure: Evidence From the Lab***  
Patricio S. Dalton and Sayantan Ghosal
- 84    *Discounting and Digit Ratio: Low 2D:4D Predicts Patience for a Sample of Females***  
Diego Aycinena and Lucas Rentschler
- 97    *A Context Dependent Interpretation of Inconsistencies in 2D:4D Findings: The Moderating Role of Status Relevance***  
Kobe Millet and Florian Buehler
- 106    *Prenatal Temperature Shocks Reduce Cooperation: Evidence From Public Goods Games in Uganda***  
Jan Duchoslav
- 118    *No Robust Association Between Static Markers of Testosterone and Facets of Socio-Economic Decision Making***  
Laura Kaltwasser, Una Mikac, Vesna Buško and Andrea Hildebrandt
- 133    *The (Null) Effect of Affective Touch on Betrayal Aversion, Altruism, and Risk Taking***  
Lina Koppel, David Andersson, India Morrison, Daniel Västfjäll and Gustav Tinghög
- 144    *Investigating Gender Differences Under Time Pressure in Financial Risk Taking***  
Zhixin Xie, Lionel Page and Ben Hardy
- 157    *Stress Induces Contextual Blindness in Lotteries and Coordination Games***  
Isabelle Brocas, Juan D. Carrillo and Ryan Kendall



- 170** *Digit Ratio (2D:4D) Predicts Self-Reported Measures of General Competitiveness, but not Behavior in Economic Experiments*  
Werner Bönte, Vivien D. Procher, Diemo Urbig and Martin Voracek
- 187** *Differential Effects of Oxytocin on Visual Perspective Taking for Men and Women*  
Tong Yue, Yuhan Jiang, Caizhen Yue and Xiting Huang
- 195** *The Effect of Testosterone Administration and Digit Ratio (2D:4D) on Implicit Preference for Status Goods in Healthy Males*  
Yin Wu, Samuele Zilioli, Christoph Eisenegger, Luke Clark and Hong Li
- 201** *Testosterone and Cortisol Jointly Predict the Ambiguity Premium in an Ellsberg-Urns Experiment*  
Giuseppe Danese, Eugénia Fernandes, Neil V. Watson and Samuele Zilioli



# Testosterone, Cortisol and Financial Risk-Taking

**Joe Herbert\***

*John van Geest Centre for Brain Repair, Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom*

Both testosterone and cortisol have major actions on financial decision-making closely related to their primary biological functions, reproductive success and response to stress, respectively. Financial risk-taking represents a particular example of strategic decisions made in the context of choice under conditions of uncertainty. Such decisions have multiple components, and this article considers how much we know of how either hormone affects risk-appetite, reward value, information processing and estimation of the costs and benefits of potential success or failure, both personal and social. It also considers how far we can map these actions on neural mechanisms underlying risk appetite and decision-making, with particular reference to areas of the brain concerned in either cognitive or emotional functions.

**Keywords:** testosterone, cortisol, finance, decision-making, risk appetite, emotion, cognition, amygdala

## INTRODUCTION

Many hormones may be able to influence financial decision-making, but two stand out as prime candidates because of their biological functions. Testosterone has well-established roles in reproduction, which embrace aggression, competitiveness and risk-taking, all essential elements of financial dealings as well as successful reproduction. Professional finance is primarily the province of males, though the situation is slowly changing; the financial world has been largely constructed by males and this reflects how hormones influence it. Cortisol is a fundamental component of the response to stress and is important for coping with unpredictable or threatening events, also a common feature or consequence of financial decisions, particularly those made under conditions of duress. Although the role of each hormone is usually considered separately, it must be recognized that under real-life conditions both will be operating together in the same individual. Because hormonal events are not apparent to the individual concerned, their influence on decision-making is covert. Furthermore, levels of hormones, the way they respond to events, and the effects these changes may have on the brain and behavior are all individually variable. So, although it is possible to define an overall action of both testosterone and cortisol on financial behavior in general, and risk-taking in particular, it is equally important to take into account those other factors, genetic or experiential, that modify endocrine responses and the effects they have in individual cases. Most of these have yet to be studied.

## WHAT IS RISK?

Risk appetite is the propensity to take risks: risk-seeking is the behavior that may, or may not, follow a given level of risk appetite. Risk occurs when there is more than one outcome when pursuing a desirable goal, in which one or more of these outcomes may be lower than the safe alternative and thus result in relative or absolute loss, danger or other undesirable consequences.

## OPEN ACCESS

### Edited by:

Nuno Sousa,  
*Instituto de Pesquisa em Ciências da  
Vida e da Saúde (ICVS), Portugal*

### Reviewed by:

Pablo Brañas-Garza,  
*Universidad Loyola Andalucía, Spain*  
Alicia Izquierdo,  
*University of California, Los Angeles,  
United States*

### \*Correspondence:

Joe Herbert  
jh24@cam.ac.uk

**Received:** 26 October 2017

**Accepted:** 27 April 2018

**Published:** 16 May 2018

### Citation:

Herbert J (2018) Testosterone,  
Cortisol and Financial Risk-Taking.  
*Front. Behav. Neurosci.* 12:101.  
doi: 10.3389/fnbeh.2018.00101

In the more restricted context of finance, risk as outcome variance contributes to the subjective value an individual attaches to that risky option. The subjective value derived from risk is typically determined by giving individuals a choice between a safe (i.e., risk-free) and a risky alternative. If one adjusts the magnitude of the safe alternative until the decision maker is indifferent between the two alternatives, one has determined the subjective value of the risk. Individuals who are risk averse give up money to avoid risk. That is, they are indifferent at safe magnitudes that are smaller than the expected value of the risky alternative. Conversely, individuals who are risk-seeking pay money in order to experience risk. The important point here is that it is the subjective, not the objective, value of the reward and the perceived (rather than the actual) probability of success that influences risk-taking.

Risky decision-making involves several distinct components. Information about the likelihood of success of a particular action is the first, and this depends on previous experience of similar situations, the amount and accuracy of current information, and the ability of the individual to assess that information. From this information, the risk-taker estimates the probability of success and the consequences of failure. The decision to take a given action depends on the subjective value of success or failure to the individual concerned (utility), which can include personal consequences directly related to the decision (e.g., immediate loss or gain of money) or secondary ones (social esteem, promotion, loss of job or livelihood). Major theoretical accounts of risk valuation include expected utility theory, prospect theory and the summary statistics approach to finance theory (reviewed in Schultz, 2006; D'Acremont and Bossaerts, 2008). One problem with many theories of economic risk-taking is that they attempt to cover all contexts and eventualities. But there are substantial differences between, say, a professional trader with much experience and specific training, dealing in millions of pounds every day upon which his salary and even his employment depends, and an average citizen, untrained and inexperienced in financial matters, making everyday financial decisions, some of which may have little consequence. Attempts to devise a more comprehensive theoretical base for economics continue (Orrell, 2018).

There are different types of risk, including liquidity risks, sovereign risks, insurance risks, business risks, default risks etc. Mathematical definitions of risk mostly assume that rewards fluctuate around the mean value (variance) but other patterns include situations in which high reward occurs only occasionally (positive skewness) or scanty reward occurs often (negative skewness; Genest et al., 2016). Most of the literature on the role of hormones in finance focuses on rapid decisions made under the artificial conditions of the laboratory that attempt to reproduce, to some extent, those made in real life within a narrow definition of risk (see below).

Financial decisions and assessments of associated risks are in many ways no different from other types of decisions (Kusev et al., 2017). In particular, decisions taken in contexts of violence or combat have many of the same properties (see below). Both may require rapid decisions, based on estimates of current information which may be available in rapidly changing amounts

and to varying degrees of accuracy. Much of the literature on risk-taking in other contexts, particularly those that include urgent and personally-important outcomes, will therefore be highly applicable to understanding the basis of financial risk-taking, even if they have not been directly tested. It should be noted that these circumstances, historically at least, have been mostly masculine ones, a point considered further below. The major difference is that financial risks involve the loss or gain of money rather than personal danger or physical assets. But money represents both potential gain of assets and alterations in social and personal status, factors which are not so different from the more traditional objectives of personal conflict or war or assets such as territory, food supply or sexual partners (Slovic, 1964). A major difference between money and these more biological rewards (based on current or anticipated need) is that gain or loss of money does not necessarily apply to any particular primary reward, such as food, drink or sex. Furthermore, unlike these primary rewards, the rewarding nature of money has to be learnt, and varies with culture and circumstance.

Both testosterone and cortisol have central roles in these behaviors. In both situations, not only the outcome but also the actions associated with risk-taking may themselves be important, since display of such behaviors may have social implications for esteem or leadership, and may therefore contribute to the decision-making process (Eckel and Grossman, 2002). It follows that the neural and endocrine mechanisms associated with neuroeconomics will resemble those in other behavioral contexts involving evaluating risks and making decisions, and the extensive psychological literature on learning and reward assessment will also have direct relevance (Camerer, 2008).

The notion that financial decisions are always taken as a result of accurate and objective assessments of risks and benefits has long since been superseded by a more nuanced approach; in particular, psychological theory realized that risk needs to be perceived and that emotional factors as well as cognitive processes can influence this perception and the decisions that follow from it (Kahneman and Tversky, 1979). Distinctions between “emotion” and “cognition” are difficult and not always clear, and the contribution of either depend not only on the current assessment of a risky choice but on such general properties as personality, emotionality and current mood as well as experience, training, and the particular properties and circumstances of the choice to be made and how they are computed (Zuckerman, 1991). We shall need to consider which components of this manifold system are controlled or influenced by hormones. There is an extensive account of the theoretical basis of risk and decisions made under conditions of uncertainty (Starcke and Brand, 2012).

This article focusses on the roles of testosterone and cortisol in acute decisions made under such uncertainty. As outlined above, such decisions are common in finance, but also in other aspects of life. We can therefore apply some of the information on the way these two hormones affect behavior to the more particular context of finance. The choice of these two hormones rests on the knowledge that they are the ones most obviously concerned with some of the fundamental aspects of behavior that occur under conditions when rewards are only obtainable if there is an

assessment of the associated risks, culminating in decisions about whether or not to take them.

## SIMILARITIES AND DIFFERENCES BETWEEN TESTOSTERONE AND CORTISOL

Though both testosterone and cortisol have powerful influences on decision-making, there are important differences as well as similarities between the hormones themselves. Both are steroids, which means that the cellular action they have on neurons is similar to the extent that both act on intracellular steroid-binding molecules, receptors, which are reasonably but not entirely specific for each hormone (Claessens et al., 2017; Gray et al., 2017; Maney, 2017). There is also evidence for a second, more rapidly acting membrane-bound receptor for both steroids (Vernocchi et al., 2013; Shiha et al., 2014). Thus the neural actions of both hormones can be both rapid (within a few minutes) via the membrane-located receptors or more prolonged (hours or days), since the intracellular receptors, when activated by a bound steroid, act directly on the genome through on different elements—either glucocorticoid or androgen receptor binding sites. In each case, there are large numbers of downstream genes that are either activated or suppressed as the result of this addressing of the genome. The respective patterns of this genomic response, and how they differ between the two steroids, have not been adequately elucidated.

Access to the brain is essential if they are to influence behavior and this is regulated in a similar way for both steroids. Secreted testosterone and cortisol bind to large plasma proteins, either sex-hormone or corticoid binding globulin (SHBG, CBG). These carrier proteins limit access to the brain because only unbound (“free”) steroid can pass through the blood-brain barrier. So alterations on either the proportion of steroid binding to its respective globulin, or the levels of that globulin, will influence how much reaches the brain irrespective of blood levels. However, as blood levels rise there will come a point at which the carrier globulin is saturated: this will result in any extra steroid being immediately available for entry to the brain, and therefore a disproportionate surge of intracerebral hormone. This may have important consequences for behavior.

There are also significant differences between testosterone and cortisol. Both steroids are secreted in a series of c. 90 min (circadian) pulses. Testosterone levels have a minor daily rhythm whose physiological significance has never been shown. Cortisol has a major rhythm, with morning levels being 4–5 higher than those in the evening (the amplitude is individually very variable; Bailey and Silver, 2014). Both the circadian and daily rhythms of cortisol have coding properties for the expression of corticoid-sensitive genes (Russell et al., 2015; Lightman, 2016; George et al., 2017). Disturbances in the daily rhythm (e.g., during episodes of stress or depressed mood) will alter this coding property, but this can be distinct from increases in overall exposure of the brain to cortisol. Both may have neurobiological consequences (Herbert et al., 2006). There are marked gender differences in testosterone levels, but much less in cortisol, though morning cortisol levels

are around 20% higher in females (Netherton et al., 2004). Adult male testosterone levels are very labile and environmental events that are very relevant to financial decisions, such as a psychological or physical challenge or success in a competitive encounter, raise levels whereas situations of persistent stress or fear lower them (Archer, 2006; Goetz et al., 2014). There is also a gradual decline with age, though this is also individually variable (O'Connor et al., 2011). Cortisol rapidly responds to stressful events, particularly those that are threatening, unpredictable and lack evident means for coping with them, including social or material support (Lucassen et al., 2014). This stress-related increase is an essential part of the response to adversity. Unlike testosterone, very low levels or absent cortisol (Addison's disease) are life-threatening.

There are well-known genetic variations in the androgen receptor which have significant consequences for its function. The length of the CAG repeat at the N-terminal has a reciprocal effect on testosterone sensitivity, and is individually variable (Morimoto et al., 1996). This will moderate the behavioral effects of testosterone in an individual manner, but has seldom been taken into account. Other, less common, variants include some that prevent testosterone from acting on the brain, resulting in a female phenotype in an XY individual (Wisniewski et al., 2000; Jääskeläinen, 2012). Genetic variants in the glucocorticoid receptor are also known; there is no coherent account of their physiological significance, though they have been implicated in vulnerability for depression (Wüst et al., 2004; van Rossum et al., 2005; Bustamante et al., 2016). Some of the behavioral effects of testosterone depend on aromatization to estrogen (Finkelstein et al., 2013). Genetic and other moderators of aromatization will therefore affect the behavioral consequences of altered testosterone. The actions of cortisol on the brain do not depend on an equivalent mechanism, but conversion to inactive cortisone by 11 $\beta$ -hydroxysteroid dehydrogenase (also genetically variable) protects mineralocorticoid receptors from its action (MacLulich et al., 2012). Altered conversion of cortisol to other metabolites in the brain is another individual difference (Alikhani-Koupaei et al., 2007; Ragnarsson et al., 2014). The areas of the brain on which the two steroids act are also different and are discussed in more detail below. There are also proposed interactions between testosterone and cortisol—the behavioral effects of changes in one depending on levels of the other (Mehta and Josephs, 2010) that will also be considered further below.

## LIFETIME TRAJECTORIES IN TESTOSTERONE AND CORTISOL

The lifetime trajectories of the two steroids differ. The human male brain is exposed to three successive waves of testosterone (Nieschlag and Behre, 2012). The first, beginning at around 10 weeks post-fertilization, has major effects on the organization of the brain, particularly sexual identity, preference and behavior and sensitivity to testosterone in adulthood, though other aspects of testosterone-related functions may also be affected. The second surge lasts around 4 months postnatally, and lasts about 16 weeks. Its function is still largely mysterious.

The third surge is responsible for puberty and its associated physical and psychological events, and lasts for the remainder of the male's life, though levels may decline with age (Lewis et al., 1976). Cortisol does not show these age-related surges, though adverse events early in life may alter subsequent levels or the way they respond to stress: labeled "re-programming" (Pearson et al., 2015; see below) and levels may increase with age (Wrosch et al., 2007; Lupien et al., 2009). Moreover, there may be significant sex differences in the way that cortisol affects decision-making (van den Bos et al., 2009) since males make most of the financial decisions under the conditions considered here, this will be our focus. However, changes in the financial industry in the future may alter this perspective.

## ASSESSING THE ROLES OF HORMONES IN RISK-TAKING

There are several methods of assessing the roles of hormones in financial decision-making, none of them entirely satisfactory (this also applies to other studies of risk appetite). The first, essentially correlational, is to relate differences in levels of testosterone or cortisol, or changes in those levels, with liability to take risks or avoid losses. The advantages of this method are that it allows observations to be made under real-life conditions: the disadvantage is that can never establish causality. The most direct method is to give steroids (e.g., testosterone or cortisol) to those engaged in finance (e.g., daily trading) and measure the outcome. This is legally, practically and ethically impossible, as is giving androgenic steroids to competitive athletes. But steroids can be administered to subjects under experimental or laboratory conditions, in which they play games that are designed to reproduce at least some of the features of real life. However, it should not be forgotten that these experimental conditions never reproduce, entirely, the conditions and consequences of real-life financial dealings.

Levels of testosterone are only one way of assessing changes in its activity: the effect it has on behavior will vary according, for example, to genetic variance in the androgen receptor or SHBG, and the pattern of other genes with which testosterone interacts, as well as factors such as the "personality" and experience of the individual concerned. Similar reservations apply to cortisol. So far, there are no studies on the genetic make-up of professional financiers (e.g., traders) or those making everyday financial decisions, and how it might be related to performance under various conditions and relate to changes in hormone levels. Similar considerations apply to investigations in which subjects play a financial game which has some similarity to real-life conditions (though usually less complex and demanding; Cueva et al., 2015; Schipper, 2014). Under these conditions it is possible to give hormones (e.g., testosterone or cortisol), though the fact that the rewards or the consequences are seldom very significant for the subjects robs such studies of an important real-life element. Since risks are a component of other activities, it should also be possible to extrapolate from non-financial studies to yield a greater understanding of financial risk-taking, after taking any special features into

account. Experimental studies on risk-taking and reward-related behavior in animals are collateral evidence, though the differential cognitive abilities of human and animal brains limit their usefulness. Nevertheless, the basic neural mechanisms may be similar, and there are greater opportunities for experimental manipulations and examination.

## TESTOSTERONE AND ADOLESCENT RISK-TAKING

The surge in testosterone that occurs at puberty and during adolescence is associated with increased appetite for risks and rewards including those related to financial gain in both sexes, particularly as these affect peer relationships and social status, perhaps most prominently in boys (Morrongiello and Rennie, 1998; Steinberg, 2008; Vermeersch et al., 2008; Cardoos et al., 2017). There is increased activation of the nucleus accumbens, an area associated with reward (see below) though this was not related to individual testosterone levels (Alarcón et al., 2017). The neuroendocrine explanation for this has focused on the role of dopamine, referring to its well-known role in the neural basis of reward (Schultz, 2006). There is experimental evidence that dopamine is necessary for testosterone-induced motivated behavior, and that testosterone also moderates dopamine transporters and receptors in the substantia nigra (Bell and Sisk, 2013; Purves-Tyson et al., 2014; Morris et al., 2015).

However, in humans there is an additional factor: the maturation of the frontal lobes. Progressive reduction in the age of puberty has resulted in a mismatch between the advent of the pubertal testosterone surge and the maturation of the brain, particularly the frontal lobes (late adolescence, early 20 s). Furthermore, the frontal lobes mature later in boys than girls (Lenroot and Giedd, 2010; Raznahan et al., 2010; Mills et al., 2014). Since this part of the brain plays an established role in the evaluation of rewards and associated risks (see below), as well as in the emotional response to them, the increasing mismatch between the endocrine and neural events now occurring at puberty may well play a crucial role in adolescent risk-taking, including those associated with financial decisions. For example, pubertal testosterone increases the responses of the frontal lobe to emotional events (Tyborowska et al., 2016). It is interesting to speculate whether increases in the utility of financial gains at puberty might be secondary to the advent of sexual motivation. Testosterone (in both sexes) heightens sexual motivation (reward). This increases the utility of money, in the sense that it may promote access to sexual objectives either directly or by increasing social status. It may be one example of how hormones, through their selective action on reward value, can alter the pattern of financial risk-taking in a setting that ostensibly has no relation to the primary action of that hormone, in this case testosterone on sexual motivation. It should also be noted that the pubertal surge of testosterone in males, unlike females, acts on a brain that has already been exposed to the same hormone prenatally, an event which may sensitize it to the pubertal surge as well as influencing the nature of the behavioral response to it (Apicella et al., 2008).



## TESTOSTERONE AND RISK-TAKING IN ADULTS

The impact of the basic reproductive function of testosterone and its influence on financial risk-taking is supported by other experiments on adult men. Heterosexual men exposed to opposite-sex stimuli take greater financial risks (Baker and Maner, 2008). This suggests sexual motivation, which is testosterone-dependent, accentuates risk-taking as part of the process of getting a mate (display, increased assets, etc.). However, images of physically-attractive men also increase risk-taking (also in heterosexual subjects), suggesting that this stimulus acts on the competitive element of sexual selection (Chan, 2015). As part of its widespread effects on behavior, all related to its fundamental role in reproduction, testosterone helps to maintain social status, and levels can reflect social or physical challenge as well as status (Booth et al., 1989; Mazur and Booth, 1998). This may influence risk-taking as part of the competition to sustain that status (Stanton and Schultheiss, 2009). However, there may also be a reciprocal interaction between social status and testosterone: men with lower testosterone put into a high status position showed poorer cognitive functioning than those with higher testosterone: the reverse occurred after being put into lower status positions (Josephs et al., 2006). There seems to be a variety of ways, all related to sex or its concomitants, but differing in proximal mechanisms, by which testosterone-related behavior could alter financial risk-taking.

Studies on the association between testosterone levels and financial trading in real-life contexts and have provided intriguing findings—traders made more money on days when their testosterone levels were highest (Coates and Herbert, 2008). This agrees with laboratory studies showing that subjects with higher testosterone levels made riskier bids in a financial game (Apicella et al., 2008), though this has not always been confirmed (Sapienza et al., 2009). These findings are associations, and unless there is considerably more information on individual strategies, supported by interventional studies, the level of analysis is limited. Interestingly, giving testosterone to traders playing an economic game that resembled real-life resulted in increased price offers (i.e., mispricing) and over-optimism about future changes in asset values (Nadler et al., 2017) and non-professional subjects showed similar effects, together with increased appetite for risk (Cueva et al., 2015). Thus, testosterone appears to increase individual willingness to take financial risks because it biases estimates of outcome. It is interesting to speculate that collective over-ambitious estimates may be one reason for the periodic “bubbles” that affect the stability of financial markets (see below). Whether the “winner” effect—increased levels of testosterone after a successful deal—has any effect on subsequent risk-taking has not been established, though it remains a possibility. Men playing with a gun (but not a children’s toy) showed increased testosterone, and were more willing to inflict discomfort to others (adding a hot sauce to food; Klinesmith et al., 2006), suggesting that similar “carry-over” effects may occur in a financial setting, and there are associations between acute

changes in testosterone following a competitive challenge and features such as subsequent competitiveness, aggression and rating faces as trustworthy (reviewed by Apicella et al., 2015) though whether these depended upon increased testosterone or on related psychological traits independent of the actual rise in testosterone remains speculative. A recent history of receiving rewards can reset estimation of future rewards (Khaw et al., 2017), though whether the response of either testosterone or cortisol to such previous rewards contributes to this effect is not yet known. Note that there have been no substantive assessments of the role of other testosterone-related features, including genetic variants of the androgen receptor, in financial risk-taking behavior.

We should not be surprised if testosterone has manifold actions on financial decision-making. A similarly wide canvas is seen in its primary role in reproduction. In order to achieve its role, testosterone has to act on both physical features, such as the growth of horns, teeth and muscles, as well as on a range of behavioral attributes such as aggressiveness, competitiveness and willingness to take risks, in addition to primary actions on sexual motivation and attractiveness (Herbert, 2017).

## GENDER DIFFERENCES IN RISK APPETITE

Gender differences in risk appetite are an indirect and incomplete way of assessing the effects of hormones, particularly testosterone. It is important to recognize that not all gender differences are testosterone-based. The Y chromosome expresses genes that directly affect behavior, and the presence of two X chromosomes in females is also important. But more significantly, environmental factors such as upbringing, expectations, opportunities and social attitudes, though directly related to gender, are an indirect effect of testosterone-dependent gender differences in the brain and its phenotype and can have potent actions on any aspect of gender-related behavior, including the perception of risk and risk-taking (Lenroot and Giedd, 2010). Nevertheless, careful assessment of gender differences in risk appetite or processing can add some information about testosterone-dependent aspects of risk-taking, though it may be difficult to separate the role of early exposure to testosterone from the action of post-pubertal hormone (see below), and to account for the effects of social attitudes and expectations.

First, the nature of the risk is important. Many studies show that males and females differ with respect to the kinds of risks they find attractive or aversive (Schubert et al., 1999; Rolison et al., 2014). A meta-analysis of risk-taking across several domains showed that males were generally more inclined to take risks than females, though the size of the effect varied with different risks. Gambling, for example, showed a greater gender difference than risky sexual behavior, but less than physical risk-taking (Bryrnes et al., 1999). More recent work has moderated this view: risky social behavior either shows no gender difference or more risks were taken by females; the greater appetite for financial risks (e.g., gambling) by males was confirmed, women being more pessimistic about a positive

outcome and enjoying it less (i.e., reward value; Harris and Jenkins, 2006).

Giving testosterone to women and then assessing the effect it has on risk-taking has dubious value if it is regarded as a test of gender differences (i.e., making women more “male-like”) or a demonstration of the action of testosterone in general, since this ignores both the gender difference in early exposure to testosterone, and the presence or absence of two X or one Y chromosomes. Bearing this in mind, exogenous testosterone increases stress reactivity in women (startle reflex; Hermans et al., 2007) and decreases empathy (which is generally greater in women than men; Hermans et al., 2006). This will impact financial decisions, since stress and empathy both affect risk appetite and concepts of fairness and are examples of the interaction between stress (cortisol) and testosterone (see below). Gender differences in risk-taking have been related to corresponding differences in the 2D:4D digit ratio, proposed to be a reliable index of exposure to early testosterone in females as well as males (van Honk et al., 2011); however, there are serious questions about the information given by the digit ratio.

Since prenatal testosterone has such a powerful effect on subsequent behavior and physiology in males, there is considerable interest in estimating its action in individual cases. It should be noted that this depends not only on levels of testosterone, but also on the sensitivity of response to it, which includes genetic variation in the androgen receptor (Vermeersch et al., 2010; Hurd et al., 2011). Direct measurement of testosterone during the critical period (c.10–20 weeks) is not possible. The 2D:4D digit ratio has been used as a proxy, but this is highly dubious. The ratio is less in males than females (though there is a considerable overlap; Manning et al., 1998; Breedlove, 2010; Knickmeyer et al., 2011); XY individuals with complete androgen insensitivity have ratios in the female range (van Hemmen et al., 2017). Prenatal testosterone thus plays a role in determining the ratio (which has no known function), but this is very different from concluding that individual differences in prenatal testosterone are reflected in individual measures of the digit ratio in males, for which there is no convincing evidence (see Ventura et al., 2013). Yet the ratio, which is easily measured, continues to be used in this way (e.g., Kim et al., 2014). Lower ratios in males have been associated with higher risk taking (the opposite was found for females), though this was attributed to greater ability for abstract reasoning as well as greater risk appetite, but only in males (Brañas-Garza and Rustichini, 2011; Branas-Garza et al., 2018). There have been both negative reports and positive ones for the association of lower digit ratios with increased risk-taking within both males and females (Branas-Garza et al., 2018) as well as with greater reflective consideration of decisions in both sexes (Bosch-Domènech et al., 2014). Lower ratios have been associated with less over-confidence in males (estimate of success in a quiz) but only when success was rewarded, suggesting that this might be related to adult surges of testosterone responding to challenge and acting on a brain pre-conditioned by pre-natal testosterone—though this was not measured (Neyse et al., 2016). This seems incompatible with a report that administration of testosterone to adult males increases optimism (confidence) about outcomes (Cueva

et al., 2015). Since females are not exposed to early testicular testosterone, the rationale for relating their individual digit ratios to risk-taking seems obscure. Furthermore, the variance in digit ratios for females is quite similar to males: this suggests that factors other than prenatal testosterone influences individual digit ratios; the same may apply to males. The current uncertainty about the accuracy or validity of the digit ratio as a marker of the amount of early exposure to testosterone in individual males makes interpretation of these results both difficult and tentative.

Empathy plays a role in many financial dealings, for example in the ultimatum game, and is generally greater in women than men (Auyeung et al., 2009). Generosity in this game is reduced by giving men or women testosterone (Zak et al., 2009; van Honk et al., 2011), and men with higher levels are more likely to reject low offers (Burnham, 2007). Higher testosterone is associated with less empathy and greater “utilitarianism” in decisions that require a choice that has immediate costly consequences: this would impact financial as well as other types of decisions (Carney and Mason, 2010). It should be noted that in this context, as in all others, the actions of testosterone are only one factor determining such behavior (Takahashi et al., 2012). Entrepreneurship is a form of risk-taking and challenge, in that the participant risks assets in setting up and developing his/her own business. Whether this can be related to testosterone is disputed: males setting up a new venture had higher testosterone levels, whereas those who had ever been self-employed (a different definition) did not (White et al., 2006; van der Loos et al., 2013).

## THE COMPLEXITIES OF STRESS

Stress is often used as if it is a single defined concept. This is not the case. Stress is actually a generic term for a range of situations: the only commonality is that they represent an unusual demand which, if this is to be met satisfactorily, requires an adaptive response. But an inadequate or mal-adaptive response may also occur, with corresponding consequences. There is also confusion between stressors (the nature of the demand) and the reaction to the demand (the stress response). The response to stress (often abbreviated to “stress”) is also complex. The physiological response to an acute stress involves both catecholamines as well as cortisol, and there is experimental evidence that they interact in the brain (Ferry et al., 1999; McReynolds et al., 2010; Wolf et al., 2016). This will differentiate the effects of acute from chronic stress, since catecholamines play a lesser role in the latter. There is a recent report that increased loss aversion after cortisol administration only occurred when combined with simultaneous noradrenergic activation (Margittai et al., 2018).

Most laboratory studies of the effects of stress on decision-making focus on acute stress (Starcke and Brand, 2012) at a single time point, but another complication is that the effects of stress may alter with time. For example, an initial response may be to increase risk-appetite, but this may reverse at later time periods (Bendahan et al., 2017) either because of the altered interaction between catecholamines and cortisol, or because its initial membrane-dependent actions differ from the slower

genomic ones. The nature of the stressor is also important: physical stressors, such as cold immersion (pressor test) are not the same either physiologically, cognitively or emotionally as psychological stressors such as the Trier test or cognitive overloading, such as simultaneous distractors (e.g., mathematical problems), and none of these capture all the features of the stress associated with incipient or current risky financial decisions. The latter incorporate emotional and cognitive reactions to the nature of the decision itself, which are not present in background stressors, unrelated to the risk. This may well have different consequences for decision-related behavior than other types of stress. A recent report describes distinct metabolic patterns in the hippocampus following either physical or psychological stress, emphasizing the difference between them (Liu et al., 2018). Yet all are often included in the single sobriquet of “stress” and interpreted as such. Stress is also more than elevated cortisol, though this is an important component of the stress response. A major reason for the inconsistency of reports on the effects of stress on decision-making is one result of insufficient attention to these important distinctions and variables. There is also, as already mentioned, the problem of modeling real-life situations in the laboratory.

## CORTISOL AND RISK-TAKING

It is important to recognize the different effects of raising cortisol levels and altering the shape of the daily rhythm. Both have consequences for brain function, but they can differ (see above). Experiments that give subjects cortisol several times a day will confuse the two mechanisms (e.g., Kandasamy et al., 2014). Even though persistent stress can result in both increased cortisol and altered daily rhythms, it is important to bear this distinction in mind. In contrast to testosterone, dysregulated cortisol has been implicated in the increased susceptibility of the brain to damage by toxic agents, in heightened incidence of depression, and in the risk that depression poses for decision-making as well as for subsequent Alzheimer’s disease (Herbert et al., 2006; Herbert, 2013; Herbert and Lucassen, 2016).

Persistently high levels of cortisol, such as those in Cushing’s disease, impair cognitive function and also predispose to depressed mood (Starkman et al., 1981; Newcomer et al., 1999; Hook et al., 2007). The magnitude and duration of the cortisol response to stress in a financial context depends on many factors, of which uncertainty about market movements and their volatility are the most relevant to financial decisions (Coates and Herbert, 2008; Cueva et al., 2015). Most evidence has been on the effects of short-term cortisol administration, which is certainly relevant to real-life trading conditions. However, there will be circumstances in which subjects are experiencing more persistent stress, and therefore more prolonged elevations of cortisol, and this may have different results.

There are thus indications that acute, short-term increases in cortisol may have different effects from more long-term, chronic, ones (Lucassen et al., 2014). This would differentiate the influence that cortisol has on decisions in response to a

short-term financial demand from its effect on those made during a more persistent state of stress. This separates acute responses (attention to threats, fear etc.) from those characteristic of more chronic states—which may relate to altered risk aversion (Putnam et al., 2007; van Ast et al., 2013). The intrinsic nature of the decision that has to be made is likely to be associated with a more acute cortisol response, whereas a pre-existing state, which may or may not be associated with the context of the financial risk to be taken, will result in a more prolonged cortisol reaction which may also influence that decision in a manner that is different from more acute or short-term cortisol responses (Porcelli et al., 2012). Acute administration of cortisol in other contexts increases the arousal response to stimuli, as well as enhancing the consolidation of memories of adverse events whilst reducing their recall (Abercrombie et al., 2003, 2005; Wirth et al., 2011; Wolf et al., 2016). There are similar indicators in the brain: the reaction of the amygdala (which contains profuse glucocorticoid receptors) to facial expression changes with time, an effect which has been related to its connections with the medial frontal cortex (Henckens et al., 2010). Stress has pervasive effects on cognitive functions highly relevant to finance, including selective attention, working memory, and cognitive control (Okon-Singer et al., 2015). Though it is usually assumed that the effects of stress are the result of altered corticoids, it should be recognized that there are other physiological and neurological consequences of stress that may contribute (Lucassen et al., 2014; see above). However, in one study stress increased risk-taking only in those in whom cortisol was elevated (Buckert et al., 2014), thus suggesting that it was cortisol that underpinned most of the effects of stress in this case.

Cortisol administration also impairs detection of errors (Hsu et al., 2003), a crucial element of rapid decisions made under duress. It increases appetite for risk (Cueva et al., 2015), though there is a contrary report, possibly as the result of a different regime of cortisol administration—repeated daily administration, which would alter both cortisol levels and its daily rhythm (Kandasamy et al., 2014). Note, too, that the effect of acute cortisol may be time-dependent (see above). A meta-analysis confirmed that stress increased appetite for rewards together with associated accentuated risk-taking: together, these resulted in overall disadvantageous outcomes (Starcke and Brand, 2016). Stress impairs executive functions such as attention and inhibition, task management and planning (Starcke et al., 2016). However, the exact consequences depend on the type of stress and the context in which it occurs (Starcke and Brand, 2016) since the behavioral action of cortisol is so widespread. Not all the effects of stress or cortisol are necessarily disadvantageous. Stress can be an enhancing experience, particularly if there are adequate resources for coping with it or if emotional states (e.g., anxiety) are consciously appraised (O’Connor et al., 2010; Akinola et al., 2016).

In contrast to testosterone, cortisol does not show a financial “winners” response (McCaul et al., 1992); another significant difference between the two hormones is that whilst both increased risky choices, only testosterone increased optimism



about price changes; cortisol did not (Cueva et al., 2015). This suggests that while the effect of testosterone on risk-taking might be secondary to over-optimistic assessments of possible outcomes, cortisol had a more direct action on risk-appetite itself. This may be an adaptive (or mal-adaptive) response to financial situations that are unpredictable or apparently uncontrollable, since cortisol responds so sensitively to such conditions. An uncontrolled stress response thus becomes a hindrance to the most advantageous courses of actions under these circumstances. It should be emphasized that nearly all these results have been obtained on male subjects and that there is no reason to assume that they might apply to females (Cueva et al., 2015).

## INDIVIDUAL MODULATION OF THE RESPONSE TO CORTISOL

Although, as for testosterone, it is possible to make general statements about the effects of cortisol, either acute or chronic, or immediate or delayed, on decision-making, it is important to recognize that these effects can be moderated by individual characteristics. These include impulsivity, which tends to increase risky behavior (see below; Lempert et al., 2012) and state anxiety (Lempert et al., 2012) as well as cognitive style, such as rapid (“fast”, habitual) or slower (model-based) decision-making, and thus interactions between speed and accuracy (Kahneman, 2011) as well as other aspects of personality (Nicholson et al., 2005). For example, the accumulation of lifetime stress accentuates habitual responses to risky decisions only in those with slower cognitive styles (Friedel et al., 2017). The bases for these differences, which would likely include variation in genetic constitution and/or individual experience, has not been explored adequately. Early life stress can also have effects on decision-making in adulthood, particularly altering sensitivity to loss (Birn et al., 2017). Whilst the mechanism for such an influence is not yet known, it does recall the long-lasting epigenetic changes in the glucocorticoid receptor described in other contexts of early adversity (Mazur and Booth, 1998; Meaney et al., 2007; Herbert and Lucassen, 2016; Gray et al., 2017) which would have wide-ranging effects on the pattern of cortisol secretion.

Again, as for testosterone, cortisol can alter a number of parameters associated with financial decisions, including loss aversion, but also reward sensitivity as well as a tendency to favor short-term over longer-term gains (Canale et al., 2017). This is not surprising, given the widespread action of cortisol on the brain. However, it does mean that cortisol may have different consequences on risky behavior for those engaged in short-term decisions under duress (e.g., traders) from decisions made more deliberately for the longer term (e.g., stock investments).

## IMPULSIVITY AND HERDING

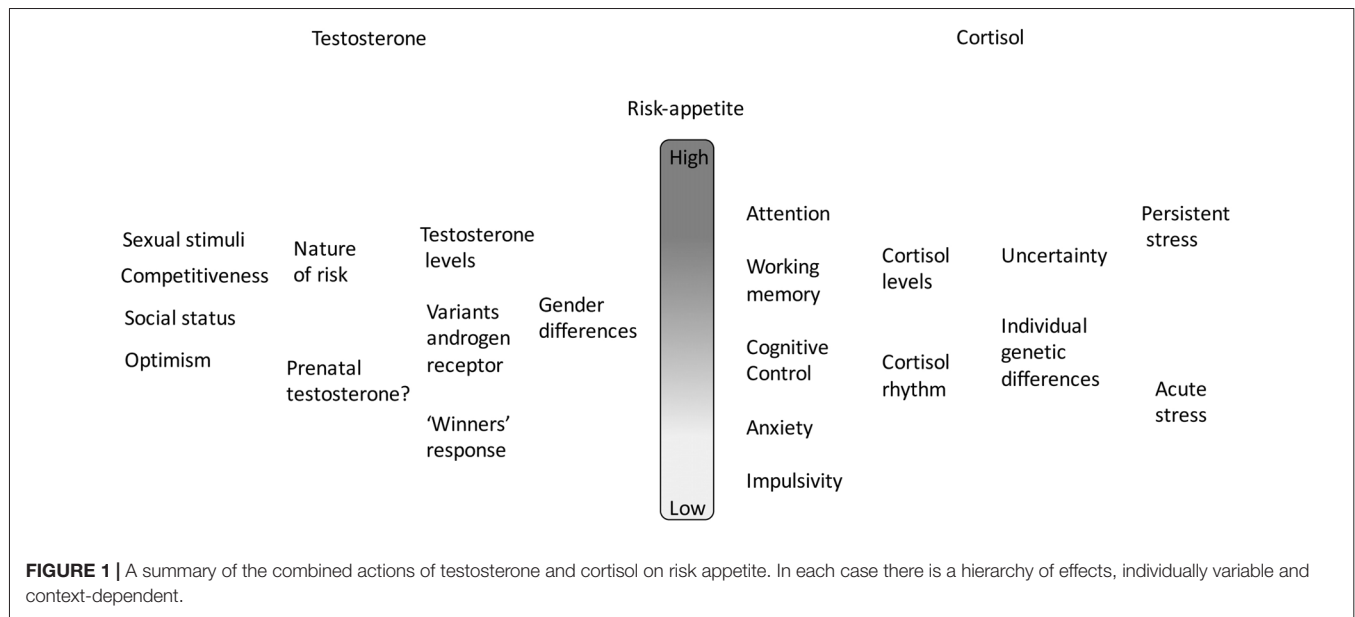
The tendency to act on impulse, characterized by little reflection or consideration of possible consequences, and its influence on risky decisions, has already been mentioned. Another aspect

is temporal discounting, the tendency to accept an immediate financial reward rather than a delayed, but greater, one. One measure of this is to increase the value of the delayed reward until the individual switches choices. The difference between this value and the immediate one is an index of temporal discounting, or impulsivity. This has to exclude circumstances that might make an immediate reward necessary (e.g., to settle a debt). Attention-deficient hyperactivity (ADHD) is a common developmental disorder characterized by impulsivity and is associated with greater risk-taking (Blomqvist et al., 2007).

Both cortisol and testosterone have been implicated in the control of impulsivity. Several studies show that the general trait of impulsivity—but particularly related to aggression—is associated with lower basal cortisol levels and a reduced response to stress (Blomqvist et al., 2007; Flegr et al., 2012; Lovallo, 2013; Brown et al., 2016). Lower levels of cortisol predicted temporal discounting in males, but this was opposite in women (Takahashi et al., 2010). Increased testosterone, or reduced cortisol/testosterone ratio, has been related to low impulse control (Pavlov et al., 2012), but rats treated with testosterone chose a larger, delayed reward compared to controls (Wood et al., 2013). However, higher testosterone was related to increased temporal discounting in males, though the opposite was recorded in women (Doi et al., 2015). This result in males is at odds with other reports linking higher testosterone with higher sensation-seeking, aggression and harmful risk-taking, though it has been suggested that impulsivity is actually a complex trait with different components (Reynolds et al., 2006; Bari and Robbins, 2013).

There is an extensive literature on the role of serotonin (but also dopamine) in impulsive behavior and the consequences this has for decision-making (Dalley and Roiser, 2012; Homberg, 2012; Bari and Robbins, 2013). There is an equivalent literature on the regulation of serotonin by cortisol (Chaouloff, 2000; Joels, 2011). Corticoids moderate the activity of tryptophan hydroxylase and thus the synthesis of serotonin, as well as the activity of several of its receptors (Hanley and Van de Kar, 2003; Mueller et al., 2011). There has been little study on whether genetic variants in serotonin-related genes could contribute to financial impulsivity, though low expression variants of the serotonin transporter (hSERT) or reduced cerebral concentrations of serotonin have been associated with an increased tendency for impulsive behavior in other contexts (Walderhaug et al., 2008; Pavlov et al., 2012; Cha et al., 2017).

Another example of socially-relevant behavior that influences risky economic decisions is “herding”, the tendency for individuals to follow a leader or trend without question. In situations of uncertainty, rational choices can be made following principles of statistical inference using Bayesian approaches and such explanations for herding lie in scenarios in which different individuals’ decisions are interdependent and reinforcing. However, a more complete approach takes into account a range of other factors from social psychology, neuroscience and even evolutionary biology (Baddeley, 2009). Herding is seen in many other species: deer run if one member of the group is startled without waiting to see the cause; if one bird takes off, the rest of the flock may



well follow almost instantly. In these instances, herding is advantageous. There may be occasions when this is also true in financial contexts (“rational herding” (Devenow and Welch, 1996)); for example, when a small number of participants, or a prominent leader, really do have private information of value.

There are no studies on the effect of hormones on the tendency to herd in a financial context, but empirical observations suggest this is more likely to occur under conditions of stress or market uncertainty, particularly in individuals of lower cognitive ability and those susceptible to “framing” effects (Tversky and Kahneman, 1974, 1981; Kahneman and Tversky, 1979; Devenow and Welch, 1996; Baddeley, 2009; Zheng et al., 2010). These, as we have seen, are exactly the conditions that result in heightened secretion of cortisol: and the effects this might have on anxiety, risk-perception etc could easily be translated into an increased tendency for herding behavior, and hence market de-stabilization. Testosterone, it seems, might also have an action on the tendency to herd. If the digit ratio is accepted as an index of prenatal exposure (but see above) then a lower ratio (male-like) might encourage a more deliberate strategy (and less imitation), and adult levels greater abstract reasoning ability (Brañas-Garza and Rustichini, 2011; Bosch-Domènech et al., 2014). However, there is a marked tendency for the males of many species (including humans) to act collectively if their group is attacked or challenged. Thus the males of a group of monkeys will combine to repel an invasion of their territory by another group, putting aside intra-group competition or rank (Wrangham and Glowacki, 2012). This is a form of herding, though whether it is a direct consequent of the actions of testosterone remains possible but speculative. An fMRI study suggested that the amygdala, well-known as important for emotion and sensitive to both cortisol and testosterone, might be implicated in individual tendencies to herd (Baddeley et al., 2012). We should not

forget that other hormones may play a role, including oxytocin, which influences “bonding” between individuals, and hence the tendency to imitate or follow an example (Panksepp, 1992; Olff et al., 2013).

The actions of both testosterone and cortisol on risk-appetite are summarized in **Figure 1**.

## MAPPING THE RESPONSES IN THE BRAIN

Mapping the actions of these hormones onto the brain presents many problems. There are differences in the distribution of androgen and corticoid receptors in the brain. Androgen receptors are located mostly in limbic structures, such as the hypothalamus, amygdala and hippocampus, though there are lesser concentrations in the brainstem and deeper layers of the cerebral cortex (Simerly et al., 1990). This points to the major sites of action of testosterone on areas known to be concerned with emotion and motivation. By contrast, glucocorticoid receptors are more widely distributed, including not only limbic structures but also the cerebral and cerebellar cortices, and brain stem nuclei (e.g., those expressing serotonin or noradrenaline; Morimoto et al., 1996). This implies a different pattern of neuronal activation or inhibition which would include both emotional and cognitive functions.

It has already been pointed out that testosterone, even though its receptors are concentrated in the limbic areas, has to influence many aspects of behavior other than sexual activity in order to fulfil its primary reproductive function (e.g., aggressiveness, competitiveness, risk-taking). This variety will be reflected in the way that testosterone influences financial decisions: the effect may vary with the situation. For example, presenting sexually-related stimuli may affect decisions and their associated risks by distinct neural mechanisms, which may vary in different individuals and from situations that are more competitive or threatening (Herbert, 2017).

There is a conundrum about the role of testosterone in the brain. One way in which risk-taking varies within an individual according to context or between individuals in the same context is related to the value of the reward on offer. Most current evidence places the brain areas that respond to, anticipate, or evaluate reward in the ventral striatum, its dopaminergic innervation, or the orbital (OFC), anterior cingulate or parietal cortex (Schultz, 2004; Hsu et al., 2009; Kang et al., 2009; Kahnt et al., 2010; Louie et al., 2011; Soutschek et al., 2017). None of these forebrain areas is notable for high concentrations of androgen receptors (Rubinow and Schmidt, 1996), though they have been discerned in the midbrain dopaminergic neurons of humans and rats (Aubele and Kritzer, 2012; Morris et al., 2015). If testosterone is to bias the reward system, then there must be a link between this system and the areas of the brain (e.g., amygdala, hypothalamus, septum) responding to testosterone, and its influence on midbrain dopaminergic neurons might be one way for this to happen. There is some experimental evidence suggesting that testosterone can modulate dopaminergic activity (Purves-Tyson et al., 2014) though whether this accounts for all its actions on reward remains uncertain. This also applies to the principal action of testosterone on sexual behavior or motivation, which, as we have seen, may influence financial risk-taking. Emotion and cognition are closely interwoven, so there must be a corresponding neural representation of this association (Okon-Singer et al., 2015). Though profuse connections between, for example, the amygdala and OFC are known (Cavada et al., 2000), there is as yet no coherent account of how these bias the reward system.

The glucocorticoid receptors, having a wider distribution in the brain than androgen receptors, enable cortisol to access directly a wider neural network, hence its more general actions on cognitive and emotional functions associated with risk. But this raises questions about which particular function will predominate in a given financial situation.

A second problem is how much of the experimental work on the neural mechanisms underlying reward and choice, or the effects that stress or hormones have on these behaviors, have direct relevance to financial decisions and their associated risks humans (see above). The use of money as an asset involves cognitive and emotional processes that are not really observable in animals. Studies on the latter rely on primary rewards, such as food or palatable juice (Schultz, 2016). So much of the information on humans has to come either from studies on those with defined areas of damage to the brain, or on techniques, such as scanning, that give limited information on neural function and the way it varies both in different contexts and between individuals.

A third difficulty is that the process of risk assessment and subsequent decision-making involves a series of neural processes (as already mentioned). The perception, processing and assessment of information concerning the nature of the decision will involve several regions of the brain. Estimation of the reward value of success, or the consequences of failure involves a further process. Then comes the emotional response to the perceived risk or anticipation of success or failure. All this takes place on the background of neural states representing

personality, learning, experience and knowledge of the context in which the decision is made (see above). Each stage is potentially sensitive either directly to these steroids or indirectly to their action elsewhere in the brain. Nevertheless, we would expect the actions of either testosterone or cortisol on financial decisions to reflect their primary functions: for testosterone, its central role in promoting reproductive success; for cortisol, its role in coping with stress.

The amount of information on regions of the brain involved in risk assessment and decision-making is too large to allow anything more than a summary here, with particular emphasis on whether it sheds light on the action of either testosterone or cortisol on financial risk-taking. It is generally agreed that the prefrontal cortex and its associated connections with the striatum (and its dopaminergic innervation) play a central part in recognizing risk, and deciding what action to take (Hsu et al., 2005; Holper et al., 2014; Goh et al., 2016; Ouerchefani et al., 2017). Acute stress activates a neural network that includes fronto-insular, dorsal anterior cingulate, inferior-temporal, and temporo-parietal and amygdala, thalamus, hypothalamus and midbrain (Hermans et al., 2011). The anterior insular cortex, strongly implicated in emotional expression, and with plentiful connections to the limbic brain, is also activated by risk (Mohr et al., 2010).

Risk-taking implies uncertainty about outcome. fMRI studies have suggested separate brain areas that react to uncertainty (e.g., the amygdala and orbital frontal lobe) and expected reward or its valuation (the striatum; Hsu et al., 2005, 2009; Christopoulos et al., 2009; Tobler et al., 2009; Burke and Tobler, 2011). Direct action of corticoids has been implicated in the impairment of the frontal lobes by stress (McKlveen et al., 2013, 2016); this may include alterations in dopamine release, and hence the signaling of either reward or reward errors (Butts and Phillips, 2013). Serotonin neurons also respond to reward, but differently from dopaminergic ones: dopamine may signal the relative value of a reward, whereas serotonin neurons signal its absolute value, and are inhibited by stress (Zhong et al., 2017). Though cortisol has not been directly implicated, as already mentioned there is an extensive literature on the regulation of serotonin in the brain by corticoids (Chaouloff, 2000).

Perceptual learning, and hence appraisal of risk, may also be impaired (Dinse et al., 2017). Testosterone, either acting directly or indirectly, by contrast alters the activity of the anterior insula and inferior frontal lobe (more closely associated with emotional states and the integration of risk with returns, respectively), and this is associated with increased risk taking (Tobler et al., 2009; Burke and Tobler, 2011). Both effects were moderated by genetic variants of MAOA (Wagels et al., 2017), a gene implicated in impulsivity and aggression (Dorfman et al., 2014). However, the blurred boundary between cognition and emotion is emphasized by the fact that the ventral prefrontal cortex is also concerned with reward (Juechems et al., 2017).

But the frontal lobes are not the only part of the cortex implicated in risky decision-making. The cingulate cortex, insula, temporo-parietal lobe as well as subcortical areas (e.g., ventral striatum), may respond to value according to the way it is assessed or objective features of choice alternatives (Clithero

et al., 2009; Fitzgerald et al., 2010; Kahnt et al., 2010; Kahnt and Tobler, 2013). Age-related changes in the parietal cortex have been associated with age-dependent changes in risk perception (Grubb et al., 2016) and with the time-related processing of uncertain information (de Lange et al., 2010; Bode et al., 2012). All these areas (particularly the frontal lobes) have plentiful connections with subcortical structures such as the amygdala. Stress could therefore impair the process of decision-making by actions on this system, rather than on individual components of it (Maier et al., 2015).

The amygdala has been implicated in both cognitive and emotional components of risk-taking, another example of the blurred boundary between them (Bhatt et al., 2012). Since there is a profusion of androgen and glucocorticoid receptors in the amygdala, this is one avenue by which either hormone could influence financial risk-taking in a variety of ways, including the influence of testosterone on estimations of trustworthiness (in women; Bos et al., 2012). The amygdala is concerned with the regulation of loss aversion (Sokol-Hessner et al., 2013), and damage to it reduces this aversion though without impairing the ability to recognize changes in monetary value (De Martino et al., 2010). It can only be surmised that testosterone, which has a similar action, may operate through the amygdala and its connections with the orbital frontal cortex. Prediction of outcomes, and hence the risk associated with them, is also a function of this system (Dolan, 2007); there is as yet no clear evidence that either steroid alters the way this information is obtained or used, though since both alter risk appetite and, in the case of testosterone, estimates of expected outcome, it is highly likely that information processing is affected. Incidentally, although the hippocampus has high concentrations of glucocorticoid receptors (Gray et al., 2017), it has not, so far, been implicated in neural processes affecting risk appetite. Corticoids have a pronounced suppressive action on the formation of new neurons in the hippocampus (Cameron and Gould, 1994; Pinnock et al., 2007), though whether this influences financial decisions in the longer-term is also unknown.

## FINANCIAL DECISIONS AS CONFLICT

As already mentioned, there are striking parallels between the modern situation in which acute and highly significant decisions have to be taken in a financial context (e.g., by day traders, who are mostly male) and an older biological one in which males are required to make equally rapid decisions in the context of personal competition (for mates) or collaborative conflict (war). In both, the outcome of a wrong decision may be either personal loss (finance: money; conflict: loss of assets, wounds, death), whereas success brings not only personal gain but social acclaim and heightened status, or gain of corporate assets (finance: profits for the company, conflict: territory, access to mates and other assets). In both situations, current information on which decisions are made or risks taken is likely to be complex, rapidly changing and incomplete. In both, experience and temperament will contribute to the behavioral response to a current acute and risk-laden situation. Whilst these considerations apply most

obviously to rapid and possibly life-changing decisions in both contexts, more deliberate assessment of risks also occur in both conflicts and finance; for example, decisions on strategy, usually taken by older males (generals in war, managers in finance) than those who do the trading or the fighting. Much of the information on the factors that guide decisions made under the more primeval conditions of conflict will also apply to the more modern situations of finance. For example, testosterone reduces males' tendency to reflect on decisions, a property which might be advantageous during fights as well as bond trading (Nave et al., 2017). It also increases confrontational decisions in a competitive financial encounter (Mehta et al., 2017). Though many studies focus on one or other steroid, it should be noted that both testosterone and cortisol do not act alone, but in the context of many factors, including interactions between the two hormones themselves. For example, the action of testosterone may depend on coincident levels or changes in cortisol, and vice versa (Mehta and Josephs, 2010; Mehta et al., 2015). Both testosterone and cortisol, the former implicated in the (male) involvement in competition, aggression and war, the latter in the stress response to urgent need and demand, and the areas of the brain on which they act, will thus play roles in finance that are foreshadowed by a more ancient biological imperative (Herbert, 2017).

Comparing physical conflicts with financial exchanges suggests another parallel: that a beneficial outcome for a group may not always be the same as for an individual, and this will affect not only processing of information, risk-assessment and decision-making, but also the individual endocrine response to a given situation. For example, in war it may be that the sacrifice of an individual works to the group's advantage; in financial terms, risk-taking by an individual, though detrimental to that individual's success, may provide information that benefits the group (company). This may be reflected in the function of both testosterone and cortisol. For example, since testosterone increases risk-appetite, it may be that there are situations in which this is related to in over-ambitious actions that result in individual loss, but future gain for the group. Group interactions, as well as personal characteristics, will therefore influence risk-taking. Similar ideas apply to cortisol. Excessive stress may impair individual performance, but provide corporate benefits in terms of heightened caution or inter-personal learning. It is thus difficult to define "optimal" levels of either hormone: this will depend both on the qualities of the individual and of the group of which he is a member (females will need a separate analysis). That is not to say that non-optimal levels of either cortisol or testosterone may not occur, and which contribute to disadvantageous outcomes both for the individual and the group. The lack of information on hormonal responses and correlations in real-life situations, and ignorance about the background on which they act (context, experience, personality, genetic variations etc.) means that we are currently unable to assess these factors with any certainty.

## NEUROSCIENCE VS. ECONOMICS

The focus of neuroscience is primarily on the role of hormones in the way that individuals respond to financial risks. This



includes a wide range of related disciplines associated with decision-making, including psychological and social factors that influence such decisions. Neuroscience is thus concerned mostly with individual variation in risk assessment and decisions consequent on this neural process. Economists, on the other hand, are primarily interested in the way this affects the price of assets, and the occurrence of bubbles and crashes (i.e., market stability). Their concern is not so much with individuals and how they might vary, but with the results that corporate decisions might have on the market. The recent realization that there is considerable overlap between the two approaches has given rise to the relatively new topic of neuroeconomics (Camerer and Fehr, 2006; Camerer, 2008).

An example is the action of either testosterone or cortisol on financial decisions. This will have a median (average) effect on individuals, but modulated by genetic constitution, early and recent experience, and social context. Whereas individual behavior is unlikely to influence asset prices, large-scale median action may well do so. This may be one result of a general effect on, say, assessment of risk or optimism about outcome, but also on socially-determined responses such as “herding” (see above). It is important to distinguish simultaneous actions, prompted by equivalent information, from concerted actions that occur in the absence of new information or even despite it, driven by imitation or false (irrational) belief in private information held by others (herding). This may result in a cascade in which progressively more members of a particular financial community (e.g., a trading floor) follow each other (Bikhchandani and Sharma, 2001).

Collective decisions made independently, but influenced overall by either testosterone or cortisol (or both), may also have a de-stabilizing effect on markets. It is therefore relevant that those concerned with trading should pay attention to the effects these two steroids (as well as the numerous other factors) have on individual or collective responses to a given market situation. Alterations in biases, emotions, risk-assessment and cognitive appraisal (Kahneman, 2011), all influenced by hormones, can be powerful drivers of markets. But they will not necessarily be the same in everyone, as repeatedly emphasized in this article. So, in addition to knowledge

about the overall effects of hormones, the financial world also needs to understand how these may be moderated individually. Despite the current interest in neuroeconomics, financiers would do well to take greater interest in the way that individual decisions are made, including the powerful effects of hormones and their actions on emotion and cognition, whereas neuroscience needs to understand better the impact on the financial world of risk-laden decisions taken under duress and the consequences these may have for an economy.

## CONCLUSION

Much of this review has been concerned with experimental or laboratory studies on the role of testosterone or cortisol in risky financial decisions. Though these have been, to an extent, informative, there is a great need for two further lines of enquiry: studies on the effects of either hormone in real-life situations, difficult but not impossible, and the contribution that individual genetic variations make to the effects that either hormone has in situations in which they may play a part, or to propensities for individuals to engage in risky financial behavior either as a profession or in everyday life. Although it is always possible to characterize the roles of hormones on the basis of mean or median effects, another aspect of equal interest is the extent to which the financial behavior of individuals varies in their response to their own hormones, and the ways this comes about.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## ACKNOWLEDGMENTS

I am deeply grateful to my colleagues Phillipe Tobler (University of Zurich), Wolfram Schultz (University of Cambridge) and Raghavendra (Raghu) Rau (University of Cambridge) for their help with this article, and to several referees for some thoughtful comments.

## REFERENCES

- Abercrombie, H. C., Kalin, N. H., and Davidson, R. J. (2005). Acute cortisol elevations cause heightened arousal ratings of objectively nonarousing stimuli. *Emotion* 5, 354–359. doi: 10.1037/1528-3542.5.3.354
- Abercrombie, H. C., Kalin, N. H., Thuro, M. E., Rosenkranz, M. A., and Davidson, R. J. (2003). Cortisol variation in humans affects memory for emotionally laden and neutral information. *Behav. Neurosci.* 117, 505–516. doi: 10.1037/0735-7044.117.3.505
- Akinola, M., Fridman, I., Mor, S., Morris, M. W., and Crum, A. J. (2016). Adaptive appraisals of anxiety moderate the association between cortisol reactivity and performance in salary negotiations. *PLoS One* 11:e0167977. doi: 10.1371/journal.pone.0167977
- Alarcón, G., Cservénka, A., and Nagel, B. J. (2017). Adolescent neural response to reward is related to participant sex and task motivation. *Brain Cogn.* 111, 51–62. doi: 10.1016/j.bandc.2016.10.003
- Alikhani-Koupaei, R., Fouladkou, F., Fustier, P., Cenni, B., Sharma, A. M., Deter, H. C., et al. (2007). Identification of polymorphisms in the human 11 $\beta$ -hydroxysteroid dehydrogenase type 2 gene promoter: functional characterization and relevance for salt sensitivity. *FASEB J.* 21, 3618–3628. doi: 10.1096/fj.07-8140com
- Apicella, C. L., Carré, J. M., and Dreber, A. (2015). Testosterone and economic risk taking: a review. *Adapt. Hum. Behav. Physiol.* 1, 358–385. doi: 10.1007/s40750-014-0020-2
- 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
- Archer, J. (2006). Testosterone and human aggression: an evaluation of the challenge hypothesis. *Neurosci. Biobehav. Rev.* 30, 319–345. doi: 10.1016/j.neubiorev.2004.12.007
- Aubele, T., and Kritzer, M. F. (2012). Androgen influence on prefrontal dopamine systems in adult male rats: localization of cognate intracellular receptors in medial prefrontal projections to the ventral tegmental area and effects of gonadectomy and hormone replacement on glutamate-stimulated extracellular dopamine level. *Cereb. Cortex* 22, 1799–1812. doi: 10.1093/cercor/bhr258
- Auyeung, B., Wheelwright, S., Allison, C., Atkinson, M., Samarawickrema, N., and Baron-Cohen, S. (2009). The children’s empathy quotient and systemizing

- quotient: sex differences in typical development and in autism spectrum conditions. *J. Autism Dev. Disord.* 39, 1509–1521. doi: 10.1007/s10803-009-0772-x
- Baddeley, M. (2009). Herding, social influence and economic decision-making: socio-psychological and neuroscientific analyses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 281–290. doi: 10.1098/rstb.2009.0169
- Baddeley, M., Burke, C. J., Schultz, W., and Tobler, P. N. (2012). Herding in financial behaviour: a behavioural and neuroeconomic analysis of individual differences. Available online at: <http://www.econ.cam.ac.uk/research-files/repec/cam/pdf/cwpe1225.pdf>
- Bailey, M., and Silver, R. (2014). Sex differences in circadian timing systems: implications for disease. *Front. Neuroendocrinol.* 35, 111–139. doi: 10.1016/j.yfrne.2013.11.003
- Baker, M. D., and Maner, J. K. (2008). Risk-taking as a situationally sensitive male mating strategy. *Evol. Hum. Behav.* 29, 391–395. doi: 10.1016/j.evolhumbehav.2008.06.001
- Bari, A., and Robbins, T. W. (2013). Inhibition and impulsivity: behavioral and neural basis of response control. *Prog. Neurobiol.* 108, 44–79. doi: 10.1016/j.pneurobio.2013.06.005
- Bell, M. R., and Sisk, C. L. (2013). Dopamine mediates testosterone-induced social reward in male Syrian hamsters. *Endocrinology* 154, 1225–1234. doi: 10.1210/en.2012-2042
- Bendahan, S., Goette, L., Thoresen, J., Loued-Khenissi, L., Hollis, F., and Sandi, C. (2017). Acute stress alters individual risk taking in a time-dependent manner and leads to anti-social risk. *Eur. J. Neurosci.* 45, 877–885. doi: 10.1111/ejn.13395
- Bhatt, M. A., Lohrenz, T., Camerer, C. F., and Montague, P. R. (2012). Distinct contributions of the amygdala and parahippocampal gyrus to suspicion in a repeated bargaining game. *Proc. Natl. Acad. Sci. U S A* 109, 8728–8733. doi: 10.1073/pnas.1200738109
- Bikhchandani, S., and Sharma, S. (2001). Herd behavior in financial markets. *IMF Staff Pap.* 47, 279–310.
- Birn, R. M., Roeber, B. J., and Pollak, S. D. (2017). Early childhood stress exposure, reward pathways and adult decision making. *Proc. Natl. Acad. Sci. U S A* 114, 13549–13554. doi: 10.1073/pnas.1708791114
- Blomqvist, M., Holmberg, K., Lindblad, F., Fernell, E., Ek, U., and Dahllöf, G. (2007). Salivary cortisol levels and dental anxiety in children with attention deficit hyperactivity disorder. *Eur. J. Oral Sci.* 115, 1–6. doi: 10.1111/j.1600-0722.2007.00423.x
- Bode, S., Bogler, C., Soon, C. S., and Haynes, J. D. (2012). The neural encoding of guesses in the human brain. *Neuroimage* 59, 1924–1931. doi: 10.1016/j.neuroimage.2011.08.106
- Booth, A., Shelley, G., Mazur, A., Tharp, G., and Kittock, R. (1989). Testosterone, and winning and losing in human competition. *Horm. Behav.* 23, 556–571. doi: 10.1016/0018-506x(89)90042-1
- Bos, P. A., Hermans, E. J., Ramsey, N. F., and van Honk, J. (2012). The neural mechanisms by which testosterone acts on interpersonal trust. *Neuroimage* 61, 730–737. doi: 10.1016/j.neuroimage.2012.04.002
- Bosch-Domènech, A., Brañas-Garza, P., and Espín, A. M. (2014). Can exposure to prenatal sex hormones (2D:4D) predict cognitive reflection? *Psychoneuroendocrinology* 43, 1–10. doi: 10.1016/j.psyneuen.2014.01.023
- Branas-Garza, P., Galizzi, M. M., and Nieboer, J. (2018). Experimental and self-reported measures of risk taking and digit ratio (2D:4D): evidence from a large, systematic study. *Int. Econ. Rev.* Available online at: <https://onlinelibrary.wiley.com/doi/abs/10.1111/iere.12299>
- 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
- Breedlove, S. M. (2010). Minireview: organizational hypothesis: instances of the fingerpost. *Endocrinology* 151, 4116–4122. doi: 10.1210/en.2010-0041
- Brown, T. G., Ouimet, M. C., Eldeb, M., Tremblay, J., Vingilis, E., Nadeau, L., et al. (2016). Personality, executive control, and neurobiological characteristics associated with different forms of risky driving. *PLoS One* 11:e0150227. doi: 10.1371/journal.pone.0150227
- Brynes, J. P., Miller, D. C., and Schafer, W. D. (1999). Gender differences in risk taking: a meta-analysis. *Psychol. Bull.* 125, 367–383. doi: 10.1037/0033-2909.125.3.367
- Buckert, M., Schwieren, C., Kudiella, B. M., and Fiebach, C. J. (2014). Acute stress affects risk taking but not ambiguity aversion. *Front. Neurosci.* 8:82. doi: 10.3389/fnins.2014.00082
- Burke, C. J., and Tobler, P. N. (2011). Reward skewness coding in the insula independent of probability and loss. *J. Neurophysiol.* 106, 2415–2422. doi: 10.1152/jn.00471.2011
- Burnham, T. C. (2007). High-testosterone men reject low ultimatum game offers. *Proc. Biol. Sci.* 274, 2327–2330. doi: 10.1098/rspb.2007.0546
- Bustamante, A. C., Aiello, A. E., Galea, S., Ratanatharathorn, A., Noronha, C., Wildman, D. E., et al. (2016). Glucocorticoid receptor DNA methylation, childhood maltreatment and major depression. *J. Affect. Disord.* 206, 181–188. doi: 10.1016/j.jad.2016.07.038
- Butts, K. A., and Phillips, A. G. (2013). Glucocorticoid receptors in the prefrontal cortex regulate dopamine efflux to stress via descending glutamatergic feedback to the ventral tegmental area. *Int. J. Neuropsychopharmacol.* 16, 1799–1807. doi: 10.1017/s1461145713000187
- Camerer, C. F. (2008). Neuroeconomics: opening the gray box. *Neuron* 60, 416–419. doi: 10.1016/j.neuron.2008.10.027
- Camerer, C. F., and Fehr, E. (2006). When does “economic man” dominate social behavior? *Science* 311, 47–52. doi: 10.1126/science.1110600
- Cameron, H. A., and Gould, E. (1994). Adult neurogenesis is regulated by adrenal steroids in the dentate gyrus. *Neuroscience* 61, 203–209. doi: 10.1016/0306-4522(94)90224-0
- Canale, N., Rubaltelli, E., Vieno, A., Pittarello, A., and Billieux, J. (2017). Impulsivity influences betting under stress in laboratory gambling. *Sci. Rep.* 7:10668. doi: 10.1038/s41598-017-10745-9
- Cardoos, S. L., Ballonoff Suleiman, A., Johnson, M., van den Bos, W., Hinshaw, S. P., and Dahl, R. E. (2017). Social status strategy in early adolescent girls: testosterone and value-based decision making. *Psychoneuroendocrinology* 81, 14–21. doi: 10.1016/j.psyneuen.2017.03.013
- Carney, D. R., and Mason, M. F. (2010). Decision making and testosterone: when the ends justify the means. *J. Exp. Soc. Psychol.* 46, 668–671. doi: 10.1016/j.jesp.2010.02.003
- Cavada, C., Compañy, T., Tejedor, J., Cruz-Rizzolo, R. J., and Reinoso-Suárez, F. (2000). The anatomical connections of the macaque monkey orbitofrontal cortex. A review. *Cereb. Cortex* 10, 220–242. doi: 10.1093/cercor/10.3.220
- Chan, E. Y. (2015). Physically-attractive males increase men’s financial risk-taking. *Evol. Hum. Behav.* 36, 407–413. doi: 10.1016/j.evolhumbehav.2015.03.005
- Cha, J., Guffanti, G., Gingrich, J., Talati, A., Wickramaratne, P., Weissman, M., et al. (2017). Effects of serotonin transporter gene variation on impulsivity mediated by default mode network: a family study of depression. *Cereb. Cortex* doi: 10.1093/cercor/bhx097 [Epub ahead of print].
- Chauloff, F. (2000). Serotonin, stress and corticoids. *J. Psychopharmacol.* 14, 139–151. doi: 10.1177/026988110001400414
- Christopoulos, G. I., Tobler, P. N., Bossaerts, P., Dolan, R. J., and Schultz, W. (2009). Neural correlates of value, risk and risk aversion contributing to decision making under risk. *J. Neurosci.* 29, 12574–12583. doi: 10.1523/JNEUROSCI.2614-09.2009
- Claessens, F., Joniau, S., and Helsen, C. (2017). Comparing the rules of engagement of androgen and glucocorticoid receptors. *Cell. Mol. Life Sci.* 74, 2217–2228. doi: 10.1007/s00018-017-2467-3
- Clithero, J. A., Carter, R. M., and Huettel, S. A. (2009). Local pattern classification differentiates processes of economic valuation. *Neuroimage* 45, 1329–1338. doi: 10.1016/j.neuroimage.2008.12.074
- Coates, J. M., and Herbert, J. (2008). Endogenous steroids and financial risk taking on a London trading floor. *Proc. Natl. Acad. Sci. U S A* 105, 6167–6172. doi: 10.1073/pnas.0704025105
- Cueva, C., Roberts, R. E., Spencer, T., Rani, N., Tempest, M., Tobler, P. N., et al. (2015). Cortisol and testosterone increase financial risk taking and may destabilize markets. *Sci. Rep.* 5:11206. doi: 10.1038/srep11206
- D’Acromont, M., and Bossaerts, P. (2008). Neurobiological studies of risk assessment: a comparison of expected utility and mean-variance approaches. *Cogn. Affect. Behav. Neurosci.* 8, 363–374. doi: 10.3758/CABN.8.4.363
- Dalley, J. W., and Roiser, J. P. (2012). Dopamine, serotonin and impulsivity. *Neuroscience* 215, 42–58. doi: 10.1016/j.neuroscience.2012.03.065
- de Lange, F. P., Jensen, O., and Dehaene, S. (2010). Accumulation of evidence during sequential decision making: the importance of top-down factors. *J. Neurosci.* 30, 731–738. doi: 10.1523/JNEUROSCI.4080-09.2010

- De Martino, B., Camerer, C. F., and Adolphs, R. (2010). Amygdala damage eliminates monetary loss aversion. *Proc. Natl. Acad. Sci. U S A* 107, 3788–3792. doi: 10.1073/pnas.0910230107
- Devenow, A., and Welch, I. (1996). Rational herding in financial economics. *Eur. Econ. Rev.* 40, 603–615. doi: 10.1016/0014-2921(95)00073-9
- Dinse, H. R., Kattenstroth, J. C., Lenz, M., Tegenthoff, M., and Wolf, O. T. (2017). The stress hormone cortisol blocks perceptual learning in humans. *Psychoneuroendocrinology* 77, 63–67. doi: 10.1016/j.psyneuen.2016.12.002
- Doi, H., Nishitani, S., and Shinohara, K. (2015). Sex difference in the relationship between salivary testosterone and inter-temporal choice. *Horm. Behav.* 69, 50–58. doi: 10.1016/j.yhbeh.2014.12.005
- Dolan, R. J. (2007). The human amygdala and orbital prefrontal cortex in behavioural regulation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 787–799. doi: 10.1098/rstb.2007.2088
- Dorfman, H. M., Meyer-Lindenberg, A., and Buckholtz, J. W. (2014). Neurobiological mechanisms for impulsive-aggression: the role of MAOA. *Curr. Top. Behav. Neurosci.* 17, 297–313. doi: 10.1007/7854\_2013\_272
- Eckel, C. E., and Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evout. Hum. Behav.* 23, 281–295. doi: 10.1016/s1090-5138(02)00097-1
- Ferry, B., Roozendaal, B., and McGaugh, J. L. (1999). Role of norepinephrine in mediating stress hormone regulation of long-term memory storage: a critical involvement of the amygdala. *Biol. Psychiatry* 46, 1140–1152. doi: 10.1016/s0006-3223(99)00157-2
- Finkelstein, J. S., Yu, E. W., and Burnett-Bowie, S. A. (2013). Gonadal steroids and body composition, strength, and sexual function in men. *N. Engl. J. Med.* 369:2457. doi: 10.1056/NEJM1313169
- Fitzgerald, T. H., Seymour, B., Bach, D. R., and Dolan, R. J. (2010). Differentiable neural substrates for learned and described value and risk. *Curr. Biol.* 20, 1823–1829. doi: 10.1016/j.cub.2010.08.048
- Flegr, J., Hampl, R., Černochová, D., Preiss, M., Bičíková, M., Sieger, L., et al. (2012). The relation of cortisol and sex hormone levels to results of psychological, performance, IQ and memory tests in military men and women. *Neuro Endocrinol. Lett.* 33, 224–235.
- Friedel, E., Sebold, M., Kuitunen-Paul, S., Nebe, S., Veer, I. M., Zimmermann, U. S., et al. (2017). How accumulated real life stress experience and cognitive speed interact on decision-making processes. *Front. Hum. Neurosci.* 11:302. doi: 10.3389/fnhum.2017.00302
- Genest, W., Stauffer, W. R., and Schultz, W. (2016). Utility functions predict variance and skewness risk preferences in monkeys. *Proc. Natl. Acad. Sci. U S A* 113, 8402–8407. doi: 10.1073/pnas.1602217113
- George, C. L., Birnie, M. T., Flynn, B. P., Kershaw, Y. M., Lightman, S. L., and Conway-Campbell, B. L. (2017). Ultradian glucocorticoid exposure directs gene-dependent and tissue-specific mRNA expression patterns *in vivo*. *Mol. Cell. Endocrinol.* 439, 46–53. doi: 10.1016/j.mce.2016.10.019
- Goetz, S. M., Tang, L., Thomason, M. E., Diamond, M. P., Hariri, A. R., and Carre, J. M. (2014). Testosterone rapidly increases neural reactivity to threat in healthy men: a novel two-step pharmacological challenge paradigm. *Biol. Psychiatry* 76, 324–331. doi: 10.1016/j.biopsych.2014.01.016
- Goh, J. O., Su, Y. S., Tang, Y. J., McCarrey, A. C., Tereshchenko, A., Elkins, W., et al. (2016). Frontal, striatal, and medial temporal sensitivity to value distinguishes risk-taking from risk-averse older adults during decision making. *J. Neurosci.* 36, 12498–12509. doi: 10.1523/JNEUROSCI.1386-16.2016
- Gray, J. D., Kogan, J. F., Marrocco, J., and McEwen, B. S. (2017). Genomic and epigenomic mechanisms of glucocorticoids in the brain. *Nat. Rev. Endocrinol.* 13, 661–673. doi: 10.1038/nrendo.2017.97
- Grubb, M. A., Tymula, A., Gilaie-Dotan, S., Glimcher, P. W., and Levy, I. (2016). Neuroanatomy accounts for age-related changes in risk preferences. *Nat. Commun.* 7:13822. doi: 10.1038/ncomms13822
- Hanley, N. R., and Van de Kar, L. D. (2003). Serotonin and the neuroendocrine regulation of the hypothalamic–pituitary–adrenal axis in health and disease. *Vitam. Horm.* 66, 189–255. doi: 10.1016/s0083-6729(03)01006-9
- Harris, C. R., and Jenkins, M. (2006). Gender differences in risk assessment: why do women take fewer risks than men? *Judgm. Decis. Mak.* 1, 48–63.
- Henckens, M. J. A. G., van Wingen, G. A., Joëls, M., and Fernández, G. (2010). Time-dependent effects of corticosteroids on human amygdala processing. *J. Neurosci.* 30, 12725–12732. doi: 10.1523/JNEUROSCI.3112-10.2010
- Herbert, J. (2013). Cortisol and depression: three questions for psychiatry. *Psychol. Med.* 43, 449–469. doi: 10.1017/s0033291712000955
- Herbert, J. (2017). *Testosterone: The Molecule Behind Power, Sex and the Will to Win*. Oxford: Oxford University Press.
- Herbert, J., Goodyer, I. M., Grossman, A. B., Hastings, M. H., de Kloet, E. R., Lightman, S. L., et al. (2006). Do corticosteroids damage the brain? *J. Neuroendocrinol.* 18, 393–411. doi: 10.1111/j.1365-2826.2006.01429.x
- Herbert, J., and Lucassen, P. J. (2016). Depression as a risk factor for Alzheimer's disease: genes, steroids, cytokines and neurogenesis—What do we need to know? *Front. Neuroendocrinol.* 41, 153–171. doi: 10.1016/j.yfrne.2015.12.001
- Hermans, E. J., Putman, P., and van Honk, J. (2006). Testosterone administration reduces empathetic behavior: a facial mimicry study. *Psychoneuroendocrinology* 31, 859–866. doi: 10.1016/j.psyneuen.2006.04.002
- Hermans, E. J., Putman, P., Baas, J. M., Geckes, N. M., Kenemans, J. L., and van Honk, J. (2007). Exogenous testosterone attenuates the integrated central stress response in healthy young women. *Psychoneuroendocrinology* 32, 1052–1061. doi: 10.1016/j.psyneuen.2007.08.006
- Hermans, E. J., van Marle, H. J., Ossewaarde, L., Henckens, M. J., Qin, S., van Kesteren, M. T., et al. (2011). Stress-related noradrenergic activity prompts large-scale neural network reconfiguration. *Science* 334, 1151–1153. doi: 10.1126/science.1209603
- Holper, L., Wolf, M., and Tobler, P. N. (2014). Comparison of functional near-infrared spectroscopy and electrodermal activity in assessing objective versus subjective risk during risky financial decisions. *Neuroimage* 84, 833–842. doi: 10.1016/j.neuroimage.2013.09.047
- Homberg, J. R. (2012). Serotonin and decision making processes. *Neurosci. Biobehav. Rev.* 36, 218–236. doi: 10.1016/j.neubiorev.2011.06.001
- Hook, J. N., Giordani, B., Schteingart, D. E., Guire, K., Giles, J., Ryan, K., et al. (2007). Patterns of cognitive change over time and relationship to age following successful treatment of Cushing's disease. *J. Int. Neuropsychol. Soc.* 13, 21–29. doi: 10.1017/s1355617707070051
- Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., and Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310, 1680–1683. doi: 10.1126/science.1115327
- Hsu, F. C., Garside, M. J., Massey, A. E., and McAllister-Williams, R. H. (2003). Effects of a single dose of cortisol on the neural correlates of episodic memory and error processing in healthy volunteers. *Psychopharmacology* 167, 431–442. doi: 10.1007/s00213-003-1413-2
- Hsu, M., Krajbich, I., Zhao, C., and Camerer, C. F. (2009). Neural response to reward anticipation under risk is nonlinear in probabilities. *J. Neurosci.* 29, 2231–2237. doi: 10.1523/JNEUROSCI.5296-08.2009
- Hurd, P. L., Vaillancourt, K. L., and Dinsdale, N. L. (2011). Aggression, digit ratio and variation in androgen receptor and monoamine oxidase a genes in men. *Behav. Genet.* 41, 543–556. doi: 10.1007/s10519-010-9404-7
- Jääskeläinen, J. (2012). Molecular biology of androgen insensitivity. *Mol. Cell. Endocrinol.* 352, 4–12. doi: 10.1016/j.mce.2011.08.006
- Joels, M. (2011). Impact of glucocorticoids on brain function: relevance for mood disorders. *Psychoneuroendocrinology* 36, 406–414. doi: 10.1016/j.psyneuen.2010.03.004
- Josephs, R. A., Sellers, J. G., Newman, M. L., and Mehta, P. H. (2006). The mismatch effect: when testosterone and status are at odds. *J. Pers. Soc. Psychol.* 90, 999–1013. doi: 10.1037/0022-3514.90.6.999
- Juechems, K., Balaguer, J., Ruz, M., and Summerfield, C. (2017). Ventromedial prefrontal cortex encodes a latent estimate of cumulative reward. *Neuron* 93, 705.e4–714.e4. doi: 10.1016/j.neuron.2016.12.038
- Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux.
- Kahneman, D., and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–291. doi: 10.2307/1914185
- Kahnt, T., Heinze, J., Park, S. Q., and Haynes, J. D. (2010). The neural code of reward anticipation in human orbitofrontal cortex. *Proc. Natl. Acad. Sci. U S A* 107, 6010–6015. doi: 10.1073/pnas.0912831107
- Kahnt, T., and Tobler, P. N. (2013). Salience signals in the right temporoparietal junction facilitate value-based decisions. *J. Neurosci.* 33, 863–869. doi: 10.1523/JNEUROSCI.3531-12.2013
- Kandasamy, N., Hardy, B., Page, L., Schaffner, M., Graggaber, J., Powlson, A. S., et al. (2014). Cortisol shifts financial risk preferences. *Proc. Natl. Acad. Sci. U S A* 111, 3608–3613. doi: 10.1073/pnas.1317908111



- Kang, M. J., Hsu, M., Krajchich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T., et al. (2009). The wick in the candle of learning: epistemic curiosity activates reward circuitry and enhances memory. *Psychol. Sci.* 20, 963–973. doi: 10.1111/j.1467-9280.2009.02402.x
- Khaw, M. W., Glimcher, P. W., and Louie, K. (2017). Normalized value coding explains dynamic adaptation in the human valuation process. *Proc. Natl. Acad. Sci. U S A* 114, 12696–12701. doi: 10.1073/pnas.1715293114
- Kim, Y., Kim, K., and Kim, T. H. (2014). Domain specific relationships of 2D:4D digit ratio in risk perception and risk behavior. *J. Gen. Psychol.* 141, 373–392. doi: 10.1080/00221309.2014.950188
- Klinesmith, J., Kasser, T., and McAndrew, F. T. (2006). Guns, testosterone and aggression: an experimental test of a mediational hypothesis. *Psychol. Sci.* 17, 568–571. doi: 10.1111/j.1467-9280.2006.01745.x
- Knickmeyer, R. C., Woolson, S., Hamer, R. M., Konneker, T., and Gilmore, J. H. (2011). 2D:4D ratios in the first 2 years of life: stability and relation to testosterone exposure and sensitivity. *Horm. Behav.* 60, 256–263. doi: 10.1016/j.yhbeh.2011.05.009
- Kusev, P., Purer, H., Hellman, R., Cooke, A. J., van Schalk, P., Baranova, V., et al. (2017). Understanding risky behavior: the influence of cognitive, emotional and hormonal factors on decision-making under risk. *Front. Psychol.* 8:102. doi: 10.3389/fpsyg.2017.00102
- Lempert, K. M., Porcelli, A. J., Delgado, M. R., and Tricomi, E. (2012). Individual differences in delay discounting under acute stress: the role of trait perceived stress. *Front. Psychol.* 3, 251–263. doi: 10.3389/fpsyg.2012.00251
- Lenroot, R. K., and Giedd, J. N. (2010). Sex differences in the adolescent brain. *Brain Cogn.* 72, 46–55. doi: 10.1016/j.bandc.2009.10.008
- Lewis, J. G., Ghanadian, R., and Chisholm, G. D. (1976). Serum 5 $\alpha$ -dihydrotestosterone and testosterone changes with age in man. *Acta Endocrinol.* 82, 444–448. doi: 10.1530/acta.0.0820444
- Lightman, S. (2016). “Rhythms within rhythms: the importance of oscillations for glucocorticoid hormones,” in *A Time for Metabolism and Hormones*, eds P. Sassone-Corsi and Y. Christen (Cham: Springer International Publishing), 87–99.
- Liu, L., Zhou, X., Zhang, Y., Pu, J., Yang, L., Yuan, S., et al. (2018). Hippocampal metabolic differences implicate distinctions between physical and psychological stress in four rat models of depression. *Transl. Psychiatry* 8:4. doi: 10.1038/s41398-017-0018-1
- Louie, K., Grattan, L. E., and Glimcher, P. W. (2011). Reward value-based gain control: divisive normalization in parietal cortex. *J. Neurosci.* 31, 10627–10639. doi: 10.1523/JNEUROSCI.1237-11.2011
- Lovallo, W. R. (2013). Early life adversity reduces stress reactivity and enhances impulsive behavior: implications for health behaviors. *Int. J. Psychophysiol.* 90, 8–16. doi: 10.1016/j.ijpsycho.2012.10.006
- Lucassen, P. J., Pruessner, J., Sousa, N., Almeida, O. F., Van Dam, A. M., Rajkowska, G., et al. (2014). Neuropathology of stress. *Acta Neuropathol.* 127, 109–135. doi: 10.1007/s00401-013-1223-5
- Lupien, S. J., McEwen, B. S., Gunnar, M. R., and Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nat. Rev. Neurosci.* 10, 434–445. doi: 10.1038/nrn2639
- MacLulich, A. M., Ferguson, K. J., Reid, L. M., Deary, I. J., Starr, J. M., Wardlaw, J. M., et al. (2012). 11 $\beta$ -hydroxysteroid dehydrogenase type 1, brain atrophy and cognitive decline. *Neurobiol. Aging* 33, 207.e1–208.e8. doi: 10.1016/j.neurobiolaging.2010.09.010
- Maier, S. U., Makwana, A. B., and Hare, T. A. (2015). Acute stress impairs self-control in goal-directed choice by altering multiple functional connections within the Brain’s decision circuits. *Neuron* 87, 621–631. doi: 10.1016/j.neuron.2015.07.005
- Maney, D. L. (2017). Polymorphisms in sex steroid receptors: from gene sequence to behavior. *Front. Neuroendocrinol.* 47, 47–65. doi: 10.1016/j.yfrne.2017.07.003
- 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
- Margittai, Z., Nave, G., Van Wingerden, M., Schnitzler, A., Schwabe, L., and Kalenscher, T. (2018). Combined effects of glucocorticoid and noradrenergic activity on loss aversion. *Neuropsychopharmacology* 43, 334–341. doi: 10.1038/npp.2017.75
- Mazur, A., and Booth, A. (1998). Testosterone and dominance in men. *Behav. Brain Sci.* 21, 353–363; discussion 363–397. doi: 10.1017/s0140525x98001228
- McCaul, K. D., Gladue, B. A., and Joppa, M. (1992). Winning, losing, mood, and testosterone. *Horm. Behav.* 26, 486–504. doi: 10.1016/0018-506x(92)90016-o
- McKlveen, J. M., Morano, R. L., Fitzgerald, M., Zoubovsky, S., Cassella, S. N., Scheimann, J. R., et al. (2016). Chronic stress increases prefrontal inhibition: a mechanism for stress-induced prefrontal dysfunction. *Biol. Psychiatry* 80, 754–764. doi: 10.1016/j.biopsych.2016.03.2101
- McKlveen, J. M., Myers, B., Flak, J. N., Bundzikova, J., Solomon, M. B., Seroogy, K. B., et al. (2013). Role of prefrontal cortex glucocorticoid receptors in stress and emotion. *Biol. Psychiatry* 74, 672–679. doi: 10.1016/j.biopsych.2013.03.024
- McReynolds, J. R., Donowho, K., Abdi, A., McGaugh, J. L., Roozendaal, B., and McIntyre, C. K. (2010). Memory-enhancing corticosterone treatment increases amygdala norepinephrine and Arc protein expression in hippocampal synaptic fractions. *Neurobiol. Learn. Mem.* 93, 312–321. doi: 10.1016/j.nlm.2009.11.005
- Meaney, M. J., Szyf, M., and Seckl, J. R. (2007). Epigenetic mechanisms of perinatal programming of hypothalamic-pituitary-adrenal function and health. *Trends Mol. Med.* 13, 269–277. doi: 10.1016/j.molmed.2007.05.003
- Mehta, P. H., and Josephs, R. A. (2010). Testosterone and cortisol jointly regulate dominance: evidence for a dual-hormone hypothesis. *Horm. Behav.* 58, 898–906. doi: 10.1016/j.yhbeh.2010.08.020
- Mehta, P. H., Lawless DesJardins, N. M., van Vugt, M., and Josephs, R. A. (2017). Hormonal underpinnings of status conflict: testosterone and cortisol are related to decisions and satisfaction in the hawk-dove game. *Horm. Behav.* 92, 141–154. doi: 10.1016/j.yhbeh.2017.03.009
- Mehta, P. H., Mor, S., Yap, A. J., and Prasad, S. (2015). Dual-hormone changes are related to bargaining performance. *Psychol. Sci.* 26, 866–876. doi: 10.1177/0956797615572905
- Mills, K. L., Goddings, A. L., Clasen, L. S., Giedd, J. N., and Blakemore, S. J. (2014). The developmental mismatch in structural brain maturation during adolescence. *Dev. Neurosci.* 36, 147–160. doi: 10.1159/000362328
- Mohr, P. N., Biele, G., and Heekeren, H. R. (2010). Neural processing of risk. *J. Neurosci.* 30, 6613–6619. doi: 10.1523/JNEUROSCI.0003-10.2010
- Morimoto, M., Morita, N., Ozawa, H., Yokoyama, K., and Kawata, M. (1996). Distribution of glucocorticoid receptor immunoreactivity and mRNA in the rat brain: an immunohistochemical and in situ hybridization study. *Neurosci. Res.* 26, 235–269. doi: 10.1016/s0168-0102(96)01105-4
- Morris, R. W., Purves-Tyson, T. D., Weickert, C. S., Rothmond, D., Lenroot, R., and Weickert, T. W. (2015). Testosterone and reward prediction-errors in healthy men and men with schizophrenia. *Schizophr. Res.* 168, 649–660. doi: 10.1016/j.schres.2015.06.030
- Morrongioello, B. A., and Rennie, H. (1998). Why do boys engage in more risk taking than girls? The role of attributions, beliefs and risk appraisals. *J. Pediatr. Psychol.* 23, 33–43. doi: 10.1093/jpepsy/23.1.33
- Mueller, A., Armbruster, D., Moser, D. A., Canli, T., Lesch, K. P., Brocke, B., et al. (2011). Interaction of serotonin transporter gene-linked polymorphic region and stressful life events predicts cortisol stress response. *Neuropsychopharmacology* 36, 1332–1339. doi: 10.1038/npp.2011.11
- Nadler, A., Jiao, P., Johnson, C. J., Alexander, V., and Zak, P. J. (2017). The bull of wall street: experimental analysis of testosterone and asset trading. *Manage. Sci.* doi: 10.1287/mnsc.2017.2836 [Epub ahead of print].
- Nave, G., Nadler, A., Zava, D., and Camerer, C. (2017). Single-dose testosterone administration impairs cognitive reflection in men. *Psychol. Sci.* 28, 1398–1407. doi: 10.1177/0956797617709592
- Netherton, C., Goodyer, I., Tamplin, A., and Herbert, J. (2004). Salivary cortisol and dehydroepiandrosterone in relation to puberty and gender. *Psychoneuroendocrinology* 29, 125–140. doi: 10.1016/s0306-4530(02)00150-6
- Newcomer, J. W., Selke, G., Melson, A. K., Hershey, T., Craft, S., Richards, K., et al. (1999). Decreased memory performance in healthy humans induced by stress-level cortisol treatment. *Arch. Gen. Psychiatry* 56, 527–533. doi: 10.1001/archpsyc.56.6.527
- Neyse, L., Bosworth, S., Ring, P., and Schmidt, U. (2016). Overconfidence, incentives and digit ratio. *Sci. Rep.* 6:23294. doi: 10.1038/srep23294
- Nicholson, N., Soane, E., Fenton-O’Creavy, M., and Willman, P. (2005). Personality and domain specific risk taking. *J. Risk Res.* 8, 157–176. doi: 10.1080/1366987032000123856



- Nieschlag, E., and Behre, H. M. (2012). *Testosterone. Action, Deficiency, Substitution*. Cambridge, MA: Cambridge University Press.
- O'Connor, K. M., Arnold, J. A., and Maurizio, A. M. (2010). The prospect of negotiating: stress, cognitive appraisal, and performance. *J. Exp. Soc. Psychol.* 46, 729–735. doi: 10.1016/j.jesp.2010.04.007
- O'Connor, D. B., Lee, D. M., Corona, G., Forti, G., Tajar, A., O'Neill, T. W., et al. (2011). The relationships between sex hormones and sexual function in middle-aged and older European men. *J. Clin. Endocrinol. Metab.* 96, E1577–E1587. doi: 10.1210/jc.2010-2216
- Okon-Singer, H., Hendler, T., Pessoa, L., and Shackman, A. J. (2015). The neurobiology of emotion-cognition interactions: fundamental questions and strategies for future research. *Front. Hum. Neurosci.* 9:58. doi: 10.3389/fnhum.2015.00058
- Olff, M., Frijling, J. L., Kubzansky, L. D., Bradley, B., Ellenbogen, M. A., Cardoso, C., et al. (2013). The role of oxytocin in social bonding, stress regulation and mental health: an update on the moderating effects of context and interindividual differences. *Psychoneuroendocrinology* 38, 1883–1894. doi: 10.1016/j.psyneuen.2013.06.019
- Orrell, D. (2018). Does economics need a quantum revolution? *Aeon Magazine*. Available online at: <https://aeon.co/essays/has-the-time-come-for-a-quantum-revolution-in-economics>
- Ouerchefani, R., Ouerchefani, N., Allain, P., Ben Rejeb, M. R., and Le Gall, D. (2017). Contribution of different regions of the prefrontal cortex and lesion laterality to deficit of decision-making on the Iowa Gambling Task. *Brain Cogn.* 111, 73–85. doi: 10.1016/j.bandc.2016.06.010
- Panksepp, J. (1992). Oxytocin effects on emotional processes: separation distress, social bonding, and relationships to psychiatric disorders. *Ann. N Y Acad. Sci.* 652, 243–252. doi: 10.1111/j.1749-6632.1992.tb34359.x
- Pavlov, K. A., Chistiakov, D. A., and Chekhonin, V. P. (2012). Genetic determinants of aggression and impulsivity in humans. *J. Appl. Genet.* 53, 61–82. doi: 10.1007/s13353-011-0069-6
- Pearson, J., Tarabulsi, G. M., and Bussi eres, E. L. (2015). Foetal programming and cortisol secretion in early childhood: a meta-analysis of different programming variables. *Infant Behav. Dev.* 40, 204–215. doi: 10.1016/j.infbeh.2015.04.004
- Pinnock, S. B., Balendra, R., Chan, M., Hunt, L. T., Turner-Stokes, T., and Herbert, J. (2007). Interactions between nitric oxide and corticosterone in the regulation of progenitor cell proliferation in the dentate gyrus of the adult rat. *Neuropsychopharmacology* 32, 493–504. doi: 10.1038/sj.npp.1301245
- Porcelli, A. J., Lewis, A. H., and Delgado, M. R. (2012). Acute stress influences neural circuits of reward processing. *Front. Neurosci.* 6:157. doi: 10.3389/fnins.2012.00157
- Purves-Tyson, T. D., Owens, S. J., Double, K. L., Desai, R., Handelsman, D. J., and Weickert, C. S. (2014). Testosterone induces molecular changes in dopamine signaling pathway molecules in the adolescent male rat nigrostriatal pathway. *PLoS One* 9:e91151. doi: 10.1371/journal.pone.0091151
- Putnam, P., Hermans, E. J., Koppeschaar, H., van Schijndel, A., and van Honk, J. (2007). A single administration of cortisol acutely reduces preconscious attention for fear in anxious young men. *Psychoneuroendocrinology* 32, 793–802. doi: 10.1016/j.psyneuen.2007.05.009
- Ragnarsson, O., Glad, C. A., Berglund, P., Berghthorsdottir, R., Eder, D. N., and Johannsson, G. (2014). Common genetic variants in the glucocorticoid receptor and the 11 -hydroxysteroid dehydrogenase type 1 genes influence long-term cognitive impairments in patients with Cushing's syndrome in remission. *J. Clin. Endocrinol. Metab.* 99, E1803–E1807. doi: 10.1210/jc.2014-1906
- Raznahan, A., Lee, Y., Stidd, R., Long, R., Greenstein, D., Clasen, L., et al. (2010). Longitudinally mapping the influence of sex and androgen signaling on the dynamics of human cortical maturation in adolescence. *Proc. Natl. Acad. Sci. U S A* 107, 16988–16993. doi: 10.1073/pnas.1006025107
- Reynolds, B., Ortengren, A., Richards, J. B., and de Wit, H. (2006). Dimensions of impulsive behavior: personality and behavioral measures. *Pers. Individ. Dif.* 40, 305–315. doi: 10.1016/j.paid.2005.03.024
- Rolison, J. J., Hanoch, Y., Wood, S., and Liu, P. J. (2014). Risk-taking differences across the adult life span: a question of age and domain. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 69, 870–880. doi: 10.1093/geronb/gbt081
- Rubinow, D. R., and Schmidt, P. J. (1996). Androgens, brain, and behavior. *Am. J. Psychiatry* 153, 974–984. doi: 10.1176/ajp.153.8.974
- Russell, G. M., Kalafatakis, K., and Lightman, S. (2015). The importance of biological oscillators for hypothalamic-pituitary-adrenal activity and tissue glucocorticoid response: coordinating stress and neurobehavioural adaptation. *J. Neuroendocrinol.* 27, 378–388. doi: 10.1111/jne.12247
- Sapienza, P., Zingales, L., and Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proc. Natl. Acad. Sci. U S A* 106, 15268–15273. doi: 10.1073/pnas.0907352106
- Schipper, B. C. (2014). Sex hormones and competitive bidding. *Manage. Sci.* 61, 249–266. doi: 10.1287/mnsc.2014.1959
- Schubert, R., Brown, M., Gysler, M., and Brachinger, H. W. (1999). Financial decision-making: are women really more risk-averse? *Am. Econ. Rev.* 89, 381–385. doi: 10.1257/aer.89.2.381
- Schultz, W. (2004). Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioural ecology. *Curr. Opin. Neurobiol.* 14, 139–147. doi: 10.1016/j.conb.2004.03.017
- Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annu. Rev. Psychol.* 57, 87–115. doi: 10.1146/annurev.psych.56.091103.070229
- Schultz, W. (2016). Reward functions of the basal ganglia. *J. Neural Transm.* 123, 679–693. doi: 10.1007/s00702-016-1510-0
- Shih, M., Buldan, A., and Scheiner-Bobis, G. (2014). Non-classical testosterone signaling is mediated by a G-protein-coupled receptor interacting with G  11. *Biochim. Biophys. Acta* 1843, 1172–1181. doi: 10.1016/j.bbamcr.2014.03.002
- Simerly, R. B., Chang, C., Muramatsu, M., and Swanson, L. W. (1990). Distribution of androgen and estrogen receptor mRNA-containing cells in the rat brain: an in situ hybridization study. *J. Comp. Neurol.* 294, 76–95. doi: 10.1002/cne.902940107
- Slovic, P. (1964). Assessment of risk taking behavior. *Psychol. Bull.* 61, 220–233. doi: 10.1037/h0043608
- Sokol-Hessner, P., Camerer, C. F., and Phelps, E. A. (2013). Emotion regulation reduces loss aversion and decreases amygdala responses to losses. *Soc. Cogn. Affect. Neurosci.* 8, 341–350. doi: 10.1093/scan/nss002
- Soutschek, A., Burke, C. J., Beharelle, A. R., Schreiber, R., Weber, S. C., Karipidis, I. I., et al. (2017). The dopaminergic reward system underpins gender differences in social preferences. *Nat. Hum. Behav.* 11, 819–827. doi: 10.1038/s41562-017-0226-y
- Stanton, S. J., and Schultheiss, O. C. (2009). The hormonal correlates of implicit power motivation. *J. Res. Pers.* 43, 942–949. doi: 10.1016/j.jrp.2009.04.001
- Starcke, K., and Brand, M. (2012). Decision making under stress: a selective review. *Neurosci. Biobehav. Rev.* 36, 1228–1248. doi: 10.1016/j.neubiorev.2012.02.003
- Starcke, K., and Brand, M. (2016). Effects of stress on decisions under uncertainty: a meta-analysis. *Psychol. Bull.* 142, 909–933. doi: 10.1037/bul0000068
- Starcke, K., Wiesen, C., Trotzke, P., and Brand, M. (2016). Effects of acute laboratory stress on executive functions. *Front. Psychol.* 7:461. doi: 10.3389/fpsyg.2016.00461
- Starkman, M. N., Schteingart, D. E., and Schork, M. A. (1981). Depressed mood and other psychiatric manifestations of Cushing's syndrome: relationship to hormone levels. *Psychosom. Med.* 43, 3–18. doi: 10.1097/00006842-198102000-00002
- Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Dev. Rev.* 28, 78–106. doi: 10.1016/j.dr.2007.08.002
- Takahashi, H., Takano, H., Camerer, C. F., Ideno, T., Okubo, S., Matsui, H., et al. (2012). Honesty mediates the relationship between serotonin and reaction to unfairness. *Proc. Natl. Acad. Sci. U S A* 109, 4281–4284. doi: 10.1073/pnas.1204895109
- Takahashi, T., Shinada, M., Inukai, K., Tanida, S., Takahashi, C., Mifune, N., et al. (2010). Stress hormones predict hyperbolic time-discount rates six months later in adults. *Neuro Endocrinol. Lett.* 31, 616–621.
- Tobler, P. N., Christopoulos, G. I., O'Doherty, J. P., Dolan, R. J., and Schultz, W. (2009). Risk-dependent reward value signal in human prefrontal cortex. *Proc. Natl. Acad. Sci. U S A* 106, 7185–7190. doi: 10.1073/pnas.0809599106
- Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131. doi: 10.1126/science.185.4157.1124
- Tversky, A., and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science* 211, 453–458. doi: 10.1126/science.7455683
- Tyborowska, A., Volman, I., Smeekens, S., Toni, I., and Roelofs, K. (2016). Testosterone during puberty shifts emotional control from pulvinar to anterior prefrontal cortex. *J. Neurosci.* 36, 6156–6164. doi: 10.1523/JNEUROSCI.3874-15.2016

- van Ast, V. A., Cornelisse, S., Meeter, M., Joëls, M., and Kindt, M. (2013). Time-dependent effects of cortisol on the contextualization of emotional memories. *Biol. Psychiatry* 74, 809–816. doi: 10.1016/j.biopsych.2013.06.022
- van den Bos, R., Harteveld, M., and Stoop, H. (2009). Stress and decision-making in humans: performance is related to cortisol reactivity, albeit differently in men and women. *Psychoneuroendocrinology* 34, 1449–1458. doi: 10.1016/j.psyneuen.2009.04.016
- van der Loos, M. J., Haring, R., Rietveld, C. A., Baumeister, S. E., Groenen, P. J., Hofman, A., et al. (2013). Serum testosterone levels in males are not associated with entrepreneurial behavior in two independent observational studies. *Physiol. Behav.* 119, 110–114. doi: 10.1016/j.physbeh.2013.06.003
- 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., Schutter, D. J., Bos, P. A., Kruijt, A. W., Lentjes, E. G., and Baron-Cohen, S. (2011). Testosterone administration impairs cognitive empathy in women depending on second-to-fourth digit ratio. *Proc. Natl. Acad. Sci. U S A* 108, 3448–3452. doi: 10.1073/pnas.1011891108
- van Rossum, E. F., Russcher, H., and Lamberts, S. W. (2005). Genetic polymorphisms and multifactorial diseases: facts and fallacies revealed by the glucocorticoid receptor gene. *Trends Endocrinol. Metab.* 16, 445–450. doi: 10.1016/j.tem.2005.10.009
- Ventura, T., Gomes, M. C., Pita, A., Neto, M. T., and Taylor, A. (2013). Digit ratio (2D:4D) in newborns: influences of prenatal testosterone and maternal environment. *Early Hum. Dev.* 89, 107–112. doi: 10.1016/j.earhumdev.2012.08.009
- Vermeersch, H., T'Sjoen, G., Kaufman, J. M., and Vincke, J. (2008). The role of testosterone in aggressive and non-aggressive risk-taking in adolescent boys. *Horm. Behav.* 53, 463–471. doi: 10.1016/j.yhbeh.2007.11.021
- Vermeersch, H., T'Sjoen, G., Kaufman, J. M., Vincke, J., and Van Houtte, M. (2010). Testosterone androgen receptor gene CAG repeat length, mood and behaviour in adolescent males. *Eur. J. Endocrinol.* 163, 319–328. doi: 10.1530/eje-10-0090
- Vernocchi, S., Battello, N., Schmitz, S., Revets, D., Billing, A. M., Turner, J. D., et al. (2013). Membrane glucocorticoid receptor activation induces proteomic changes aligning with classical glucocorticoid effects. *Mol. Cell. Proteomics* 12, 1764–1779. doi: 10.1074/mcp.m112.022947
- Wagels, L., Votinov, M., Radke, S., Clemens, B., Montag, C., Jung, S., et al. (2017). Blunted insula activation reflects increased risk and reward seeking as an interaction of testosterone administration and the MAOA polymorphism. *Hum. Brain Mapp.* 38, 4574–4593. doi: 10.1002/hbm.23685
- Walderhaug, E., Landro, N. I., and Magnusson, A. (2008). A synergic effect between lowered serotonin and novel situations on impulsivity measured by CPT. *J. Clin. Exp. Neuropsychol.* 30, 204–211. doi: 10.1080/13803390701346311
- White, R. E., Thornhill, S., and Hampson, E. (2006). Entrepreneurs and evolutionary biology: the relationship between testosterone and new venture creation. *Orag. Behav. Hum. Decis. Process.* 100, 21–34. doi: 10.1016/j.obhdp.2005.11.001
- Wirth, M. M., Scherer, S. M., Hoks, R. M., and Abercrombie, H. C. (2011). The effect of cortisol on emotional responses depends on order of cortisol and placebo administration in a within-subject design. *Psychoneuroendocrinology* 36, 945–954. doi: 10.1016/j.psyneuen.2010.11.010
- Wisniewski, A. B., Migeon, C. J., Meyer-Bahlburg, H. F., Gearhart, J. P., Berkovitz, G. D., Brown, T. R., et al. (2000). Complete androgen insensitivity syndrome: long-term medical, surgical, and psychosexual outcome. *J. Clin. Endocrinol. Metab.* 85, 2664–2669. doi: 10.1210/jcem.85.8.6742
- Wolf, O. T., Atsak, P., de Quervain, D. J., Roozendaal, B., and Wingenfeld, K. (2016). Stress and memory: a selective review on recent developments in the understanding of stress hormone effects on memory and their clinical relevance. *J. Neuroendocrinol.* 28:8. doi: 10.1111/jne.12353
- Wood, R. I., Armstrong, A., Fridkin, V., Shah, V., Najafi, A., and Jakowec, M. (2013). Roid rage in rats? Testosterone effects on aggressive motivation, impulsivity and tyrosine hydroxylase. *Physiol. Behav.* 110–111, 6–12. doi: 10.1016/j.physbeh.2012.12.005
- Wrangham, R. W., and Glowacki, L. (2012). Intergroup aggression in chimpanzees and war in nomadic hunter-gatherers: evaluating the chimpanzee model. *Hum. Nat.* 23, 5–29. doi: 10.1007/s12110-012-9132-1
- Wrosch, C., Schulz, R., Miller, G. E., Lupien, S., and Dunne, E. (2007). Physical health problems, depressive mood and cortisol secretion in old age: buffer effects of health engagement control strategies. *Health Psychol.* 26, 341–349. doi: 10.1037/0278-6133.26.3.341
- Wüst, S., Van Rossum, E. F., Federenko, I. S., Koper, J. W., Kumsta, R., and Hellhammer, D. H. (2004). Common polymorphisms in the glucocorticoid receptor gene are associated with adrenocortical responses to psychosocial stress. *J. Clin. Endocrinol. Metab.* 89, 565–573. doi: 10.1210/jc.2003-031148
- Zak, P. J., Kurzban, R., Ahmadi, S., Swerdloff, R. S., Park, J., Efremidze, L., et al. (2009). Testosterone administration decreases generosity in the ultimatum game. *PLoS One* 4:e8330. doi: 10.1371/journal.pone.0008330
- Zheng, H., Wang, X. T., and Zhu, L. (2010). Framing effects: behavioral dynamics and neural basis. *Neuropsychologia* 48, 3198–3204. doi: 10.1016/j.neuropsychologia.2010.06.031
- Zhong, W., Li, Y., Feng, Q., and Luo, M. (2017). Learning and stress shape the reward response patterns of serotonin neurons. *J. Neurosci.* 37, 8863–8875. doi: 10.1523/JNEUROSCI.1181-17.2017
- Zuckerman, M. (1991). *Psychobiology of Personality*. Cambridge, MA: Cambridge University Press.

**Conflict of Interest Statement:** The author declares 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 © 2018 Herbert. 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) and the copyright owner 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.



# Endogenous Oxytocin Release Eliminates In-Group Bias in Monetary Transfers With Perspective-Taking

Elizabeth T. Terris<sup>1</sup>, Laura E. Beavin<sup>1</sup>, Jorge A. Barraza<sup>1</sup>, Jeff Schloss<sup>2</sup> and Paul J. Zak<sup>1\*</sup>

<sup>1</sup>Center for Neuroeconomics Studies, Claremont Graduate University, Claremont, CA, United States, <sup>2</sup>Department of Biology, Westmont College, Santa Barbara, CA, United States

## OPEN ACCESS

### Edited by:

Ulrich Schmidt,  
Institut für Weltwirtschaft, Germany

### Reviewed by:

Valery Grinevich,  
Deutsches Krebsforschungszentrum  
(DKFZ), Germany  
Gábor B. Makara,  
Hungarian Academy of Sciences  
(MTA), Hungary

### \*Correspondence:

Paul J. Zak  
paul@neuroeconomicstudies.org

**Received:** 27 September 2017

**Accepted:** 15 February 2018

**Published:** 05 March 2018

### Citation:

Terris ET, Beavin LE, Barraza JA,  
Schloss J and Zak PJ  
(2018) Endogenous Oxytocin Release  
Eliminates In-Group Bias in Monetary  
Transfers With Perspective-Taking.  
*Front. Behav. Neurosci.* 12:35.  
doi: 10.3389/fnbeh.2018.00035

Oxytocin (OT) has been shown to facilitate trust, empathy and other prosocial behaviors. At the same time, there is evidence that exogenous OT infusion may not result in prosocial behaviors in all contexts, increasing in-group biases in a number of studies. The current investigation seeks to resolve this inconsistency by examining if endogenous OT release is associated with in-group bias. We studied a large group of participants ( $N = 399$ ) in existing groups and randomly formed groups. Participants provided two blood samples to measure the change in OT after a group salience task and then made computer-mediated monetary transfer decisions to in-group and out-group members. Our results show that participants with an increase in endogenous OT showed no bias in monetary offers in the ultimatum game (UG) to out-group members compared to in-groups. There was also no bias in accepting UG offers, though in-group bias persisted for a unilateral monetary transfer. Our analysis shows that the strength of identification with one's group diminished the effects that an increase in OT had on reducing bias, but bias only recurred when group identification reached 87% of its maximum value. Our results indicate that the endogenous OT system appears to reduce in-group bias in some contexts, particularly those that require perspective-taking.

**Keywords:** prosociality, neuroendocrinology, selfishness, monetary exchange, bias

## INTRODUCTION

As with all social animals, it is the nature of humans to form groups. People more readily affiliate with those who share common traits or behaviors (Prentice et al., 1994). Group bonding can benefit members in a group by promoting cooperation and altruism (Penner et al., 2005; Hein et al., 2010; Weller and Hansen Lagattuta, 2013), but it may also lead to discrimination or derogation of non-group members (Brewer, 1999). The biological mechanisms that drive in-group favoritism and out-group prejudice are just beginning to be studied (Amodio et al., 2004; Knutson et al., 2007; Van Bavel et al., 2008). Some of this research has focused on the neuropeptide oxytocin (OT) because it facilitates attachment, social approach, and prosocial behaviors like trust and cooperation, as well as maternal defense (e.g., Zak et al., 2004; Kosfeld et al., 2005; Huffmeijer et al., 2013; Carter, 2014; Hostinar et al., 2014; Algoe et al., 2017).

## In-Group Bias

OT's prosocial effects are likely to be depend to social context (e.g., Bartz et al., 2011; Shamay-Tsoory and Abu-Akel, 2016). OT has been shown to facilitate social recognition in human and non-human animals (Bielsky and Young, 2004) and to enhance the saliency of social cues

(Pfundmair et al., 2017). Social salience, in turn, can increase prosocial behaviors that are facilitated through negative emotions like anger, leading to punishment of non-cooperative behaviors like free-riding (Aydogan et al., 2017). Social salience is the likely cause of the so-called “dark side” of OT, namely bias of one’s preferences toward in-group members (Shamay-Tsoory and Abu-Akel, 2016). Studies indicate that exogenous OT infusion promotes in-group (parochial) altruism (De Dreu et al., 2010; Ten Velden et al., 2017), ethnic in-group preference (De Dreu et al., 2011), protection of vulnerable in-group members (De Dreu et al., 2012), and the promotion of in-group norms (Daughters et al., 2017). Taken together, these studies show that OT promotes in-group preference rather than out-group derogation or hate (De Dreu, 2012; Shamay-Tsoory and Abu-Akel, 2016).

When drawing these conclusions, though, one needs to consider studies that question whether OT induces a bias against out-groups. For instance, OT given to Jewish Israelis increased empathy for pain experienced by Palestinian Arabs (Shamay-Tsoory et al., 2013). Notably, OT did not impact in-group empathy toward fellow Jewish Israelis. More generally, OT infusion appears to produce either prosocial or defensive behaviors depending on context, consistent with findings in animal studies (Bartz et al., 2011). Situational context is known to influence in-group/out-group behaviors (Mackie and Hamilton, 1993; Goette et al., 2012; LaBouff et al., 2012). Yet, studies using exogenous OT often pit an in-group against an out-group by asking people make decisions that explicitly benefit their group (De Dreu et al., 2010, 2011; De Dreu, 2012). These studies claim that OT preserves group membership by avoiding or possibly punishing out-groups (De Dreu, 2012). However, studies that do not stimulate group competition report that OT administration is associated with an increase in benefits for both in- and out-group members compared to placebo (Israel et al., 2012; Shamay-Tsoory et al., 2013; Huang et al., 2015). In a similar vein, a meta-analysis of OT infusion and trust found that OT increases in-group trust but does not reduce trust toward out-group members (Van Ijzendoorn and Bakermans-Kranenburg, 2012). The balance of evidence in the OT infusion and group literature indicates that exogenous OT increases the effect of primed group competition by intensifying a situational feature in the experiment. Absent a competition prime, OT is more likely to amplify what appears to be a moderate predilection for prosocial behaviors in humans.

Another factor that can affect how OT impacts group behavior is the use of groups formed in the laboratory, rather than studying existing groups. OT infusion appears to have a different effect when interacting with a known other compared to a stranger (Declerck et al., 2010, 2014). Using only randomly-formed groups to study biases may be another contextual feature that impacts extant OT findings. Further, a larger OT signal may be needed to motivate social interactions among strangers compared to known individuals (Wacker and Ludwig, 2012). Studies that examine endogenous OT release have only reported prosocial effects in psychologically healthy populations (Zak et al., 2005; Gonzaga et al., 2006; Morhenn et al., 2008; Barraza and Zak, 2009; Israel et al., 2009; Hurlemann et al., 2010;

Crockford et al., 2014). In animals and humans, endogenous OT appears to be a response to a positive social stimulus and causes most people to reciprocate in a positive manner (reviewed in Zak, 2012).

## Endogenous Oxytocin

OT infusion studies seldom test if endogenous OT responds to the experimental stimulus. If we want to understand how the brain processes social information, best practice is to measure the response of endogenous OT and then confirm such a finding using exogenous OT. To date, studies examining the role of OT on in-group/out-group behavior have almost exclusively utilized exogenous OT infusion, with a few notable exceptions using less reliable endogenous OT analytes (urine, saliva). Urinary OT has been observed to increase before and during intergroup conflict in wild chimpanzees (Samuni et al., 2017). The increase in reactive OT was positively associated with greater group cohesion during intergroup conflict, but not the degree of out-group threat. A study examining Jewish-Israeli and Arab-Palestinian adolescents found a positive correlation between salivary OT concentrations and the extent of in-group bias (Levy et al., 2016). However, the positive correlation for OT and in-group bias only came from the Jewish-Israeli participants, and only for what the authors termed “neural in-group bias” defined as the amount of alpha modulation in the somatosensory cortex while empathizing with vicarious pain from in-group and out-group members. No results were reported on social behavior or self-reported bias toward the out-group and OT. Blood draws, if done rapidly because of OT’s approximately 3 min half-life, are the most effective way to capture the release of OT after a stimulus (Rydén and Sjöholm, 1969). While there are many ways to induce OT release, in every experiment with healthy adults, none generate this effect in every participant for a variety of reasons (Zak, 2012).

## Current Study

The studies of bias and OT do not provide a clear prediction on whether endogenous OT release will be associated with an in-group bias. Moreover, emerging research reveals a concern with the reliability and replicability OT infusion studies (Nave et al., 2015; Lane et al., 2016) and disagreements regarding how intranasal OT research should be interpreted (Churchland and Winkelman, 2012; Leng and Ludwig, 2016; Walum et al., 2016). These concerns show the need for a comprehensive approach to studying OT and social phenomena. We seek to do this in the present study by measuring the change in endogenous OT following interactions with group members, including both males and females in non-competitive tasks (i.e., allocations toward one group do not impact the other group), using a large sample size, and studying both previously-formed and randomly-formed groups.

## MATERIALS AND METHODS

This study used group activities to stimulate endogenous OT release and relate the change in OT to in- and out-group bias. While basal plasma OT and central OT are unrelated, after



stimulation, the change in OT in plasma and cerebral spinal fluid are positively correlated across several studies (Neumann et al., 2013; Valstad et al., 2017). Taking this into account, the analysis here only uses the percent change in OT in plasma to reflect the effects of central OT. In more than a decade of research measuring endogenous OT, we have found that social interactions that stimulate OT will only do so for a subset of participants (Zak, 2012). Our approach uses this finding to compare the behavior of participants who had an increase OT (OT+) to those for whom the interaction did not increase OT (OT-).

## Participants and Recruitment

Three hundred and ninety-nine participants were recruited from Claremont Graduate University, Westmont College, and local organizations within the Claremont and Santa Barbara communities. The sample size was based on size effects for OT release during monetary transfer tasks (Zak et al., 2005; Barraza and Zak, 2013). Two locations were used to increase the diversity of participants and group membership. Randomly formed groups were made up of 176 Claremont College students and 66 Westmont College students. These participants were randomly assigned to members of either “red” or “blue” groups (based on the minimal groups paradigm, Brewer, 1979; Lemyre and Smith, 1985; Ford and Stangor, 1992; Dunham et al., 2011). Previously formed groups included a group of local Claremont Colleges Reserve Officer Training Corps (ROTC) members ( $N = 30$ ), a group of individuals from a student-led Claremont Colleges Christian organization ( $N = 27$ ), a group of students from Westmont College ( $N = 56$ ), and a group of Pentecostal church members recruited in Santa Barbara ( $N = 44$ ). Sixty-four percent of the participants were Caucasian, 14% were Asian, 7% were Hispanic, 3% were African American, 3% described themselves as multi-ethnic, 7% described themselves as other, and 2% did not reveal their race. Participants were between the ages of 18 and 67 (with 82% between 18 and 22;  $M = 22.76$ ,  $SD = 8.61$ ). Fifty-three percent of participants were females. Recruitment for those in previously-formed groups (P) used target groups, and recruitment for randomly-assigned groups (R) focused on the broader population of students from the Claremont Colleges and Westmont College. This study was carried out in accordance with the recommendations of institutional review boards with written informed consent from all participants. All participants gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the institutional review boards at Claremont Graduate University and Westmont College.

## Procedures

After assignment to the red or blue groups, participants were given a random identification number on a paper badge in either blue or red ink to place on their chests for visibility. Color assignment was counterbalanced. After color assignment, participants completed trait surveys and provided a 12 ml blood sample obtained by a qualified phlebotomist to establish basal levels of OT.

After blood samples were obtained, groups were led into rooms segregated by color. Participants completed pre-task surveys, and a research assistant explained the group task. We did not want our findings to depend on a particular group task so we designed tasks that were ecologically valid for different groups. We expected that by making group membership salient, these tasks would stimulate OT release. R participants engaged in one of three group tasks. The first involved playing the game Scribblish; this game was chosen because it is noncompetitive, fun, and something people of all ages can do. Other R participants were asked to have a group conversation to get to know each other, or to sing folk songs with a leader who was not a participant. Tasks for those in P groups were also designed to reinforce group membership. These included marching for 15 min for the ROTC group, singing religious songs for 15 min with a song leader in the student Christian organization, and participating in a typical worship ceremony with a leader for 15 min for the Pentecostal church members. After the group task, participants completed post-task surveys and then provided a second 12 ml blood sample. Group tasks were staggered to reduce waiting time for the second blood draw (Zak et al., 2005). This allowed blood samples to be obtained from all participants within 5 min after the group task concluded. Next, participants were seated in a large computer lab with partitioned stations where they were instructed in and made monetary decisions. Once the decision tasks were finished, participants completed post-experiment surveys, were informed of their earnings in private, and were paid and released from the experiment.

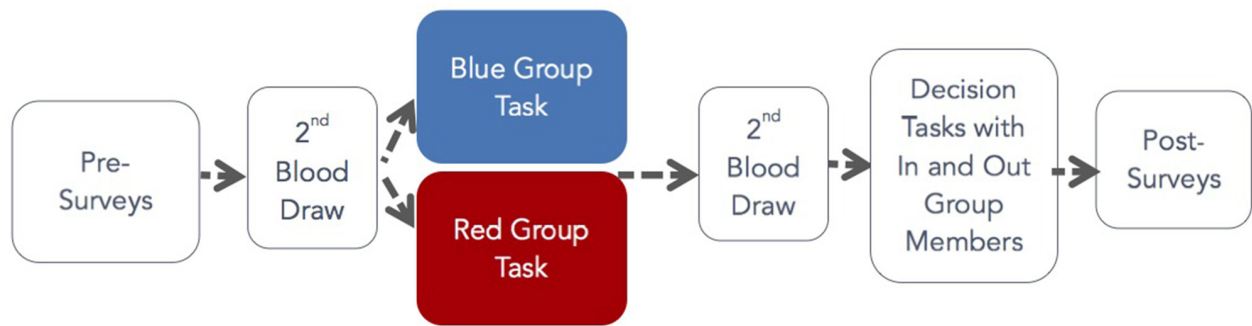
## Materials

### Pre-task Surveys

Participants were asked to complete a demographic survey that included questions on age, ethnicity and religious affiliation. Two surveys measured closeness to others and mood using the Inclusion of Other in Self (IOS; Aron et al., 1992) and the Positive Affect and Negative Affect Scale (PANAS; Watson et al., 1988). The PANAS asked participants to rate their current affective state on a scale from 1 to 5 (1 meaning they were currently feeling the emotion *very slightly or not at all*, and 5 meaning they were currently feeling the emotion *extremely*). The IOS asked participants about how close they felt to: (1) others in their group (red or blue); (2) something bigger than themselves; and (3) to their previously formed group when appropriate.

### Post-task Surveys

The IOS, PANAS, Religious Commitment Inventory that refers to how much an individual is involved in religious activities (RCI, Worthington et al., 2003) and a survey we created on the context of one's identification with their in-group (GROUPID) based on related research (Hogg et al., 1998) were given after the group task. The GROUPID survey asked participants to rate how much they favored their group on a scale from 1 to 5 (1 being *not very favorable* and 5 being *very favorable*) on seven dimensions (e.g., belonging, fit with one's values) that were summed to create a GROUPID score.



**FIGURE 1** | Experiment flow with randomization to group tasks.

### Decision Making Tasks

To measure in-group and out-group bias, participants made monetary decisions involving people from both groups. In these tasks, participants made choices by computer in two rounds of the ultimatum game (UG), and dictator game (DG) as Decision-Maker 1 (DM1) and as Decision-Maker 2 (DM2). Participants were fully and identically instructed in each task, all decisions were double-blind, and there was no deception of any kind. Before each decision, participants were informed via software if their decision partner was a member of the red or blue group (i.e., was an in- or out-group member). All participants made choices in each task with both an in-group member and an out-group member and decisions were made in private in partitioned computer stations. Random assignment determined whether a participant was DM1 or DM2, and dyads were determined by random assignment. Pairings were not sustained across decision tasks to remove the effect of reputation and tasks were counterbalanced across sessions. Participants were informed that they would be paid 50 cents for every dollar they earned in the decision tasks described below.

In the UG, DM1 was endowed with \$10 USD, while DM2 had nothing. The instructions stated that DM1 would be prompted to offer a split of the \$10 to DM2. If DM2 accepted the split, the money would be paid to both DMs. If DM2 rejected the split, both DMs would receive \$0. Both DMs were informed of this structure. After instruction and a chance to ask questions, DM1 was prompted by computer to enter the split proposal. At the same time, DM2 was prompted to report the minimum amount of money she/he was willing to accept from DM1. The software tallied the payoffs but these were not revealed to DMs so as to reduce possible experience effects. The UG requires the use of theory of mind (Camerer, 2003) and is used to measure selfishness and generosity (Zak et al., 2007).

In the DG, DM1 was endowed with \$10 and DM2 had \$0. The endowment amounts were common knowledge. After instruction, DM1 was prompted by computer choose how much, if any, of his or her \$10 to transfer to the DM2 in the dyad. DM2 made no decision in this task. The DM1 transfer is thought to measure altruism (Smith, 1998). **Figure 1** shows the flow of the experiment.

### Blood Handling

Blood was drawn from an antecubital vein using an EDTA (ethylenediaminetetraacetic acid) whole blood tube while maintaining a sterile field and using a Vacutainer® (BD, Franklin Lakes, NJ, USA). Following the draw, blood tubes were rocked to facilitate mixing and prevent coagulation and were immediately placed on ice. Within 15 min, tubes were centrifuged at 1500 rpm for 12 min at 4°C following our published protocol (Zak et al., 2005). Plasma was removed from the tubes with disposable pipettes and placed into 2 ml microtubes with screw caps. These tubes were immediately placed on dry ice and stored at −80°C until assays were performed.

OT was assayed from plasma using an RIA (radioimmunoassay) kit produced by Bachem, Incorporation (Torrance, CA, USA) in duplicate including an extraction step. The RIA has been shown to be more reliable at detecting OT than an ELISA (enzyme-linked immunosorbent assay), with extraction as a necessary step in the process (McCullough et al., 2013; Christensen et al., 2014). The inter- and intra-assay coefficients of variation for OT were 4.58% and 4.01%, and detection levels were 0.5 pg/ml. OT was assayed at the Reproductive Endocrine Research Laboratory at the University of Southern California (Los Angeles, CA, USA). Ten outliers (>3SD over mean) in basal OT or stimulated OT were removed from the sample and on inspection the percent change in OT was normally distributed.

### Statistical Analysis

Independent *t*-test were utilized to examine the extent of bias shown toward the in-group and out-group for decision tasks and how OT release affected this decision. We examined the context of decisions using independent *t*-test to examine differences between those from previously formed groups vs. randomly formed groups. We analyzed the overall impact of group type (P or R), OT (OT+, OT−), and group identification (GROUPID) using a linear regression model. This model was also used to determine the extent that personality traits affected bias.

**TABLE 1** | Descriptive statistics for Oxytocin + (OT+) and OT– groups and for previously (P) and randomly (R) formed groups.

Variable	OT+	OT–	P	R
N	205	180	157	239
Age	22.41 (8.73)	22.95 (8.49)	23.70 (10.18)	22.13 (7.35)
Gender	50% female	58% female	55% female	53% female
In-group UG DM1	5.24 (1.82)	5.29 (1.94)	5.75 (2.19)	4.93 (1.52)
Out-group UG DM1	5.26 (1.90)	4.98 (1.75)	5.54 (2.08)	4.83 (1.58)
In-group UG DM2	2.38 (1.76)	2.34 (1.85)	2.19 (18.6)	2.50 (1.75)
Out-group UG DM2	2.36 (1.70)	2.45 (1.98)	2.31 (1.92)	2.49 (1.77)
In-group DG DM1	4.34 (2.75)	4.21 (2.65)	4.97 (2.85)	3.79 (2.53)
Out-group UG DM1	3.88 (2.84)	3.72 (2.74)	4.40 (2.97)	3.40 (2.61)

Values in parentheses are standard deviations.

## RESULTS

Of the 399 participants, 11 did not have complete blood data, 17 did not complete the monetary decisions tasks, and 53 were missing survey data for the GROUPEID questionnaire. Participants with missing data were used in all analyses except for in cases where their data was missing. **Table 1** has descriptive statistics for the sample.

### Overall Bias

When considering the entire sample, more money was transferred to in-group members compared to the out-group participants for all DM1 decisions (UG DM1: in-group  $M = 5.26$ ,  $SD = 1.87$ , out-group  $M = 5.12$ ,  $SD = 1.84$ , paired  $t_{(381)} = 2.26$ ,  $p = 0.025$ , 95% CI [0.08, 0.25]; DG: in-group  $M = 4.26$ ,  $SD = 2.72$ , out-group  $M = 3.82$ ,  $SD = 2.80$ , paired  $t_{(381)} = 5.51$ ,  $p < 0.001$ , 95% CI [0.28, 0.60]).

### Bias by Group Type

As we expected, P participants gave more to their in-group compared to their out-group in all decisions except as DM2 in the UG. Those in the R group gave more money to their in-group in the DG, but not as DM1 and DM2 in the UG ( $p > 0.05$ ). These biases are partially attributable to a stronger contextual identification (GROUPEID) for P vs. R participants (P: 3.84,  $SD = 0.83$ ; R: 3.44,  $SD = 0.69$ ,  $t_{(194.13)} = -4.48$ ,  $p < 0.001$ , 95% CI [−0.58, 0.23]). GROUPEID was positively correlated with in-group bias by DM1s in both decision tasks (UG:  $r = 0.12$ ,  $p = 0.034$ ; DG:  $r = 0.12$ ,  $p = 0.035$ ). Bias was unrelated to group closeness (IOS) or changes in mood (PANAS).

### Oxytocin Stimulation

Average basal OT was in the expected range ( $M = 5.97$  pg/ml,  $SD = 12.75$ ) and the average percentage change in OT was positive ( $M = 116.09\%$ ,  $SD = 452.40\%$ ,  $t_{(387)} = 5.06$ ,  $p = 0.004$ , 95% CI [70.93, 161.25]). Consistent with our hypothesis, the percentage change in OT for those in randomly-formed groups ( $M = 156.59\%$ ,  $SD = 567.80\%$ ,  $N = 231$ ) showed a significantly larger increase than for those in the previously formed groups ( $M = 56.50\%$ ,  $SD = 162.49\%$ ,  $N = 157$ ,  $t_{(282.72)} = 2.53$ ,  $p = 0.012$ ,  $d = 0.22$ , 95% CI [22.24, 177.93], see **Figure 2**).

Fifty-two percent ( $N = 205$ ) of participants showed an increase in OT (OT+) following the group task. Among these individuals, the average increase was 251.57%, which was significantly different from zero ( $t_{(207)} = 6.20$ ,  $p < 0.001$ , 95%

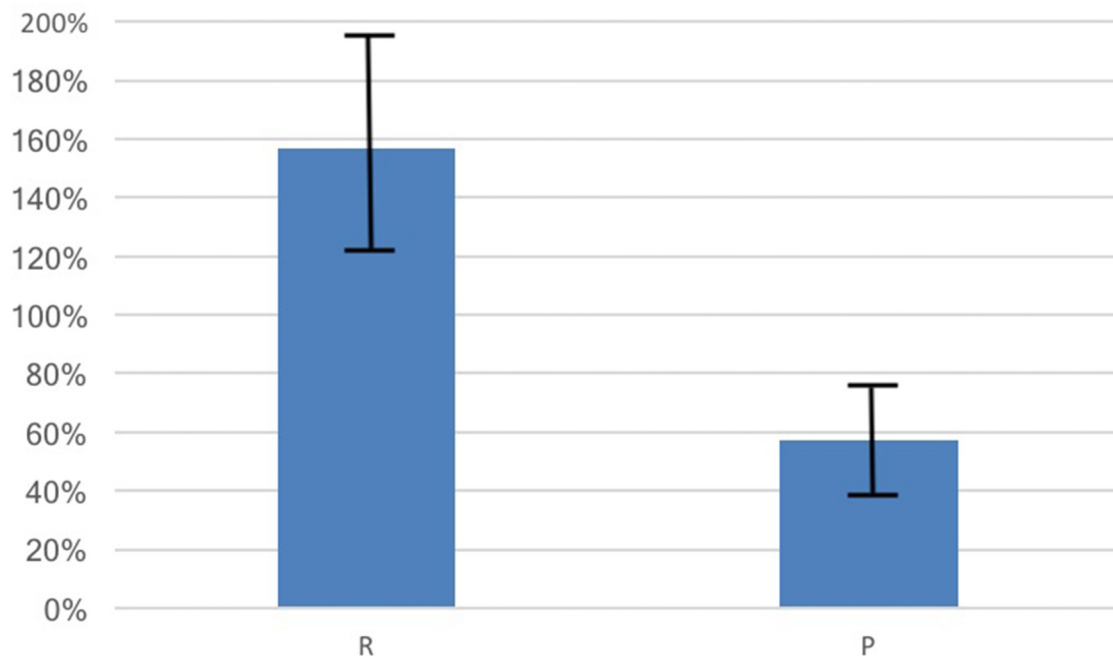
CI [171.62, 331.53]). As above, OT+ participants in randomly-formed groups had a larger increase in OT than those in previously formed groups (R:  $M = 526.65\%$ ,  $SD = 1089.51$ ; P:  $M = 273.28\%$ ,  $699.07$ ;  $t_{(203.62)} = 2.04$ ,  $p = 0.043$ , 95% CI [8.47, 498.27]).

### Oxytocin and Bias

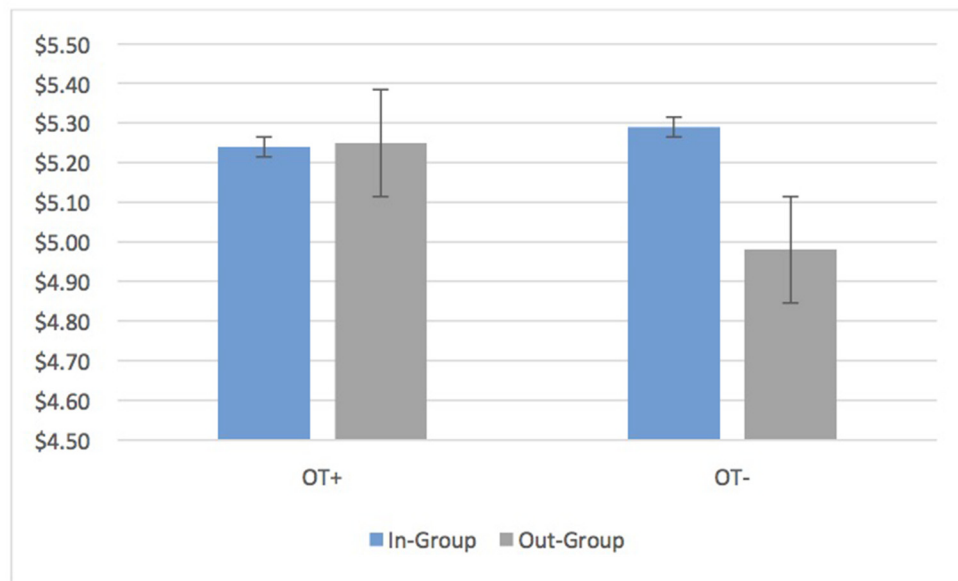
Average transfers by OT+ as DM1s in the UG showed no bias at all (OT+ In: 5.24,  $SD = 1.82$  Out: 5.25,  $SD = 1.90$ ,  $t_{(202)} = -0.20$ ,  $p = 0.84$ , 95% CI [−0.16, 0.13]). OT– participants continued to have in-group bias in the UG and DG (DM1 UG In: 5.29,  $SD = 1.94$  Out: 4.98,  $SD = 1.77$ ;  $t_{(173)} = 3.09$ ,  $p = 0.002$ , 95% CI [0.11, 0.50]; DG DM1 In: 4.21,  $SD = 2.65$ , Out: 3.78,  $SD = 2.73$ ,  $t_{(171)} = 3.50$ ,  $p = 0.001$ , 95% CI [0.19, 0.67]; **Figure 3**). Put differently, the relative in-group bias in the UG (In-group transfer—Out-group transfer) disappeared for OT+ while it was sustained for OT– (OT+:  $M = -0.015$ ,  $SD = 1.30$ , OT–  $M = 0.31$ ,  $SD = 1.61$ ;  $t_{(331.68)} = 2.59$ ,  $p = 0.01$ , 95% CI [0.08, 0.56]). Nevertheless, an in-group bias continued to appear for OT+ for unilateral transfers in the DG (In: \$4.34,  $SD = 2.75$ , Out: \$3.88,  $SD = 2.84$ ;  $p < 0.001$ , 95% CI [0.25, 0.68]). When it came to reciprocation (UG DM2), there was no bias in the minimum acceptable offer for OT+ and OT– (OT+:  $M = 0.015$ ,  $SD = 0.952$ ; OT–:  $M = -0.139$ ,  $SD = 0.750$ ;  $t_{(374)} = -1.72$ ,  $p = 0.087$ , 95% CI [−0.33, 0.02]).

To isolate the effects of OT, a linear regression model using group type (previously-formed or randomly-formed) and binary indicator for OT+ or OT– to explain DM1 in-group bias (in-group transfer minus out-group transfer) was estimated for both decisions tasks. Age and gender were included as covariates. OT+ was negatively related to in-group bias across both tasks ( $R^2 = 0.03$ ,  $F_{(4,368)} = 3.06$ ,  $p = 0.017$ ;  $b = -0.293$ ,  $\beta = -0.12$ ,  $t_{(368)} = -2.40$ ,  $p = 0.017$ ). Age and gender were not significant and the OT+ indicator continues to be significant without their inclusion. Group type was also insignificant ( $p = 0.31$ ).

When GROUPEID was added to the regression model, it significantly increased in-group bias ( $R^2 = 0.04$ ,  $F_{(5,324)} = 2.32$ ,  $p = 0.043$ ;  $b = 0.174$ ,  $\beta = 0.12$ ,  $t_{(324)} = 2.18$ ,  $p = 0.03$ ) even though GROUPEID and the OT indicator are not correlated ( $r = -0.069$ ,  $p = 0.209$ ). We also tested the role of religion on bias since some of the previously-formed groups had religious members. We created the indicator variable REL that took the value of 1 if the participant's score on the RCI exceeded the median. The group-type indicator was



**FIGURE 2 |** OT increased from baseline due to social interactions by 157% for those in the randomly-formed group (R) while participants in the previously-formed group (P) had an OT increase of 57%. The change in OT for the R group is significantly larger than for the P group ( $p = 0.012$ ). Bars shown are standard errors.



**FIGURE 3 |** OT+ participants have identical average transfers to in-group and out-group members in the ultimatum game (UG) while OT- participants show an average bias of 6.2% (\$0.31) towards in-group members. Bars are standard errors.

dropped from model because of its high correlation with REL ( $r = 0.796$ ,  $p < 0.001$ ) and the model was re-estimated. REL was insignificant ( $\beta = 0.07$ ,  $p = 0.163$ ) while the OT indicator remained significant ( $\beta = -0.13$ ,  $p = 0.015$ ). Average

values for GROUPID, REL, closeness to those in one's group (IOS) or mood (PANAS) at baseline, after the group task, or pre-to-post change showed no differences when comparing OT+ participants to OT- ones. We examined the degree of



group identification required to overwhelm the impact of a positive change in OT producing a bias towards one's in-group. Using the regression of in-group bias on the OT+ indicator and GROUPID, in-group bias occurs when GROUPID is one standard deviation above the mean, or 87% of its maximum value.

We also tested if personality traits might vary across the OT+ and OT− groups and might affect our findings. We found that, on average, those in the OT+ group were less agreeable (OT+:  $M = 4.01$ ,  $SD = 0.60$ , OT−:  $M = 4.16$ ,  $SD = 0.63$ ;  $t_{(375)} = 2.29$ ,  $p = 0.022$ ), were less neurotic (OT+:  $2.41$ , OT−:  $2.67$ ;  $t_{(375)} = 3.16$ ,  $p = 0.002$ ), reported less empathic concern (OT+:  $M = 3.85$ ,  $SD = 0.64$ , OT−:  $M = 4.04$ ,  $SD = 0.60$ ;  $t_{(374)} = 2.87$ ,  $p = 0.004$ ), and more personal distress (OT+:  $M = 2.46$ ,  $SD = 0.70$ , OT−:  $M = 2.64$ ,  $SD = 0.70$ ;  $t_{(376)} = 2.40$ ,  $p = 0.017$ ). When traits were added to the linear regression model, none of the trait variables were significant ( $ps > 0.15$ ) and OT+ and GROUPID continued to be significant and had similar beta coefficients to the regression without the trait measures.

## CONCLUSIONS AND DISCUSSION

The present study investigated the relationship between in-group bias and endogenous OT in a non-competitive environment using previously-established groups and randomly-formed groups. Research using exogenous OT administration has suggested that OT increases in-group bias in competitive contexts (De Dreu et al., 2011; De Dreu, 2012) but may decrease bias when competition is not explicit (Israel et al., 2012; Shamay-Tsoory et al., 2013; Huang et al., 2015). Whether the endogenous release of OT affects group bias was an open question, with only a few studies on the topic (Levy et al., 2016; Samuni et al., 2017). We found that half of the 399 participants had a positive increase in endogenous OT after a group activity and OT+ participants showed no bias as DM1 or DM2 in the UG, though they did show bias in the DG. OT− participants were biased as DM1 in both decision tasks. While the UG is a bilateral social interaction in which both parties make choices, in the DG only one person makes a decision. Indeed, transfers in the DG do not appear to be affected by OT infusion (Zak et al., 2007; Barraza et al., 2011) perhaps because the other person's needs do not need to be considered in relation to the self.

Our results show that the effect of OT on bias is context-dependent (Bartz et al., 2011). Endogenous OT, even when group membership was made salient across the various types of groups we studied, seems to generally reduce group differences, although not fully eliminate bias when group identification was high (87% of maximum value or higher). As argued by others (Shamay-Tsoory and Abu-Akel, 2016), OT may benefit out-group members when there is a strong social cue, or when group status is highly-charged as in the Israeli-Palestinian conflict (Shamay-Tsoory et al., 2013). Consistent with a large literature on the prosocial effects of OT, we showed that an increase in endogenous OT eliminated bias in the UG, a task that motivates others to think about the

other player whether in-group or out-group. This was true for both previously-formed groups and randomly-formed groups. Behaviorally, those in P groups had a larger in-group bias than R participants because they identified more strongly with people they already knew or a group to which they belonged. Yet, when OT increased, the bias from being a member of a previously-formed group largely disappeared even though the strength of group identification diminished out-group transfers. This result held even when accounting for personality traits. The motivation for perspective-taking is relatively absent in the DG and bias in the DG was unrelated to OT reactivity. Future studies should examine whether similar social cuing impacts group biases. A related study has shown that there are no in-group/out-group saliency differences during the early stages of information processing (Pfundmair et al., 2017).

There are two caveats when considering research utilizing peripheral plasma measures of OT. First, much like methods in OT administration, OT plasma assays methods have come under criticism. Commercially available immunoassays have been questioned on their validity due to high variability (e.g., McCullough et al., 2013; Christensen et al., 2014; Rutigliano et al., 2016). These same authors report that using the methods utilized in this study (radioimmunoassay along with an extraction step) reduces this high variability (e.g., Szeto et al., 2011; Christensen et al., 2014). A second concern in measuring peripheral plasma OT is in attributing the levels to central OT (McCullough et al., 2013). The most recent meta-analysis has found that peripheral and central OT concentrations are positively correlated, but only after an environmental stimulus and not under basal conditions (Valstad et al., 2017). Future research is needed to identify the types of environmental stimuli that lead to a connection between peripheral and central OT concentrations.

The present study also advances knowledge about group bias by using a large and diverse participant population, tested in two locations, and using ecologically valid group tasks to make group membership salient. This approach increases the likelihood that our results will replicate. This is especially important given the small effect sizes noted in exogenous OT infusion studies (Walum et al., 2016). Additional research should also test participants from non-Western societies to see how OT modulates group biases because of differences found in the behavioral expression of the OT receptor system across ethnicities (Kim et al., 2010).

## AUTHOR CONTRIBUTIONS

PJZ and JS: funded, designed, analyzed and wrote the study. ETT and LEB: executed study, analyzed data and wrote the findings. JAB: executed, designed, analyzed and wrote the study.

## ACKNOWLEDGMENTS

This research was supported by grant #153751 from the John Templeton Foundation to JS and PJZ. We also thank Hillary Lenfesty for technical support.

## REFERENCES

- Algoe, S. B., Kurtz, L. E., and Grewen, K. (2017). Oxytocin and social bonds: the role of oxytocin in perceptions of romantic partners' bonding behavior. *Psychol. Sci.* 28, 1763–1772. doi: 10.1177/0956797617716922
- Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartley, S. L., and Covert, A. E. (2004). Neural signals for the detection of unintentional race bias. *Psychol. Sci.* 15, 88–93. doi: 10.1111/j.0963-7214.2004.01502003.x
- Aron, A., Aron, E. N., and Smollan, D. (1992). Inclusion of other in the self-scale and the structure of interpersonal closeness. *J. Pers. Soc. Psychol.* 63, 596–612. doi: 10.1037/0022-3514.63.4.596
- Aydogan, G., Jobst, A., D'Ardenne, K., Müller, N., and Kocher, M. G. (2017). The detrimental effects of oxytocin-induced conformity on dishonesty in competition. *Psychol. Sci.* 28, 751–759. doi: 10.1177/0956797617695100
- Barraza, J. A., McCullough, M. E., Ahmadi, S., and Zak, P. J. (2011). Oxytocin infusion increases charitable donations regardless of monetary resources. *Horm. Behav.* 60, 148–151. doi: 10.1016/j.yhbeh.2011.04.008
- Barraza, J. A., and Zak, P. J. (2009). Empathy toward strangers triggers oxytocin release and subsequent generosity. *Ann. N Y Acad. Sci.* 1167, 182–189. doi: 10.1111/j.1749-6632.2009.04504.x
- Barraza, J., and Zak, P. J. (2013). "Oxytocin instantiates empathy and produces prosocial behaviors, Chapter 18," in *Oxytocin, Vasopressin and Related Peptides in the Regulation of Behavior*, eds E. Choleris, D. W. Pfaff and M. Kavaliers (Cambridge, MA: Cambridge University Press), 331–342.
- Bartz, J. A., Zaki, J., Bolger, N., and Ochsner, K. N. (2011). Social effects of oxytocin in humans: context and person matter. *Trends Cogn. Sci.* 15, 301–309. doi: 10.1016/j.tics.2011.05.002
- Bielsky, I. F., and Young, L. J. (2004). Oxytocin, vasopressin, and social recognition in mammals. *Peptides* 25, 1565–1574. doi: 10.1016/j.peptides.2004.05.019
- Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: a cognitive-motivational analysis. *Psychol. Bull.* 86, 307–324. doi: 10.1037/0033-2909.86.2.307
- Brewer, M. B. (1999). The psychology of prejudice: in-group love or out-group hate? *J. Soc. Issues* 55, 429–444. doi: 10.1111/0022-4537.00126
- Camerer, C. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.
- Carter, C. S. (2014). Oxytocin pathways and the evolution of human behavior. *Annu. Rev. Psychol.* 65, 17–39. doi: 10.1146/annurev-psych-010213-115110
- Christensen, J. C., Shiyonov, P. A., Estepp, J. R., and Schlager, J. J. (2014). Lack of association between human plasma oxytocin and interpersonal trust in a prisoner's dilemma paradigm. *PLoS One* 9:e116172. doi: 10.1371/journal.pone.0116172
- Churchland, P. S., and Winkielman, P. (2012). Modulating social behavior with oxytocin: how does it work? What does it mean? *Horm. Behav.* 61, 392–399. doi: 10.1016/j.yhbeh.2011.12.003
- Crockford, C., Deschner, T., Ziegler, T. E., and Wittig, R. M. (2014). Endogenous peripheral oxytocin measures can give insight into the dynamics of social relationships: a review. *Front. Behav. Neurosci.* 8:68. doi: 10.3389/fnbeh.2014.00068
- Daughters, K., Manstead, A. S., Ten Velden, F. S., and De Dreu, C. K. (2017). Oxytocin modulates third-party sanctioning of selfish and generous behavior within and between groups. *Psychoneuroendocrinology* 77, 18–24. doi: 10.1016/j.psyneuen.2016.11.039
- De Dreu, C. K. (2012). Oxytocin modulates cooperation within and competition between groups: an integrative review and research agenda. *Horm. Behav.* 61, 419–428. doi: 10.1016/j.yhbeh.2011.12.009
- De Dreu, C. K., Greer, L. L., Handgraaf, M. J., Shalvi, S., Van Kleef, G. A., Baas, M., et al. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science* 328, 1408–1411. doi: 10.1126/science.1189047
- De Dreu, C. K., Greer, L. L., Van Kleef, G. A., Shalvi, S., and Handgraaf, M. J. (2011). Oxytocin promotes human ethnocentrism. *Proc. Natl. Acad. Sci. U S A* 108, 1262–1266. doi: 10.1073/pnas.1015316108
- De Dreu, C. K. W., Shalvi, S., Greer, L. L., Van Kleef, G. A., and Handgraaf, M. J. J. (2012). Oxytocin motivates non-cooperation in intergroup conflict to protect vulnerable in-group members. *PLoS One* 7:e46751. doi: 10.1371/journal.pone.0046751
- Declerck, C. H., Boone, C., and Kiyonari, T. (2010). Oxytocin and cooperation under conditions of uncertainty: the modulating role of incentives and social information. *Horm. Behav.* 57, 368–374. doi: 10.1016/j.yhbeh.2010.01.006
- Declerck, C. H., Boone, C., and Kiyonari, T. (2014). The effect of oxytocin on cooperation in a prisoner's dilemma depends on the social context and a person's social value orientation. *Soc. Cogn. Affect. Neurosci.* 9, 802–809. doi: 10.1093/scan/nst040
- Dunham, Y., Baron, A. S., and Carey, S. (2011). Consequences of "minimal" group affiliations in children. *Child Dev.* 82, 793–811. doi: 10.1111/j.1467-8624.2011.01577.x
- Ford, T. E., and Stangor, C. (1992). The role of diagnosticity in stereotype formation: perceiving group means and variances. *J. Pers. Soc. Psychol.* 63, 356–367. doi: 10.1037/0022-3514.63.3.356
- Goette, L., Huffman, D., and Meier, S. (2012). The impact of social ties on group interactions: evidence from minimal groups and randomly assigned real groups. *Am. Econ. J. Microecon.* 4, 101–115. doi: 10.1257/mic.4.1.101
- Gonzaga, G. C., Turner, R. A., Keltner, D., Campos, B., and Altemus, M. (2006). Romantic love and sexual desire in close relationships. *Emotion* 6, 163–179. doi: 10.1037/1528-3542.6.2.163
- Hein, G., Silani, G., Preuschoff, K., Batson, C. D., and Singer, T. (2010). Neural responses to in-group and out-group members' suffering predict individual differences in costly helping. *Neuron* 68, 149–160. doi: 10.1016/j.neuron.2010.09.003
- Hogg, M. A., Hains, S. C., and Mason, I. (1998). Identification and leadership in small groups: salience, frame of reference, and leader stereotypicality effects on leader evaluations. *J. Pers. Soc. Psychol.* 75, 1248–1263. doi: 10.1037/0022-3514.75.5.1248
- Hostinar, C. E., Sullivan, R. M., and Gunnar, M. R. (2014). Psychobiological mechanisms underlying the social buffering of the hypothalamic-pituitary-adrenocortical axis: a review of animal models and human studies across development. *Psychol. Bull.* 140, 256–282. doi: 10.1037/a0032671
- Huang, Y., Kendrick, K. M., Zheng, H., and Yu, R. (2015). Oxytocin enhances implicit social conformity to both in-group and out-group opinions. *Psychoneuroendocrinology* 60, 114–119. doi: 10.1016/j.psyneuen.2015.06.003
- Huffmeijer, R., Van Ijzendoorn, M. H., and Bakermans-Kranenburg, M. J. (2013). Ageing and oxytocin: a call for extending human oxytocin research to ageing populations—a mini-review. *Gerontology* 59, 32–39. doi: 10.1159/000341333
- Hurlemann, R., Patin, A., Onur, O. A., Cohen, M. X., Baumgartner, T., Metzler, S., et al. (2010). Oxytocin enhances amygdala-dependent, socially reinforced learning and emotional empathy in humans. *J. Neurosci.* 30, 4999–5007. doi: 10.1523/JNEUROSCI.5538-09.2010
- Israel, S., Lerer, E., Shalev, I., Uzevovsky, F., Riebold, M., Laiba, E., et al. (2009). The oxytocin receptor (OXTR) contributes to prosocial fund allocations in the dictator game and the social value orientations task. *PLoS One* 4:e5535. doi: 10.1371/journal.pone.0005535
- Israel, S., Weisel, O., Ebstein, R. P., and Bornstein, G. (2012). Oxytocin, but not vasopressin, increases both parochial and universal altruism. *Psychoneuroendocrinology* 37, 1341–1344. doi: 10.1016/j.psyneuen.2012.02.001
- Kim, H. S., Sherman, D. K., Sasaki, J. Y., Xu, J., Chu, T. Q., Ryu, C., et al. (2010). Culture, distress and oxytocin receptor polymorphism (OXTR) interact to influence emotional support seeking. *Proc. Natl. Acad. Sci. U S A* 107, 15717–15721. doi: 10.1073/pnas.1010830107
- Knutson, K. M., Mah, L., Manly, C. F., and Grafman, J. (2007). Neural correlates of automatic beliefs about gender and race. *Hum. Brain Mapp.* 28, 915–930. doi: 10.1002/hbm.20320
- Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., and Fehr, E. (2005). Oxytocin increases trust in humans. *Nature* 435, 673–676. doi: 10.1038/nature03701
- LaBouff, J. P., Rowatt, W. C., Johnson, M. K., and Finkle, C. (2012). Differences in attitudes toward outgroups in religious and nonreligious contexts in a multinational sample: a situational context priming study. *Int. J. Psychol. Relig.* 22, 1–9. doi: 10.1080/10508619.2012.634778
- Lane, A., Luminet, O., Nave, G., and Mikolajczak, M. (2016). Is there a publication bias in behavioural intranasal oxytocin research on humans? Opening the

- file drawer of one laboratory. *J. Neuroendocrinol.* 28:4. doi: 10.1111/jne.12384
- Lemyre, L., and Smith, P. M. (1985). Intergroup discrimination and self-esteem in the minimal group paradigm. *J. Pers. Soc. Psychol.* 49, 660–670. doi: 10.1037/0022-3514.49.3.660
- Leng, G., and Ludwig, M. (2016). Intranasal oxytocin: myths and delusions. *Biol. Psychiatry* 79, 243–250. doi: 10.1016/j.biopsych.2015.05.003
- Levy, J., Goldstein, A., Inliss, M., Masalha, S., Zagoory-Sharon, O., and Feldman, R. (2016). Adolescents growing up amidst intractable conflict attenuate brain response to pain of out-group. *Proc. Nat. Acad. Sci. U S A* 113, 13696–13701. doi: 10.1073/pnas.1612903113
- Mackie, D. M., and Hamilton, D. L. (Eds.). (1993). *Affect, Cognition and Stereotyping: Interactive Processes in Group Perception*. San Diego, CA: Academic Press, Inc.
- McCullough, M. E., Churchland, P. S., and Mendez, A. J. (2013). Problems with measuring peripheral oxytocin: can the data on oxytocin and human behavior be trusted? *Neurosci. Biobehav. Rev.* 37, 1485–1492. doi: 10.1016/j.neubiorev.2013.04.018
- Morhenn, V. B., Park, J. W., Piper, E., and Zak, P. J. (2008). Monetary sacrifice among strangers is mediated by endogenous oxytocin release after physical contact. *Evol. Hum. Behav.* 29, 375–383. doi: 10.1016/j.evolhumbehav.2008.04.004
- Nave, G., Camerer, C., and McCullough, M. (2015). Does oxytocin increase trust in humans? A critical review of research. *Perspect. Psychol. Sci.* 10, 772–789. doi: 10.1177/1745691615600138
- Neumann, I. D., Maloumy, R., Beiderbeck, D. I., Lukas, M., and Landgraf, R. (2013). Increased brain and plasma oxytocin after nasal and peripheral administration in rats and mice. *Psychoneuroendocrinology* 38, 1985–1993. doi: 10.1016/j.psyneuen.2013.03.003
- Penner, L. A., Dovidio, J. F., Piliavin, J. A., and Schroeder, D. A. (2005). Prosocial behavior: multilevel perspectives. *Annu. Rev. Psychol.* 56, 365–392. doi: 10.1146/annurev.psych.56.091103.070141
- Pfundmair, M., Zwarg, C., Paulus, M., and Rimpel, A. (2017). Oxytocin promotes attention to social cues regardless of group membership. *Horm. Behav.* 90, 136–140. doi: 10.1016/j.yhbeh.2017.03.006
- Prentice, D. A., Miller, D., and Lightdale, J. R. (1994). Asymmetries in attachment to groups and to their members: distinguishing between common-identity and common-bond groups. *Pers. Soc. Psychol. Bull.* 20, 484–493. doi: 10.1177/0146167294205005
- Rutigliano, G., Rocchetti, M., Paloyelis, Y., Gillean, J., Sardella, A., Cappucciati, M., et al. (2016). Peripheral oxytocin and vasopressin: biomarkers of psychiatric disorders? A comprehensive systematic review and preliminary meta-analysis. *Psychiatry Res.* 241, 207–220. doi: 10.1016/j.psychres.2016.04.117
- Rydén, G., and Sjöholm, I. (1969). Half-life of oxytocin in blood of pregnant and non-pregnant women. *Acta Endocrinol.* 61, 425–431. doi: 10.1530/acta.0.0610425
- Samuni, L., Preis, A., Mundry, R., Deschner, T., Crockford, C., and Wittig, R. M. (2017). Oxytocin reactivity during intergroup conflict in wild chimpanzees. *Proc. Natl. Acad. Sci. U S A* 114, 268–273. doi: 10.1073/pnas.1616812114
- Shamay-Tsoory, S. G., and Abu-Akel, A. (2016). The social salience hypothesis of oxytocin. *Biol. Psychiatry* 79, 194–202. doi: 10.1016/j.biopsych.2015.07.020
- Shamay-Tsoory, S. G., Abu-Akel, A., Palgi, S., Suleiman, R., Fischer-Shofty, M., Levkovitz, Y., et al. (2013). Giving peace a chance: oxytocin increases empathy to pain in the context of the Israeli-Palestinian conflict. *Psychoneuroendocrinology* 38, 3139–3144. doi: 10.1016/j.psyneuen.2013.09.015
- Smith, J. M. (1998). The origin of altruism. *Nature* 393, 639–640. doi: 10.1038/31383
- Szeto, A., McCabe, P. M., Nation, D. A., Tabak, B. A., Rossetti, M. A., McCullough, M. E., et al. (2011). Evaluation of enzyme immunoassay and radioimmunoassay methods for the measurement of plasma oxytocin. *Psychosom. Med.* 73, 393–400. doi: 10.1097/psy.0b013e31821df0c2
- Ten Velden, F. S., Daughters, K., and De Dreu, C. K. W. (2017). Oxytocin promotes intuitive rather than deliberated cooperation with the in-group. *Horm. Behav.* 92, 164–171. doi: 10.1016/j.yhbeh.2016.06.005
- Valstad, M., Alvares, G. A., Egknud, M., Matziorinis, A. M., Andreassen, O. A., Westlye, L. T., et al. (2017). The correlation between central and peripheral oxytocin concentrations: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* 78, 117–124. doi: 10.1016/j.neubiorev.2017.04.017
- Van Bavel, J. J., Packer, D. J., and Cunningham, W. A. (2008). The neural substrates of in-group bias: a functional magnetic resonance imaging investigation. *Psychol. Sci.* 19, 1131–1139. doi: 10.1111/j.1467-9280.2008.02214.x
- Van Ijzendoorn, M. H., and Bakermans-Kranenburg, M. J. (2012). A sniff of trust: meta-analysis of the effects of intranasal oxytocin administration on face recognition, trust to in-group and trust to out-group. *Psychoneuroendocrinology* 37, 438–443. doi: 10.1016/j.psyneuen.2011.07.008
- Wacker, D. W., and Ludwig, M. (2012). Vasopressin, oxytocin and social odor recognition. *Horm. Behav.* 61, 259–265. doi: 10.1016/j.yhbeh.2011.08.014
- Walum, H., Waldman, I. D., and Young, L. J. (2016). Statistical and methodological considerations for the interpretation of intranasal oxytocin studies. *Biol. Psychiatry* 79, 251–257. doi: 10.1016/j.biopsych.2015.06.016
- Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070. doi: 10.1037/0022-3514.54.6.1063
- Weller, D., and Hansen Lagattuta, K. (2013). Helping the in-group feels better: children's judgments and emotion attributions in response to prosocial dilemmas. *Child Dev.* 84, 253–268. doi: 10.1111/j.1467-8624.2012.01837.x
- Worthington, E. L. Jr., Wade, N. G., Hight, T. L., Ripley, J. S., McCullough, M. E., Berry, J. W., et al. (2003). The religious commitment inventory–10: development, refinement, and validation of a brief scale for research and counseling. *J. Couns. Psychol.* 50, 84–96. doi: 10.1037/0022-0167.50.1.84
- Zak, P. J. (2012). *The Moral Molecule: the Source of Love and Prosperity*. New York, NY: Random House.
- Zak, P. J., Kurzban, R., and Matzner, W. T. (2005). Oxytocin is associated with human trustworthiness. *Horm. Behav.* 48, 522–527. doi: 10.1016/j.yhbeh.2005.07.009
- Zak, P. J., Matzner, W. T., and Kurzban, R. (2004). The neurobiology of trust. *Ann. N Y Acad. Sci.* 1032, 224–227. doi: 10.1196/annals.1314.025
- Zak, P. J., Stanton, A. A., and Ahmadi, S. (2007). Oxytocin increases generosity in humans. *PLoS One* 2:e1128. doi: 10.1371/journal.pone.0001128

**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 © 2018 Terris, Beavin, Barraza, Schloss and Zak. 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) and the copyright owner 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.



# No Evidence for a Relationship Between Hair Testosterone Concentrations and 2D:4D Ratio or Risk Taking

Richard Ronay<sup>1\*</sup>, Leander van der Meij<sup>2</sup>, Janneke K. Oostrom<sup>3</sup> and Thomas V. Pollet<sup>4</sup>

<sup>1</sup>Department of Leadership and Management, Faculty of Economics and Business, University of Amsterdam, Amsterdam, Netherlands, <sup>2</sup>Department of Industrial Engineering, Eindhoven University of Technology, Eindhoven, Netherlands,

<sup>3</sup>Department of Management & Organization, School of Business and Economics, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, <sup>4</sup>Department of Psychology, Faculty of Health and Life Sciences, Northumbria University, Newcastle, United Kingdom

Using a recently developed alternative assay procedure to measure hormone levels from hair samples, we examined the relationships between testosterone, cortisol, 2D:4D ratio, overconfidence and risk taking. A total of 162 (53 male) participants provided a 3 cm sample of hair, a scanned image of their right and left hands from which we determined 2D:4D ratios, and completed measures of overconfidence and behavioral risk taking. While our sample size for males was less than ideal, our results revealed no evidence for a relationship between hair testosterone concentrations, 2D:4D ratios and risk taking. No relationships with overconfidence emerged. Partially consistent with the Dual Hormone Hypothesis, we did find evidence for the interacting effect of testosterone and cortisol on risk taking but only in men. Hair testosterone concentrations were positively related to risk taking when levels of hair cortisol concentrations were low, in men. Our results lend support to the suggestion that endogenous testosterone and 2D:4D ratio are unrelated and might then exert diverging activating vs. organizing effects on behavior. Comparing our results to those reported in the existing literature we speculate that behavioral correlates of testosterone such as direct effects on risk taking may be more sensitive to state-based fluctuations than baseline levels of testosterone.

**Keywords:** testosterone, cortisol, hair samples, 2D:4D ratio, risk taking, dual hormone hypothesis

## OPEN ACCESS

### Edited by:

Levent Neyse,  
Institut für Weltwirtschaft, Germany

### Reviewed by:

Siegfried Dewitte,  
KU Leuven, Belgium  
Roel Van Veldhuizen,  
Social Science Research Center  
Berlin, Germany

### \*Correspondence:

Richard Ronay  
r.ronay@uva.nl

**Received:** 11 September 2017

**Accepted:** 08 February 2018

**Published:** 05 March 2018

### Citation:

Ronay R, van der Meij L, Oostrom JK and Pollet TV (2018) No Evidence for a Relationship Between Hair Testosterone Concentrations and 2D:4D Ratio or Risk Taking. *Front. Behav. Neurosci.* 12:30. doi: 10.3389/fnbeh.2018.00030

## INTRODUCTION

Although studies have documented a positive relationship between testosterone and risky economic decisions, the evidence has been inconsistent, with linear (Apicella et al., 2008), non-linear (Stanton et al., 2011) and null relationships (Zethraeus et al., 2009). One explanation for these inconsistencies could be the failure to distinguish between measurements of state-based levels of testosterone and the measurement of more trait-like (baseline) levels of testosterone. The majority of studies exploring the relationships between testosterone and risk taking have measured state-based levels of testosterone via saliva samples. This lends itself to experimental studies seeking to test the contextual role of fluctuations in testosterone on behavior. However, studies that aim to test for relationships between baseline endogenous testosterone levels are potentially confounded by these same contextually bound fluctuations when using saliva samples.



In the current study we measure testosterone using a recently developed alternative assay procedure in which hormone levels are assayed from hair samples. Hair samples should provide a stronger test of the relationship between baseline levels of testosterone and risk taking, as hair samples indicate average fluctuating testosterone levels across 3 months and thus filter out contextual noise in hormone measurements. As per the Dual Hormone Hypothesis (Mehta and Josephs, 2010), we test both the direct effect of hair testosterone concentrations on risk taking and its interaction effect with hair cortisol concentrations. Contributing to the research on the relationships between different hormone measurements, we also examine the relationship between hair sample testosterone and an often used measure of prenatal testosterone, the 2D:4D ratio—the relative length of the index finger (2D) and the ring finger (4D) (Manning, 2002).

Two influential and complementary theoretical models that have been offered as explanatory frameworks for understanding the dynamic relationship between testosterone and social behavior are the Challenge Hypothesis (Wingfield et al., 1990; Archer, 2006) and the Biosocial Model of Status (Mazur, 1985; Mazur and Booth, 1998). The Challenge Hypothesis posits that testosterone motivates resource and mate-seeking behaviors, including those associated with aggression and competition, when the social context deems such behaviors as *reproductively beneficial* for the organism. Similarly, the Biosocial Model of Status states that testosterone encourages competitive behaviors that serve the function of increasing status. In support of these frameworks, testosterone has been repeatedly linked to competitive, dominance- and status-seeking behaviors in human and non-human males. For instance, the males of many species show increased competitive behaviors during breeding season when testosterone levels are known to peak (Harding, 1981; Balthazart, 1983; Wingfield et al., 1990; Denson et al., 2013), with similar hormonal (Van der Meij et al., 2010) and behavioral (Ronay and von Hippel, 2010) responses to mating competition among human males (for a review in humans, see Eisenegger et al., 2011).

One way in which testosterone might fuel competition is via an increased tolerance for risk. Although the literature does not offer a consistent picture of the relationship between endogenous testosterone and risk taking, a number of studies have reported positive relationships. For instance, Apicella et al. (2008) reported a positive linear relationship between testosterone and financial risk taking in a sample of Harvard undergraduate men. Similarly, Coates and Herbert (2008) reported a positive relationship between testosterone and the day to day returns of London financial traders. Sapienza et al. (2009) found a positive relationship between testosterone and risk taking for women, though not men. Ronay and von Hippel (2010) reported that adult male skateboarders' testosterone levels, measured in the context of sexual competition primed by the presence of an attractive female experimenter, are positively associated with physical risk taking. Last, Stanton et al. (2011) found a non-linear relationship—both low and high testosterone predicted greater risk taking—among men and women. Taken together, the empirical evidence suggests an

intriguing but inconsistent relationship between testosterone and risk taking.

Similarly, the published work exploring the relationship between exogenously administered testosterone and risk taking consists of a small collection of intriguing but inconsistent findings. Although two administration studies involving only women found no evidence for a causal relationship between testosterone and economic risk preferences (Zethraeus et al., 2009; Boksem et al., 2013), testosterone administration has been shown to increase women's risk taking on the Iowa gambling task (Van Honk et al., 2003). However, another study involving pharmacological manipulations in men found that higher testosterone levels were associated with increased risk seeking as measured via the balloon analog risk task (BART; Lejuez et al., 2002), but not in the Iowa gambling task or a dice task (Goudriaan et al., 2010).

Although results are mixed, the theoretical foundations (Mazur, 1985; Wingfield et al., 1990; Mazur and Booth, 1998; Archer, 2006) that have inspired these empirical tests seem sound, and comparative studies among non-human animals (Rose et al., 1971; Rada et al., 1976; Harding, 1981; Schwabl and Kriner, 1991; Wingfield and Hahn, 1994) provide corroborating support for a relationship between testosterone and competitive behaviors in general. Ancillary evidence is also suggestive of such a positive relationship. For instance, men's higher testosterone levels relative to women (e.g., Pollet et al., 2011; Ronay and Carney, 2013), and a robust age-related decline in testosterone (Harman et al., 2001) map onto reliable sex differences in risk taking (Byrnes et al., 1999; Ronay and Kim, 2006), and age-related declines in risk taking (Kaufman and Vermeulen, 2005). The inconsistency of the empirical work therefore represents something of a puzzle for researchers seeking to understand the behavioral effects of testosterone.

Testosterone not only has activating effects that emerge from endogenous circulating levels of the hormone, but prenatal testosterone also manifests organizing effects that shape how the brain and body develop (Manning, 2002). One putative marker of *in utero* androgen exposure is the 2D:4D ratio, with lower ratios indicating exposure to higher levels of androgens during prenatal development (Manning, 2002). Lutchmaya et al. (2004) examined the relationship between the 2D:4D ratios of 33 children at age two, and the level of fetal testosterone (measured via amniocentesis) they were exposed to during the second trimester of their gestation. They reported a strong negative relationship between digit ratios and fetal testosterone levels.

Evidence for a negative relationship between 2D:4D ratio and endogenous levels of circulating testosterone during adulthood is less persuasive. Although Manning et al. (1998) report a significant negative relationship between 2D:4D ratio and endogenous testosterone levels of 58 men, further investigations (Campbell et al., 2010; Sanchez-Pages and Turiegano, 2010) have been unable to reproduce this effect and a meta-analysis (Hönekopp et al., 2007) also suggests no robust effect.

Nonetheless, the conceptual overlap between the two measures has motivated a number of researchers to examine the behavioral effects of 2D:4D ratio in contexts where theory suggests testosterone should play a role, with conceptually consistent results (Bailey and Hurd, 2005; Van den Bergh and Dewitte, 2006; Voracek et al., 2006; Millet and Dewitte, 2009; Ronay and von Hippel, 2010; Ronay and Galinsky, 2011; Ronay et al., 2012). Irrespective of the likely surfeit of failed studies in this vein that remain buried in file drawers, the conceptual consistency between the effects of 2D:4D ratio and testosterone on behavior, coupled with the lack of empirical support for a reliable relationship between the two produces yet another puzzle of interest. To explore one possible solution to this puzzle, we turned our attention to the method by which testosterone levels are most commonly measured.

Testosterone levels vary across the day (Granger et al., 1999) as well as in response to a range of social contextual factors (Mehta and Josephs, 2006; Van der Meij et al., 2008, 2010). Endogenous testosterone levels vary even in response to partisan alignment following presidential election outcomes (Stanton et al., 2009), and football team affiliation following match day (Van der Meij et al., 2012). This has obvious advantages for researchers seeking to test the contextual role of fluctuations in testosterone on behavior (e.g., Ronay and von Hippel, 2010; Apicella et al., 2014), such as would be predicted by The Challenge Hypothesis (Wingfield et al., 1990; Archer, 2006) and the Biosocial Model of Status (Mazur, 1985; Mazur and Booth, 1998). However, studies seeking to test the relationships between *baseline* endogenous testosterone levels and other variables—such as 2D:4D ratio and risk taking—are disadvantaged by these same contextually bound fluctuations. This problem is exacerbated by the fact that much of the published research, samples testosterone levels at a single time point, rather than via multiple measures that might lead to a more accurate and stable measure of baseline testosterone. Thus, one possible contributing factor to the inconsistent effects of testosterone on risk taking, and the relationship between 2D:4D ratio and circulating testosterone, may be the failure to distinguish between measurements of state-based levels of testosterone—such as are derived from single time point measures—and the more stable, trait-like levels of testosterone—such as might be captured by aggregating across multiple time points.

Mehta and Josephs (2010) have proposed the Dual Hormone Hypothesis, which posits that testosterone's role in status-relevant behavior should depend on concentrations of cortisol, a hormone that is released in response to physical and/or psychological stress. Specifically, the Dual Hormone Hypothesis predicts that behavioral effects follow from an interaction between testosterone and cortisol—testosterone should be positively related to status-seeking behaviors only when cortisol concentrations are low. According to the model, when cortisol concentrations are high, status-seeking behaviors should be inhibited. The predictions of the model have been demonstrated on a range of dependent variables including risk taking (Mehta et al., 2015), self-reported aggression (Popma et al., 2007; Denson et al., 2013) and retrospectively in

juvenile crime (Dabbs et al., 1991). However, in keeping with the majority of the endocrinological literature, these tests of the Dual Hormone Hypothesis have relied upon isolated single time point measures of both testosterone and cortisol.

The goal of the current research was to reexamine the relationships between baseline testosterone, 2D:4D ratios, and risk taking, using a recently developed alternative assay procedure in which testosterone levels are assayed from hair samples using an liquid chromatography tandem mass spectrometry method (LC-MS/MS)-based method. We measured cortisol simultaneously so as to test for possible interacting effects of testosterone and cortisol on risk taking, as per the Dual Hormone Hypothesis (Mehta and Josephs, 2010). As testosterone (Johnson et al., 2006; Ronay et al., 2017) has been suggested to facilitate higher levels of overconfidence, and overconfidence has been linked to risk taking (Miller and Byrnes, 1997; Camerer and Lovo, 1999; Campbell et al., 2004; Malmendier and Tate, 2008) we also measured participants' overconfidence in order to examine the possibility of these relationships with hair testosterone concentrations.

## MATERIALS AND METHODS

### Participants

Participants were 162 non-psychology students (53 male, 109 female;  $M_{\text{age}} = 22.05$ ,  $SD_{\text{age}} = 2.85$ ) from the Vrije Universiteit Amsterdam. Participants received 8 € for their participation. Prior to analysis we made a decision to exclude 14 participants due to incomplete measures or measurement error. Initial analysis of the hair samples revealed five cases to be outside of known measurement limits, suggesting unacceptable noise in the assaying, and so these cases were excluded from further analyses. Three further cases reported medical histories known to directly affect hormones (Polycystic ovary syndrome, Betamethason medication and cancer treatment), and so these too were excluded from further analyses (Granger et al., 2009). This yielded a final sample of 140 participants (43 male, 97 female;  $M_{\text{age}} = 21.93$ ,  $SD_{\text{age}} = 2.88$ ). We acknowledge that our final sample size for males is less than our initial goal of 100 males and 100 females, thus tempering the strength of our conclusions.

### Procedure

The study was approved by the Scientific and Ethical Review Board (VCWE) of the Vrije Universiteit Amsterdam. Participants first read an informed consent form and provided written consent for their participation. Participants then provided demographic and health information. To assess risk taking, participants completed the BART (Lejuez et al., 2002). In addition, they completed measures on self-esteem, personality, and sexual behavior, which are not the focus of the current research and thus not discussed here. Participants were then asked to position their hands palm down on a flatbed scanner so as to allow us to capture images of both hands for determining 2D:4D ratios. Finally, hair samples were taken and participants were debriefed and paid.

## Measures

### Hair Samples

Testosterone and cortisol concentrations were determined from hair samples with a LC-MS/MS. This method is considered to be a reliable and precise way to measure testosterone and cortisol concentrations (Gao et al., 2013). Specifically, for these hormones, intra- and inter-assay coefficients of variation are between 3.1% and 8.8% and the limits of quantification (LOQ) are below 0.1 pg/mg (Gao et al., 2013). Hair sampling was done according to the instructions of the laboratory of Biological Psychology at the Technical University of Dresden. Three hair strands were cut with scissors as close as possible from the scalp from a posterior vertex position and tied with a thread. Hair strands were placed in aluminum foils that were put in envelopes. The envelopes were placed in a specially prepared box and sent to the laboratory of biological psychology at the Technical University of Dresden (Germany) for analyses. Steroid concentrations were determined from hair segments 3 cm closest to the scalp, which represents hair grown over the last 3 months prior to sampling when assuming an average hair growth of 1 cm per month (Wennig, 2000).

### 2D:4D Ratio

The lengths of the second and fourth digits were independently measured by two master's students, from the ventral proximal crease of the digit to the tip of the finger using the "Measure" tool in Adobe Photoshop. Digit ratios were calculated by dividing the length of the 4th digit on the hand by the length of the 2nd digit on the same hand (Manning et al., 1998). Measurements were computed in the absence of any other information about the participant. The correlation between the measurers was  $>0.99$ .

### Risk Taking

Risk taking was assessed via the BART (Lejuez et al., 2002). The BART has been shown to possess good test-retest reliability (White et al., 2008) and has been validated against self-reported correlates of risk taking, including psychopathy (Hunt et al., 2005), impulsivity and sensation seeking (Lejuez et al., 2002). Critically, the BART has also been shown to predict a number of real-world risk taking behaviors including cigarette smoking, alcohol use, illicit drug use, gambling and sexual risk taking (Lejuez et al., 2002, 2003; Hopko et al., 2006).

The BART is a computer task in which participants are presented with a series of 30 onscreen balloons and a virtual "pump" that when clicked incrementally expands the size of the current balloon until a randomly determined pop point is reached and the balloon explodes. Participants were presented with a series of 30 balloons and not just a single balloon to increase the reliability of our measurement. Participants were instructed that with each additional pump they would earn 1 cent that would accumulate in a temporary bank, also on screen. However, when a balloon was inflated past its pop point, the balloon exploded and all money earned on that particular balloon would be lost. To guard against this risk, participants could choose to stop at any point by clicking on a "Collect \$\$\$" button, also onscreen, at which point the money in the

temporary bank would be transferred to a permanent bank. The probability that a balloon would explode increased incrementally with each pump—1/128 for the first pump, 1/127 for the second pump, etc., the probability of an explosion on the 128th pump was therefore 1/1. According to this algorithm, the average breakpoint was 64 pumps (Lejuez et al., 2002). Participants received onscreen instruction before the test started but did not receive any information about the probability of the explosion, neither at the start or during the task. Thus, the game creates a tension between securing one's accumulated winnings, against the pursuit of further, albeit diminishing relative returns. As our goal was to measure risk taking behavior and not hypothetical or self-reported risk attitudes, which might capture diverging aspects of risk taking (Battalio et al., 1990; Holt and Laury, 2005; Harrison, 2006; Branas-Garza et al., in press), participants were informed that they would be paid 10 percent of their winnings at the conclusion of the experiment ( $M_{\text{euro}} = 0.76$ ,  $SD = 0.21$ ). However, as this is a rather minimal stake, which may incentivise riskier decisions than in real life (Holt and Laury, 2005), we decided to also inform participants that the participant who accumulated the most money on the BART (30 balloons, across all sessions) would receive a cash prize of 50 € once testing was concluded. Together, these incentives were intended to parallel real world risk taking decisions in which risk taking is rewarded up until a point, after which further riskiness results in poorer outcomes. All participants were paid accordingly. Each participant was presented with 30 virtual balloons and as recommended (Lejuez et al., 2002) the average number of pumps on all unexploded balloons served as our dependent variable.

### Overconfidence

Overconfidence was operationalized as overestimation of one's actual performance (Fischhoff et al., 1977; Kruger and Dunning, 1999; Kruger and Mueller, 2002; Larrick et al., 2007; Moore and Healy, 2008) on an existing General Knowledge Questionnaire (GKQ; Michailova, 2010). We used a previously adapted version (Ronay et al., 2017) of the GKQ (Michailova, 2010; Michailova and Katter, 2014), taking the 18 items from Michailova's (2010) original measure (e.g., *How many days does a hen need to incubate an egg?*) and adding six further items (Ronay et al., 2017). Participants were instructed to choose the correct answer from three alternatives and to provide a number between 33% (chance) and 100% (absolute certainty) indicating their confidence in the accuracy of that answer. Consistent with previous work and as many scholars recommend<sup>1</sup>, we computed overconfidence by regressing participants' confidence scores (i.e., mean confidence ratings) onto their accuracy (i.e., percentage of correctly answered items) and saving the standardized residual scores (DuBois, 1957; Cronbach and Furby, 1970; John and Robins, 1994; Cohen et al., 2003; Anderson et al., 2012). This approach isolates the variance

<sup>1</sup>The use of difference scores has received widespread criticism as difference scores are unreliable and tend to be confounded with variables that constitute the index (e.g., Cronbach and Furby, 1970; Cohen et al., 2003). Scholars have suggested regressing participants' actual performance onto their self-evaluations and retaining the residuals of the self-evaluations (e.g., John and Robins, 1994).

in participants' confidence while controlling for variance in accuracy—i.e., confidence over and above accuracy.

## Statistical Analyses

Our analysis plan was registered on [osf.io: 4h3cd](https://osf.io/4h3cd). We analyzed male and female data separately as the distribution markedly differs between the sexes (Stanton, 2011). Given the skewness we performed a log transformation for testosterone and cortisol concentrations for our core analyses. The analysis plan fully details the analytical strategy as well as the robustness checks employed. Our key analyses are Bayesian Regression Models via the “BRMS” package in R (Buerkner, 2015). The estimation was based on four chains, each containing 2000 iterations (1000 for burn-in) using non-informative priors on all model parameters. We examined convergence via Rhat (close to 1; see ESM) and evaluated model fits via information criteria (WAIC, LOOIC) compared to a null model (intercept only; Vehtari et al., 2017). These differences between models in terms of fit can be roughly interpreted according to the following rules of thumb: with a difference ( $\Delta$ ) of 1–2 units offering little to no support over a null, between 4–7 units offering considerable support for an alternative model, and those with  $>10$  units offer full support for the alternative model (Raftery, 1996; Burnham and Anderson, 2002, 2004). For the final model, we report parameter estimates and 95% credible interval. Other models, additional analyses,

and further details of the robustness checks are reported in the ESM.

## RESULTS

### Descriptive Statistics

The key descriptive statistics and baseline correlations can be found in **Tables 1, 2**. **Figure 1** shows histograms for raw testosterone and cortisol levels. The medians were different between men and women for T (Mood's median test:  $p < 0.0001$ ), but not for C (Mood's median test:  $p = 1$ ). There were no extreme cases in hair testosterone concentrations for men, based on Tukey's interquartile's range (IQR) criterion (Tukey, 1977; Pollet and van der Meij, 2017). Whereas for women there were three extreme cases ( $>3 * \text{IQR}$ ) in hair testosterone concentrations. For hair cortisol concentrations, there was one extreme value in the male data and three extreme values in the female data. Where relevant we reported the results with and without these extreme cases. **Figure 2** shows the distribution of the BART scores.

### 2D:4D Ratio and Testosterone

None of the models provided substantial support for an effect of 2D:4D ratio on hair testosterone concentrations (all models  $\Delta\text{WAIC}$  and  $\Delta\text{LOOIC} < 2.1$ ). In both males ( $r_{\text{left hand}} = -0.25$ ;  $r_{\text{right hand}} = -0.28$ ) and females

**TABLE 1 |** Means, standard deviations and correlations with confidence intervals for male sample.

Variable	M	SD	1	2	3	4	5
1. Testosterone	1.10	0.47					
2. Cortisol	5.64	3.49	0.29 [−0.01, 0.54]				
3. Left hand 2D:4D	0.96	0.03	−0.25 [−0.51, 0.05]	0.01 [−0.29, 0.31]			
4. Right hand 2D:4D	0.95	0.03	−0.28 [−0.54, 0.02]	−0.37* [−0.60, −0.08]	0.49** [0.23, 0.69]		
5. Overconfidence	12.10	12.52	−0.08 [−0.37, 0.23]	−0.06 [−0.35, 0.25]	0.09 [−0.22, 0.38]	−0.20 [−0.47, 0.10]	
6. Risk taking	43.48	12.46	−0.28 [−0.54, 0.02]	−0.12 [−0.41, 0.19]	0.08 [−0.23, 0.38]	0.21 [−0.10, 0.48]	−0.12 [−0.41, 0.19]

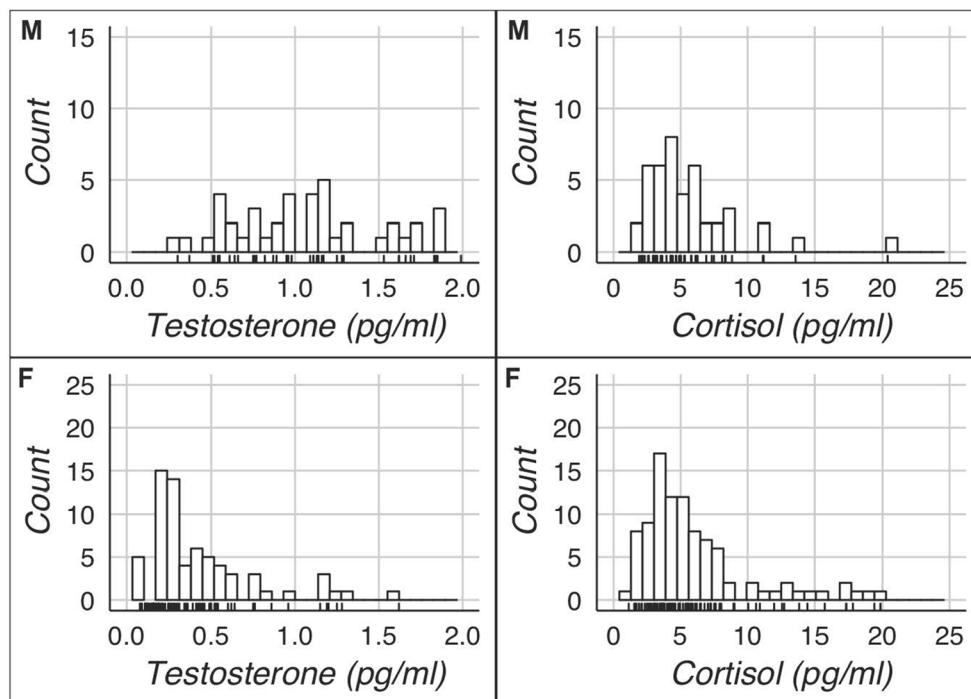
Note. M and SD are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. \* $p < 0.05$ , \*\* $p < 0.01$ .

**TABLE 2 |** Means, standard deviations and correlations with confidence intervals for female sample.

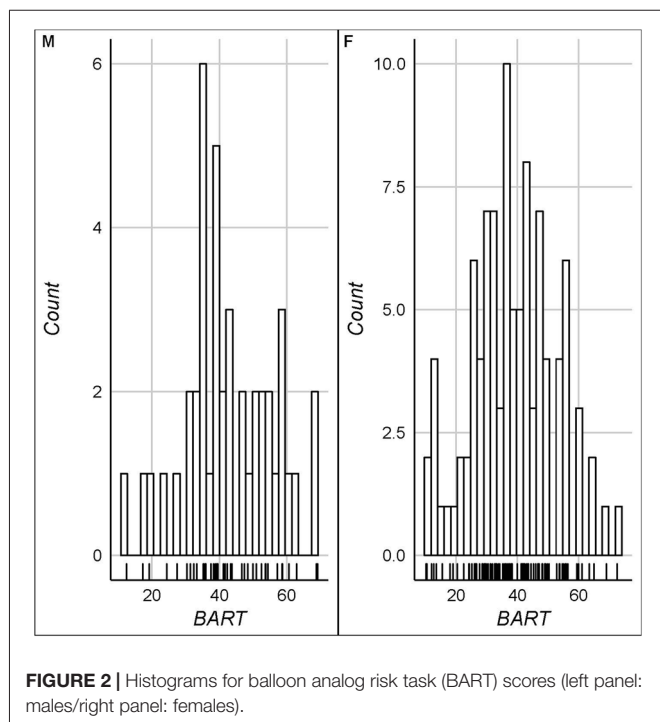
Variable	M	SD	1	2	3	4	5	6
1. Testosterone	0.35	0.31						
2. Cortisol	6.28	4.86	0.07 [−0.13, 0.26]					
3. Left hand 2D:4D	0.98	0.04	−0.05 [−0.25, 0.15]	0.03 [−0.17, 0.23]				
4. Right hand 2D:4D	0.97	0.03	−0.11 [−0.30, 0.09]	−0.06 [−0.26, 0.14]	0.76** [0.66, 0.83]			
5. Overconfidence	11.69	11.67	0.07 [−0.13, 0.27]	−0.15 [−0.34, 0.05]	−0.04 [−0.24, 0.16]	0.06 [−0.14, 0.26]		
6. Risk taking	38.52	13.50	−0.03 [−0.23, 0.17]	−0.00 [−0.20, 0.20]	−0.02 [−0.22, 0.18]	−0.09 [−0.28, 0.12]	0.01 [−0.19, 0.21]	
7. Hormonal contraception use	0.70	0.46	−0.22* [−0.40, −0.02]	−0.09 [−0.28, 0.12]	0.03 [−0.18, 0.22]	0.11 [−0.10, 0.30]	−0.09 [−0.28, 0.11]	−0.23* [−0.41, −0.03]

Note. M and SD are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. \* $p < 0.05$ , \*\* $p < 0.01$ .





**FIGURE 1** | Hormonal distributions of men (M) and women (F).



**FIGURE 2** | Histograms for balloon analog risk task (BART) scores (left panel: males/right panel: females).

( $r_{\text{left hand}} = -0.05$ ;  $r_{\text{right hand}} = -0.11$ ), our data thus offer no support for a digit ratio effect on baseline testosterone. We acknowledge that the size of our male sample limits the robustness of this test and we cannot rule out the possibility

of a small to moderate effect being undetected in our analysis. The correlations for both females and males are directionally consistent with such expectations.

### Bart Scores

In women, none of the models substantially supported an effect beyond the null model. The only exception was a model containing an effect of oral contraceptive use ( $\Delta\text{WAIC}$ : 3.52 and  $\Delta\text{LOOIC}$ : 3.52). This model suggests that those who take hormonal contraceptives have lower BART scores ( $B = -6.65 \pm 2.85$ ; 95%CI:  $-12.24$  to  $-1.00$ ).

In men, a model with a testosterone by cortisol interaction on BART scores is supported above the null ( $\Delta\text{WAIC} = 3.73$  and  $\Delta\text{LOOIC} = 3.45$ ). No other models were supported beyond the null. The parameter estimates, SE, and 95%CI for the testosterone by cortisol interaction model are reported in **Table 3** (see ESM for further details on the model). The interaction effect is plotted in **Figure 3**. For those men low in cortisol, testosterone had a positive effect on their BART scores. In contrast, for those men high in cortisol, testosterone was negatively related to BART scores ( $\beta_{\text{interaction}} = -0.44 \pm 0.16$ ; 95%CI:  $-0.76$  to  $-0.11$ ). For women, there is no evidence for such an interaction effect (women:  $B = 21.04$ , 95%CI:  $-13.6$  to  $55.19$ ) and, if anything, it runs in the opposite direction of the male effect (men:  $B = -135.84$ , 95%CI:  $-234.07$  to  $-35.61$ ).

We performed numerous, pre-specified robustness checks to further examine the results for men. Exclusion of the extreme case for cortisol did not alter our conclusions

**TABLE 3 |** Parameter estimates, standard errors and 95%CI for T\*C interaction model.

	Estimate	SE	95% lower	95% upper
Intercept	42.01	6.22	29.74	54.09
Log T	77.82	34.09	10.29	145.24
Log C	4.95	8.73	-11.96	21.93
Log T*Log C	-135.84	49.5	-234.07	-35.61

Note. Log T = log transformed testosterone and log C = log transformed cortisol.

( $B_{\text{interaction}} = -148.85 \pm 50.97$ ; 95%CI:  $-246.51$  to  $-48.12$ ). Similarly, controlling for age, BMI, or sexual orientation lead to the same conclusions (respective 95%CI for the interaction effect:  $-230.79$  to  $-36.60$ ;  $-251.18$  to  $-31.66$ ;  $-231.18$  to  $-46.95$ ). Excluding four cases due to excessive alcohol consumption or hard drug use, also upheld the effect (95%CI:  $-260.97$  to  $-55.76$ ). Neither accounting for how often participants washed their hair in a week, nor the method of hair drying affected this conclusion (respectively 95%CI:  $-239.55$  to  $-40.97$  and  $-233.93$  to  $-38.65$ ). Finally, controlling for certain medication usage, use of allergy medication or a history of psychological disorder also did not alter the statistical conclusion (respectively: 95%CI:  $-232.86$  to  $-37.62$ ;  $-237.04$  to  $-38.96$ ; and  $-231.72$  to  $-40.27$ ). Thus, after a range of checks we find consistent support for a testosterone by cortisol interaction effect on BART in men.

## Overconfidence

None of the models provided substantial support for a relationship between overconfidence and hair testosterone ( $r_{\text{males}} = -0.08$ ;  $r_{\text{females}} = 0.07$ ) or cortisol concentrations ( $r_{\text{males}} = -0.06$ ;  $r_{\text{females}} = -0.15$ ), nor overconfidence and risk taking ( $r_{\text{males}} = -0.12$ ;  $r_{\text{females}} = 0.01$ ).

## DISCUSSION

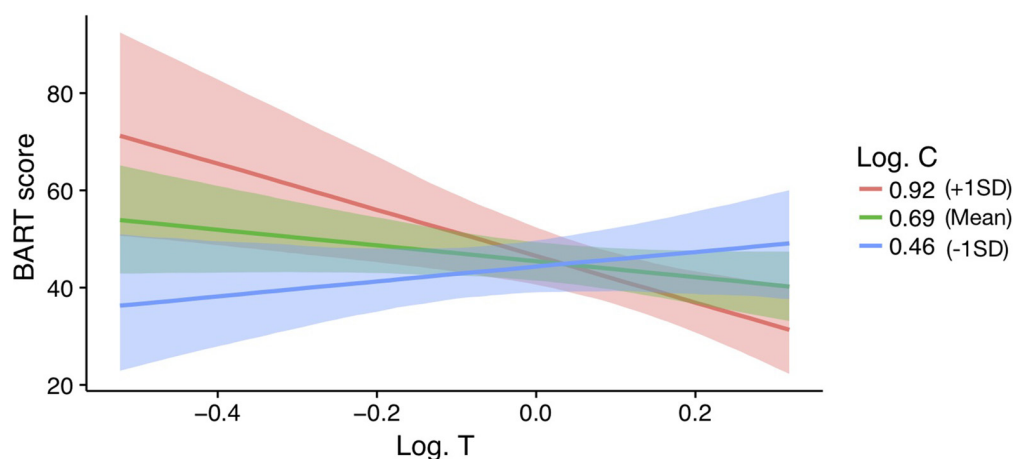
The present study reexamined the relationships between testosterone and risk taking, using an alternative assay procedure in which testosterone levels are assayed from hair samples.

We did not find evidence for a relationship between hair testosterone concentrations, 2D:4D ratios, and risk taking. However, we did find evidence for the interacting effect of hair testosterone and cortisol concentrations on risk taking in men, albeit in a small sample. We acknowledge that our final sample size for males imposes limitations on our statistical power, thus tempering the strength of our conclusions<sup>2</sup>.

## Theoretical Implications

Our findings did not support a relationship between hair testosterone concentrations and risk taking. As our testosterone sampling aggregated across approximately 3 months of participants' testosterone levels, this finding provides necessary (but insufficient) support for the predictions of the Challenge Hypothesis (Wingfield et al., 1990; Archer, 2006) and the Biosocial Model of Status (Mazur, 1985; Mazur and Booth, 1998), both of which specify dynamic bidirectional relationships between socially driven fluctuations in testosterone and behavior. Consistent with these theoretical perspectives, previous reports have focused on context driven

<sup>2</sup>One reviewer requested a "traditional" frequentist power analysis (see ESM-Supplementary analysis: frequentist power analysis). This analysis showed that based on our sample size and with a power of 0.80 and a  $p$  level of 0.05, we were able to detect estimates of  $f^2 = 0.099$  and  $0.254$  for the female and male sample respectively. Cohen (1988) suggests interpretations of 0.02, 0.15 and 0.35 as small, moderate and large.  $f^2$  is a standardized measure of effect size.



**FIGURE 3 |** Effect of log transformed testosterone (Log T) on BART scores for varying levels of Alog transformed cortisol (Log C) in men (z-scored). Bands represent 95% confidence intervals.

relationships between testosterone and risk taking (Coates and Herbert, 2008; Ronay and von Hippel, 2010), and while other studies have not specifically identified context as a factor, they have nonetheless measured testosterone and risk taking at a single time point, and examined the relationship between them at that moment in time (Apicella et al., 2008; Sapienza et al., 2009; Stanton et al., 2011). Previous results have been inconsistent, with positive (Apicella et al., 2008) and null relationships (Zethraeus et al., 2009). While it is possible that the positive effects in these studies are due to false positives, and the null effects perhaps the result of a weak relationship that is not captured by small sample sizes, or inconsistencies in the operationalization of risk taking, we speculate that the evidence for a relationship between testosterone and risk taking appears to be bound to the *activating* effects of the hormone within a specific context.

However, qualifying this speculative conclusion, we did find evidence in support of the Dual (hair) Hormone Hypothesis (Mehta and Josephs, 2010), albeit only in men and with a relatively small sample size ( $n = 53$ ). Mehta and Josephs (2010) first articulated the possibility that the moderating role of cortisol might be due to low cortisol facilitating social approach, thus allowing for the overt expression of dominant (and perhaps risky) behaviors. However, due to cortisol's effects on stress and social inhibition, higher testosterone may decrease dominance (and perhaps risky) behavior when cortisol is high. Those interested in reviewing the existing evidence for the Dual Hormone Hypothesis might read Mehta and Prasad (2015). In the current study we found that for men, hair testosterone concentrations were positively related to risk taking, only when levels of hair cortisol concentrations were low. When hair cortisol concentrations were high, we observed a negative relationship between testosterone and risk taking. Thus, although it has been suggested that one possibility for the few null findings surrounding the Dual Hormone Hypothesis might be that such effects emerge in response to social contextual primes (Mehta and Prasad, 2015), our data suggest this is not the case. Specifically, our data help clarify the Dual Hormone Hypothesis by demonstrating that the relationship between risk taking and the combination of high testosterone and low cortisol is not isolated to a time specific social context. Rather, we find that hormone levels, synthesized across a period of 3 months prior to completing a behavioral measure of risk taking, interact to predict risk taking behavior in a theory consistent manner.

Contributing to the lack of evidence for a relationship between circulating testosterone and 2D:4D ratio, we find no evidence for a relationship between hair testosterone concentrations and 2D:4D ratio. While further research is warranted before strong conclusions are drawn, we suggest this is an important null effect within the context of the ongoing discussion in the literature regarding the relationship between second to fourth digit ratio and circulating testosterone (Hönekopp et al., 2007). Aggregating testosterone levels across 3 months via hair samples filters out contextual

noise in hormone measurements, so providing a stronger test of the relationship between testosterone and 2D:4D ratio. Taken together, the evidence suggests that both state-based levels of testosterone—such as are derived from single time point measures—and more stable aggregated levels of baseline testosterone—such as we captured via hair sampling—appear to be unrelated to second to fourth digit ratios. Future research might however explore the possibility of an interaction between 2D:4D ratio and hair testosterone concentrations, as previous research has reported that the effects of testosterone administration on women's cognitive empathy are moderated by 2D:4D ratio (Van Honk et al., 2011).

Furthermore, despite theoretical suggestions of a relationship between testosterone and overconfidence (Johnson et al., 2006), we find no empirical support for this relationship with hair testosterone concentrations. This null effect is consistent with previous research (Ronay et al., 2017) that assayed testosterone concentrations from saliva samples.

Finally, we also found that hair cortisol concentrations were unrelated to overconfidence and risk taking. This finding is in line with other research showing that hair cortisol concentrations were unrelated to risk taking in behavioral tasks (Chumbley et al., 2014; Ceccato et al., 2016). However, only in men, Ceccato et al. (2016) did find a trend between higher hair cortisol concentrations and more investment in a gambling task. Furthermore, our null findings are not in line with research showing that high levels of conscientious, which are related to less risk taking behavior (Strickhouser et al., 2017), were related to smaller hair cortisol concentrations (Steptoe et al., 2017).

## Limitations and Future Directions

We acknowledge several limitations that serve as avenues for future research. First, although the total sample size is relatively large compared to other hair sample studies (e.g., Iglesias et al., 2015; Dettenborn et al., 2016), the number of men in our sample was relatively small. As the behavioral effects of testosterone are known to differ between men and women (e.g., Turanovic et al., 2017), future studies should replicate our findings in a more balanced gender sample. Second, Ribeiro et al. (2016) have shown that indirect finger length measures (from scans or photos) result in lower 2D:4D ratio scores than direct measures. Further work is needed in order to clarify whether the effect sizes of 2D:4D ratios are dependent on measurement protocol. Third, although the BART measure is an often used measure of risk taking (Lejuez et al., 2002), the measure could be confounded with participants' beliefs about the choices and outcomes of others in the experiment (because of the cash prize). Although no computer task can perfectly simulate naturally occurring risk taking behaviors, the BART does simulate risk situations in a natural environment and has been shown to predict a number of real-world risk taking behaviors (Lejuez et al., 2002, 2003; Hopko et al., 2006). Furthermore, it allows for the assessment of an overall propensity for risk taking rather than the likelihood of engaging in a particular type of risk

taking behavior, as is often case with self-report measures of risk-related constructs. Nevertheless, future studies should test the generalizability of the results to real-world situations. Fourth, our evidence suggests that both state-based levels of testosterone and baseline testosterone appear to be unrelated to 2D:4D ratios. This does not, however, rule against the possibility that 2D:4D is indeed a putative marker of prenatal testosterone exposure, and so lends itself to exploring the organizing effects of testosterone on behavior (Hönekopp et al., 2007).

## REFERENCES

Anderson, C., Brion, S., Moore, D. A., and Kennedy, J. A. (2012). A status-enhancement account of overconfidence. *J. Pers. Soc. Psychol.* 103, 718–735. doi: 10.1037/a0029395

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

Apicella, C. L., Dreber, A., and Möllerstrom, J. (2014). Salivary testosterone change following monetary wins and losses predicts future financial risk-taking. *Psychoneuroendocrinology* 39, 58–64. doi: 10.1016/j.psyneuen.2013.09.025

Archer, J. (2006). Testosterone and human aggression: an evaluation of the challenge hypothesis. *Neurosci. Biobehav. Rev.* 30, 319–345. doi: 10.1016/j.neubiorev.2004.12.007

Bailey, A. A., and Hurd, P. L. (2005). Finger length ratio (2D:4D) correlates with physical aggression in men but not in women. *Biol. Psychol.* 68, 215–222. doi: 10.1016/j.biopsycho.2004.05.001

Balthazart, J. (1983). “Hormonal correlates of behavior,” in *Avian Biology*, (Vol. 7) eds D. S. Farner, J. R. King and K. C. Parkes (New York, NY: Academic Press), 221–365.

Battalio, R. C., Kagel, J. C., and Jiranyakul, K. (1990). Testing between alternative models of choice under uncertainty: some initial results. *J. Risk Uncertain.* 3, 25–50. doi: 10.1007/bf00213259

Boksem, M. A., Mehta, P. H., Van den Bergh, B., van Son, V., Trautmann, S. T., Roelofs, K., et al. (2013). Testosterone inhibits trust but promotes reciprocity. *Psychol. Sci.* 24, 2306–2314. doi: 10.1177/0956797613495063

Branas-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.*

Buerkner, P. C. (2015). *BRMS: Bayesian Regression Models Using Stan*. Muenster: R package version 0.6.0.

Burnham, K. P., and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York, NY: Springer.

Burnham, K. P., and Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociol. Methods Res.* 33, 261–304. doi: 10.1177/0049124104268644

Byrnes, J. P., Miller, D. C., and Schafer, W. D. (1999). Gender differences in risk taking: a meta-analysis. *Psychol. Bull.* 125, 367–383. doi: 10.1037/0033-2909.125.3.367

Camerer, C., and Lovallo, D. (1999). Overconfidence and excess entry: an experimental approach. *Am. Econ. Rev.* 89, 306–318. doi: 10.1257/aer.89.1.306

Campbell, B. C., Dreber, A., Apicella, C. L., Eisenberg, D. T., Gray, P. B., Little, A. C., et al. (2010). Testosterone exposure, dopaminergic reward, and sensation-seeking in young men. *Physiol. Behav.* 99, 451–456. doi: 10.1016/j.physbeh.2009.12.011

Campbell, W. K., Goodie, A. S., and Foster, J. D. (2004). Narcissism, confidence, and risk attitude. *J. Behav. Decis. Mak.* 17, 297–311. doi: 10.1002/bdm.475

Ceccato, S., Kudielka, B. M., and Schwioren, C. (2016). Increased risk taking in relation to chronic stress in Adults. *Front. Psychol.* 6:2036. doi: 10.3389/fpsyg.2015.02036

Chumbley, J. R., Krajchich, I., Engelmann, J. B., Russell, E., Van Uum, S., Koren, G., et al. (2014). Endogenous cortisol predicts decreased loss aversion

## AUTHOR CONTRIBUTIONS

RR and LM designed the study and collected the data. TVP conducted the analyses. All authors (RR, LM, JKO and TVP) contributed to the writing of the manuscript.

## FUNDING

This project was funded by an internal grant from Vrije Universiteit Amsterdam.

in young men. *Psychol. Sci.* 25, 2102–2105. doi: 10.1177/0956797614546555

Coates, J. M., and Herbert, J. (2008). Endogenous steroids and financial risk taking on a London trading floor. *Proc. Natl. Acad. Sci. U S A* 105, 6167–6172. doi: 10.1073/pnas.0704025105

Cohen, J. (1988). *Statistical Power for The Social Sciences*. Hillsdale, NJ: LEA and Associates.

Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Mahwah, NJ: Erlbaum.

Cronbach, L. J., and Furby, L. (1970). How we should measure “change”: or should we? *Psychol. Bull.* 74, 68–80. doi: 10.1037/h0029382

Dabbs, J. M. Jr., Jurkovic, G. J., and Frady, R. L. (1991). Salivary testosterone and cortisol among late adolescent male offenders. *J. Abnorm. Child Psychol.* 19, 469–478. doi: 10.1007/bf00919089

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

Dettenborn, L., Kirschbaum, C., Gao, W., Spitzer, C., Roepke, S., Otte, C., et al. (2016). Increased hair testosterone but unaltered hair cortisol in female patients with borderline personality disorder. *Psychoneuroendocrinology* 71, 176–179. doi: 10.1016/j.psyneuen.2016.05.026

DuBois, P. H. (1957). *Multivariate Correlational Analysis*. Oxford, UK: Harper.

Eisenegger, C., Haushofer, J., and Fehr, E. (2011). The role of testosterone in social interaction. *Trends Cogn. Sci.* 15, 263–271. doi: 10.1016/j.tics.2011.04.008

Fischhoff, B., Slovic, P., and Lichtenstein, S. (1977). Knowing with certainty: the appropriateness of extreme confidence. *J. Exp. Psychol. Hum. Percept. Perform.* 3, 552–564. doi: 10.1037/0096-1523.3.4.552

Gao, W., Stalder, T., Foley, P., Rauh, M., Deng, H., and Kirschbaum, C. (2013). Quantitative analysis of steroid hormones in human hair using a column-switching LC-APCI-MS/MS assay. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 928, 1–8. doi: 10.1016/j.jchromb.2013.03.008

Goudriaan, A. E., Lapauw, B., Ruige, J., Feyen, E., Kaufman, J. M., Brand, M., et al. (2010). The influence of high-normal testosterone levels on risk-taking in healthy males in a 1-week letrozole administration study. *Psychoneuroendocrinology* 35, 1416–1421. doi: 10.1016/j.psyneuen.2010.04.005

Granger, D. A., Hibell, L. C., Fortunato, C. K., and Kapelewski, C. H. (2009). Medication effects on salivary cortisol: tactics and strategy to minimize impact in behavioral and developmental science. *Psychoneuroendocrinology* 34, 1437–1448. doi: 10.1016/j.psyneuen.2009.06.017

Granger, D. A., Schwartz, E. B., Booth, A., and Arentz, M. (1999). Salivary testosterone determination in studies of child health and development. *Horm. Behav.* 35, 18–27. doi: 10.1006/hbeh.1998.1492

Harding, C. F. (1981). Social modulation of circulating hormone levels in the male. *Am. Zool.* 21, 223–231. doi: 10.1093/icb/21.1.223

Harman, S. M., Metter, E. J., Tobin, J. D., Pearson, J., and Blackman, M. R. (2001). Longitudinal effects of aging on serum total and free testosterone levels in healthy men. *J. Clin. Endocrinol. Metab.* 86, 724–731. doi: 10.1210/jc.86.2.724

Harrison, G. W. (2006). “Hypothetical bias over uncertain outcomes,” in *Using Experimental Methods in Environmental and Resource Economics*, ed. J. List (Northampton, MA: Elgar), 41–69.

Holt, C. A., and Laury, S. K. (2005). Risk aversion and incentive effects: new data without order effects. *Am. Econ. Rev.* 95, 902–904. doi: 10.1257/0002828054201459



- 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
- Hopko, D. R., Lejuez, C. W., Daughters, S. B., Aklin, W. M., Osborne, A., Simmons, B. L., et al. (2006). Construct validity of the balloon analogue risk task (BART): relationship with MDMA use by inner-city drug users in residential treatment. *J. Psychopathol. Behav. Assess.* 28, 95–101. doi: 10.1007/s10862-006-7487-5
- Hunt, M. K., Hopko, D. R., Bare, R., Lejuez, C. W., and Robinson, E. V. (2005). Construct validity of the balloon analog risk task (BART) associations with psychopathy and impulsivity. *Assessment* 12, 416–428. doi: 10.1177/1073191105278740
- Iglesias, S., Jacobsen, D., Gonzalez, D., Azzara, S., Repetto, E. M., Jamardo, J., et al. (2015). Hair cortisol: a new tool for evaluating stress in programs of stress management. *Life Sci.* 141, 188–192. doi: 10.1016/j.lfs.2015.10.006
- John, O. P., and Robins, R. W. (1994). Accuracy and bias in self-perception: individual differences in self-enhancement and the role of narcissism. *J. Pers. Soc. Psychol.* 66, 206–219. doi: 10.1037//0022-3514.66.1.206
- Johnson, D. D. P., McDermott, R., Barrett, E. S., Cowden, J., Wrangham, R., McIntyre, M. H., et al. (2006). Overconfidence in wargames: experimental evidence on expectations, aggression, gender and testosterone. *Proc. R. Soc. B* 273, 2513–2520. doi: 10.1098/rspb.2006.3606
- Kaufman, J. M., and Vermeulen, A. (2005). The decline of androgen levels in elderly men and its clinical and therapeutic implications. *Endocr. Rev.* 26, 833–876. doi: 10.1210/er.2004-0013
- Kruger, J., and Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J. Pers. Soc. Psychol.* 77, 1121–1134. doi: 10.1037/0022-3514.77.6.1121
- Kruger, J., and Mueller, R. A. (2002). Skilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of one's own performance. *J. Pers. Soc. Psychol.* 82, 180–188. doi: 10.1037/0022-3514.82.2.180-188
- Larrick, R. P., Burson, K. A., and Soll, J. B. (2007). Social comparison and confidence: when thinking you're better than average predicts overconfidence (and when it does not). *Organ. Behav. Hum. Decis. Process.* 102, 76–94. doi: 10.1016/j.obhdp.2006.10.002
- Lejuez, C. W., Aklin, W. M., Zvolensky, M. J., and Pedulla, C. M. (2003). Evaluation of the Balloon Analogue Risk Task (BART) as a predictor of adolescent real-world risk-taking behaviours. *J. Adolesc.* 26, 475–479. doi: 10.1016/s0140-1971(03)00036-8
- Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., et al. (2002). Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). *J. Exp. Psychol. Appl.* 8, 75–84. doi: 10.1037/1076-898x.8.2.75
- 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
- Malmendier, U., and Tate, G. (2008). Who makes acquisitions? CEO overconfidence and the market's reaction. *J. Financ. Econ.* 89, 20–43. doi: 10.1016/j.jfineco.2007.07.002
- Manning, J. T. (2002). *Digit Ratio: A Pointer to Fertility, Behavior, and Health*. New Jersey, NJ: Rutgers University Press.
- 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
- Mazur, A. (1985). A biosocial model of status in face-to-face primate groups. *Soc. Forces* 64, 377–402. doi: 10.2307/2578647
- Mazur, A., and Booth, A. (1998). Testosterone and dominance in men. *Behav. Brain Sci.* 21, 353–363; discussion 363–397. doi: 10.1017/s0140525x9801228
- Mehta, P. H., and Josephs, R. A. (2006). Testosterone change after losing predicts the decision to compete again. *Horm. Behav.* 50, 684–692. doi: 10.1016/j.yhbeh.2006.07.001
- Mehta, P. H., and Josephs, R. A. (2010). Testosterone and cortisol jointly regulate dominance: evidence for a dual-hormone hypothesis. *Horm. Behav.* 58, 898–906. doi: 10.1016/j.yhbeh.2010.08.020
- Mehta, P. H., and Prasad, S. (2015). The dual-hormone hypothesis: a brief review and future research agenda. *Curr. Opin. Behav. Sci.* 3, 163–168. doi: 10.1016/j.cobeha.2015.04.008
- Mehta, P. H., Welker, K. M., Zilioli, S., and Carré, J. M. (2015). Testosterone and cortisol jointly modulate risk-taking. *Psychoneuroendocrinology* 56, 88–99. doi: 10.1016/j.psyneuen.2015.02.023
- Michailova, J. (2010). *Development of the Overconfidence Measurement Instrument for the Economic Experiment*. MPRA paper no. 26384. Germany: Christian Albrechts University of Kiel.
- Michailova, J., and Katter, J. K. Q. (2014). Quantifying overconfidence in experimental finance. *J. Behav. Account. Finance* 4, 351–367. doi: 10.1504/ijbaf.2014.067629
- Miller, D. C., and Byrnes, J. P. (1997). The role of contextual and personal factors in children's risk taking. *Dev. Psychol.* 33, 814–823. doi: 10.1037/0012-1649.33.5.814
- Millet, K., and Dewitte, S. (2009). The presence of aggression cues inverts the relation between digit ratio (2D:4D) and prosocial behaviour in a dictator game. *Br. J. Psychol.* 100, 151–162. doi: 10.1348/000712608x324359
- Moore, D. A., and Healy, P. J. (2008). The trouble with overconfidence. *Psychol. Rev.* 115, 502–517. doi: 10.1037/0033-295X.115.2.502
- Pollet, T. V., and van der Meij, L. (2017). To remove or not to remove: the impact of outlier handling on significance testing in testosterone data. *Adapt. Hum. Behav. Physiol.* 3, 43–60. doi: 10.1007/s40750-016-0050-z
- Pollet, T. V., van der Meij, L., Cobey, K. D., and Buunk, A. P. (2011). Testosterone levels and their associations with lifetime number of opposite sex partners and remarriage in a large sample of American elderly men and women. *Horm. Behav.* 60, 72–77. doi: 10.1016/j.yhbeh.2011.03.005
- Popma, A., Vermeiren, R., Geluk, C. A., Rinne, T., van den Brink, W., Knol, D. L., et al. (2007). Cortisol moderates the relationship between testosterone and aggression in delinquent male adolescents. *Biol. Psychiatry* 61, 405–411. doi: 10.1016/j.biopsych.2006.06.006
- Rada, R. T., Kellner, R., and Winslow, W. W. (1976). Plasma testosterone and aggressive behavior. *Psychosomatics* 17, 138–142. doi: 10.1016/s0033-3182(76)71132-0
- Raftery, A. E. (1996). Approximate Bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika* 83, 251–266. doi: 10.1093/biomet/83.2.251
- Ribeiro, E., Neave, N., Morais, R. N., and Manning, J. T. (2016). Direct versus indirect measurement of digit ratio (2D:4D) a critical review of the literature and new data. *Evol. Psychol.* 14, 1–8. doi: 10.1177/1474704916632536
- Ronay, R., and Carney, D. R. (2013). Testosterone's negative relationship with empathic accuracy and perceived leadership ability. *Soc. Psychol. Pers. Sci.* 4, 92–99. doi: 10.1177/1948550612442395
- Ronay, R., and Galinsky, A. D. (2011). Lex talionis: testosterone and the law of retaliation. *J. Exp. Soc. Psychol.* 47, 702–705. doi: 10.1016/j.jesp.2010.11.009
- Ronay, R., and Kim, D. Y. (2006). Gender differences in explicit and implicit risk attitudes: a socially facilitated phenomenon. *Br. J. Soc. Psychol.* 45, 397–419. doi: 10.1348/014466005x66420
- Ronay, R., Tybur, J., Van Huijstee, D., and Morssinkhof, M. (2017). Embodied power, testosterone and overconfidence as a causal pathway to risk taking. *Compr. Results Soc. Psychol.* 2, 1–18. doi: 10.1080/23743603.2016.1248081
- Ronay, R., and von Hippel, W. (2010). The presence of an attractive woman elevates testosterone and physical risk taking in young men. *Soc. Psychol. Pers. Sci.* 1, 57–64. doi: 10.1177/1948550609352807
- Ronay, R., Greenaway, K., Anicich, E. M., and Galinsky, A. D. (2012). The path to glory is paved with hierarchy: when hierarchical differentiation increases group effectiveness. *Psychol. Sci.* 23, 669–677. doi: 10.1177/0956797611433876
- Rose, R. M., Holaday, J. W., and Bernstein, I. S. (1971). Plasma testosterone, dominance rank and aggressive behaviour in male rhesus monkeys. *Nature* 231, 366–368. doi: 10.1038/231366a0
- 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
- Sapientza, P., Zingales, L., and Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proc. Natl. Acad. Sci. U S A* 106, 15268–15273. doi: 10.1073/pnas.0907352106

- Schwabl, H., and Kriner, E. (1991). Territorial aggression and song of male European robins (*Erithacus rubecula*) in autumn and spring: effects of antiandrogen treatment. *Horm. Behav.* 25, 180–194. doi: 10.1016/0018-506x(91)90049-n
- Stanton, S. J. (2011). The essential implications of gender in human behavioral endocrinology studies. *Front. Behav. Neurosci.* 5:9. doi: 10.3389/fnbeh.2011.00009
- Stanton, S. J., Beehner, J. C., Saini, E. K., Kuhn, C. M., and LaBar, K. S. (2009). Dominance, politics, and physiology: voters' testosterone changes on the night of the 2008 United States presidential election. *PLoS One* 4:e7543. doi: 10.1371/journal.pone.0007543
- Stanton, S. J., Mullette-Gillman, O. A., McLaurin, R. E., Kuhn, C. M., LaBar, K. S., Platt, M. L., et al. (2011). Low- and high-testosterone individuals exhibit decreased aversion to economic risk. *Psychol. Sci.* 22, 447–453. doi: 10.1177/0956797611401752
- Stephoe, A., Easterlin, E., and Kirschbaum, C. (2017). Conscientiousness, hair cortisol concentration and health behaviour in older men and women. *Psychoneuroendocrinology* 86, 122–127. doi: 10.1016/j.psyneuen.2017.09.016
- Strickhouser, J. E., Zell, E., and Krizan, Z. (2017). Does personality predict health and well-being? A metasynthesis. *Health Psychol.* 36, 797–810. doi: 10.1037/hea0000475
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wellesley.
- Turanovic, J. J., Pratt, T. C., and Piquero, A. R. (2017). Exposure to fetal testosterone, aggression and violent behavior: a meta-analysis of the 2D:4D digit ratio. *Aggress. Violent Behav.* 33, 51–61. doi: 10.1016/j.avb.2017.01.008
- Van den Bergh, B., and Dewitte, S. (2006). Digit ratio (2D:4D) moderates the impact of sexual cues on men's decisions in ultimatum games. *Proc. Biol. Sci.* 273, 2091–2095. doi: 10.1098/rspb.2006.3550
- Van der Meij, L., Almela, M., Hidalgo, V., Villada, C., Ijzerman, H., Van Lange, P. A., et al. (2012). Testosterone and cortisol release among Spanish soccer fans watching the 2010 World Cup final. *PLoS One* 7:e34814. doi: 10.1371/journal.pone.0034814
- Van der Meij, L., Buunk, A. P., Almela, M., and Salvador, A. (2010). Testosterone responses to competition: the opponent's psychological state makes it challenging. *Biol. Psychol.* 84, 330–335. doi: 10.1016/j.biopsycho.2010.03.017
- Van der Meij, L., Buunk, A. P., Van de Sande, J. P., and Salvador, A. (2008). The presence of a woman increases testosterone in aggressive dominant men. *Horm. Behav.* 54, 640–644. doi: 10.1016/j.yhbeh.2008.07.001
- Van Honk, J., Schutter, D. J., Bos, P. A., Kruijt, A. W., Lentjes, E. G., and Baron-Cohen, S. (2011). Testosterone administration impairs cognitive empathy in women depending on second-to-fourth digit ratio. *Proc. Natl. Acad. Sci. U S A* 108, 3448–3452. doi: 10.1073/pnas.1011891108
- Van Honk, J., Schutter, D. J., Hermans, E. J., and Putman, P. (2003). Low cortisol levels and the balance between punishment sensitivity and reward dependency. *Neuroreport* 14, 1993–1996. doi: 10.1097/01.wnr.0000091690.72892.ec
- Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* 27, 1413–1432. doi: 10.1007/s11222-016-9696-4
- Voracek, M., Reimer, B., Ertl, C., and Dressler, S. G. (2006). Digit ratio (2D:4D), lateral preferences and performance in fencing. *Percept. Mot. Skills* 103, 427–446. doi: 10.2466/pms.103.6.427-446
- Wennig, R. (2000). Potential problems with the interpretation of hair analysis results. *Forensic Sci. Int.* 107, 5–12. doi: 10.1016/s0379-0738(99)00146-2
- White, T. L., Lejuez, C. W., and de Wit, H. (2008). Test-retest characteristics of the Balloon Analogue Risk Task (BART). *Exp. Clin. Psychopharmacol.* 16, 565–570. doi: 10.1037/a0014083
- Wingfield, J. C., and Hahn, T. P. (1994). Testosterone and territorial behaviour in sedentary and migratory sparrows. *Anim. Behav.* 47, 77–89. doi: 10.1006/anbe.1994.1009
- Wingfield, J. C., Hegner, R. E., Dufty, A. M. Jr., and Ball, G. F. (1990). The “challenge hypothesis”: theoretical implications for patterns of testosterone secretion, mating systems and breeding strategies. *Am. Nat.* 136, 829–846. doi: 10.1086/285134
- Zethraeus, N., Kocoska-Maras, L., Ellingsen, T., Von Schoultz, B., Hirschberg, A. L., and Johannesson, M. (2009). A randomized trial of the effect of estrogen and testosterone on economic behavior. *Proc. Natl. Acad. Sci. U S A* 106, 6535–6538. doi: 10.1073/pnas.0812757106

**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 © 2018 Ronay, van der Meij, Oostrom and Pollet. 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) and the copyright owner 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.



# The Dopamine Receptor D4 Gene (*DRD4*) and Financial Risk-Taking: Stimulating and Instrumental Risk-Taking Propensity and Motivation to Engage in Investment Activity

Rafał Muda<sup>1\*</sup>, Mariusz Kicia<sup>1</sup>, Małgorzata Michalak-Wojnowska<sup>2</sup>, Michał Ginszt<sup>3</sup>, Agata Filip<sup>2</sup>, Piotr Gawda<sup>3</sup> and Piotr Majcher<sup>3</sup>

<sup>1</sup>Faculty of Economics, Maria Curie-Skłodowska University, Lublin, Poland, <sup>2</sup>Department of Cancer Genetics with Cytogenetics Laboratory, Medical University of Lublin, Lublin, Poland, <sup>3</sup>Department of Rehabilitation and Physiotherapy, Medical University of Lublin, Lublin, Poland

## OPEN ACCESS

### Edited by:

Monica Capra,  
Claremont Graduate University,  
United States

### Reviewed by:

Walter Adriani,  
Istituto Superiore di Sanità, Italy  
Claudio Lucchiari,  
Università degli Studi di Milano, Italy

### \*Correspondence:

Rafał Muda  
rafal.muda@umcs.pl

**Received:** 13 October 2017

**Accepted:** 14 February 2018

**Published:** 02 March 2018

### Citation:

Muda R, Kicia M, Michalak-Wojnowska M, Ginszt M, Filip A, Gawda P and Majcher P (2018) The Dopamine Receptor D4 Gene (*DRD4*) and Financial Risk-Taking: Stimulating and Instrumental Risk-Taking Propensity and Motivation to Engage in Investment Activity. *Front. Behav. Neurosci.* 12:34. doi: 10.3389/fnbeh.2018.00034

The Dopamine receptor D4 gene (*DRD4*) has been previously linked to financial risk-taking propensity. Past works demonstrated that individuals with a specific variant of the *DRD4* gene (7R+) are more risk-seeking than people without it (7R−). The most prominent explanation for this effect is the fact that 7R+ individuals are less sensitive to dopamine and thus seek more stimulation to generate “normal” dopaminergic activity and feel pleasure. However, results about this relationship have not been conclusive, and some revealed a lack of the relationship. In the current work, we tested if those unclear results might be explained by the motivation that underlies the risk-taking activity; i.e., if people take risks to feel excitement or if they take risk to obtain a specific goal. In our study we tested the differences in risk-taking between 7R+ and 7R− among people who are experienced in financial risk-taking (113 investors) and non-experienced financial decision makers (104 non-investors). We measured risk-taking propensity with the Holt-Laury test and the Stimulating-Instrumental Risk Inventory. Moreover, we asked investors about their motivations for engaging in investment activity. Our study is the next one to report a lack of differences in risk-taking between 7R+ and 7R− individuals. As well, our results did not indicate any differences between the 7R+ and 7R− investors in motivation to engage in investment activity. We only observed that risk-taking propensity was higher among investors than non-investors and this was noticed for all measures. More research is needed to better understand the genetic foundations of risk-taking, which could answer the question about the substantial variation in the domain of risky financial decisions.

**Keywords:** *DRD4* gene, financial risk-taking, investors, dopamine, genetic determinants, risk preferences

## INTRODUCTION

As previous studies have indicated, the dopamine gene *Dopamine Receptor D4 Gene (DRD4)* is one of the most promising candidates that can be associated with risk-taking propensity (Carpenter et al., 2011; Dreber et al., 2011). The *DRD4*, a dopamine D4 receptor gene, is located near the telomere of chromosome 11p and contains a 48-bp Variable Number Tandem Repeat (VNTR) polymorphism in the third exon, repeated between 2 and 11 times (Grady et al., 2003). Moreover, the 48-bp repeat is thought to reside in the third cytoplasmic loop of the receptor protein and seems to affect the function of the D4 receptor (Ptáček et al., 2011). It was discovered that a variant with 7 or more VNTR repeats (7R+) is connected with the decreased binding of dopamine (Asghari et al., 1995). 7R+ individuals are less sensitive to dopamine and thus require a higher level of stimuli to produce a similar response as compared with people with the 7R− variant (with less than 7 VNTR repeats; Schoots and Van Tol, 2003). The site of dopamine's release seems to determine the role that it plays. Four major dopamine-rich pathways have been identified within the brain (mesolimbic, mesocortical, nigrostriatal, and tuberoinfundibular pathways). These pathways arise from two regions of the midbrain: the ventral tegmental area (VTA) and the substantia nigra, which primarily projects to the striatal complex—ventral striatum (VS) and dorsal striatum (Ernst and Luciana, 2015). Several studies have shown that dopaminergic projection from the VTA to the VS is particularly important in reward processing (McBride et al., 1999; Pierce and Kumaresan, 2006).

As a gene responsible for the regulation of the dopaminergic system and in turn reward processing (Wise, 2002), the *DRD4* gene may contribute to the behaviors connected with dopamine levels, e.g., risk-taking. The role of dopamine in reward processing and risk taking has been investigated in animal studies. For example, rats with an over-expressed dopamine transporter showed increased impulsivity for smaller and sooner rewards, and increased risk proneness (Adriani et al., 2009). Moreover, release of dopamine reinforces particular behaviors (especially those related to the expectation of reward), causes feelings of joy, and increases physiological arousal (Berridge and Robinson, 1998). As Schwarz (2012) noticed, bodily experiences like physiological arousal might inform us about physical states of the organism that, in turn, may be perceived as a source of information and influence decision-making. Moreover, through the activation of the nucleus accumbens, which is activated during the anticipation of monetary gains and positively correlates with a positive affect, dopamine is related to risk-taking behavior (Kuhnen and Knutson, 2005). Taking this into account, we should expect that the *DRD4* gene plays a moderating role in risk-taking propensity and 7R+ individuals should take more risks.

Indeed, previous studies about behavioral traits and the *DRD4* gene revealed that 7R+ individuals are prone to take more risks in specific situations that may cause positive stimulation, i.e., gambling or drinking alcohol. Researchers indicated that the presence of the 7R allele is connected to alcoholism (Laucht et al., 2007), impulsivity (Eisenberg et al., 2007), pathological gambling

(Pérez de Castro et al., 1997), or novelty-seeking (Ebstein et al., 1996).

Also in the domain of financial risk-taking, so far, four studies have revealed that 7R+ individuals make more risky decisions than 7R− individuals (Dreber et al., 2009, 2011; Kuhnen and Chiao, 2009; Carpenter et al., 2011). More precisely, Dreber et al. (2009) showed that the 7R+ polymorphism is associated with higher financial risk-taking and explains roughly 20% of the variance in financial risk-taking. In their next article, Dreber et al. (2011) confirmed the previous result. However, they found that the 7R+ variant is related to higher risk-taking propensity only among men but not among women. Also, Kuhnen and Chiao (2009) noticed a significant relationship between 7R+ and risk taking—in their study, 7R+ individuals invested 25% more assets in risky options than 7R− individuals.

However, some findings revealed a lack of differences. Another four studies failed to find significant differences between 7R+ and 7R− individuals in the domain of financial risk-taking (Eisenegger et al., 2010; Frydman et al., 2011; Dreber et al., 2012; Anderson et al., 2015). For example, Frydman et al. (2011) asked subjects to make choices between 140 pairs of monetary gambles. In each pair, subjects decided if they preferred the certain non-negative option involving a payout of  $x$  with 100% chance or a risky option involving a gain  $\$y$  and a loss  $\$z$  with equal probability. The results revealed that 7R+ individuals chose risky options in 39% of cases, while 7R− chose risky options in 38% of cases. No differences were also shown between the group of 7R+ and 7R− individual investors in both financial risk-taking task (choices between a certain payoff ranging from \$140 to \$1000 and a 50:50 gamble between the gain of \$1000 or nothing) and measures of equity holdings (based on national registry data on detailed asset holdings; Anderson et al., 2015). A lack of differences in risk-taking between 7R+ and 7R− was also observed in a group of owners, presidents and managers of large companies who performed the investment task. In this task participants started with \$250 and decided how much money they allocated in a risky investment which gave a 50% chance to multiply the invested amount 2.5 times, and a 50% chance to lose the allocated amount (Dreber et al., 2012). Surprisingly, in two other studies that used the same investment task, differences in risk-taking between 7R+ and 7R− were observed (Dreber et al., 2009, 2011).

The aim of our study is to verify if the previous inconclusive results about the *DRD4* gene and financial risk-taking might be explained by different needs that motivate risk-taking behavior. In the financial domain, risky behaviors might depend on motives that stimulate risk-taking. We can distinguish two kinds of risk preference that could potentially moderate the association between *DRD4* gene and risk-taking: (1) stimulating; and (2) instrumental risk-taking (Zaleśkiewicz, 2001). The motivation behind stimulating risk is to take action due to need for excitement seeking and to provide positive emotional arousal. Such experiences motivate to seek stimuli that provide pleasant feelings, and thus one is more prone to engage in risky activities. On the other hand, instrumental risk-taking is driven by motives that are oriented on achieving a specific



goal and analytic information processing instead of arousal seeking. For example, consider one who has \$1000 and desperately needs an additional \$1000 for medical treatment by the end of the day. After analyzing every possibility how to collect the money, one concludes that the only option is to play in the casino. Although, one engages in risky activity, this is due to a rational decision motivated by the need to achieve a particular economic goal (i.e., gain an additional \$1000 for medical treatment), not due to the need for experiencing pleasant feelings connected with gambling (Zaleśkiewicz, 2001).

In our study, we want to test if the *DRD4* gene is connected with financial risk-taking propensity in general, or if it is associated only with a specific risk-taking propensity that is oriented toward the search for stimulation and arousal. Taking into account that: (1) 7R+ individuals are more prone to engage in risky behaviors that increase arousal (e.g., gambling or drinking alcohol), as well as; (2) they need more stimuli to overcome the blunted response to dopamine to function “normally”, we might expect that, in the financial domain, we will notice the differences between 7R+ and 7R− individuals in stimulating risk-taking propensity but not in instrumental risk-taking propensity.

Additionally, in our study, we wanted to test the differences in risk-taking between 7R+ and 7R− among people who are experienced in financial decision-making and risk-taking (i.e., stock market investors). So far, only three studies have focused on different groups than students (Dreber et al., 2011, 2012; Anderson et al., 2015), and testing such a group could give more reliable results than testing just undergraduate students. Moreover, as Dorn and Sengmueller (2009) revealed, investors who have a tendency to trade excessively (which implies higher costs and in turn increases the risk) report enjoying investing or gambling<sup>1</sup>. This result suggests that investors who enjoy investing are more prone to accept risk for other reasons than monetary incentives (e.g., looking for excitement). This seems to be in line with our hypothesis that people who seek stimulation (i.e., 7R+ individuals) might take more risks in the financial domain than others.

## MATERIALS AND METHODS

### Participants

We conducted our study on two groups: (1) a group of private investors ( $n = 120$ , mean age = 33.63 [three subjects missing data for age], standard deviation [SD] = 9.85; we successfully genotyped 113 investors, mean age = 33.70 [one subject missing data for age], SD = 9.95, mean years of investing [missing data for four subjects] = 10.27, SD = 7.34, for 20 subjects investment activity was a main source of income, for 89 subjects it was additional income [missing data for four subjects]); and (2) a group of non-investors ( $n = 112$ , mean age = 32.46 SD = 10.14; we successfully genotyped 104 non-investors, mean age = 32.34 [missing data for age for one subject], SD = 10.00). We defined an investor/non-investor as a person who invests/has

never invested assets in the stock market or allocates/has never allocated money in an investment fund. Moreover, we controlled for academic major (financial/economics vs. others) and found no differences between group of investors and non-investors ( $\chi^2[1, n = 224] = 1.03, p = 0.348, \phi = 0.068$ ).

### Data Collection

The study was conducted during the Wall Street Conference—the biggest conference in Poland about the practice of investment, organized by the Society of Individual Investors. Before the event, all conference participants were informed about the study and invited to participate via email. Subjects were also recruited by flyers distributed at the conference place. For data collection, we invited subjects to a dedicated location in the conference place. The experiment was done with paper and pencil and tasks referred to non-incentivized decisions. At the beginning, we informed participants about the study protocol and collected their written consent to take part in the experiment. Next, we asked participants to provide two salivary samples. Cotton swab-derived buccal cells were scraped from the inner side of the cheeks. Prior to the sample collection, each of participants vigorously rinsed their mouth with water for about 30 s to remove food particles. They were given two cotton swabs and two test tubes labeled with a participant number. Then each of the participants was asked to give a buccal swab from each side of the cheek by scraping the inside of their cheek with the swab firmly for 30 s. Donors were reminded to turn the swabs to utilize both sides of the swab. In order to maximize the buccal cell yield, the samples were brought back to the laboratory in an ice-filled cooler. Afterward, subjects completed a sociodemographic survey and two risk-taking tasks.

### Risk-Taking Tasks

We measured the risk-taking propensity in three ways. The first one was the Holt-Laury test (Holt and Laury, 2002), which is one of the most widely used tests to measure risk-taking propensity in experimental economics. The Holt-Laury test is a measure based on choices between paired lotteries that involve only gains (see **Table 1**). In each pair (all pairs are presented in advance), the participant makes a decision between Lottery A and Lottery B. For each decision, lotteries give the possibility to win a fixed amount: Lottery A: 100 PLN or 80 PLN (which is about 25 USD and 20 USD), Lottery B: 185 PLN and 5 PLN. The subsequent lottery pairs differ on the probability of obtaining particular amount. In the first pair, the probability of winning the larger payoff (100 PLN and 185 PLN, respectively) is relatively low (i.e., a 10% chance), whereas the probability of winning a smaller payoff (i.e., 80 PLN and 5 PLN) is relatively high (i.e., a 90% chance). With each new pair, the probability of getting the higher reward increases by 10 percentage points, and in the last decision the chance for a higher gain is 100%.

Notice that the larger gain in Lottery B (i.e., 185 PLN) is higher than the larger gain in Lottery A (i.e., 100 PLN), whereas a smaller gain in Lottery A (i.e., 80 PLN) is larger than a smaller gain in Lottery B (i.e., 5 PLN). Thus, depending on the participant's

<sup>1</sup>This result is robust, controlling for gender and overconfidence.

**TABLE 1** | Choices between paired lotteries in the Holt-Laury test (Holt and Laury, 2002; polish version Tyska, 2010).

Lottery A	Lottery B	Expected value of Lottery A	Expected value of Lottery B
10% 100 PLN; 90% 80 PLN	10% 185 PLN; 90% 5 PLN	82	23
20% 100 PLN; 80% 80 PLN	20% 185 PLN; 80% 5 PLN	84	41
30% 100 PLN; 70% 80 PLN	30% 185 PLN; 70% 5 PLN	86	59
40% 100 PLN; 60% 80 PLN	40% 185 PLN; 60% 5 PLN	88	77
<b>50% 100 PLN; 50% 80 PLN</b>	<b>50% 185 PLN; 50% 5 PLN</b>	<b>90</b>	<b>95</b>
60% 100 PLN; 40% 80 PLN	60% 185 PLN; 40% 5 PLN	92	113
70% 100 PLN; 30% 80 PLN	70% 185 PLN; 30% 5 PLN	94	131
80% 100 PLN; 20% 80 PLN	80% 185 PLN; 20% 5 PLN	96	149
90% 100 PLN; 10% 80 PLN	90% 185 PLN; 10% 5 PLN	98	167
100% 100 PLN; 0% 80 PLN	100% 185 PLN; 0% 5 PLN	100	185

Bold text indicates the first lottery pair where expected value of Lottery B is higher than Lottery A. During the experiment, the text was not bolded and participants were presented only with first two columns and the place to indicate the response (columns with expected values were not presented).

risk-taking propensity, the switch from Lottery A to Lottery B will occur at different points. Someone who is an extreme risk-seeker might decide to take a chance to win the highest payoff and choose Lottery B in the first step, whereas one who is extremely risk averse and does not want to risk “losing” a moderate payoff might choose Lottery A until the last step.

The next two risk-taking measures were stimulating and instrumental risk-taking. Both were from the Stimulating-Instrumental Risk Inventory (Zaleśkiewicz, 2001). The Stimulating-Instrumental Risk Inventory is a questionnaire composed of 17 questions: 10 questions measure stimulating risk-taking (e.g., I often take risk just for fun; Gambling seems something very exciting to me), and seven questions instrumental risk-taking (e.g., At work I would prefer a position with a high salary which could be lost easily to a stable position but with a lower salary). In the Stimulating-Instrumental Risk Inventory each statement is scored on a five-point scale with end-points described as 1—*does not describe me at all*; to 5—*describes me very well*.

Moreover, we asked private investors about their motivations for engaging in investment activity. Asset allocation in the stock market is a risky activity itself. Thus, by asking investors what motives underlay their decision to start investing, we wanted to test on the basis of real-life behavior the assumption that 7R+ individuals take more risk because of their need for stimuli. After the study, three independent judges evaluated the answers and grouped them into two categories: (1) the instrumental motivation category in which judges included all motives focused on achieving a specific goal, e.g., multiplying capital, saving for retirement; and (2) stimulating motivation category in which judges included all motives focused on achieving excitement and stimulation, e.g., the need for competition, curiosity. If discrepancies between judges occurred, the fourth independent judge made the final decision.

## Genotyping

For all subjects, we also performed genotyping for the *DRD4* gene. Genomic DNA was extracted from mucosal swabs with the Swab Extract GeneMATRIX DNA Purification Kit (EURx, Gdansk, Poland). Genotyping was performed by the use of amplified fragment length polymorphism (AFLP). The PCR primer sequences and thermal profiles of the reaction were identical to those published by Dmitrieva et al. (2011). The

PCR reaction was conducted in a volume of 20  $\mu$ l with 0.75  $\mu$ l (0.75 U) of Color Perpetual Taq DNA Polymerase, 3  $\mu$ l buffer B, 0.8  $\mu$ l dNTP mix (5 mM each; EURx, Gdansk, Poland), 1.5  $\mu$ l DMSO (DNA Gdansk, Poland), 1.5  $\mu$ l of each primer (10  $\mu$ M), and 150 ng of genomic DNA. PCR products were visualized on 2% agarose gel stained with SimplySafe (EURx, Gdansk, Poland). This study was carried out in accordance with the recommendations of Ethical Committee of the Medical University of Lublin. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol, the procedures of the study and the genotyping was approved by the Ethical Committee of the Medical University of Lublin.

The results of genotyping revealed that among the successfully genotyped group ( $n = 217$ ), 177 individuals were homozygous (10 were 7+/7+, 167 were 7−/7−) and 40 individuals were heterozygous (7+/7−). Fifty participants (24 investors and 26 non-investors) were classified as 7R+ individuals and 167 participants (89 investors and 78 non-investors) were 7R− individuals. The frequencies of the gene variants (7R+ vs. 7R−) did not differ significantly between groups ( $\chi^2[1, n = 217] = 0.43, p = 0.511, \phi = -0.045$ ). Deviations from Hardy-Weinberg equilibrium were determined using the chi-square test. Genotype frequencies were consistent with the Hardy-Weinberg equilibrium (non-investors,  $p = 0.38$ ; investors,  $p = 0.42$ ).

## Statistics

Before the main analysis, we checked the pairwise correlation for the three risk-taking measures that we used (see **Table 2**). Results revealed that there is: (1) a moderate correlation between stimulating and instrumental risk-taking propensity ( $r = 0.45, p < 0.001$ )—this result is consistent with initial results observed by Zaleśkiewicz (2001); and (2) a weak correlation between the Holt-Laury test and instrumental risk-taking propensity

**TABLE 2** | Pairwise correlation between risk-taking measures ( $p$  value in parentheses).

	HLT	SRT	IRT
HLT	1		
SRT	0.111 (0.114)	1	
IRT	<b>0.178 (0.009)</b>	<b>0.448 (&lt; 0.001)</b>	1

HLT, risk-taking propensity measured with Holt-Laury test; SRT, stimulating risk-taking; IRT, instrumental risk-taking.

( $r = 0.18$ ,  $p = 0.009$ ). Thus, we can conclude that the risk-taking measures we used examine different aspects of risk taking.

### Holt-Laury Test

As we mentioned before, the point at which participant decides to switch from Lottery A to Lottery B can indicate one's risk preferences. Usually, participants make their decisions in a way that for the first four lottery pairs they prefer Lottery A (it has higher expected value and also guarantees the safer reward), whereas when making decision about the last four lottery pairs, participants prefer Lottery B (it has clearly higher expected value; Holt and Laury, 2002). The crucial point in the Holt-Laury test is a fifth lottery pair at which higher expected value switches from Lottery A to Lottery B. For this pair, expected values of each option are quite similar (Lottery A: 90 vs. Lottery B: 95). Thus one who is risk averse still prefers Lottery A, where risk seeker switches to Lottery B.

What does this tell us about risk preferences? We might conclude that one who chooses Lottery B during the first five lottery pairs is a risk seeker, whereas one who still prefers Lottery A during the last five lottery pair is risk averse. Thus, we analyzed both halves of lottery pairs as separate variables and each participant was checked for two variables: (1) score for Lottery B choices for lottery pairs

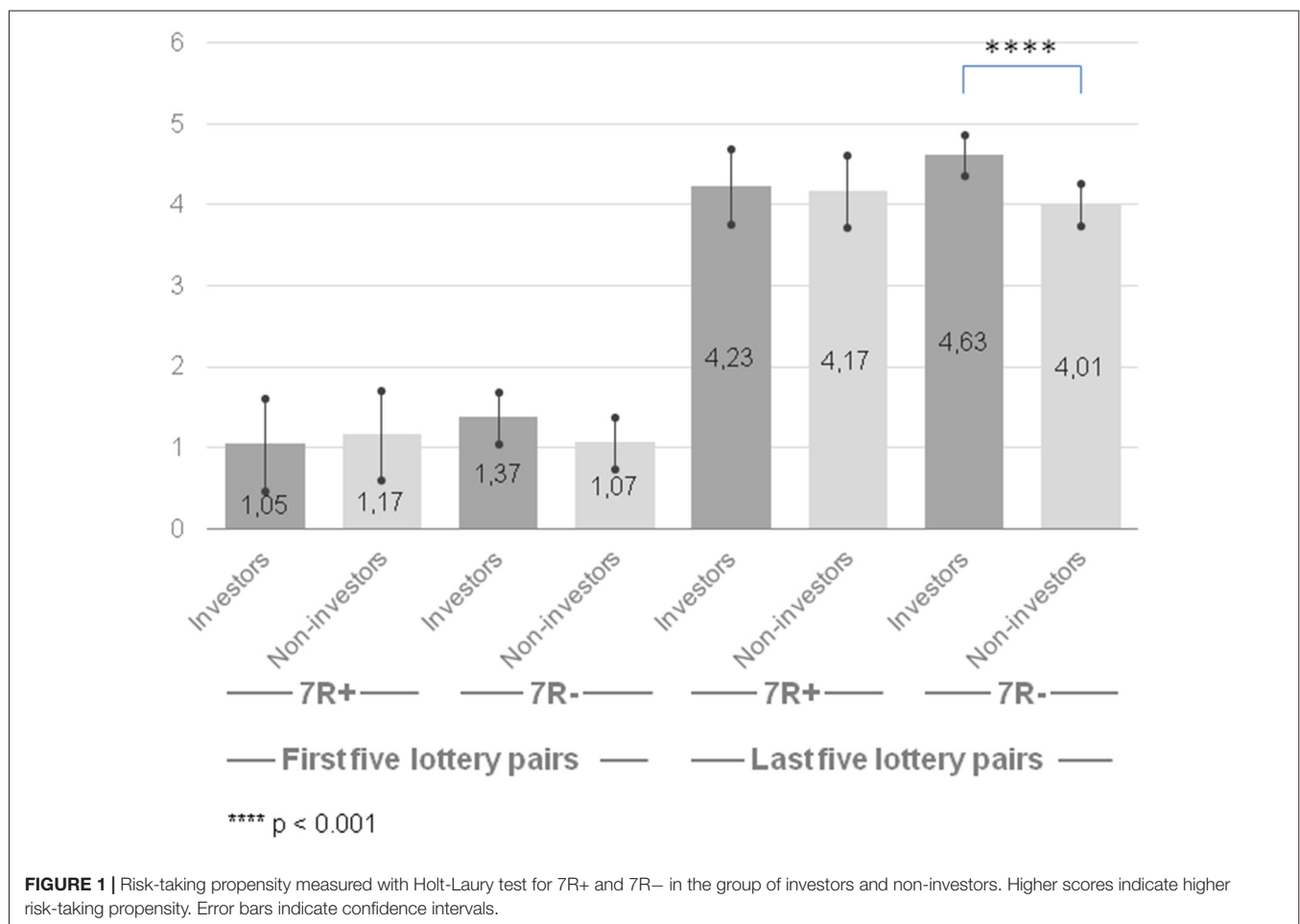
1–5; and (2) score for Lottery B choices for lottery pairs 6–10<sup>2</sup>.

To verify if the specific variant of *DRD4* gene (i.e., 7R+) is associated with higher risk-taking propensity measured with the Holt-Laury test and whether it is moderated by experience in financial risk-taking activity (i.e., being an investor or not), we analyzed our data using a 2 (gene: 7R+ vs. 7R-)  $\times$  2 (group: investors vs. non-investors) univariate analysis of variance (ANOVA; both factors between-subject), separately for: (1) first half of the test (first five lottery pairs); and (2) second half (last five lottery pairs) as dependent variables.

### Instrumental Risk-Taking

To assess if the specific variant of *DRD4* gene (i.e., 7R+) is connected with higher instrumental risk-taking propensity (dependent variable) and whether it is moderated by experience in financial risk-taking activity (i.e., being an investor or not), we analyzed our data using a 2 (gene: 7R+ vs. 7R-)  $\times$  2 (group: investors vs. non-investors) univariate ANOVA (both factors between-subject).

<sup>2</sup>We would like to thank to our Reviewer for suggesting this analysis.



## Stimulating Risk-Taking

To verify if the specific variant of *DRD4* gene (i.e., 7R+) is connected with higher stimulating risk-taking propensity (dependent variable) and whether it is moderated by experience in financial risk-taking activity (i.e., being an investor or not), we analyzed our data using a 2 (gene: 7R+ vs. 7R-)  $\times$  2 (group: investors vs. non-investors) univariate ANOVA (both factors between-subject).

## RESULTS

### Holt-Laury Test

For the Holt-Laury test, we scored each choice of Lottery B (with higher possible payoff and higher variance) as 1 point. Thus, the ultimate risk-seeker who chose in each pair the riskier lottery could achieve the maximum 10-point score. Taking into account that in the last lottery pair, higher payoffs in both lotteries are certain, we decided to exclude participants ( $n = 19$ ) who chose lottery A in the last pair (with a lower payoff)—we suspect this might suggest that they did not understand the task or answered randomly. Hence, the minimum score in the Holt-Laury test was 1. Eventually, we conducted our analysis on a group of 97 investors and 95 non-investors (six participants did not indicate their choices in each lottery pair).

The results of analysis for first five lottery pairs revealed no significant effects. Neither a main effect of gene ( $F_{(1,188)} = 0.25$ ,  $p = 0.618$ ,  $\eta_p^2 = 0.001$ ) nor a main effect of group ( $F_{(1,188)} = 0.15$ ,  $p = 0.695$ ,  $\eta_p^2 = 0.001$ ) was significant. As well, we did not observe significant group  $\times$  sequence ( $F_{(1,188)} = 0.84$ ,  $p = 0.361$ ,  $\eta_p^2 = 0.004$ ) interaction.

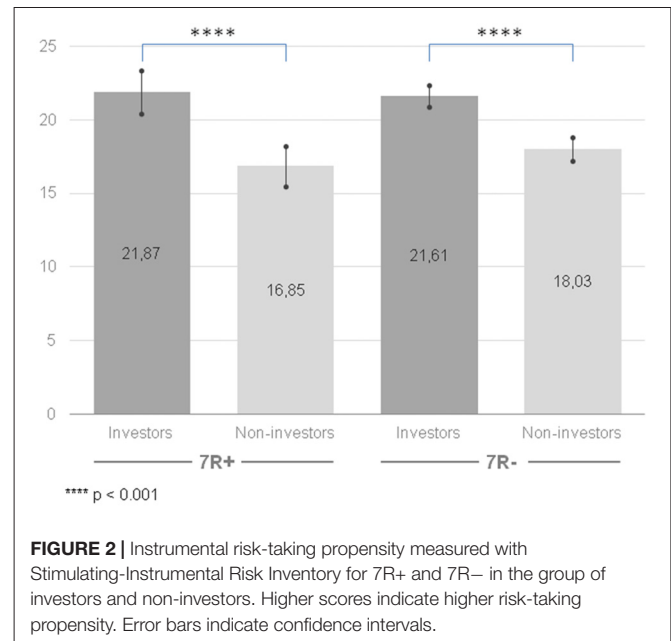
The results of analysis for last five lottery pairs revealed that only a main effect of group was marginally significant ( $F_{(1,188)} = 3.24$ ,  $p = 0.074$ ,  $\eta_p^2 = 0.017$ ). The group of the investors was more risk-taking ( $M = 4.43$ , CI [4.16, 4.69]) than the group of non-investors ( $M = 4.09$ , CI [3.83, 4.35]), however this pattern was observed only for 7R- individuals ( $F_{(1,188)} = 11.20$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.056$ ). Neither a main effect of gene ( $F_{(1,188)} = 0.44$ ,  $p = 0.510$ ,  $\eta_p^2 = 0.002$ ) nor a gene  $\times$  group interaction ( $F_{(1,188)} = 2.18$ ,  $p = 0.142$ ,  $\eta_p^2 = 0.011$ ) was significant (Figure 1).

### Instrumental Risk-Taking

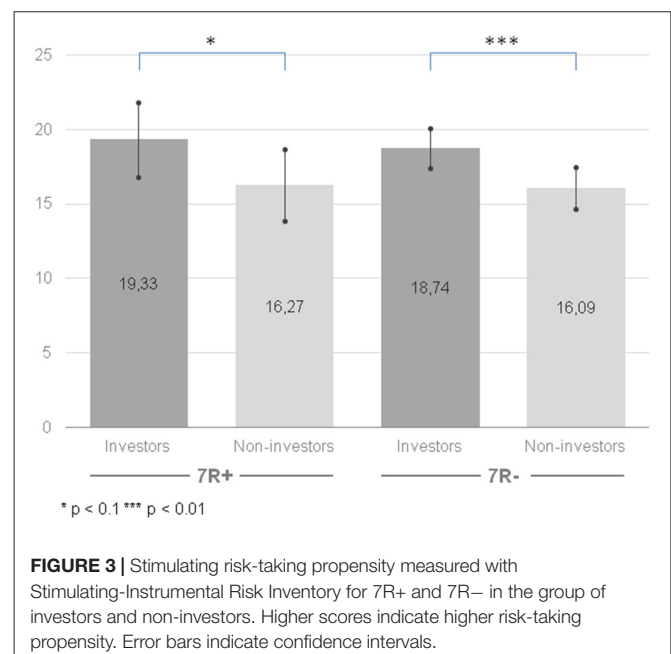
Once again, we observed a significant difference for a main effect of the group factor ( $F_{(1,210)} = 55.43$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.209$ ): the group of investors achieved higher results in instrumental risk-taking propensity ( $M = 21.74$ , CI [20.92, 22.56]) than the group of non-investors ( $M = 17.47$ , CI [16.64, 18.23]). The effect existed when investors and non-investors were compared regardless of their *DRD4* gene variant (see Figure 2). However, there were no differences for a main effect of the gene factor ( $F_{(1,210)} = 0.63$ ,  $p = 0.429$ ,  $\eta_p^2 = 0.003$ ). The interaction of group  $\times$  gene was also not significant ( $F_{(1,210)} = 1.561$ ,  $p = 0.213$ ,  $\eta_p^2 = 0.007$ ).

### Stimulating Risk-Taking

Similarly, like in the case of instrumental risk-taking, our analysis indicated significant differences for a main effect of the group



( $F_{(1,208)} = 8.022$ ,  $p = 0.005$ ,  $\eta_p^2 = 0.037$ ): the group of the investors was more prone to stimulating risk-taking ( $M = 19.04$ , CI [17.62, 20.45]) than the group of non-investors ( $M = 16.18$ , CI [14.79, 17.58]). Once again, the effect persisted when comparing 7R- investors ( $M = 18.74$ , CI [17.41, 20.06]) with 7R- non-investors ( $M = 16.09$ , CI [14.69, 17.50];  $F_{(1,208)} = 7.29$ ,  $p = 0.008$ ,  $\eta_p^2 = 0.034$ ) and was slightly significant between 7R+ investors ( $M = 19.33$ , CI [16.83, 21.84]) and 7R+ non-investors ( $M = 16.27$ , CI [13.86, 18.68];  $F_{(1,208)} = 3.03$ ,  $p = 0.083$ ,  $\eta_p^2 = 0.014$ ; see Figure 3). However, contrary to our hypothesis, we did not observe a main effect of the gene ( $F_{(1,208)} = 0.147$ ,  $p = 0.702$ ,  $\eta_p^2 = 0.001$ ).





The interaction of group  $\times$  gene ( $F_{(1,208)} = 0.043$ ,  $p = 0.836$ ,  $\eta_p^2 < 0.001$ ) was also not significant.

## Motivation to Engage in Investment Activity

We compared the frequencies of motivation to engage in investment activity between 7R+ and 7R− individuals (stimulating motivation vs. instrumental motivation). Ninety-seven investors indicated an answer to the question about their motivation to engage in investment activity (missing data  $n = 16$ ). Once more, we did not observe a significant difference ( $\chi^2[1, n = 97] = 1.35$ ,  $p = 0.245$ ,  $\phi = 0.118$ ) between the 7R+ (10 of 22 investors indicated the stimulating motivation) and 7R− individuals (24 of 75 investors indicated the stimulating motivation).

## DISCUSSION

We hypothesized that the previous inconclusive results about the *DRD4* gene might be explained by the moderating role of the motivation to take risk. Namely, if the dopamine gene *DRD4* is associated with a blunted response to dopamine and 7R+ individuals need to seek higher stimulation to feel the same activation in the dopamine reward pathway compared to 7R− individuals, then 7R+ individuals should be more motivated to engage in risky activities that deliver arousal. However, we failed to notice any differences between the 7R+ and 7R− individuals on: (1) the stimulating risk-taking scale; (2) the instrumental risk-taking scale; (3) their indicated motivation to engage in investment activity; and (4) the experimental task—the Holt-Laury test. We observed no differences between neither 7R+ and 7R− investors nor 7R+ and 7R− non-investors. On the other hand, we found evidence that investors are more prone to take risk than non-investors. This result was present for the stimulating and instrumental risk-taking scales (Zaleśkiewicz,

2001). For the Holt-Laury test (Holt and Laury, 2002) we noticed that only 7R− investors were more risk-seeking than non-investors. This might suggest that we used appropriate risk-taking measures, which might distinguish groups with different levels of risk-taking propensity.

Nevertheless, our study is another one to report a lack of differences between 7R+ and 7R− individuals in the domain of financial risk-taking. To our knowledge, our study is the second one that focused on a group of active investors who are experienced in financial decisions and risk-taking. In a previous study, conducted by Anderson et al. (2015), a sample of 140 active investors were examined, and there was no significant relationship between the *DRD4* gene and risk-taking in three risk-taking measures: measures of equity holdings, multiple price listing, and the survey risk measure. Also, Dreber et al. (2012) failed to find differences between 7R+ and 7R− when the subject pool was composed of professional decision-makers (i.e., owners, presidents, and managers of large companies). Only one study (Dreber et al., 2011) where participants were not undergraduate students noticed a significant association between the 7R+ variant and risk-taking (see Table 3 for a summary of previous results and tested subject pool). These findings and our results might suggest it is likely that the relationship of the *DRD4* gene with risk-taking is mediated by environmental factors, e.g., experience, familiarity with risky situations, or wealth. For example, Lo and Repin (2002) demonstrated that during live trading sessions, the autonomic responses of more experienced investors were significantly lower than less experienced traders. It is possible then that the level of experience among our subject pool was heterogeneous, and, thus, a few factors were associated with lower emotional reactions, not only the specific variant of the *DRD4* gene. This might be a reason why the 7R+ and 7R− investors did not differ in risk-taking propensity.

We are cautious with interpreting our results and do not claim that there is no relationship between the *DRD4* gene

**TABLE 3 |** Summary of the existing studies on the *DRD4* gene and risk-taking propensity.

Reference	Subject pool	Group size	Risk-taking measures	Result
Dreber et al. (2009)	Undergraduate students	94 (7R+ $n = 24$ )	Experimental investment task	7R+ more risk-taking
Kuhnen and Chiao (2009)	Undergraduate students	65 (7R+ $n = 15$ )	Experimental investment task	7R+ more risk-taking
Carpenter et al. (2011)	Mainly undergraduate students ( $n = 125$ )	140 (7R+ $n = 51$ )	Three gambling tasks—lottery choices with: 1. Known probabilities 2. Ambiguous probabilities 3. Possible loss	1. No differences 2. 7R+ more risk-taking 3. 7R+ more risk-taking ( $p = 0.10$ )
Dreber et al. (2011)	Bridge players	98 men (7R+ $n = 16$ ) 77 women (7R+ $n = 6$ )	1. Bridge risk-taking 2. Experimental investment task	7R+ more risk-taking only among men in both measures
Eisenegger et al. (2010)	No info, mean age 23.5 (SD = 3.6)	200 (7R+ $n = 42$ )	Gambling task	No differences in control (placebo administration) group
Frydman et al. (2011)	Undergraduate students	90 (no info)	Gambling task	No differences
Dreber et al. (2012)	Owners, presidents, and managers of large companies	121 (7R+ $n = 17$ )	Experimental investment task	No differences
Anderson et al. (2015)	Investors	149 (7R+ $n = 53$ )	1. Measures of equity holdings 2. Multiple price listing 3. Survey risk measure (Dohmen et al., 2011)	No significant differences

and risk-taking. There are numerous studies demonstrating that genes may determine risk preferences (e.g., Cesarini et al., 2010; Cronqvist and Siegel, 2014) and also a few studies have revealed that 7R+ individuals take more risks than 7R− individuals (Dreber et al., 2009, 2011; Kuhnen and Chiao, 2009; Carpenter et al., 2011). Nevertheless, as Benjamin et al. (1996) observed on a group of almost 10,000 subjects, the single nucleotide polymorphism across the human genome can explain a maximum 1.25% variation of any psychological trait. Moreover, the association of the *DRD4* gene and risk taking is probably a complex phenomenon and the risk-taking trait in general depends on many factors, such as individual differences, sex, age, financial knowledge, income and cognitive abilities (Hallahan et al., 2003; Bali et al., 2009; Burks et al., 2009; Mayfield and Shapiro, 2010).

Our present study reveals that the type of motivation (i.e., stimulating and instrumental) underlying the risk-taking activity is not a factor that mediates the relationship between *DRD4* and risk taking. Perhaps our main finding is evidence that 7R+ individuals might be highly heterogeneous. As we observed, 7R+ investors were significantly more prone to risk-taking than 7R+ non-investors. To our knowledge, this is the first study that reports differences between two groups of 7R+ individuals and gives strict evidence that the variation in risk-taking among 7R+ individuals is environmentally sensitive and might depend on factors like familiarity with financial risky decision-making, i.e., being an investor or not.

Of course, our study has limitations. As one of the risk-taking measures, we used the Holt-Laury test with only hypothetical payoffs. This could be perceived by our subjects (especially investors) as not engaging and thus induce responses not convergent with real-life risk-taking propensity. However, as Holt and Laury (2002) indicated, using high hypothetical payoffs (as in our study) elicits the proper level of risk aversion. Moreover, as Camerer and Hogarth (1999) noticed on the basis of 74 studies with no, low, or high real payoffs, the presence of monetary incentives does not influence the mean performance. Thus, we believe that the level of risk-taking propensity measured with the Holt-Laury test was not affected by the lack of possible winnings. Another possible limitation is that we used a questionnaire scale to assess the stimulating and instrumental risk-taking propensity. Due to self-reported estimations that highly rely on self-perception, subjects could not accurately present their real behaviors. For example, Brañas-Garza et al. (in press) observed using a large sample that the digit ratio (2D:4D—a biomarker for prenatal testosterone exposure) was significantly associated with risk preferences; however, this was noticed when risk-taking propensity was measured by the experimental task. There was no relationship between 2D:4D and risk-taking propensity as measured by the self-reported scale. As Brañas-Garza et al. (in press) noticed, this result could arise because of the complexity of risk-taking behavior and the fact that various risk-taking measures correlate only imperfectly. However, in our study, we observed a lack of differences not only in self-reported risk-taking propensity but in the experimental task as well.

Moreover, we used measures that examine different nuances of risk taking—we observed only a moderate correlation between stimulating and instrumental risk-taking scales and a weak correlation between instrumental risk-taking and the Holt-Laury test. All of this suggests that the lack of differences between 7R+ and 7R− individuals in our study is not a case of inadequate selection of methods but is rather a robust finding. Also, the lack of differences in motivation for engaging in investment activity between 7R+ and 7R− investors seems to be in line with the above assumption. As we mentioned before, asset allocation in the stock market is a risky activity itself. Thus, if the 7R+ individuals should seek more stimuli to overcome the blunted response to dopamine, we should expect that they would be more willing to engage in investment activity because for reasons of stimulation. However, one more time we observed no differences between 7R+ and 7R− individuals, which supports previous results.

As previous studies revealed inconclusive results about the association between the *DRD4* gene and risk-taking, it is worth wondering whether this relation might be moderated by some other psychological factors than instrumental and stimulating risk-taking. For example, if 7R+ are more risk-taking due to the need for stimulation and seeking for positive feelings, it is possible that individual differences in susceptibility to affect might moderate this relation. Consider a 7R+ individual who is not sensitive to changes in affect—we can imagine that in such a case two factors might work in opposite directions: the 7R+ variant increases the need for stimuli, whereas the lack of susceptibility to affect attenuates this impact. Thus, changes in arousal and emotional states might not have an impact on the behavior of individuals with low susceptibility to affect (a 7R+ individual). In our study, we wanted to avoid the issues related to multiple testing and thus, we decided to focus only on two psychological factors: instrumental and stimulating risk-taking. Hence, this explanation is only hypothetical and needs further investigation.

Moreover, in our study we focused solely on psychological factors that could potentially mediate relation between *DRD4* gene and risk-taking; and as previous studies revealed (e.g., Docherty et al., 2012) also epigenetic processes associated with e.g., methylation levels at the promoter of the *DRD4* gene may mediate genetic influences. It was revealed that methylation levels at the promoter of the *DRD4* gene are associated with schizophrenia (Cheng et al., 2014), Alzheimer's disease (Ji et al., 2016), drug addiction (Ji et al., 2018) and alcohol dependence (Zhang et al., 2013). Thus, future work is needed to verify if other than psychological factors (e.g., methylation levels) might also mediate the relation between the *DRD4* gene and financial risk-taking.

It is also worth noting that our procedure included only tasks that probably did not induce the feelings of excitement or stimuli. Perhaps a procedure with tasks that elicit arousal is needed to catch the differences between the 7R+ and 7R− individuals in the domain of financial risk-taking. A similar procedure with “cold” (less emotional) and “hot” (much more arousing) tasks

were used by Costa et al. (2014) to examine the impact of a factor that might potentially decrease emotional arousal on decision-making. What occurred was that in the case of the “hot” version, significant differences were observed. The “cold” one revealed no significant results.

At the end, note that in our study we used the traditional procedure of Holt-Laury test that is, we presented items in a fixed order starting with a very low probability of winning a higher prize that increased in subsequent lottery pairs. Such sequence could suggest the strategy of choices based on a need for consistency to avoid cognitive dissonance (Festinger, 1957): “If I chose riskier lottery (Lottery B) in the earlier pair I would also do the same in a next step (when Lottery B is less risky)”. It is possible that participants, especially investors who are familiarized with financial decision making, noticed such linear sequence what could potentially influence their choices. Thus, it would be beneficial to test how subjects respond to Holt-Laury test when presenting the items in a random way.

In sum, we still need more research to better understand the genetic foundations of risk-taking, which could answer the question about the substantial variation in the domain of risky financial decisions. However, it seems that we need to examine homogeneous groups, i.e., undergraduate students, if we want to observe substantial differences. Otherwise, the effect of the genes might be suppressed by environmental factors.

## REFERENCES

- Adriani, W., Boyer, F., Gioiosa, L., Macri, S., Dreyer, J. L., and Laviola, G. (2009). Increased impulsive behavior and risk proneness following lentivirus-mediated dopamine transporter over-expression in rats' nucleus accumbens. *Neuroscience* 159, 47–58. doi: 10.1016/j.neuroscience.2008.11.042
- Anderson, A., Dreber, A., and Vestman, R. (2015). Risk taking, behavioral biases and genes: results from 149 active investors. *J. Behav. Exp. Finance* 6, 93–100. doi: 10.1016/j.jbef.2015.04.002
- Asghari, V., Sanyal, S., Buchwaldt, S., Paterson, A., Jovanovic, V., and Van Tol, H. H. (1995). Modulation of intracellular cyclic AMP levels by different human dopamine D4 receptor variants. *J. Neurochem.* 65, 1157–1165. doi: 10.1046/j.1471-4159.1995.65031157.x
- Bali, T., Demirtas, O., Levy, H., and Wolf, A. (2009). Bond versus stocks: investors' age and risk taking. *J. Monet. Econ.* 56, 817–830. doi: 10.1016/j.jmoneco.2009.06.015
- Benjamin, J., Li, L., Patterson, C., Greenberg, B. D., Murphy, D. L., and Hamer, D. H. (1996). Population and familial association between the D4 dopamine receptor gene and measures of novelty seeking. *Nat. Genet.* 12, 81–84. doi: 10.1038/ng0196-81
- Berridge, K. C., and Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res. Rev.* 28, 309–369. doi: 10.1016/s0165-0173(98)00019-8
- Brañas-Garza, P., Galizzi, M. M., and Nieboer, J. (in press). Experimental self-reported measures of risk taking digit ratio (2D:4D): evidence from a large, systematic study. *Int. Econ. Rev.*
- Burks, S., Carpenter, J., Goette, L., and Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proc. Natl. Acad. Sci. U S A* 106, 7745–7750. doi: 10.1073/pnas.0812360106
- Camerer, C. F., and Hogarth, R. H. (1999). The effects of financial incentives in experiments: a review and capital labor production framework. *J. Risk Uncertain.* 19, 7–42. doi: 10.1023/A:1007850605129
- Carpenter, J. P., Garcia, J. R., and Lum, J. K. (2011). Dopamine receptor genes predict risk preferences, time preferences, and related economic choices. *J. Risk Uncertain.* 42, 233–261. doi: 10.1007/s11166-011-9115-3
- Cesarini, D., Johannesson, M., Lichtenstein, P., Sandewall, O., and Wallace, B. (2010). Genetic variation in financial decision making. *J. Finance* 65, 1725–1754. doi: 10.1111/j.1540-6261.2010.01592.x
- Cheng, J., Wang, Y., Zhou, K., Wang, L., Li, J., Zhuang, Q., et al. (2014). Male-specific association between dopamine receptor D4 gene methylation and schizophrenia. *PLoS One* 9:e89128. doi: 10.1371/journal.pone.0089128
- Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J., et al. (2014). Your morals depend on language. *PLoS One* 9:e94842. doi: 10.1371/journal.pone.0094842
- Cronqvist, H., and Siegel, S. (2014). The genetics of investment biases. *J. Financ. Econ.* 113, 215–234. doi: 10.1016/j.jfineco.2014.04.004
- Dmitrieva, J., Chen, C., Greenberger, E., Ogunseitan, O., and Ding, Y. C. (2011). Gender-specific expression of the DRD4 gene on adolescent delinquency, anger and thrill seeking. *Soc. Cogn. Affect. Neurosci.* 6, 82–89. doi: 10.1093/scan/nsq020
- Docherty, S. J., Davis, O. S., Haworth, C. M., Plomin, R., D'Souza, U., and Mill, J. (2012). A genetic association study of DNA methylation levels in the DRD4 gene region finds associations with nearby SNPs. *Behav. Brain Funct.* 8:31. doi: 10.1186/1744-9081-8-31
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: measurement, determinants, and behavioral consequences. *J. Eur. Econ. Assoc.* 9, 522–550. doi: 10.1111/j.1542-4774.2011.01015.x
- Dorn, D., and Sengmueller, P. (2009). Trading as entertainment? *Manage. Sci.* 55, 591–603. doi: 10.1287/mnsc.1080.0962
- Dreber, A., Apicella, C. L., Eisenberg, D. T. A., Garcia, J. R., Zamore, R. S., Lum, J. K., et al. (2009). The 7R polymorphism in the dopamine receptor D4 gene (DRD4) is associated with financial risk taking in men. *Evol. Hum. Behav.* 30, 85–92. doi: 10.1016/j.evolhumbehav.2008.11.001
- Dreber, A., Rand, D. G., Wernerfelt, N., Garcia, J. R., Vilar, M. G., Lum, J. K., et al. (2011). Dopamine and risk choices in different domains: findings among serious tournament bridge players. *J. Risk Uncertain.* 43, 19–38. doi: 10.1007/s11166-011-9119-z

## AUTHOR CONTRIBUTIONS

RM, MK and MG conceived the study. RM, MK, MM-W, MG, AF, PG and PM designed the study protocol. RM drafted the manuscript, coordinated the data gathering, and carried out the statistical analyses. MK, MM-W and MG helped draft the manuscript. MM-W conducted the genetic analyses. AF, PG and PM revised the manuscript. AF provided advice and facilities for genetic analyses. All authors gave final approval for publication.

## FUNDING

This work was supported by the National Science Centre, Poland (NCN) under Grant 2015/17/D/HS6/02684.

## ACKNOWLEDGMENTS

We would like to thank Barbara Sioma, Marlena Joniec, Łukasz Kuśmierz, Patryk Wróblewski, Ewa Hałas-Korhan, and Paweł Kozłowski for great help with data gathering. We also thank Tadeusz Tyszką and audiences at the meeting of the Center for Economic Psychology and Decision Sciences for helpful comments. Special thanks to Aleksandra Muda for her invaluable help and suggestions on an earlier version of this article.

- Dreber, A., Rand, D. G., Wernerfelt, N., Montgomery, C., and Malhotra, D. K. (2012). Genetic correlates of economic and social risk taking. *SSRN working paper* 2141601. doi: 10.2139/ssrn.2141601
- Ebstein, R. P., Novick, O., Umansky, R., Priel, B., Osher, Y., Blaine, D., et al. (1996). Dopamine D4 receptor (D4DR) exon III polymorphism associated with the human personality trait of novelty seeking. *Nat. Genet.* 12, 78–80. doi: 10.1038/ng0196-78
- Eisenberg, D. T., Mackillop, J., Modi, M., Beauchemin, J., Dang, D., Lisman, S. A., et al. (2007). Examining impulsivity as an endophenotype using a behavioral approach: a *DRD2 TaqI A* and *DRD4 48-bp VNTR* association study. *Behav. Brain Funct.* 3:2. doi: 10.1186/1744-9081-3-2
- Eisenegger, C., Knoch, D., Ebstein, R. P., Gianotti, L. R., Sandor, P. S., and Fehr, E. (2010). Dopamine receptor D4 polymorphism predicts the effect of L-DOPA on gambling behavior. *Biol. Psychiatry* 67, 702–706. doi: 10.1016/j.biopsych.2009.09.021
- Ernst, M., and Luciana, M. (2015). Neuroimaging of the dopamine/reward system in adolescent drug use. *CNS Spectr.* 20, 427–441. doi: 10.1017/s1092852915000395
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.
- Frydman, C., Camerer, C., Bossaerts, P., and Rangel, A. (2011). MAOA-L carriers are better at making optimal financial decisions under risk. *Proc. Biol. Sci.* 278, 2053–2059. doi: 10.1098/rspb.2010.2304
- Grady, D. L., Chi, H. C., Ding, Y. C., Smith, M., Wang, E., Schuck, S., et al. (2003). High prevalence of rare dopamine receptor D4 alleles in children diagnosed with attention-deficit hyperactivity disorder. *Mol. Psychiatry* 8, 536–545. doi: 10.1038/sj.mp.4001350
- Hallahan, T., Faff, R., and McKenzie, M. (2003). An exploratory investigation of the relation between risk tolerance scores and demographic characteristics. *J. Multinational Financ. Manag.* 13, 483–502. doi: 10.1016/s1042-444x(03)00022-7
- Holt, C. A., and Laury, S. K. (2002). Risk aversion and incentive effects. *Am. Econ. Rev.* 92, 1644–1655. doi: 10.2139/ssrn.893797
- Ji, H., Wang, Y., Jiang, D., Liu, G., Xu, X., Dai, D., et al. (2016). Elevated *DRD4* promoter methylation increases the risk of Alzheimer's disease in males. *Mol. Med. Rep.* 14, 2732–2738. doi: 10.3892/mmr.2016.5560
- Ji, H., Xu, X., Liu, G., Liu, H., Wang, Q., Shen, W., et al. (2018). Dopamine receptor D4 promoter hypermethylation increases the risk of drug addiction. *Exp. Ther. Med.* 15, 2128–2133. doi: 10.3892/etm.2017.5615
- Kuhnen, C., and Chiao, J. (2009). Genetic determinants of financial risk taking. *PLoS One* 4:e4362. doi: 10.1371/journal.pone.0004362
- Kuhnen, C. M., and Knutson, B. (2005). The neural basis of financial risk taking. *Neuron* 47, 763–770. doi: 10.1016/j.neuron.2005.08.008
- Laucht, M., Becker, K., Blomeyer, D., and Schmidt, M. H. (2007). Novelty seeking involved in mediating the association between the dopamine D4 receptor gene exon III polymorphism and heavy drinking in male adolescents: results from a high-risk community sample. *Biol. Psychiatry* 61, 87–92. doi: 10.1016/j.biopsych.2006.05.025
- Lo, A. W., and Repin, D. V. (2002). The psychophysiology of real-time financial risk processing. *J. Cogn. Neurosci.* 14, 323–339. doi: 10.1162/089892902317361877
- Mayfield, S., and Shapiro, M. (2010). Gender and risk: women, risk taking and risk aversion. *Gender Manag. Int. J.* 25, 586–604. doi: 10.1108/17542411011081383
- McBride, W. J., Murphy, J. M., and Ikemoto, S. (1999). Localization of brain reinforcement mechanisms: intracranial self-administration and intracranial place-conditioning studies. *Behav. Brain Res.* 101, 129–152. doi: 10.1016/s0166-4328(99)00022-4
- Pérez de Castro, I., Ibáñez, A., Torres, P., Sáiz-Ruiz, J., and Fernández-Piqueras, J. (1997). Genetic association study between pathological gambling and a functional DNA polymorphism at the D4 receptor gene. *Pharmacogenetics* 5, 345–348.
- Pierce, R. C., and Kumaresan, V. (2006). The mesolimbic dopamine system: the final common pathway for the reinforcing effect of drugs of abuse? *Neurosci. Biobehav. Rev.* 30, 215–238. doi: 10.1016/j.neubiorev.2005.04.016
- Ptáček, R., Kuzelova, H., and Stefano, G. B. (2011). Dopamine D4 receptor gene *DRD4* and its association with psychiatric disorders. *Med. Sci. Monit.* 17, RA215–RA220. doi: 10.12659/msm.881925
- Schoots, O., and Van Tol, H. H. M. (2003). The human dopamine D4 receptor repeat sequences modulate expression. *Pharmacogenomics J.* 3, 343–348. doi: 10.1038/sj.tpj.6500208
- Schwarz, N. (2012). “Feelings-as-information theory,” in *Handbook of Theories of Social Psychology*, eds P. A. M. Van Lange, A. Kruglanski and E. T. Higgins (Thousand Oaks, CA: Sage), 289–308.
- Tyszk, T. (2010). *Decyzje: Perspektywa Psychologiczna i Ekonomiczna*. Warszawa: Wydawnictwo Naukowe Scholar.
- Wise, R. (2002). Brain reward circuitry: insights from uncensored incentives. *Neuron* 36, 229–240. doi: 10.1016/S0896-6273(02)00965-0
- Zaleskiewicz, T. (2001). Beyond risk seeking and risk aversion: personality and the dual nature of economic risk taking. *Eur. J. Pers.* 15, S105–S122. doi: 10.1002/per.426
- Zhang, H., Herman, A. I., Kranzler, H. R., Anton, R. F., Zhao, H., Zheng, W., et al. (2013). Array-based profiling of DNA methylation changes associated with alcohol dependence. *Alcohol. Clin. Exp. Res.* 37, E108–E115. doi: 10.1111/j.1530-0277.2012.01928.x

**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 © 2018 Muda, Kicia, Michalak-Wojnowska, Ginszt, Filip, Gawda and Majcher. 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) and the copyright owner 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.





# Facts and Misconceptions about 2D:4D, Social and Risk Preferences

Judit Alonso<sup>1</sup>, Roberto Di Paolo<sup>1</sup>, Giovanni Ponti<sup>1,2,3</sup> and Marcello Sartarelli<sup>1\*</sup>

<sup>1</sup> Departamento de Fundamentos de Análisis Económico, Universidad de Alicante, San Vicente del Raspeig/Sant Vicent del Raspeig, Alicante, Spain, <sup>2</sup> Department of Economics, The University of Chicago, Chicago, IL, United States, <sup>3</sup> Dipartimento di Economia e Finanza, Libera Università Internazionale degli Studi Sociali Guido Carli (LUISS), Rome, Italy

We study how the ratio between the length of the second and fourth digit (2D:4D) correlates with choices in social and risk preferences elicitation tasks by building a large dataset from five experimental projects with more than 800 subjects. Our results confirm the recent literature that downplays the link between 2D:4D and many domains of economic interest, such as social and risk preferences. As for the former, we find that social preferences are significantly lower when 2D:4D is above the median value only for subjects with low cognitive ability. As for the latter, we find that a high 2D:4D is not correlated with the frequency of subjects' risky choices.

**Keywords:** 2D:4D, cognitive reflection, gender, risk, social preferences

**JEL Classification:** C91, C92, D8

## 1. INTRODUCTION

Research both in the hard sciences (e.g., Neurology and Physiology) and in the social sciences (e.g., Economics and Psychology) has increasingly focused on biological markers to improve our understanding of the biological basis of social behavior. Earlier research had claimed that prenatal exposure to sexual hormones has an effect on brain development that, in turn, influences individuals' decision making routines later in life (see for a survey Manning, 2002). Motivated by this evidence, a growing number of experimental studies has tested the relationship between the ratio between the second and fourth hand digit (2D:4D hereafter) - a marker which has been claimed to be negatively related to prenatal exposure to testosterone- and behavior in a wide variety of cognitive domains, including social and risk preferences.

Social preferences are a ubiquitous phenomenon in everyday life and have gained increasing attention in the social sciences. While there is robust evidence that shows that females exhibit more pronounced social concerns, only few studies have looked at their relationship with 2D:4D. Within this small set, Millet and Dewitte (2006) find a negative relationship between 2D:4D and giving in the dictator game. Using a variety of games, such as public good and dictator, Buser (2012) finds, instead, a positive relationship with giving. In related studies using the ultimatum game, Brañas-Garza et al. (2013) find that the relationship with giving follows an inverted U-shape while Van den Bergh and Dewitte (2006) find a negative relationship with rejection rates.

The relationship between 2D:4D and risk-taking has been widely studied experimentally to quantify the role played by innate traits in this type of decisions. Again, the evidence so far is mixed, as some studies find a negative relationship with the frequency of risky choices (e.g., Garbarino et al., 2011; Brañas-Garza et al., 2018) while others do not find any significant correlation (e.g., Apicella et al., 2008; Sapienza et al., 2009).

We contribute to this literature by assembling a meta-dataset consisting of five experimental projects involving 879 subjects in total. With this large dataset collecting

## OPEN ACCESS

### Edited by:

Levent Neyse,  
Institut für Weltwirtschaft, Germany

### Reviewed by:

Paul Smeets,  
Maastricht University, Netherlands  
Erik Bijleveld,  
Radboud University Nijmegen,  
Netherlands

### \*Correspondence:

Marcello Sartarelli  
marcellosartarelli@gmail.com

**Received:** 31 August 2017

**Accepted:** 25 January 2018

**Published:** 13 February 2018

### Citation:

Alonso J, Di Paolo R, Ponti G and  
Sartarelli M (2018) Facts and  
Misconceptions about 2D:4D, Social  
and Risk Preferences.  
*Front. Behav. Neurosci.* 12:22.  
doi: 10.3389/fnbeh.2018.00022

evidence on behavioral tasks of a different nature, we first assess the relationship between 2D:4D and *inequity aversion* (Fehr and Schmidt, 1999), a proxy for social preferences that identifies the role of “envy” (i.e., *negative inequity aversion*) in comparison with “guilt” (i.e., *positive inequity aversion*). Second, we assess the relationship between 2D:4D and risk attitudes, which were elicited using Multiple Price Lists (Holt and Laury, 2002). Finally, following some recent contributions (Brañas-Garza et al., 2015; Cueva et al., 2016), we also assess the mediating role played by cognitive ability in the relationship between 2D:4D and subjects’ decisions in both risk and distributional tasks.

We briefly summarize here our main results, that have been obtained by defining right hand 2D:4D high if it is greater than the gender-specific median value. When we look at social preferences, we find that for subjects with high 2D:4D the relationship with guilt is negative but not significant, whereas the relationship with envy is only significant and negative for subjects with low cognitive ability. If we, instead, use directly 2D:4D measures we find no significant association with social preferences. When we look at risk preferences, we find that the association between high 2D:4D and the frequency of risky choices is negative but not significant, with similar results holding if we use the raw 2D:4D index as a covariate. Overall, our empirical findings cannot but confirm some recent literature (discussed in section 2) which downplays the link between 2D:4D and behavior in experimental domains of interests, such as social and risk preferences.

The remainder of the paper is structured as follows. Section 2 reviews the related literature while section 3 describes the layout of our meta-dataset. In section 4, we report correlations between 2D:4D, gender and cognitive ability distilled from the debriefing questionnaire. In section 5 we report our findings on the relationship between 2D:4D and inequity aversion and in section 6 we look at risk attitudes. Finally, section 7 discusses our results and concludes, followed by an appendix collecting additional statistical evidence.

## 2. LITERATURE REVIEW

The ratio between the length of the second (“index” finger) and fourth (“ring” finger) digit, also called second-to-fourth digit ratio (2D:4D), has been claimed to be a proxy for prenatal exposure to testosterone, with a lower ratio indicating higher exposure both for children and for adults (Manning et al., 1998). Related studies find a positive correlation between sex hormones at birth and 2D:4D measured at age 2 (Lutchmaya et al., 2004; Ventura et al., 2013). More recently Hollier et al. (2015) have challenged this view by providing evidence that the relationship between a measure of exposure to testosterone obtained using umbilical cord blood and 2D:4D measured at age 19–22 is not significant<sup>1</sup>. However, this result may be due by the fact that testosterone peaks between 12 and 18 weeks of gestation and decreases thereafter (Xie et al., 2017). In addition, in a replication study, (Hönekopp et al., 2007) find no systematic evidence of a relationship between 2D:4D and circulating sex

hormones in adults. On the one hand, this result suggests that estimating the relationship between 2D:4D and proxies for decision-making without accounting for circulating testosterone does not lead to omitted variable bias. On the other, it suggests that additional research is awaited to obtain conclusive evidence on the relationship between 2D:4D and testosterone subjects are exposed to from gestation to adulthood.

Several studies have also shown that 2D:4D is a sexually dimorphic measure with, on average, males having lower 2D:4D than females (Putz et al., 2004). Moreover, earlier studies have reported that 2D:4D varies not only by gender, but also by ethnicity (Manning, 2002). It has also been found that these differences emerge prenatally and are stable during the developing years (Trivers et al., 2006). Voracek et al. (2007) carry out a wide replication study of published results on the relationship between 2D:4D and a variety of outcomes and, overall, confirm the results.

The literature on the relationship between 2D:4D and social preferences is scant and, again, results are mixed. Buser (2012) finds that in public good, dictator, trust and ultimatum games subjects with higher 2D:4D are more generous. By contrast, Brañas-Garza and Kovářík (2013) argue that, since 2D:4D measures in Buser (2012) are self-reported, his results may be affected by measurement error and biased if the error is correlated with one or more subjects’ characteristics.

As for the experimental evidence on the dictator game, Millet and Dewitte (2006) find, instead, a negative relationship between 2D:4D and giving. In related experimental studies using ultimatum games, Van den Bergh and Dewitte (2006) find a negative relationship between 2D:4D and rejection rates while Brañas-Garza et al. (2013) find evidence of non-linearities in the relationship, with subjects with either high or low 2D:4D giving less. A non-linear relationship is also found by Sanchez-Pages and Turiegano (2010) for the one-shot prisoner’s dilemma, with men with intermediate 2D:4D being more likely to cooperate<sup>2</sup>.

As for the relationship between 2D:4D and risk-taking behavior, results are mixed (see for a survey Apicella et al., 2015). Dreber and Hoffman (2007); Garbarino et al. (2011); Brañas-Garza et al. (2018) find a negative relationship for both genders, with Brañas-Garza et al. (2018) also finding that the relationship with a self-assessed and subjective measure of risk attitudes is not significant. Similarly, Ronay and von Hippel (2010); Brañas-Garza and Rustichini (2011); Stenstrom et al. (2011) find a negative relationship although only for males, with Brañas-Garza and Rustichini (2011) also finding that this result is mediated by a negative relationship between 2D:4D and abstract reasoning

<sup>1</sup>See Kaltwasser et al. (2017) for analogous findings.

<sup>2</sup>Related studies manipulate experimentally hormones levels and estimate their relationship with proxies for social preferences. Zak et al. (2009) increase the level of circulating testosterone and find that it decreases giving in ultimatum games. Kosfeld et al. (2005); Zak et al. (2007) increase, instead, levels of oxytocin, a hormone that is hypothesized to increase empathy in humans, and find that it has a positive impact on giving in ultimatum games but not in dictator games, which they interpret as evidence of generosity. In addition, neuroeconomic evidence shows that exposure to prenatal hormones (testosterone or estrogen) may affect the activity in specific brain areas that are associated with individuals’ behavior in several settings and with their personality (Fehr and Camerer, 2007; Lee, 2008; Fehr and Krajbich, 2009).

ability, an aspect of cognitive ability that was measured using the Raven Progressive Matrices task. In contrast, a number of studies find that the relationship is not significant at any conventional level (Apicella et al., 2008; Sapienza et al., 2009; Schipper, 2012; Aycinena et al., 2014; Drichoutis and Nayga, 2015)<sup>3</sup>.

### 3. DATA AND METHODS

We collect data from five experimental projects that were carried out at the Laboratory of Theoretical and Experimental Economics (LaTeX) of the Universidad de Alicante, from 2014 to 2017. The objects of these studies include, among others, risk and social preferences, which will be discussed in section 5 and 6 respectively. All experimental protocols are also endowed with a debriefing questionnaire from which we obtained information on subjects' gender and cognitive ability. **Table 1** lists the projects in our meta-dataset and summarizes their structure<sup>4</sup>.

#### 3.1. Behavioral Evidence

The behavioral content of the five projects is as follows. Social preferences are elicited in projects 3 and 4 (432 subjects) and risk preferences are elicited in projects 1–5 (497 subjects).

##### 3.1.1. Social Preferences

As for social preferences, the elicitation protocol consists in a sequence of 24 distributional decisions, whose basic layout is borrowed from Cabrales et al. (2010). Subjects are matched in pairs and must choose one out of four options, as shown in **Figure 1**. An option corresponds to a pair of monetary prizes, one for each subject within the pair. At the beginning of each round  $t = 1, \dots, 24$ , subjects are informed about the option set  $C_t = \{b^k\}$ ,  $k = 1, \dots, 4$ . Each option  $b^k = (b_1^k, b_2^k)$  assigns a monetary prize,  $b_i^k$ , to player  $i = 1, 2$ , with  $b_1^k \geq b_2^k$  for all  $k$ . In other words, player 1 (player 2) looks at the distributive problem associated with the choice of a specific option  $k$  from the viewpoint of the advantaged (disadvantaged) player, respectively.

Once choices are made, a “Random Dictator” protocol (Harrison and McDaniel, 2008) determines the payoff relevant decision, that is, an i.i.d. draw fixes the identity of the subject whose choice determines the monetary rewards for that pair and round. This design feature is particularly efficient when estimating inequity aversion in that, for roughly half of the observations we can identify separately, within-subject,

individuals' attitudes toward *envy* (i.e., social preferences from a disadvantageous position) and *guilt* (i.e., social preferences from an advantageous position), respectively. After subjects have selected their favorite options, all payoff relevant information is revealed, and round payoffs are distributed.

##### 3.1.2. Risk Preferences

Risk preferences have been elicited with a Multiple Price List (MPL, Holt and Laury, 2002) protocol in all projects, for a total of 497 subjects. In projects 2–5 our MPL protocol consists of a sequence of 21 binary choices. As **Figure 2** shows, “Option A” corresponds to a sure payment whose value increases along the sequence from 0 to 1000 pesetas in steps of 50 while “Option B” is constant along the sequence and corresponds to a 50/50 chance to win 1,000 pesetas. In project 1, instead, the list consists of 16 binary choices: “Option A” is increasing from 0 to 15 euros in steps of 1 while “Option B” is a fixed lottery over three prizes drawn from Hey and Orme (1994). Subjects are asked to elicit their certain equivalent for 50 such lotteries. In both protocols one of the binary choices is selected randomly for payment at the end of the experiment<sup>5</sup>.

#### 3.2. Individual Characteristics

In all studies, we scanned both hands and we measured 2D:4D following the protocol set up by Neyse and Brañas-Garza (2014). By using this procedure, we avoid measurement errors usually associated with self-reported statements (Brañas-Garza and Kovárik, 2013). The 2D:4D measure reported in what follows is a dummy equal to 1 for subjects with a right hand 2D:4D above the gender-specific median value, high 2D:4D hereafter, and equal to 0 otherwise. This choice is based on the non-linear relationship between 2D:4D and behavioral outcomes that is reported in Brañas-Garza et al. (2013) among others. Gender difference in 2D:4D, with men exhibiting a lower 2D:4D as shown in **Figure 3**, have been taken into account by defining our binary measure of high or low 2D:4D by computing median values separately by gender. An additional advantage of using a dummy to discriminate between high and low 2D:4D rather than 2D:4D, that takes values in a very small interval around 1, is that it tends to simplify the interpretation of coefficients of interactions between the high 2D:4D dummy and other covariates in regressions<sup>6</sup>.

The Cognitive Reflection Test (CRT hereafter, Frederick, 2005) was administered in our debriefing questionnaire. It is a simple test of a quantitative nature especially designed to elicit the “predominant cognitive system at work” in respondents' reasoning:

CRT1. A bat and a ball cost 1.10 dollars. The bat costs 1.00 dollars more than the ball. How much does the ball cost? (Correct answer: 5 cents).

<sup>3</sup>In a non-experimental setting Coates et al. (2009) find a negative relationship between 2D:4D, profitability and tenure on the job for a sample of 49 financial traders in the City of London. In a related although different experimental setting that involves strategic interactions among subjects, Pearson and Schipper (2012) find no significant association between 2D:4D, bids in sealed bid first-price auctions and subjects' total payoffs. A positive relationship is also found between 2D:4D, risky choices and criminality using field data, although with a low number of observations in Hanoch et al. (2012).

<sup>4</sup>Approval for the experiment was given by the LaTeX Ethics Committee. Participants gave their consent to participate in social experiments when they signed up in ORSEE (Greiner, 2004), the online recruitment tool used at LaTeX. When, before the experiment started, instructions about its content were read aloud to all participants, they were informed that they could leave the experiment at any stage. Separate approvals were obtained for each of the five experimental studies used in the paper.

<sup>5</sup>The interested reader in the estimation of risk preferences in a setting with several identical rounds, in which subjects may learn over rounds, can refer to Albarran et al. (2017).

<sup>6</sup>In section 5 we discuss the advantages and disadvantages of using the high 2D:4D dummy rather than 2D:4D itself. For the sake of robustness, we also report results of our analysis with 2D:4D in Appendix A (Supplementary Material).

**TABLE 1** | Summary of experimental projects in the meta-dataset.

Project	Reference	N	Topic	Social preferences	Risk preferences	2D:4D
1	Albarran et al., 2017	279	Risk and uncertainty	No	Yes (89)	Yes
2	Cueva et al., 2016	96	Behavioral finance	No	Yes	Yes
3	Ponti et al., 2014	288	Entrepreneurship	Yes	Yes (96)	Yes
4	Ponti et al., 2017	144	Agency	Yes	Yes	Yes
5	Zhukova, 2017	72	Investment	No	Yes	Yes
		879		432	497	879

Período 1 de 1

OPCIÓN	TU PAGO	SU PAGO
OPCIÓN A	70	30
OPCIÓN B	70	32
OPCIÓN C	44	44
OPCIÓN D	40	39

Para hacer tu selección, sólo tienes que señalar la OPCIÓN correspondiente.

**FIGURE 1** | User interface for distributional decisions in projects 3 and 4.

- CRT2. If it takes 5 machines 5 min to make 5 widgets, how long would it take 100 machines to make 100 widgets? (Correct answer: 5 min).
- CRT3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? (Correct answer: 47 days).

The CRT provides not only a measure of cognitive ability, but also of *impulsiveness* and, possibly, other individuals' unobservable characteristics. In this test, the "impulsive" answer (10, 100, and 24, respectively) is shown to be the modal answer (Frederick, 2005). These answers, although incorrect, may have been selected by those subjects who do not think carefully enough. Following Cueva et al. (2016), we partition individuals into three groups. *Impulsive* subjects answer the erroneous intuitive value at least in two questions, *reflective* ones answer correctly at least two questions, and *others* are the residual group.

## 4. RESULTS I: DESCRIPTIVE STATISTICS

In this section we report descriptive statistics of 2D:4D and estimates of its correlation with the CRT score and with CRT categories dummies, our proxies for cognitive ability by way of pairwise correlations.

**Figure 3** reports the distribution of 2D:4D in our meta-dataset for the full sample and separately for subsamples by gender. The distribution tends to be symmetric and the median value is slightly smaller than one for the full sample as well as for subsamples by gender. In addition, **Figure 3** shows that 2D:4D tends to be smaller for males, in line with evidence that 2D:4D is sexually dimorphic in related studies.

**Table 2** shows the correlations between 2D:4D, gender and proxies of cognitive ability. In addition, it report correlations using as a measure of prenatal exposure to testosterone a dummy equal to 1 if 2D:4D is greater than the gender-specific median and, also, a dummy equal to 1 if 2D:4D is either in the top or in the bottom tercile of the 2D:4D distribution by gender. The correlation between 2D:4D and the female dummy is positive and highly significant for both hands. 2D:4D is, instead, negatively and highly significantly correlated with the CRT reflective group dummy for the left hand when using the top-bottom tercile dummy. In addition, **Table 2** shows that correlations between 2D:4D and the frequency of risky choices, our proxy for risk attitudes, are negative and, hence, qualitatively in line with results in related studies. However, estimates are not significant, even when using binary measures of prenatal exposure to testosterone. Since our proxies for social preferences are estimated parameters of Fehr and Schmidt (1999) model, the estimation procedure and their relationship with prenatal exposure to testosterone are reported in section 5<sup>7,8</sup>.

## 5. RESULTS II: SOCIAL PREFERENCES

This section frames Dictators' behavior in projects 3 and 4 within the realm of Fehr and Schmidt (1999), one of the most popular models of social preferences. According to it, the Dictator's utility associated to option  $k$ ,  $u(k)$ , does not only depend on

<sup>7</sup>The interested reader can find additional statistical evidence on the relationship between 2D:4D and personality traits in Alonso et al. (2017), the working paper version of this manuscript.

<sup>8</sup>Out of our 879 subjects CRT reflective, with 2 or more correct answers are 149 (16.7%), CRT impulsive, with at least one incorrect and impulsive answers, are 531 (60.4%) and the residual group contains 199 (22.6%).



Para cada decisión, elige entre la opción A y la opción B. Cuando has terminado, confirma tus decisiones pinchando al botón Aceptar.

Decisión	Opción A	Opción B	Tu decisión
1.	0 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
2.	50 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
3.	100 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
4.	150 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
5.	200 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
6.	250 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
7.	300 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
8.	350 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
9.	400 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
10.	450 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
11.	500 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
12.	550 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
13.	600 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
14.	650 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
15.	700 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
16.	750 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
17.	800 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
18.	850 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
19.	900 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
20.	950 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>
21.	1000 pesetas seguro	1000 pesetas ó 0 pesetas a cara o cruz	A <input type="radio"/> B <input type="radio"/>

FIGURE 2 | User interface for the multiple price list in projects 2–5.

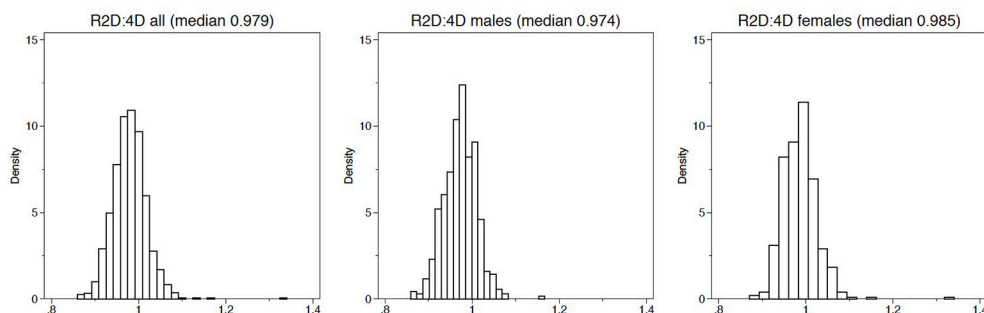


FIGURE 3 | 2D:4D histograms.

the Dictator's own monetary payoff,  $x_D^k$ , but also on that of the Recipient,  $x_R^k$ , as follows:

$$u(k) = x_D^k - \alpha \max[x_R^k - x_D^k, 0] - \beta \max[x_D^k - x_R^k, 0], \quad (1)$$

where the values of  $\alpha$  and  $\beta$  determine the Dictator's *envy* (i.e., aversion to inequality when receiving less than the Recipient) and *guilt* (i.e., aversion to inequality when receiving more than the Recipient), respectively.

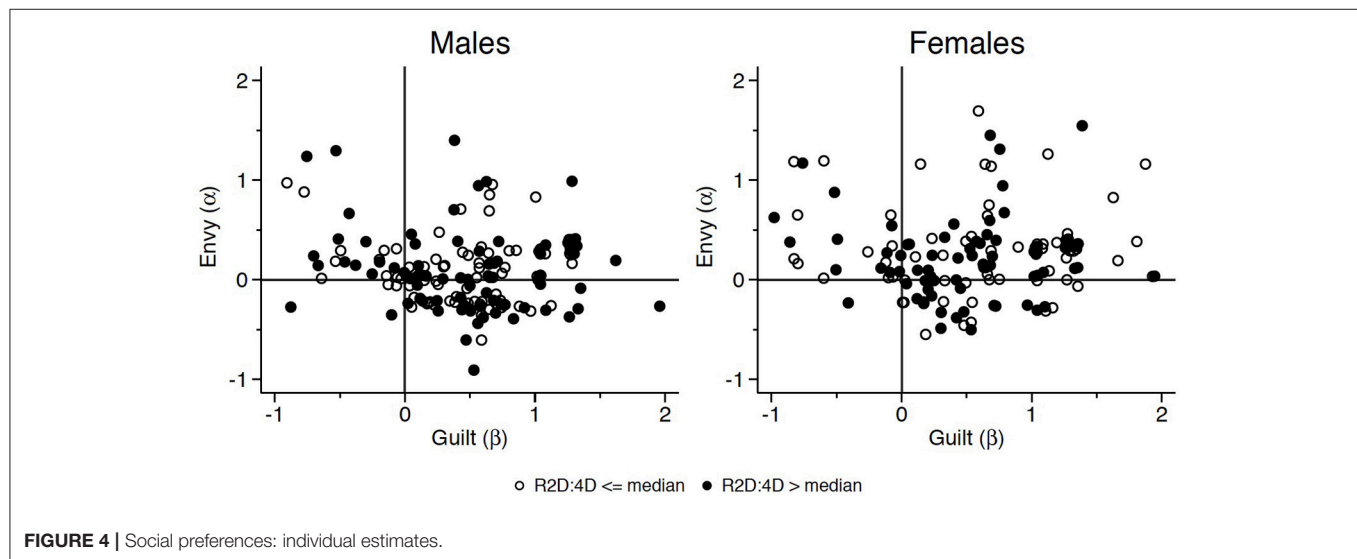
In what follows we shall estimate by maximum likelihood, for each participant, the two coefficients of Equation (1) by way of a standard multinomial logit model.

Figure 4 reports the estimated coefficients of equation (1) for each subject participating in the experiment, disaggregated by gender and by whether the right hand 2D:4D is above the gender-specific median. By conditioning on the gender-specific median, we control for the correlation between gender and 2D:4D that we detected in Table 2. As Figure 4 shows, (i) estimates for males are less dispersed with respect to the origin (corresponding to more "selfish" preferences) and (ii) inequity aversion appears to be

**TABLE 2 |** Correlations.

	2D:4D in level		Above median dummy		Top-bottom tercile dummy	
	L2D:4D	R2D:4D	LH2D:4D	HR2D:4D	TBL2D:4D	TBR2D:4D
L2D:4D	1.000	0.628***	1.000	0.456***	1.000	0.185***
Female	0.177***	0.208***	−0.001	−0.001	−0.026	0.005
CRT	−0.066**	−0.049	−0.015	−0.002	−0.073**	−0.023
CRT Impulsive	0.047	0.037	0.001	0.008	0.052	0.046
CRT Reflective	−0.068**	−0.068**	−0.036	−0.029	−0.092***	−0.026
CRT Other	0.014	0.027	0.039	0.022	0.035	−0.032
Freq. of risky choices	−0.043	−0.034	−0.004	−0.026	−0.011	0.029

\*\* $p < 0.05$ , \*\*\* $p < 0.01$ .



the modal distributional type, with specific reference to females with low 2D:4D. The pooled estimates of  $\alpha$  and  $\beta$  for the full sample (clustered at the subject level) are 0.288 (std. err. 0.001,  $p = 0.000$ ) and 0.684 (std. err. 0.008,  $p = 0.000$ ), respectively<sup>9</sup>.

In order to quantify the relationship between 2D:4D and inequity aversion, we follow a semi-parametric approach. First, for both  $\alpha$  and  $\beta$ , we partition our subject pool into three subsets, depending on whether the corresponding individual-level estimates are significantly smaller than zero (53 and 28 for  $\alpha$  and  $\beta$  respectively), not significantly different (130 and 160), or significantly greater (159 and 154). We then set up an ordered probit regression by which the probability of falling in each category is a function of high 2D:4D dummy, gender and the CRT groups, with the reflective group as omitted category. Our choice of using a dummy equal to 1 if 2D:4D is above the gender-specific median, rather than 2D:4D itself, may be subject to problems, such as a lower statistical power and a higher probability of type I or II errors (Irwin and McClelland, 2003; McClelland et al., 2015). However, by using non-linear models to estimate the relationship between 2D:4D and social preferences

in this section, our estimates are unlikely to suffer from such problems<sup>10</sup>.

**Table 3** reports the estimated coefficients, with alternative sets of covariates being used. We start estimating the relationship between social preferences and the high 2D:4D dummy (HR2D:4D) in model (1) without adding any additional control and then, in model (2) and (3) we add female and CRT categories dummies to assess if they play a mediating role. In model (4) we use an interaction term between HR2D:4D and the female dummy to account for the positive correlation between gender and 2D:4D we observed in **Table 2**. Finally, in model (5) we use an interaction term between the CRT categories dummies and HR2D:4D. In addition, we report in **Table 3** marginal effects (MFX) of HR2D:4D, evaluated at the sample mean, while MFX with respect to gender and CRT are shown in Appendix A (Supplementary Material)<sup>11</sup>.

<sup>10</sup>We also set up a bivariate ordered probit estimation in which we allow error terms in the equations of  $\alpha$  and  $\beta$  to be jointly distributed. We find that the covariance parameter is not significant.

<sup>11</sup>The number of observations shown at the bottom of **Table 3** is lower than the total number of subjects in projects 3 and 4 since we dropped those subjects for whom maximum likelihood estimation of  $\alpha$  and  $\beta$  did not converge.

**TABLE 3** | Ordered probit regressions of social preferences individual estimates.

	(1)		(2)		(3)		(4)		(5)	
	$\alpha$	$\beta$	$\alpha$	$\beta$	$\alpha$	$\beta$	$\alpha$	$\beta$	$\alpha$	$\beta$
HR2D:4D (HR)	−0.064 (0.123)	−0.235* (0.125)	−0.066 (0.124)	−0.236* (0.125)	−0.068 (0.124)	−0.213* (0.126)	0.018 (0.168)	−0.185 (0.172)	0.395 (0.249)	−0.562** (0.254)
Female (F)			0.376*** (0.124)	0.097 (0.125)	0.326*** (0.126)	0.062 (0.127)	0.423** (0.180)	0.093 (0.181)	0.326** (0.127)	0.069 (0.127)
CRT Imp (CRTI)					0.359** (0.149)	0.333** (0.151)	0.355** (0.149)	0.332** (0.151)	0.604*** (0.208)	0.111 (0.214)
CRT Others (CROTO)					0.269 (0.215)	−0.120 (0.214)	0.269 (0.215)	−0.121 (0.214)	1.034*** (0.351)	−0.441 (0.330)
HR × F							−0.189 (0.249)	−0.060 (0.251)		
HR × CRTI									−0.494* (0.294)	0.440 (0.300)
HR × CROTO									−1.298*** (0.449)	0.579 (0.433)
MFX $P(\alpha > 0)$ of HR	−0.025		−0.026		−0.027		−0.029		−0.034	
S.e.	0.049		0.049		0.049		0.049		0.050	
MFX $P(\beta > 0)$ of HR		−0.093*		−0.093*		−0.084*		−0.084*		−0.082*
S.e.		0.049		0.049		0.050		0.050		0.050
N	342		342		342		342		342	

Standard errors in parentheses. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

Table 3 shows that the relationship between HR2D:4D and negative inequity aversion, i.e., envy, is negative and the same holds for the relationship with positive inequity aversion, i.e., guilt. MFX, which are reported at the bottom of the table, show that the relationship with envy or with guilt is not significant. The table also shows that envy is higher for females while the impulsive group (CRTI) is characterized by higher envy and higher guilt than the reflective group, which is the excluded CRT category. These estimates are significant as shown by MFX in Appendix A (Supplementary Material). These results hold for the five econometric specifications reported in Table 3, as shown by MFX in Appendix A (Supplementary Material). Finally, when we interact the HR2D:4D dummy with CRT categories to assess if the influence of 2D:4D differs by subjects' cognition, we find that subjects with high 2D:4D and low cognitive ability, proxied by the CRT impulsive dummy, do not exhibit significantly lower envy than subjects with high 2D:4D in the CRT reflective group, while the relationship is significant when considering the CRT residual group dummy<sup>12,13</sup>.

<sup>12</sup>Marginal effects are the same when we estimate them using, as an alternative measure, 2D:4D in levels, except the estimated relationship with guilt.

<sup>13</sup>When we replicated our main experimental results by using a dummy equal to 1 if 2D:4D is either in the bottom tercile of the distribution or in the top one, as a sensitivity analysis, we obtained similar results, except a positive and significant relationship between envy and the top-bottom tercile dummy, as shown in Appendix A (Supplementary Material). Most of the results shown in this section on the relationship between 2D:4D and social preferences tend to lose significance

## 6. RESULTS III: RISK ATTITUDES

In this section we study the relationship between 2D:4D and proxies for risk preferences by using data on 497 subjects from all projects. Risk preferences are elicited by way of a Multiple Price List (MPL, Holt and Laury, 2002), in which individuals have to choose between two alternatives: a list of increasing sure payments and a lottery. Since the same protocol has been used in projects 2 to 5 while the number of decisions, lottery prizes, the experimental currency and their probability distribution differ in project 1, we choose two proxies for risk preferences that we believe are not affected by these differences.

Following Cueva et al. (2016), we define *consistent* those individuals whose decisions satisfy two conditions: (i) start by choosing the lottery option, as it stochastically dominates the sure payment of 0, and (ii) switch only once at some point along the price list to the sure payment and stick to it up to the end. We can use data from all projects in our empirical analysis as none of the differences between our MPL protocols has an impact on the consistency definition. We also define a dummy equal to 1 if the proportion of risky choices made by a subject, i.e., the ratio between the number of lotteries chosen in the list and the total number of decisions, is greater than the median value. By using the proportion rather than the number of risky choices, we control for the difference in the design of the MPL in project 1.

when they are obtained with the high 2D:4D dummy defined using left hand 2D:4D, as shown in Appendix A (Supplementary Material).

**TABLE 4 |** Subjects' consistency in risky choices.

	(1)	(2)	(3)	(4)	(5)
HR2D:4D	0.071*	0.072*	0.069*	0.047	−0.039
	(0.037)	(0.037)	(0.036)	(0.049)	(0.055)
Female (F)		−0.057	−0.025	−0.048	−0.023
		(0.037)	(0.038)	(0.056)	(0.038)
CRT Imp. (CRTI)			−0.164***	−0.163***	−0.240***
			(0.039)	(0.039)	(0.053)
CRT Other. (CRTO)			−0.152***	−0.152***	−0.202***
			(0.052)	(0.052)	(0.073)
HR2D:4D × F				0.047	
				(0.073)	
HR2D:4D × CRTI					0.149**
					(0.074)
HR2D:4D × CRTO					0.095
					(0.104)
Project 1	0.066	0.069	0.066	0.066	0.067
	(0.044)	(0.044)	(0.044)	(0.044)	(0.045)
Constant	0.737***	0.764***	0.879***	0.889***	0.934***
	(0.029)	(0.032)	(0.034)	(0.038)	(0.036)
<hr/>					
MFX of F				−0.025	
S.e.				0.039	
MFX of CRTI					−0.166***
S.e.					0.039
MFX of CRTO					−0.155***
S.e.					0.052
MFX of HR				0.069*	0.069*
S.e.				0.037	0.036
<hr/>					
N	497	497	497	497	497

Robust standard errors in parentheses. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

**Table 4** shows linear probability estimates of subjects' consistency dummy. In addition to the high 2D:4D dummy, our covariates include dummies for females and for the CRT groups, as well as for the interaction between the high 2D:4D dummy, female and CRT groups dummies. The top panel of the table shows regression estimates while the bottom one marginal effects (MFX) for those specifications in which we used interaction terms, evaluated at the sample mean. Because of the differences in the experimental protocol of project 1 with respect to the others, we also include a dummy equal to 1 for subjects in project 1 in order to absorb project-specific effects.

When we look at estimates in **Table 4**, we find that the probability of being consistent in their decisions is higher for subjects with a high 2D:4D but the difference is not significant, that there is no significant gender difference and that it is significantly lower for subjects in the impulsive (CRTI) or in the residual (CRTO) group than for the reflective group. We see no changes when we include the interaction between female and the high 2D:4D variable, suggesting that they do not play any mediating role. When we add interaction terms

**TABLE 5 |** Consistent subjects' relative frequency of risky choices above median.

	(1)	(2)	(3)	(4)	(5)
HR2D:4D (HR)	0.005	0.006	0.006	−0.009	−0.041
	(0.017)	(0.017)	(0.017)	(0.020)	(0.031)
Female (F)		−0.058***	−0.056***	−0.073***	−0.056***
		(0.017)	(0.018)	(0.025)	(0.018)
CRT Imp. (CRTI)			−0.007	−0.006	−0.036
			(0.020)	(0.020)	(0.028)
CRT Other. (CRTO)			0.007	0.007	−0.030
			(0.025)	(0.025)	(0.037)
HR2D:4D × F				0.033	
				(0.034)	
HR2D:4D × CRTI					0.058
					(0.038)
HR2D:4D × CRTO					0.072
					(0.050)
Project 1	−0.064***	−0.058***	−0.058***	−0.057***	−0.056***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Constant	0.453***	0.478***	0.480***	0.487***	0.503***
	(0.012)	(0.013)	(0.018)	(0.019)	(0.023)
<hr/>					
MFX of F				−0.057***	
S.e.				0.018	
MFX of CRTI					−0.007
S.e.					0.020
MFX of CRTO					0.006
S.e.					0.025
MFX of HR				0.007	0.008
S.e.				0.017	0.017
<hr/>					
N	390	390	390	390	390

Robust standard errors in parentheses. \*\*\* $p < 0.01$ .

between the high 2D:4D dummy and the female dummy, we find no significant gender differences in the relationship between 2D:4D and consistency. When we add interactions between high 2D:4D and cognitive ability dummies, the high 2D:4D dummy coefficient is no longer significant while the coefficient of the interaction with the CRTI dummy is positive and significant, suggesting that subjects in the CRT impulsive group and with high 2D:4D are more consistent. When looking at MFX, we find that consistency is significantly lower for subjects with low cognitive ability, it is higher for subjects with a high 2D:4D although the difference is not significant<sup>14</sup>.

**Table 5** shows linear probability estimates for consistent subjects of a dummy equal to 1 if the proportion of risky choices is greater than the median. We find no

<sup>14</sup>Estimates of the same regression except for using, rather than the high 2D:4D dummy, 2D:4D itself or the top-bottom tercile dummy are reported in Appendix A (Supplementary Material). We can see some differences depending on the measure used: the probability of consistency is lower for females when we use 2D:4D and also when we use the top-bottom tercile dummy, although the estimates are not significant.



significant relationship with the high 2D:4D dummy while the probability is significantly lower for females. Results are unchanged when using 2D:4D or the top-bottom tercile dummy, as shown in Appendix A (Supplementary Material)<sup>15,16,17</sup>.

## 7. DISCUSSION

When we look at social preferences, we contribute to the literature that has almost entirely focused on giving as a proxy for social preferences in a variety of experimental settings (e.g., Buser, 2012; Brañas-Garza et al., 2013) by isolating two aspects underlying the incentives to give, that is, envy and guilt. Finding a negative and significant relationship between 2D:4D and envy, i.e., less generous behavior by subjects when they play in the disadvantaged role, only for subjects with low cognitive ability and non-significant results for guilt suggests that individual heterogeneity may play a role in reconciling the mixed evidence on the relationship between 2D:4D and giving in the literature. However, giving and inequity aversion are not fully comparable proxies for social preferences as they are used in different experimental settings.

Although evidence of heterogeneity by ability in the relationship between 2D:4D and subjects' decision-making has been documented in risky choices (Brañas-Garza and Rustichini, 2011), we are the first to do so in the realm of social preferences, to the best of our knowledge. Finding that subjects with high 2D:4D and low cognitive ability exhibit significantly lower envy than subjects with low 2D:4D and high cognitive ability shows evidence of heterogeneity by ability in the relationship between social preferences and 2D:4D. This result, by suggesting an attenuating role of low cognitive ability and high 2D:4D on inequity aversion contributes to related studies, for example Cueva et al. (2016) and Ponti and Rodríguez-Lara (2015), who find that the CRT impulsive category exhibits higher inequity aversion.

When we look at risk attitudes, we find that the relationship between 2D:4D and the probability that the number of risky decisions is above the median, shows a mixed sign, it is quantitatively small and never significant. These results contribute to the related literature as the sign and significance

of the relationship is not conclusive. Overall, this may be due to the fact that there is genuinely no relationship between 2D:4D and risky decisions or, alternatively, to differences across studies. The composition of the subject pool may play a role if the willingness to participate in an experiment correlates with subjects' socio-economic background and risk aversion. In addition, the type of risk preferences elicitation task may also matter. For example, studies that, including ours, use a task in which subjects can choose a risk-free option tend to find a non-significant association while studies in which subjects choose between two lotteries tend to find a negative and significant association.

After discussing our results relative to those in related studies, we now critically assess them in the light of potential methodological issues, that we believe all researchers wanting to contribute to this interdisciplinary literature should bear in mind. Studies in hard sciences of the relationship between direct measures of prenatal exposure to testosterone and 2D:4D find mixed results, whose sign and significance seem to depend critically on whether direct measures are obtained in an early stage *in utero* or, instead, close to the birth. Studies in social sciences on the relationship between 2D:4D and decision-making find mixed results that may depend on the accuracy of 2D:4D measurement and, in addition, to the experimental tasks used to elicit subjects' preferences. Overall, this suggests both that additional research is awaited to reconcile existing differences across studies in the literature and that caution is used in the interpretation of results before these differences are better understood.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

Instituto Valenciano de Investigaciones Económicas (IVIE).

## ACKNOWLEDGMENTS

Financial support from the Spanish Ministerio de Economía y Competitividad (ECO2013-43119, ECO2015-65820-P and ECO2016-77200-P), Universidad de Alicante (GRE 13-04), Generalitat Valenciana (Research Projects Grupos 3/086) and Instituto Valenciano de Investigaciones Económicas (IVIE) is gratefully acknowledged.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnbeh.2018.00022/full#supplementary-material>

<sup>15</sup>Estimates of Table 5 obtained using the full sample are not reported as they are in line with those obtained using only observations of consistent subjects.

<sup>16</sup>Results are qualitatively unchanged when using a logit model or when the dummy equal to 1 if the frequency of risky choices is above the median, one of the dependent variables, is defined using median values separately for projects 1 since the certain equivalent is different from projects 2 to 5. They are not reported although they are available upon request. As a sensitivity analysis, we replicated our main experimental results by using 2D:4D and a dummy equal to 1 if 2D:4D is either in the bottom tercile of the distribution or in the top one and obtained similar results and obtain similar results. This seems to suggest that, at least in our case, estimates of regressions using the high 2D:4D dummy are not severely biased, as suggested by Irwin and McClelland (2003); McClelland et al. (2015).

<sup>17</sup>Most of the results shown in this section on the relationship between 2D:4D and risk attitudes do not hold when they are obtained with the high 2D:4D dummy defined using left hand 2D:4D, as shown in Appendix A (Supplementary Material).

## REFERENCES

- Albarran, P., Alonso, J., Herrero, C., Ponti, G., and Sartarelli, M. (2017). *An Experiment on Outcome Uncertainty*. Mimeo: Universidad de Alicante.
- Alonso, J., Di Paolo, R., Ponti, G., and Sartarelli, M. (2017). *Some (Mis)Facts about 2D:4D, Preferences and Personality*. Working Papers. Serie AD 2017-08, Instituto Valenciano de Investigaciones Económicas, S.A. (Ivie).
- Apicella, C. L., Carré, J. M., and Dreber, A. (2015). Testosterone and economic risk taking: a review. *Adapt. Hum. Behav. Physiol.* 1, 358–385. doi: 10.1007/s40750-014-0020-2
- 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
- Aycinena, D., Baltaduonis, R., and Rentschler, L. (2014). Risk preferences and prenatal exposure to sex hormones for ladinos. *PLoS ONE* 9:e103332. doi: 10.1371/journal.pone.0103332
- Brañas-Garza, P., Galizzi, M. M., and Nieboer, J. (2018). Digit ratio and risk taking: evidence from a large, multi-ethnic sample. *Int. Econ. Rev.* Available online at: <https://economics.sas.upenn.edu/ier/issues/accepted-papers>
- Brañas-Garza, P., and Kovárik, J. (2013). *Digit Ratios and Social Preferences: A Comment on Buser (2012)*. Working Papers 13-31, Chapman University, Economic Science Institute.
- Brañas-Garza, P., Kovárik, 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., Kujal, P., and Lenkei, B. (2015). *Cognitive Reflection Test: Whom, How, When*. Working Papers 15-25, Chapman University, Economic Science Institute.
- 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
- Buser, T. (2012). Digit ratios, the menstrual cycle and social preferences. *Games Econ. Behav.* 76, 457–470. doi: 10.1016/j.geb.2012.07.006
- Cabral, A., Miniaci, R., Piovesan, M., and Ponti, G. (2010). Social preferences and strategic uncertainty: an experiment on markets and contracts. *Am. Econ. Rev.* 100, 2261–2278. doi: 10.1257/aer.100.5.2261
- Coates, J. M., Gurnell, M., and Rustichini, A. (2009). Second-to-fourth digit ratio predicts success among high-frequency financial traders. *Proc. Natl. Acad. Sci. U.S.A.* 106, 623–628. doi: 10.1073/pnas.0810907106
- Cueva, C., Iturbe-Ormaetxe, I., Mata-Pérez, E., Ponti, G., Sartarelli, M., Yu, H., and Zhukova, V. (2016). Cognitive (ir)reflection: new experimental evidence. *J. Behav. Exp. Econ.* 64, 81–93. doi: 10.1016/j.socec.2015.09.002
- Dreber, A., and Hoffman, M. (2007). *Portfolio Selection in Utero*. Mimeo: Stockholm School of Economics.
- Drichoutis, A. C., and Nayga, R. M. (2015). Do risk and time preferences have biological roots? *South. Econ. J.* 82, 235–256. doi: 10.4284/0038-4038-2013.246
- Fehr, E., and Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn. Sci.* 11, 419–427. doi: 10.1016/j.tics.2007.09.002
- Fehr, E., and Krajbich, I. (2009). “Social preferences and the brain,” in *Neuroeconomics: Decision Making and the Brain*, Vol. 11, 1st Edn., eds P. W. Glimcher, E. Fehr, C. Camerer, and R. A. Poldrack (Amsterdam: Academic Press), 215–232.
- Fehr, E., and Schmidt, K. M. (1999). A theory of fairness, competition and cooperation. *Q. J. Econ.* 114, 817–868. doi: 10.1162/003353599556151
- Frederick, S. (2005). Cognitive reflection and decision making. *J. Econ. Perspect.* 19, 25–42. doi: 10.1257/089533005775196732
- Garbarino, E., Slonim, R., and Sydnor, J. (2011). Digit ratios (2D:4D) as predictors of risky decision making for both sexes. *J. Risk Uncert.* 42, 1–26. doi: 10.1007/s11166-010-9109-6
- Greiner, B. (2004). *An Online Recruitment System for Economic Experiments*. MPRA Paper 13513, University Library of Munich, Germany.
- Hanoch, Y., Gummerum, M., and Rolison, J. (2012). Second-to-fourth digit ratio and impulsivity: a comparison between offenders and nonoffenders. *PLoS ONE* 7:e47140. doi: 10.1371/journal.pone.0047140
- Harrison, G. W., and McDaniel, T. (2008). Voting games and computational complexity. *Oxford Econ. Pap.* 60, 546–565. doi: 10.1093/oepp/gpm045
- Hey, J. D., and Orme, C. (1994). Investigating Generalizations of expected utility theory using experimental data. *Econometrica* 62, 1291–1326. doi: 10.2307/2951750
- Hollier, L., Keelan, J., Jamnadas, E., Maybery, M., Hickey, M., and Whitehouse, A. (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
- Holt, C. A., and Laury, S. K. (2002). Risk aversion and incentive effects. *Am. Econ. Rev.* 92, 1644–1655. doi: 10.1257/000282802762024700
- 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.psychoneu.2007.01.007
- Irwin, J. R., and McClelland, G. H. (2003). Negative Consequences of dichotomizing continuous predictor variables. *J. Market. Res.* 40, 366–371. doi: 10.1509/jmkr.40.3.366.19237
- Kaltwasser, L., Mikac, U., Busko, 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
- Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., and Fehr, E. (2005). Oxytocin increases trust in humans. *Nature* 435:673. doi: 10.1038/nature03701
- Lee, D. (2008). Game theory and neural basis of social decision making. *Nat. Neurosci.* 11, 404–409. doi: 10.1038/nn2065
- Lutchmaya, 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
- Manning, J. T. (2002). *Digit Ratio: A Pointer to Fertility, Behavior, and Health*. New Brunswick, NJ: Rutgers University Press.
- 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/hmrp/13.11.3000
- McClelland, G. H., Lynch, J. G., Irwin, J. R., Spiller, S. A., and Fitzsimons, G. J. (2015). Median splits, type II errors, and false-positive consumer psychology: don't fight the power. *J. Consum. Psychol.* 25, 679–689. doi: 10.1016/j.jcps.2015.05.006
- Millet, K., and Dewitte, S. (2006). Second to fourth digit ratio and cooperative behavior. *Biol. Psychol.* 71, 111–115. doi: 10.1016/j.biopsycho.2005.06.001
- Neyse, L., and Brañas-Garza, P. (2014). *Digit Ratio Measurement Guide*. MPRA Paper 54134, University Library of Munich, Germany.
- Pearson, M., and Schipper, B. C. (2012). The visible hand: finger ratio (2D:4D) and competitive bidding. *Exp. Econ.* 15, 510–529. doi: 10.1007/s10683-011-9311-7
- Ponti, G., and Rodriguez-Lara, I. (2015). Social preferences and cognitive reflection: evidence from a dictator game experiment. *Front. Behav. Neurosci.* 9:146. doi: 10.3389/fnbeh.2015.00146
- Ponti, G., Sartarelli, M., Sikora, I., and Zhukova, V. (2014). *The Price of Entrepreneurship. Evidence from the Lab*. Mimeo: Universidad de Alicante.
- Ponti, G., Sartarelli, M., Sikora, I., and Zhukova, V. (2017). *Do Profit Opportunities and Experience Influence Principals' Efficiency? Evidence from the Lab*. Mimeo: Universidad de Alicante.
- Putz, D. A., Gaulin, S. J., Sporter, R. J., and McBurney, D. H. (2004). Sex hormones and finger length. *Evol. Hum. Behav.* 25, 182–199. doi: 10.1016/j.evolhumbehav.2004.03.005
- Ronay, R., and von Hippel, W. (2010). Power, testosterone, and risk-taking. *J. Behav. Dec. Making* 23, 473–482. doi: 10.1002/bdm.671
- 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
- Sapienza, P., Zingales, L., and Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15268–15273. doi: 10.1073/pnas.0907352106
- Schipper, B. (2012). *Sex Hormones and Choice under Risk*. Working Paper Series 12-7, University of California, Davis.

- Stenstrom, E., Saad, G., Nepomuceno, M. V., and Mendenhall, Z. (2011). Testosterone and domain-specific risk: digit ratios (2D: 4D and Rel2) as predictors of recreational, financial, and social risk-taking behaviors. *Pers. Indiv. Diff.* 51, 412–416. doi: 10.1016/j.paid.2010.07.003
- Trivers, R., Manning, J., and Jacobson, A. (2006). A longitudinal study of digit ratio (2D:4D) and other finger ratios in jamaican children. *Horm. Behav.* 49, 150–156. doi: 10.1016/j.yhbeh.2005.05.023
- Van den Bergh, B., and Dewitte, S. (2006). Digit ratio (2D: 4D) moderates the impact of sexual cues on men's decisions in ultimatum games. *Proc. R. Soc. Lond. B. Biol. Sci.* 273, 2091–2095. doi: 10.1098/rspb.2006.3550
- Ventura, T., Gomes, M., Pita, A., Neto, M., and Taylor, A. (2013). Digit ratio (2D:4D) in newborns: influences of prenatal testosterone and maternal environment. *Early Hum. Dev.* 89, 107–112. doi: 10.1016/j.earlhumdev.2012.08.009
- 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
- Xie, Z., Page, L., and Hardy, B. (2017). Investigating gender differences under time pressure in financial risk taking. *Front. Behav. Neurosci.* 11:246. doi: 10.3389/fnbeh.2017.00246
- Zak, P. J., Kurzban, R., Ahmadi, S., Swerdloff, R. S., Park, J., Efremidze, L., et al. (2009). Testosterone administration decreases generosity in the ultimatum game. *PLoS ONE* 4:e8330. doi: 10.1371/journal.pone.0008330
- Zak, P. J., Stanton, A. A., and Ahmadi, S. (2007). Oxytocin increases generosity in humans. *PLoS ONE* 2:e1128. doi: 10.1371/journal.pone.0001128
- Zhukova, V. (2017). *The Role of Firm's Trustworthiness in Business Angels' Investment. Experimental Evidence from 3-Player Trust Game.* Mimeo: Universidad de Alicante.

**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 © 2018 Alonso, Di Paolo, Ponti and Sartarelli. 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) and the copyright owner 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.



# Risk Preferences and Predictions about Others: No Association with 2D:4D Ratio

Katharina Lima de Miranda<sup>1</sup>, Levent Neyse<sup>1,2\*</sup> and Ulrich Schmidt<sup>1,3,4</sup>

<sup>1</sup> Kiel Institute for the World Economy, Kiel, Germany, <sup>2</sup> SOEP at German Institute for Economic Research (DIW), Berlin, Germany, <sup>3</sup> Department of Economics and Econometrics, University of Johannesburg, Johannesburg, South Africa,

<sup>4</sup> Department of Economics, University of Kiel, Kiel, Germany

Prenatal androgen exposure affects the brain development of the fetus which may facilitate certain behaviors and decision patterns in the later life. The ratio between the lengths of second and the fourth fingers (2D:4D) is a negative biomarker of the ratio between prenatal androgen and estrogen exposure and men typically have lower ratios than women. In line with the typical findings suggesting that women are more risk averse than men, several studies have also shown negative relationships between 2D:4D and risk taking although the evidence is not conclusive. Previous studies have also reported that both men and women believe women are more risk averse than men. In the current study, we re-test the relationship between 2D:4D and risk preferences in a German student sample and also investigate whether the 2D:4D ratio is associated with people's perceptions about others' risk preferences. Following an incentivized risk elicitation task, we asked all participants their predictions about (i) others' responses (without sex specification), (ii) men's responses, and (iii) women's responses; then measured their 2D:4D ratios. In line with the previous findings, female participants in our sample were more risk averse. While both men and women underestimated other participants' (non sex-specific) and women's risky decisions on average, their predictions about men were accurate. We also found evidence for the false consensus effect, as risky choices are positively correlated with predictions about other participants' risky choices. The 2D:4D ratio was not directly associated either with risk preferences or the predictions of other participants' choices. An unexpected finding was that women with mid-range levels of 2D:4D estimated significantly larger sex differences in participants' decisions. This finding needs further testing in future studies.

## OPEN ACCESS

### Edited by:

Oliver T. Wolf,  
Ruhr University Bochum, Germany

### Reviewed by:

Martin G. Köllner,  
Friedrich-Alexander-Universität  
Erlangen-Nürnberg, Germany  
Katrin Starcke,  
University of Duisburg-Essen,  
Germany

### \*Correspondence:

Levent Neyse  
levent.neyse@ifw-kiel.de

**Received:** 22 September 2017

**Accepted:** 15 January 2018

**Published:** 01 February 2018

### Citation:

Lima de Miranda K, Neyse L and  
Schmidt U (2018) Risk Preferences  
and Predictions about Others: No  
Association with 2D:4D Ratio.  
*Front. Behav. Neurosci.* 12:9.  
doi: 10.3389/fnbeh.2018.00009

**Keywords:** risk, decision making, prenatal testosterone, 2D:4D, stereotypes, gender

## INTRODUCTION

Human behavior and decision making are closely connected to individuals' social environment as well as their beliefs about other people's behaviors, actions, preferences, and characteristics. According to Social Comparison Theory, humans tend to continuously compare themselves with others (Festinger, 1954) and their social identity is connected to these comparisons (see Hogg, 2000). As these comparisons are often made under the influence of erroneous reference points and



social stereotypes (Katz and Braly, 1965), inaccurate stereotyping is an inevitable consequence.<sup>1</sup> Although stereotypes typically affect certain social groups externally, individuals may also influence their own self-concept through self-stereotyping (Latrofa et al., 2010) or stereotype threat (Steele and Aronson, 1995). This means that stereotypes may shape human behavior through diverse social and psychological channels. Alongside numerous types of stereotypes such as ethnic, political or religious, gender has been a significant research topic in various fields of social science, such as psychology, sociology, and economics. Examples include gender stereotypes in management (Powell et al., 2002), social inferences (Berndt and Heller, 1986), negotiation performance (Kray and Thompson, 2004) and risk preference predictions (Siegrist et al., 2002). In the field of economics in particular, gender stereotypes have been the focus of attention as numerous gender gaps are observed in both macroeconomic and microeconomic indices. Typical examples show that the balance is tipped in the favor of men; in income, education, health, political and labor force participation as well as occupied managerial positions as documented in the Global Gender Gap Report 2016 (Leopold et al., 2016).

While gender discrimination plays a major role in gender gaps in economics, there also exists a vast literature pointing out various gender differences in economic behavior. These differences might also have an impact on gender gaps or they may correlate with gender stereotypes, although the extent of causality is vague. One common finding in this regard is the higher risk aversion of women (Byrnes et al., 1999; Croson and Gneezy, 2009; Charness and Gneezy, 2012). According to existing literature, gender stereotypes are attached to gender effects in risk preferences. In Siegrist et al. (2002) for example, participants were asked to estimate other people's answers in a questionnaire on risk attitudes. Their results show that both men and women overestimated men's risk preferences; which was a clear sign of being biased by common stereotypes. Ball et al. (2010) also confirmed that the perception of others' risk attitudes reflected common stereotypes.

That women are found to be more risk averse than men on average has, in recent years, led to curiosity about the biological roots of gender differences in risk attitudes. The role of the steroid hormone testosterone (hereafter T) has been one of the most widely investigated biological foundations. As higher T is associated with more masculine behavior and personality characteristics, the association between T and risk taking has been a common inquiry. Yet, the results are not entirely conclusive due to the complexity of both human endocrinology and decision making processes. The methods used to investigate the relationship between T and financial risk taking are clustered in three categories. First method is to study *circulating T* which has a systematic impact on decision making. However, as it is a continuously fluctuating hormone, the studies focusing on circulating T are mostly limited to correlational findings. *Manipulating the circulating T* is a method of identifying

causality. The third method is to study the *organizational role of T* through indirect measurements, such as the 2D:4D ratio of hands. We investigate the association between 2D:4D and risk preferences and also the relationship between 2D:4D and one's perceptions about other people's risk preferences. Apicella et al. (2015) reviews the financial risk taking and T literature, while Nadler and Zak (2016) review the role of T in economic behavior in depth.

## Background Literature

### Stereotyping and Estimating Risk Preferences

While Social Role Theory suggests that the gender differences in behavior and gender stereotypes originate from separate social roles of men and women in society (Eagly and Steffen, 1984; Eagly et al., 2000), a stereotype itself may also drive the target group to confirm that stereotype, even if it is an inaccurate one. This phenomenon, called the *stereotype threat*, may consequently contribute to the persistence of a gender role in society. A common example is mathematical ability. Primed by the gender stereotype suggesting the higher numerical ability of men, female participants perform worse in math tests than their actual potential (Brown and Josephs, 1999; Shih et al., 1999; Spencer et al., 1999). In line with stereotype threat examples in performance, the stereotype suggesting that *men are risk-takers* was also confirmed by women in previous studies.

Siegrist et al. (2002) asked their participants to make sex specific predictions about risk preferences with hypothetical questions. While both men and women made accurate predictions about women's risk preferences, both overestimated the number of risky choices by men. Interestingly, women's predictions about the number of risky choices men would make were higher than men's predictions about their own sex. The seminal study of Eckel and Grossman (2008) experimentally confirmed that both sexes predict male peers would take higher risks than female peers. Although this prediction was accurate, it is an evidence of stereotyping in both sexes. Roszkowski and Grable (2005), Daruvala (2007), and Grossman (2013) support the existence of gender stereotyping in risk attitude predictions in the same direction.

Although the predictions were not sex-specific, the preceding studies investigated predictions about others' risk preferences. For example, Hsee and Weber (1997) argued that people's risk preferences are affected by their emotional reactions to risk and that their predictions about others are related to common (cultural) stereotypes. Wallach and Wing (1968) and Levinger and Schneider (1969) showed that people typically believe they are themselves more risk taking than others. This finding was replicated in numerous studies (Clark et al., 1971; Lamm et al., 1972) with the exception of Hsee and Weber (1997) where participants estimated higher risk taking for others than themselves. One explanation for this common finding is the *risk-as-value hypothesis* (Brown, 1965), according to which individuals perceive risk seeking as a culturally more admirable value and therefore their beliefs about themselves and others are biased accordingly. Beliefs about others' risk preferences also reflect one's own risk preferences. This effect was termed the *false*

<sup>1</sup> Even though they are often inaccurate, stereotypes may serve as facilitators in social cognition similar to heuristics and biases in decision making (Tversky and Kahneman, 1975). Judd and Park (1993) provide a thorough discussion of the definition and accuracy of stereotypes.

*consensus effect* and is also a commonly observed prediction bias (Ross et al., 1977).

## 2D:4D Ratio

The fetus' brain development and endocrine system are influenced by prenatal T exposure and the decision making patterns and personality traits of humans are also partially effected by it (Manning, 2002). Digit ratio (2D:4D) is the ratio between the index and ring fingers and it is employed as an indirect bio-marker of prenatal androgen exposure. A lower 2D:4D ratio indicates a higher level of prenatal T to estradiol ratio (Lutchmaya et al., 2004) and men typically have lower 2D:4D ratios (Hönekopp and Watson, 2010). The negative relationship between prenatal androgen exposure and 2D:4D was confirmed via various methods. For example, Lutchmaya et al. (2004) and Ventura et al. (2013) studied the relationship by taking direct evidence from amniotic fluid samples during pregnancy and linking the endogenous T and estradiol ratio data to the finger ratios of newborns and infants. Along with previous correlational approaches, the experimental study of Zheng and Cohn (2011) also observed lower 2D:4D ratios in rodents administrated androgen *in utero*. They conclude that sexually dimorphic 2D:4D is caused by androgen and estrogen signaling. In a twin study van Anders et al. (2006) showed that women with male twins have lower 2D:4D than those with female twins. Typically, 2D:4D shows greater sex differences in the right hand (Hönekopp and Watson, 2010). This is why a large majority of the 2D:4D literature is based on samples gathered from right hands. It should also be noted that circulating T and prenatal T do not necessarily correlate. No significant relationship between 2D:4D and adult sex hormones has been observed in the meta-analytical study of Hönekopp et al. (2007).

A number of studies have shown that several typical gender effects in economics were also observed between low and high 2D:4D individuals. Examples include negative relationship between 2D:4D and overconfidence (Dalton and Ghosal, 2014; Neyse et al., 2016), higher success among high-frequency traders (Coates et al., 2009), earnings in economic games (Buser, 2012) and lower degrees of loss aversion (Hermann, 2017). Note that the last two studies, Buser (2012) and Hermann (2017), use self-reported 2D:4D as a measurement method which was criticized in Brañas-Garza and Kovářík (2013).

In the domain of risk preferences, numerous studies also point out negative relationships. Dreber and Hoffman (2007) and Garbarino et al. (2011) show negative associations in both sexes, while Ronay and von Hippel (2010) only for men with incentivized tasks. Brañas-Garza and Rustichini (2011) and Stenstrom et al. (2011) also showed negative relationship for men without incentivized risk elicitation tasks. These results have been confirmed in a recent study with a large sample size and with an incentivized risk elicitation task (Brañas-Garza et al., 2017). However, there are also studies which did not report any significant associations (Apicella et al., 2008; Schipper, 2012; Aycinena et al., 2014; Drichoutis and Nayga, 2015).

One reason behind the conflicting results of these studies can be heterogeneity among (i) risk elicitation methods, (ii) sample sizes and ethnic backgrounds, (iii) incentive mechanisms, and

(iv) 2D:4D measurements methods. Above mentioned studies use different risk elicitation tasks such as the Holt and Laury (2005) method Brañas-Garza and Rustichini (2011), Schipper (2012), Aycinena et al. (2014), Drichoutis and Nayga (2015), the Gneezy and Potters (1997) method (Dreber and Hoffman, 2007; Apicella et al., 2008), multiple price lists (Garbarino et al., 2011) or the Balloon Analog Risk Task (Lejuez et al., 2002) method (Ronay and von Hippel, 2010). For example, Filippin and Crosetto (2016) reported that risk elicitation tasks, such as the Holt and Laury method, may fail to detect gender effects. Since 2D:4D is a sexually dimorphic measure, studies using this method may have failed to find a relationship. Furthermore, most of these tasks were employed with real monetary incentives while some (Brañas-Garza and Rustichini, 2011; Stenstrom et al., 2011) were not.

Other possible challenges may be the varying sizes and ethnic backgrounds of the samples. While some of the studies gathered their data from mixed samples, others used Caucasians or non-Caucasians only as the 2D:4D ratio is also reported to be sensitive to ethnic differences (Manning et al., 2004). In addition, using different 2D:4D measurement methods might have had an effect on 2D:4D distributions of the samples. Using scanners, photocopies, calipers, and rulers are the most common methods.

To the best of our knowledge, the relationship between 2D:4D and stereotyping has not been investigated to this date. In the account of circulating T, Josephs et al. (2003) showed that the participants with higher circulating T were more responsive to signals that reminded them of their social status than those with lower T. In their study, participants were primed negatively or positively depending on their sex prior to a math test. Women with higher circulating T who were primed by the low-numerical-ability stereotype performed lower in the math test than their low circulating T peers. Men with higher circulating T on the other hand, performed better when they were primed by high-numerical-ability stereotype than their low T peers. Josephs et al. (2003) suggest that a stereotype is a statement about one's dominance and status and therefore the effect of circulating T might have been moderated by status concerns. Similar to this finding, Millet and Dewitte (2008) showed that when men with low 2D:4D learn that they are in a subordinate position, they react strongly to excel in their social status. Millet (2009) also highlights that individuals with lower 2D:4D would have a higher need for achievement. Thus, lower 2D:4D may also be associated with a higher level of gender bias about risk preferences.

The current study initially tests the relationship between risk preferences and 2D:4D, using an incentivized Eckel and Grossman risk elicitation method (Eckel and Grossman, 2002). Furthermore, the participants of the study were also asked to make both sex-free and sex-specific predictions about other participants' choices.

## Main Hypotheses

When making predictions about other people's preferences, individuals typically base their predictions on their own preferences and on stereotypes. In this regard, several studies have found that people typically believe that they are themselves

more risk taking than others (Wallach and Wing, 1968; Levinger and Schneider, 1969; Clark et al., 1971; Lamm et al., 1972), resulting in the finding that the predictions of other people's risk taking is lower than own risk taking. One explanation for this common finding is the risk-as-value hypothesis (Brown, 1965), according to which, individuals perceive risk taking as a cultural value and therefore their beliefs about themselves and others are also biased accordingly.

*Hypothesis 1: Participants take higher risk than they estimate others to take.*

Another commonly observed phenomenon is that people rely on their own risk preferences when making predictions about others. This implies a positive relationship between risk preferences and the predictions about other people's risk preferences (false consensus effect, e.g., Ross et al., 1977).

*Hypothesis 2: Participants' risk preferences correlate positively with their estimations about others.*

In keeping with the wealth of such findings in the literature (Byrnes et al., 1999; Croson and Gneezy, 2009; Charness and Gneezy, 2012) we expect to observe higher levels of risk aversion in women. Although Filippin and Crosetto (2016) report that the magnitude and importance of this gender effect is debatable and seems to be task-specific, the task employed in this study has resulted in consistent gender differences in earlier studies.

*Hypothesis 3: Men's choices are less risk averse than women's.*

Considering the previously discussed inconclusive results on the association between 2D:4D and risk preferences we re-examine whether lower 2D:4D ratios are associated with higher risk taking.

*Hypothesis 4: 2D:4D is negatively correlated with risk taking.*

While the relationship between risk taking and 2D:4D has been tested in a number of studies, the relationship between 2D:4D and the perception of other people's risk preferences has not been examined so far, to the best of our knowledge. To predict other people's preferences, individuals often rely on their personal preferences as well as stereotypes. Stereotypically women should be risk averse and the opposite holds for men. Following the earlier discussion, we examine if participants with lower 2D:4D react more strongly to sex information than people with high 2D:4D ratios and, therefore, over-estimate women's risk aversion as well as men's risk taking.

*Hypothesis 5: The difference between predictions about men and women is negatively correlated with 2D:4D.*

## MATERIALS AND METHODS

### Participants and Procedures

The experiment was carried out in June 2017 at the Experimental Lab of Kiel University. 150 students from Kiel University participated in a total of 10 sessions and each participant participated only in one session of the experiment. Given the

mixed evidence on the relation between 2D:4D and risk taking, the sample size was chosen in order to assure sufficient power to determine a relatively small effect size. Our correlation power analysis suggested a minimum sample size of 125 ( $\alpha = 0.05$  – type I error,  $\beta = 0.20$  – type II error,  $r = 0.25$ ). Participants were recruited from the subject pool of the Experimental Lab Kiel with the software package hroot (Bock et al., 2014). Students from different faculties took part in the experiment with the majority (37%) studying economics, followed by students from the philosophy faculty (27%) and STEM fields (21%). The experiment as such was paper based and each session lasted approximately 30 min and had on average 17 participants (minimum 12 and maximum 20 participants per session). Participants received a show-up fee of €3.00 and could additionally win up to €13.00 depending on their responses. Gender distribution was almost balanced with 72 participants who indicated they were male and 74 female, while four participants did not specify their sex. Average age was 26 years ( $SD = 3.17$  and 95% confidence interval [25.30; 26.33]).

All participants of the experiment were informed with a written form about the content and the protocol of the study before participation. Participation and the hand scanning were completely voluntary and the participants were free to leave the experiment with their participation fee any time they wanted. Opting out from the hand scanning did not affect participants' pay. Anonymity was preserved by assigning the participants a randomly generated code that cannot be associated with any personal information or decision, either in the experiment or in the hand scanning. An ethical review and approval was not required for this study in accordance with the local legislation and institutional guidelines. As is standard in economics experiments, no ethical concerns were involved other than preserving the anonymity of the participants. Each participant signed a receipt of his/her payment at the end of the experiment. The whole protocol was performed in accordance with the ethical guidelines of the Kiel University Experimental Economics Lab, where it was approved by the lab manager.

### Risk Preferences and Predictions

To elicit risk preferences, the method developed by Eckel and Grossman (2002) was used (hereafter EG). Participants were confronted with six lotteries and had to choose one of them (Table 1). Each lottery had a 50% chance to win and a 50% chance to lose. The expected value of the lotteries increased from lottery 1 to 5 as well as the variance, lottery 6 had the same

TABLE 1 | EG risk elicitation task.

Lotteries (50/50 chance)	Low payoff	High payoff	Expected value
Lottery 1	€ 4.00	€ 4.00	€ 4.00
Lottery 2	€ 3.50	€ 5.00	€ 4.25
Lottery 3	€ 3.00	€ 6.00	€ 4.50
Lottery 4	€ 2.50	€ 7.00	€ 4.75
Lottery 5	€ 2.00	€ 8.00	€ 5.00
Lottery 6	€ 1.00	€ 9.00	€ 5.00



expected value as lottery 5 but a higher variance.<sup>2</sup> The higher the EG choice, the lower is the degree of risk aversion (reflected by the increase in variance from lottery 1 to 6). The participants were informed that their decision would be pay-out relevant, as at the end of the experiment a coin would be thrown and depending on the result the higher or lower amount would be paid out.

After this incentivized risk elicitation, participants were asked to estimate which lottery was chosen on average by other participants, which lottery men chose on average and which lottery women chose on average. In addition, the participants filled out a short questionnaire about general demographic information, life satisfaction, mindfulness, social comparison, and cooperation. At the end of the protocol participants were anonymously paid and their hands were scanned for 2D:4D measurement.

### 2D:4D Ratio

At the end of the protocol, both hands of each participant were scanned with a flatbed scanner. All participants were individually briefed about the scanning procedure and 2D:4D literature prior to the scans. The scanning was voluntary and one participant chose to opt out from the hand-scan. We followed Neyse and Brañas-Garza (2014) scanning and measuring protocol precisely. The scans were measured two times in GIMP software blindly (by generated participation numbers) and in a random order by a trained research assistant. There were 2 weeks between the first and the second measurements and we ensured that the measurements were recorded on blank paper to avoid framing effects and post-measure corrections. Both measurements were highly correlated ( $>0.95$ ). The mean of the two measures was taken as the main 2D:4D variable.

The average right hand 2D:4D is 0.964 ( $SD = 0.031$ ). Men have an average 2D:4D of 0.957 ( $SD = 0.030$ ) and women of 0.971 ( $SD = 0.032$ ). A classic  $t$ -test rejects equality ( $p = 0.012$ ,  $t_{143} = -2.553$ ;  $d = -0.424$ ). The left hand 2D:4D is 0.964 ( $SD = 0.039$ ). Men's average left 2D:4D is 0.960 ( $SD = 0.029$ ) and women's is 0.966 ( $SD = 0.046$ ). The difference is lower for the left hand but in the typical direction ( $p = 0.379$ ,  $t_{142} = -0.8821$ ;  $d = -0.147$ ). As men usually have lower 2D:4D ratios than women, these differences are in line with the previous literature (see Hönekopp and Watson, 2010 for a meta-analysis of sex differences in 2D:4D). The meta-analysis of Hönekopp and Watson (2010) also concludes that 2D:4D shows a greater difference on the right hand. This is why a big majority of the previous studies based their analysis on right hand measures. Although our main analysis is also based on the right hand, we also report the identical analysis for the left hand in tables and in the Appendix.

As ethnicity plays an important role in 2D:4D (Manning et al., 2004), many studies base their analysis on single-ethnicities. The follow-up questionnaire included an item where participants were asked to indicate their ethnicities. According to the results 134 reported themselves as Caucasian (90.54%), 7 mixed

(4.73%), and 3 Asian (2.03%). The remaining participants either did not fill in the item or belonged to different ethnicities. As our robustness checks with only Caucasian participants did not significantly differ from the results with the whole sample, the reported analysis includes the whole sample without any ethnicity restrictions. The statistical analysis of 2D:4D is based on 145 participants as 1 participant had a hand injury and another 4 did not fill in the sex item in the questionnaire. Among the latter, one participants opted out from the hand-scan.

## RESULTS

We will first present our correlation analysis of risk preferences and predictions. Further, we will compare the choices of men and women with  $t$ -tests. The relationships between 2D:4D and participants' choices will be investigated both with correlation and regression analyses. Finally, we will test the association between participants' 2D:4D and their predictions about sex differences in the task with both correlation and regression analyses. In line with the majority of previous studies, our analyses will be based on right hand ratios. However, we will also present the same analysis for the left hand in tables and the Appendix. Complete distributions of the variables can also be found in the Appendix.

### Descriptive Analysis of Risk Taking and Predictions

**Table 2** presents the descriptive statistics of the main variables for all participants in the study. The participants on average chose 3.080 in the six item Eckel and Grossman task. Their predictions about other participants were on average 2.160. The difference between the two variables is significant ( $t_{149} = 6.132$ ;  $p < 0.001$ ;  $d = 0.598$ ). This supports *Hypothesis 1* which postulated that participants take higher risk than they estimate others to take. Pairwise correlations show a significant positive correlation between participants' own choices and their predictions about others ( $r = 0.304$ ,  $p < 0.01$ ). This result supports *Hypothesis 2*. The average prediction about men was 3.873 and about women it was 1.740. Sex-specific predictions correlate both with EG choices ( $p < 0.01$  for both) and sex-free predictions ( $p < 0.01$  for both).

### Descriptive Analysis of Risk Taking and Predictions by Sex

**Tables 3, 3A and 3B** present the descriptive statistics for men and women separately while **Figure 1** shows the mean values of choices in the EG Task and predictions by sex. In the EG task, men chose 3.736 on average and women's mean choice was 2.432. This difference, suggesting that women are more risk averse than men, is statistically significant ( $p < 0.001$ ). This finding confirms *Hypothesis 3*.

Men's mean predictions about other participants (2.319) was slightly higher than women's mean predictions (1.959;  $p = 0.039$ ). On the one hand, men's average prediction for other men was 3.694 and women's average prediction for men was 4.054.

<sup>2</sup>The participants' choice of lottery number will be referred to as the "EG choice."



The equality between men's and women's predictions for men cannot be rejected ( $p = 0.125$ ). On the other hand, men's average prediction for women's choices was 1.847 and women's average prediction for other women was 1.568. The equality between the two cannot be rejected ( $p = 0.1$ ).

The equality between men's actual choices and their predictions about men's risk preferences cannot be rejected either ( $p = 0.839$ ). This result is also valid for women's predictions for men ( $p = 0.234$ ). However men's predictions for women were significantly lower than women's actual choices ( $p = 0.011$ ) and the same holds for women ( $p < 0.001$ ).

## Analysis of 2D:4D and Risk Preferences

In *Hypothesis 4*, we proposed a negative correlation between the two variables concerning the relationship between risk taking and 2D:4D. Our correlation analysis presented in **Table 2** failed to detect any significant relationship between right (left) 2D:4D and risk ( $r = -0.102$ ,  $p = 0.215$  and  $r = -0.066$ ,  $p = 0.429$ ). Furthermore, we did not observe any significant linear relationship between 2D:4D and our three prediction variables in either of the sexes. Therefore, *Hypothesis 4* is rejected.

To further assess the relationship between 2D:4D and risk taking we ran a series of regression models

**TABLE 2 |** Descriptive statistics.

Variable	Obs	Mean	SD	Pairwise correlation coefficients					
				1	2	3	4	5	6
1 Risk	150	3.080	1.774	1.000					
2 Prediction others	150	2.160	1.259	0.304*	1.000				
3 Prediction men	150	3.873	1.420	0.222*	0.282*	1.000			
4 Prediction women	150	1.740	1.089	0.330*	0.696*	0.187	1.000		
5 Right 2D:4D	148	0.964	0.031	-0.102	-0.021	-0.063	-0.019	1.000	
6 Left 2D:4D	148	0.964	0.039	-0.066	0.013	-0.118	0.066	0.502*	1.000

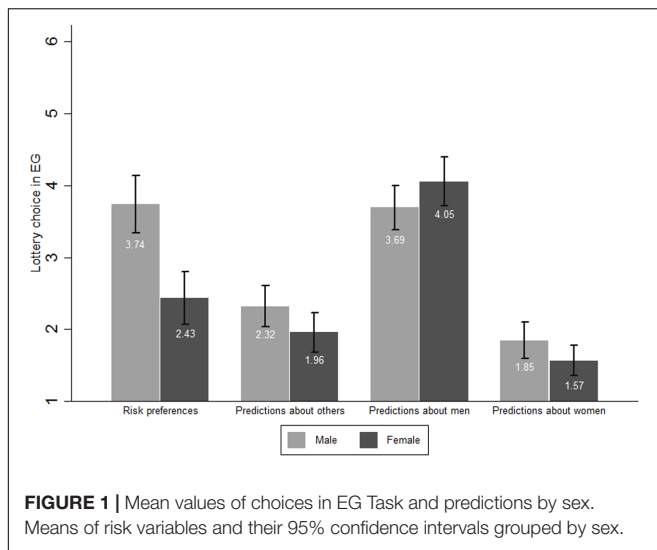
\* $p < 0.01$ .

**TABLE 3 |** Descriptive statistics by sex.

Variable	Men			t-tests	Women		
	Obs	Mean	SD		Obs	Mean	SD
Risk	72	3.736	1.728	$p < 0.001$ , $t_{144} = 4.736$ ; $d = 0.784$	74	2.432	1.597
Prediction others	72	2.319	1.243	$p = 0.039$ , $t_{144} = 1.782$ ; $d = 0.077$	74	1.959	1.199
Prediction men	72	3.694	1.328	$p = 0.1248$ , $t_{144} = -1.544$ , $d = -0.256$	74	4.054	1.479
Prediction women	72	1.847	1.109	$p = 0.1$ , $t_{144} = 1.658$ , $d = 0.275$	74	1.568	0.923
Right 2D:4D	72	0.958	0.029	$p = 0.012$ , $t_{143} = 2.553$ ; $d = -0.424$	73	0.971	0.032
Left 2D:4D	71	0.960	0.028	$p = 0.379$ , $t_{142} = -0.882$ ; $d = -0.147$	73	0.966	0.046

					Pairwise correlation coefficients					
Variable		Obs	Mean	SD	1	2	3	4	5	6
(A) Descriptive statistics men										
1	Risk	71	3.746	1.738	1.000					
2	Prediction others	71	2.324	1.251	0.315*	1.000				
3	Prediction men	71	3.704	1.335	0.382*	0.427*	1.000			
4	Prediction women	71	1.845	1.117	0.309*	0.700*	0.389*	1.000		
5	Right 2D:4D	71	0.957	0.030	−0.068	−0.056	−0.001	−0.138	1.000	
6	Left 2D:4D	71	0.960	0.029	−0.140	0.061	−0.153	0.021	0.687*	1.000
(B) Descriptive statistics women										
1	Risk	73	2.425	1.607	1.000					
2	Prediction others	73	1.959	1.207	0.181	1.000				
3	Prediction men	73	4.068	1.484	0.170	0.164	1.000			
4	Prediction women	73	1.562	0.928	0.259	0.628*	−0.033	1.000		
5	Right 2D:4D	73	0.971	0.032	0.001	0.092	−0.192	0.172	1.000	
6	Left 2D:4D	73	0.966	0.046	0.012	0.004	−0.137	0.100	0.389*	1.000

\* $p < 0.01$ .



(see Supplementary Table A1). To test the non-monotonic associations we included the quadratic form of 2D:4D in the regression analysis and controlled our models for gender effects. The results remained insignificant for both hands and also for 2D:4D-squared ( $p > 0.1$  for all 2D:4D variables).

## Correlation Analysis of 2D:4D and Gender Biases

We relate 2D:4D to the difference between predictions about men and women. To do so, we generated a *gender bias* variable by subtracting predictions about women from predictions about men. Looking at the raw correlations we observe a slight but insignificant correlation between right (left) 2D:4D and the difference in predictions about men and women ( $r = -0.042$ ,  $p = 0.609$  and  $r = -0.148$ ,  $p = 0.074$ ) and therefore we reject *Hypothesis 5* which postulated a negative correlation between 2D:4D and gender biases.

## Regression Analysis of Predictions and 2D:4D

Following our correlation analyses, we also ran an additional exploratory OLS regression analysis to investigate non-monotonic associations between predictions that participants made about other people's risk preferences and their right hand 2D:4D ratios. The dependent variable is sex-free predictions in the first four models. The latter four models investigate the association between participants' 2D:4D ratios and their predictions about the risk preference difference between the two sexes. The dependent variable is *gender bias*. First independent variable is *risk* which captures the risk preference of each participant measured by choices in the EG task. Second independent variable is 2D:4D and the third is the square of 2D:4D to observe non-monotonic relationship between 2D:4D and dependent variables. Sexes of the participants are controlled for with the dummy variable *female*. The interaction variable *2D:4Dxfemale* is also included in the models to disentangle the impact of sex on the findings about 2D:4D.

The results are shown in **Table 4**. In Models 1–4 we look at the relationship between predictions about other people's risk preferences without specifying sex. Neither 2D:4D, nor 2D:4D-squared are significant in the first four models ( $p > 0.1$  in all of them). Therefore, we may conclude that no monotonic or non-monotonic association between 2D:4D and sex-free predictions is observed. The *female* variable is also not statistically significant in any of these models. The positive and significant coefficients for personal risk taking show that participants base their predictions about others on their personal preferences ( $p < 0.01$  in all four models).

This is further assessed in **Table 4** for Models 5–8. The significant coefficients for female participants show that female participants tend to predict a higher difference between men's and women's risk taking than male participants ( $p < 0.005$  in Models 5 and 7 and  $p = 0.023$  in Model 6). As for raw correlations we do not observe a significant coefficient for 2D:4D in Models 5 and 6 ( $p$ -values are 0.297 and 0.235 respectively). Models 7 and 8, however, show that there seems to be an inverted U-shaped relationship between 2D:4D and sex difference in predictions. 2D:4D has significant and positive coefficients in both models ( $p$ -values = 0.001 and 0.012 respectively). 2D:4D-squared on the other hand has significant, negative coefficients ( $p$ -values = 0.001 and 0.013 respectively). In **Figures 2A–C** scatter plots are shown with the difference between predictions about men and women on the y-axis and right hand 2D:4D on the x-axis.<sup>3</sup> The dashed lines represent fitted quadratic models. It becomes clear that the quadratic relationship is driven by female participants where low and high 2D:4D women seem to predict a smaller difference in risk taking than women with mid-range 2D:4D ratios. The complete regression analysis on sex specific predictions can be found in Supplementary Table A2 and regressions with left hand measures in Supplementary Table A3.

## DISCUSSION

The main objective of this study was to shed light on the relationship between 2D:4D, risk taking and also predictions about risk taking of other individuals. We initially tested three common findings in the risk literature and found support for all three: (i) The (sex-free) predictions about other participants' choices were significantly lower than own choices (Wallach and Wing, 1968; Levinger and Schneider, 1969; Clark et al., 1971; Lamm et al., 1972), (ii) participants' predictions positively correlated with their own choices, which is a finding in support of the false-consensus effect (Krueger and Clement, 1994), (iii) men's choices were more risk seeking than women (Byrnes et al., 1999; Croson and Gneezy, 2009; Charness and Gneezy, 2012). These findings support our first three hypotheses.

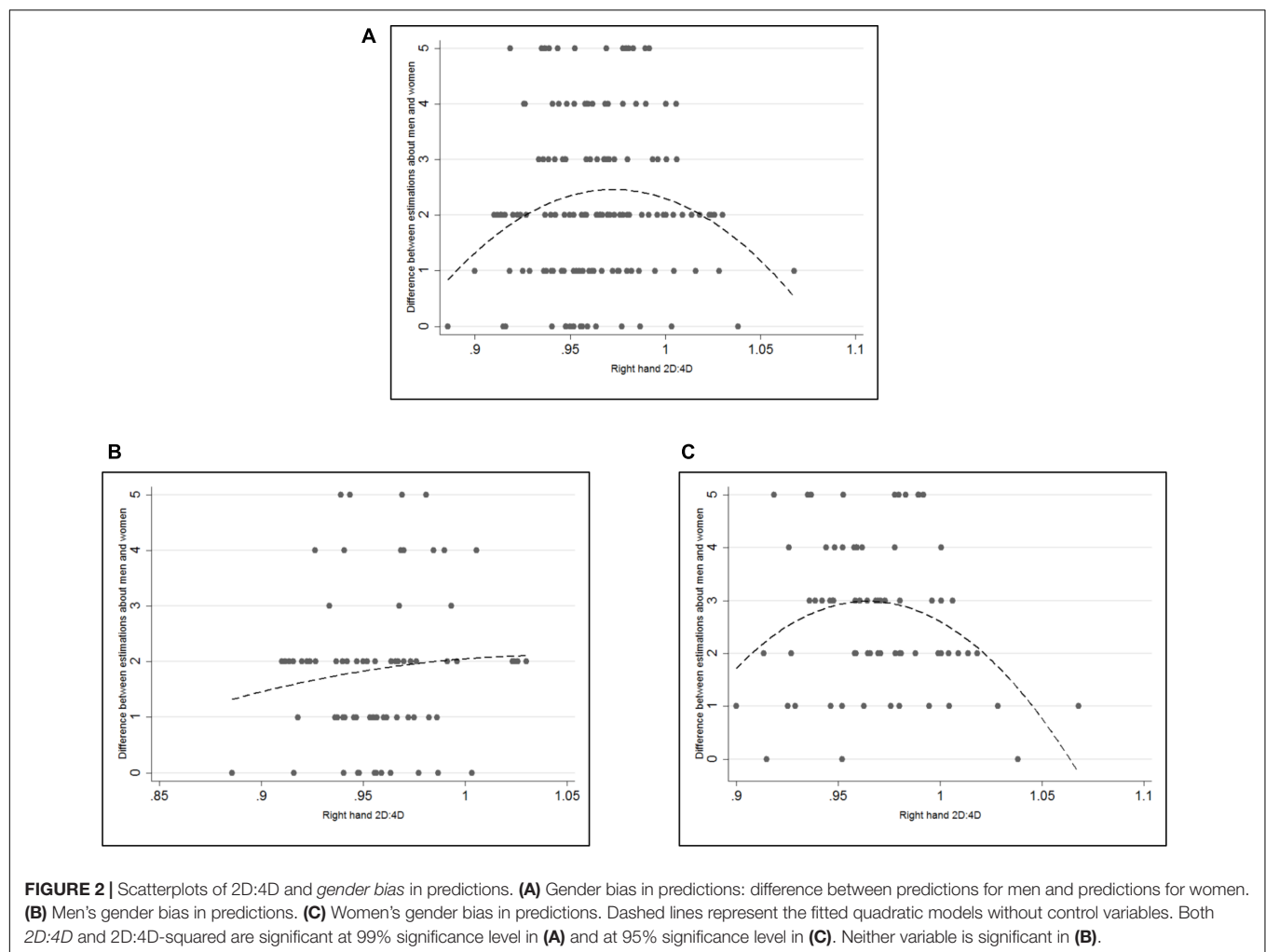
The participants also stated their predictions about men's and women's choices in the task. The results show that both men and women estimated the choices of men correctly whereas the

<sup>3</sup>For better representation, three observations with negative differences between prediction about men and women's risk taking were omitted. These observations are however included in the regression analysis in **Table 4** and the inclusion or omission of the observations makes no difference to the qualitative results.

**TABLE 4 |** Regression analysis of right hand 2D:4D, risk predictions and gender bias in predictions.

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable	predictions	predictions	predictions	predictions	gender bias	gender bias	gender bias	gender bias
Risk	0.185 (0.005)	0.183 (0.006)	0.185 (0.005)	0.183 (0.006)	0.053 (0.585)	0.060 (0.521)	0.047 (0.611)	0.053 (0.559)
2D:4D	1.137 (0.715)	-1.621 (0.699)	-3.334 (0.980)	26.189 (0.847)	-5.105 (0.297)	5.368 (0.235)	490.956 (0.001)	415.175 (0.012)
2D:4D <sup>2</sup>	—	—	2.306 (0.973)	-14.444 (0.837)	—	—	-255.847 (0.001)	-212.852 (0.013)
Female	-0.133 (0.512)	-5.073 (0.404)	-0.132 (0.514)	-5.425 (0.377)	0.795 (0.004)	19.559 (0.023)	0.784 (0.004)	14.369 (0.092)
2D:4D X female	—	5.124 (0.417)	—	5.489 (0.388)	—	-19.464 (0.029)	—	-14.090 (0.112)
Constant	0.539 (0.856)	3.186 (0.430)	2.704 (0.966)	-10.185 (0.877)	6.537 (0.174)	-3.516 (0.424)	-233.639 (0.002)	-200.554 (0.011)
R squared	0.084	0.088	0.084	0.088	0.055	0.089	0.103	0.119
N	145	145	145	145	145	145	145	145

OLS regressions, dependent variables are predictions about others [1,6] in the first four models and sex differences in predictions (gender bias = predictions about men – predictions about women). 2D:4D<sup>2</sup> is the square of 2D:4D for quadratic models and 2D:4D X female is the interaction variable for 2D:4D and female. *p*-Values are given in parentheses.



predictions about women were significantly lower than women's actual choices. Underestimation of women's risk taking behavior is commonly observed in the existing literature (Roszkowski and Grable, 2005; Daruvala, 2007; Eckel and Grossman, 2008; Grossman, 2013).

We then re-tested the connection between participants' 2D:4D and their own risk taking. No significant relationship between 2D:4D and risk taking were observed in the current study as in Apicella et al. (2008), Schipper (2012), Aycinena et al. (2014), and Drichoutis and Nayga (2015). Due to this result we reject our *Hypothesis 4*. We did not observe a significant relationship between 2D:4D and sex-free predictions either.

As gender biases may be connected to one's perceptions about others, a possible relationship between 2D:4D and biased predictions were also tested. While 2D:4D did not correlate with predictions, we also ran the same analysis with quadratic models to investigate possible non-monotonic associations between 2D:4D and sex-free predictions. Yet, no non-monotonic association was observed either. However, our *gender bias* variable showed significant, non-monotonic results for women. The inverted U-Shape pattern suggests that female participants with mid-range 2D:4D ratios estimated a higher difference between men and women's risk preferences than those with high or low 2D:4D ratios. This unanticipated non-monotonic result calls for further investigation as the relationship between 2D:4D and beliefs about other people's risk preferences has not been investigated before.

There are several studies on 2D:4D that showed non-monotonic results in various contexts. Brañas-Garza et al. (2013) observed an inverted U-Shape pattern between altruism and 2D:4D in both sexes, where the results were more consistent for men than women. This pattern showed that the participants with low and high values of 2D:4D decided to give less money in the dictator game than those with mid-range values of 2D:4D. The same inverted U-Shape pattern between altruism and 2D:4D was also confirmed for both sexes with a larger and multi-ethnic sample in Galizzi and Nieboer (2015). Moreover, in Sanchez-Pages and Turiegano (2010) the individuals with mid-range levels of 2D:4D cooperated more often in the Prisoner's Dilemma Game. Nye et al. (2012) also showed non-linear associations between 2D:4D and academic performances in samples from Manila and Moscow. In the account for circulating T, Stanton et al. (2011) showed that individuals with low or high levels of circulating T were risk and ambiguity neutral, whereas those with mid-range levels of T were more risk and ambiguity averse. Sapienza et al. (2009) also discussed non-linear associations between risk preferences and circulating T. Furthermore, non-linear associations between salivary T concentrations and visuospatial performance were found in Moffat and Hampson (1996), and between salivary T concentrations and cardiovascular health in Laughlin et al. (2010).

One possible explanation behind non-monotonic relationships between 2D:4D and certain types of behavior may be evolutionary optimization (Alexander, 1996; Sutherland, 2005). Laughlin et al. (2010) discusses the mechanisms behind the non-linear effects of T through the relationship between androgen receptor density and neurotransmitter receptor

GABA-A, which has been associated with decision patterns in humans (Lane and Gowin, 2009). As Manning et al. (2003) have shown associations between 2D:4D and androgen receptor gene, the androgen receptor density argument may also be an alternative explanation for non-linearities in 2D:4D studies. McFadden (2002) discusses the non-monotonic impacts of androgen exposure on both humans and animals in detail.

While our results support the conventional findings in the economics literature, we did not find any clear relationship between 2D:4D and risk preferences. The novelty of the current study was its inclusion of perceptions about other people's risk preferences in the analysis and controlling for sex-specific predictions. We did not find any significant linear relationship between 2D:4D and any of the prediction variables. An unanticipated finding was the inverted U-shaped pattern between 2D:4D and our generated *gender bias* variable for only women in the sample. According to this result women with low or high levels of 2D:4D predicted a smaller difference between men and women's risk preferences than women with mid-range levels of 2D:4D. Although this relationship has not been investigated before in the literature, it may initiate a new discussion on the link between 2D:4D and decision making under the impact of stereotypes.

As discussed earlier, studies examining the relationship between 2D:4D and risk preferences lack methodological consistency. Several studies use self-reported risk elicitation methods, while some others employ incentivized risk elicitation tasks. Neyse et al. (2016) and Brañas-Garza et al. (2017) showed that the behavior effected by 2D:4D is highly sensitive to monetary incentives. Thus, altering incentives may be one of the reasons behind the lack of consensus. While analyzing decision making under risk, Prospect Theory and Cumulative Prospect Theory (Tversky and Kahneman, 1975, 1992) take into account reference dependence, rank dependence and sign dependence; as risk-taking is closely connected with several other concepts such as loss aversion, ambiguity aversion, or non-linearity in utility. However, risk elicitation tasks used in previous studies have been unable to identify the association between 2D:4D and risky decisions. This is also one of the shortcomings of the current study.

Our results contribute to the growing literature on the biological underpinnings of economic behavior. Since the association between 2D:4D and risk preferences is still not clear, more detailed and systematic investigation on the connection between T and decision making under risk is needed. In this regard, we provide evidence on the gender biased predictions about others' risk taking. Several studies have pointed out that social comparisons shape risk preferences (Hill and Buss, 2010) and knowledge of income inequality has a higher impact on risk taking than the income itself (Schmidt et al., 2015). In keeping with this evidence, social underpinnings of risk preferences may also be associated with 2D:4D. As stereotypes shape economic life and decisions (see for example Fershtman and Gneezy, 2001; Andreoni and Petrie, 2008) studying the biological roots of stereotyping could also help explain important economic phenomena.



Another limitation of our study is the representativeness bias in student samples. Although a majority of experimental studies are conducted with university students, the representativeness problem is still considered a major drawback in economics experiments. See Levitt and List (2007) for a detailed discussion on laboratory experiments and also Exadaktylos et al. (2013) for a representativeness analysis of self-selected student samples. Although, the findings in 2D:4D literature give important insights into the biological factors of human behavior, the results are both context and sample dependent. Therefore, one should be careful about drawing general conclusions from these findings. Last but not least, the majority of the studies in the literature suffer from small sample sizes and lack of ethnic diversity; limitations which also apply to the current study.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## REFERENCES

- Alexander, R. M. (1996). *Optima for Animals*. Princeton, NJ: Princeton University Press.
- Andreoni, J., and Petrie, R. (2008). Beauty, gender and stereotypes: Evidence from laboratory experiments. *J. Econ. Psychol.* 29, 73–93. doi: 10.1016/j.joep.2007.07.008
- Apicella, C. L., Carré, J. M., and Dreber, A. (2015). Testosterone and economic risk taking: a review. *Adapt. Hum. Behav. Physiol.* 1, 358–385. doi: 10.1007/s40750-014-0020-2
- 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
- Aycinena, D., Baltaduonis, R., and Rentschler, L. (2014). Risk preferences and prenatal exposure to sex hormones for ladinos. *PLOS ONE* 9:e103332. doi: 10.1371/journal.pone.0103332
- Ball, S., Eckel, C. C., and Heracleous, M. (2010). Risk aversion and physical prowess: prediction, choice and bias. *J. Risk Uncertain.* 41, 167–193. doi: 10.1007/s11166-010-9105-x
- Berndt, T. J., and Heller, K. A. (1986). Gender stereotypes and social inferences: a developmental study. *J. Pers. Soc. Psychol.* 50:889. doi: 10.1037/0022-3514.50.5.889
- Bock, O., Baetge, I., and Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *Eur. Econ. Rev.* 71, 117–120. doi: 10.1016/j.eurocorev.2014.07.003
- Brañas-Garza, P., Galizzi, M. M., and Nieboer, J. (2017). *Digit Ratio and Risk Taking: Evidence from a Large, Multi-Ethnic Sample*. Orange, CA: Chapman University.
- Brañas-Garza, P., and Kovářík, J. (2013). *Digit Ratios and Social Preferences: a Comment on Buser (2012)*. Orange, CA: Chapman University.
- 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
- Brown, R. (1965). *Social Psychology*. New York, NY: Free Press.
- Brown, R. P., and Josephs, R. A. (1999). A burden of proof: stereotype relevance and gender differences in math performance. *J. Pers. Soc. Psychol.* 76:246. doi: 10.1037/0022-3514.76.2.246
- Buser, T. (2012). Digit ratios, the menstrual cycle and social preferences. *Games Econ. Behav.* 76, 457–470. doi: 10.1016/j.geb.2012.07.006
- Byrnes, J. P., Miller, D. C., and Schafer, W. D. (1999). Gender differences in risk taking: a meta-analysis. *Psychol. Bull.* 125, 367–383. doi: 10.1037/0033-2909.125.3.367
- Charness, G., and Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *J. Econ. Behav. Organ.* 83, 50–58. doi: 10.1016/j.jebo.2011.06.007
- Clark, R. D., Crockett, W. H., and Archer, R. L. (1971). Risk-as-value hypothesis: the relationship between perception of self, others, and the risky shift. *J. Pers. Soc. Psychol.* 20, 425–429. doi: 10.1037/h0031933
- Coates, J. M., Gurnell, M., and Rustichini, A. (2009). Second-to-fourth digit ratio predicts success among high-frequency financial traders. *Proc. Natl. Acad. Sci. U.S.A.* 106, 623–628. doi: 10.1073/pnas.0810907106
- Crosen, R., and Gneezy, U. (2009). Gender differences in preferences. *J. Econ. Lit.* 47, 448–474. doi: 10.1257/jel.47.2.448
- Dalton, P. S., and Ghosal, S. (2014). Self-confidence, overconfidence and prenatal testosterone exposure: evidence from the lab. *Paper Presented at the Center Discussion Paper Series No. 2014–014*, Tilburg.
- Daruvala, D. (2007). Gender, risk and related stereotypes. *J. Risk Uncertain.* 35, 265–283. doi: 10.1007/s11166-007-9024-7
- Dreber, A., and Hoffman, M. (2007). *Portfolio Selection in Utero*. Stockholm: Stockholm School of Economics.
- Drichoutis, A. C., and Nayga, R. M. (2015). Do risk and time preferences have biological roots? *South. Econ. J.* 82, 235–256. doi: 10.4284/0038-4038-2013.246
- Eagly, A. H., and Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *J. Pers. Soc. Psychol.* 46:735. doi: 10.1037/0022-3514.46.4.735
- Eagly, A. H., Wood, W., and Diekmann, A. B. (2000). “Social role theory of sex differences and similarities: a current appraisal,” in *The Developmental Social Psychology of Gender*, eds T. Eckes and H. M. Trautner (Mahwah, NJ: Lawrence Erlbaum Associates), 123–174.
- Eckel, C. C., and Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evol. Hum. Behav.* 23, 281–295. doi: 10.1016/S1090-5138(02)00097-1
- Eckel, C. C., and Grossman, P. J. (2008). Forecasting risk attitudes: an experimental study using actual and forecast gamble choices. *J. Econ. Behav. Organ.* 68, 1–17. doi: 10.1016/j.jebo.2008.04.006
- Exadaktylos, F., Espín, A. M., and Branas-Garza, P. (2013). Experimental subjects are not different. *Sci. Rep.* 3:1213. doi: 10.1038/srep01213
- Fershtman, C., and Gneezy, U. (2001). Discrimination in a segmented society: an experimental approach. *Q. J. Econ.* 116, 351–377. doi: 10.1162/003355301556338
- Festinger, L. (1954). A theory of social comparison processes. *Hum. Relat.* 7, 117–140. doi: 10.1177/001872675400700202

## FUNDING

This study was funded by the Kiel Institute for the World Economy.

## ACKNOWLEDGMENTS

We would like to thank Christian Diestel for his great assistance with research and Carsten Schröder and Antonio M. Espín for their help and support. We also want to thank the two referees for their helpful comments and remarks which significantly improved this paper.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnbeh.2018.00009/full#supplementary-material>

- Filippin, A., and Crosetto, P. (2016). A reconsideration of gender differences in risk attitudes. *Manag. Sci.* 62, 3138–3160. doi: 10.1287/mnsc.2015.2294
- 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
- Garbarino, E., Slonim, R., and Sydnor, J. (2011). Digit ratios (2D:4D) as predictors of risky decision making for both sexes. *J. Risk Uncertain.* 42, 1–26. doi: 10.1007/s11166-010-9109-6
- Gneezy, U., and Potters, J. (1997). An experiment on risk taking and evaluation periods. *Q. J. Econ.* 112, 631–645. doi: 10.1162/003355397555217
- Grossman, P. J. (2013). Holding fast: the persistence and dominance of gender stereotypes. *Econ. Inq.* 51, 747–763. doi: 10.1111/j.1465-7295.2012.00479.x
- Hermann, D. (2017). Determinants of financial loss aversion: the influence of prenatal androgen exposure (2D: 4D). *Pers. Individ. Diff.* 117, 273–279. doi: 10.1016/j.paid.2017.06.016
- Hill, S. E., and Buss, D. M. (2010). Risk and relative social rank: positional concerns and risky shifts in probabilistic decision-making. *Evol. Hum. Behav.* 31, 219–226. doi: 10.1016/j.evolhumbehav.2010.01.002
- Hogg, M. A. (2000). "Social identity and social comparison," in *Handbook of Social Comparison*, eds J. Suls and L. Wheeler (New York, NY: Springer), 401–421.
- Holt, C. A., and Laury, S. K. (2005). Risk aversion and incentive effects: new data without order effects. *Am. Econ. Rev.* 95, 902–904. doi: 10.1257/0002828054201459
- 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 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
- Hsee, C. K., and Weber, E. U. (1997). A fundamental prediction error: self–others discrepancies in risk preference. *J. Exp. Psychol.* 126:45. doi: 10.1037/0096-3445.126.1.45
- Josephs, R. A., Newman, M. L., Brown, R. P., and Beer, J. M. (2003). Status, testosterone, and human intellectual performance: stereotype threat as status concern. *Psychol. Sci.* 14, 158–163. doi: 10.1111/1467-9280.t01-1-01435
- Judd, C. M., and Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychol. Rev.* 100:109. doi: 10.1037/0033-295X.100.1.109
- Katz, D., and Braly, K. (1965). "Verbal stereotypes and racial prejudice," in *Basic Studies in Social Psychology*, eds H. Proshansky and B. Seidenberg (New York, NY: Rinehart and Winston), 266–272.
- Kray, L. J., and Thompson, L. (2004). Gender stereotypes and negotiation performance: an examination of theory and research. *Res. Organ. Behav.* 26, 103–182. doi: 10.1016/S0191-3085(04)26004-X
- Krueger, J., and Clement, R. W. (1994). The truly false consensus effect: an ineradicable and egocentric bias in social perception. *J. Pers. Soc. Psychol.* 67, 596–610. doi: 10.1037/0022-3514.67.4.596
- Lamm, H., Trommsdorff, G., and Rost-Schaude, E. (1972). Self-image, perception of peers' risk acceptance and risky shift. *Eur. J. Soc. Psychol.* 2, 255–272. doi: 10.1002/ejsp.2420020304
- Lane, S. D., and Gowin, J. L. (2009). GABAergic modulation of human social interaction in a prisoner's dilemma model via acute administration of alprazolam. *Behav. Pharmacol.* 20, 657–661. doi: 10.1097/FBP.0b013e32832ec62b
- Latrofa, M., Vaes, J., Cadinu, M., and Carnaghi, A. (2010). The cognitive representation of self-stereotyping. *Pers. Soc. Psychol. Bull.* 36, 911–922. doi: 10.1177/0146167210373907
- Laughlin, G. A., Goodell, V., and Barrett-Connor, E. (2010). Extremes of endogenous testosterone are associated with increased risk of incident coronary events in older women. *J. Clin. Endocrinol. Metab.* 95, 740–747. doi: 10.1210/jc.2009-1693
- Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., and Stuart, G. L. (2002). Evaluation of a behavioral measure of risk taking: the balloon analogue risk task (BART). *J. Exp. Psychol.* 8, 75–84. doi: 10.1037/1076-898X.8.2.75
- Leopold, T. A., Ratcheva, V., and Zahidi, S. (2016). *Global Gender Gap Report 2016, World Economic Forum*. Available at: [http://www3.weforum.org/docs/GGGR16/WEF\\_Global\\_Gender\\_Gap\\_Report\\_2016.pdf](http://www3.weforum.org/docs/GGGR16/WEF_Global_Gender_Gap_Report_2016.pdf)
- Levinger, G., and Schneider, D. J. (1969). Test of the "risk is a value" hypothesis. *J. Pers. Soc. Psychol.* 11, 165–169. doi: 10.1037/h0026966
- 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
- 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
- Manning, J. T. (2002). *Digit Ratio: A Pointer to Fertility, Behavior, and Health*. New Brunswick NJ: Rutgers University Press.
- Manning, J. T., Bundred, P. E., Newton, D. J., and Flanagan, B. F. (2003). The second to fourth digit ratio and variation in the androgen receptor gene. *Evol. Hum. Behav.* 24, 399–405. doi: 10.1016/S1090-5138(03)00052-7
- Manning, J. T., Stewart, A., Bundred, P. E., and Trivers, R. L. (2004). Sex and ethnic differences in 2nd to 4th digit ratio of children. *Early Hum. Dev.* 80, 161–168. doi: 10.1016/j.earlhumdev.2004.06.004
- McFadden, D. (2002). Masculinization effects in the auditory system. *Arch. Sex. Behav.* 31, 99–111. doi: 10.1023/A:1014087319682
- Millet, K. (2009). Low second-to-fourth-digit ratio might predict success among high-frequency financial traders because of a higher need for achievement. *Proc. Natl. Acad. Sci. U.S.A.* 106:E30. doi: 10.1073/pnas.0900396106
- Millet, K., and Dewitte, S. (2008). A subordinate status position increases the present value of financial resources for low 2D: 4D men. *Am. J. Hum. Biol.* 20, 110–115. doi: 10.1002/ajhb.20689
- Moffat, S. D., and Hampson, E. (1996). A curvilinear relationship between testosterone and spatial cognition in humans: possible influence of hand preference. *Psychoneuroendocrinology* 21, 323–337. doi: 10.1016/0306-4530(95)00051-8
- Nadler, A., and Zak, P. J. (2016). "Hormones and economic decisions," in *Neuroeconomics*, eds M. Reuter and C. Montag (Berlin: Springer), 41–66.
- Neyse, L., Bosworth, S., Ring, P., and Schmidt, U. (2016). Overconfidence, incentives and digit ratio. *Sci. Rep.* 6:23294. doi: 10.1038/srep23294
- Neyse, L., and Brañas-Garza, P. (2014). *Digit Ratio Measurement Guide (No. 54134)*. Munich: University Library of Munich.
- Nye, J. V., Androusach, G., Desierto, D., Jones, G., and Yudkevich, M. (2012). 2D:4D asymmetry and gender differences in academic performance. *PLOS ONE* 7:e46319. doi: 10.1371/journal.pone.0046319
- Powell, G. N., Butterfield, D. A., and Parent, J. D. (2002). Gender and managerial stereotypes: have the times changed? *J. Manag.* 28, 177–193. doi: 10.1177/014920630202800203
- Ronay, R., and von Hippel, W. (2010). Power, testosterone, and risk-taking. *J. Behav. Decis. Mak.* 23, 473–482. doi: 10.1002/bdm.671
- Ross, L., Greene, D., and House, P. (1977). The "false consensus effect": an egocentric bias in social perception and attribution processes. *J. Exp. Soc. Psychol.* 13, 279–301. doi: 10.1016/0022-1031(77)90049-X
- Roszkowski, M. J., and Grable, J. (2005). Gender stereotypes in advisors' clinical judgments of financial risk tolerance: objects in the mirror are closer than they appear. *J. Behav. Finance* 6, 181–191. doi: 10.1207/s15427579jpfm0604\_2
- 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
- Sapienza, P., Zingales, L., and Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15268–15273. doi: 10.1073/pnas.0907352106
- Schipper, B. (2012). *Sex Hormones and Choice under Risk*. No. 127. Davis, CA: University of California, Davis.
- Schmidt, U., Neyse, L., and Aleknonyte, M. (2015). *Income Inequality and Risk Taking*. Kiel Working Paper, 2000. Kiel: Kiel Institute for the World Economy.
- Shih, M., Pittinsky, T. L., and Ambady, N. (1999). Stereotype susceptibility: identity salience and shifts in quantitative performance. *Psychol. Sci.* 10, 80–83. doi: 10.1111/1467-9280.00111
- Siegrist, M., Cvetkovich, G., and Gutscher, H. (2002). Risk preference predictions and gender stereotypes. *Organ. Behav. Hum. Decis. Process.* 87, 91–102. doi: 10.1006/obhd.2001.2956

- Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype threat and women's math performance. *J. Exp. Soc. Psychol.* 35, 4–28. doi: 10.1006/jesp.1998.1373
- Stanton, S. J., Mullette-Gillman, O. D. A., McLaurin, R. E., Kuhn, C. M., LaBar, K. S., Platt, M. L., et al. (2011). Low-and high-testosterone individuals exhibit decreased aversion to economic risk. *Psychol. Sci.* 22, 447–453. doi: 10.1177/0956797611401752
- Steele, C. M., and Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *J. Pers. Soc. Psychol.* 69:797. doi: 10.1037/0022-3514.69.5.797
- Stenstrom, E., Saad, G., Nepomuceno, M. V., and Mendenhall, Z. (2011). Testosterone and domain-specific risk: Digit ratios (2D:4D and rel2) as predictors of recreational, financial, and social risk-taking behaviors. *Pers. Individ. Diff.* 51, 412–416. doi: 10.1016/j.paid.2010.07.003
- Sutherland, W. J. (2005). The best solution. *Nature* 435:569. doi: 10.1038/435569a
- Tversky, A., and Kahneman, D. (1975). "Judgment under uncertainty: Heuristics and biases," in *Utility, Probability, and Human Decision Making*, eds D. Wendt and C. A. Vlek (Dordrecht: Springer), 141–162.
- Tversky, A., and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk uncertain.* 5, 297–323.
- van Anders, S. M., Vernon, P. A., and Wilbur, C. J. (2006). Finger-length ratios show evidence of prenatal hormone-transfer between opposite-sex twins. *Horm. Behav.* 49, 315–319. doi: 10.1016/j.yhbeh.2005.08.003
- Ventura, T., Gomes, M. C., Pita, A., Neto, M. T., and Taylor, A. (2013). Digit ratio (2D: 4D) in newborns: influences of prenatal testosterone and maternal environment. *Early Hum. Dev.* 89, 107–112. doi: 10.1016/j.earlhumdev.2012.08.009
- Wallach, M. A., and Wing, C. W. Jr. (1968). Is risk a value? *J. Pers. Soc. Psychol.* 9, 101–106. doi: 10.1037/h0025719
- Zheng, Z., and Cohn, M. J. (2011). Developmental basis of sexually dimorphic digit ratios. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16289–16294. doi: 10.1073/pnas.1108312108

**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 © 2018 Lima de Miranda, Neyse and Schmidt. 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) and the copyright owner 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.



# Self-confidence, Overconfidence and Prenatal Testosterone Exposure: Evidence from the Lab

Patricio S. Dalton<sup>1\*</sup> and Sayantan Ghosal<sup>2</sup>

<sup>1</sup> Economics, Tilburg University, Tilburg, Netherlands, <sup>2</sup> University of Glasgow, Glasgow, United Kingdom

This paper examines whether foetal testosterone exposure predicts the extent of confidence and over-confidence in own absolute ability in adulthood. To study this question, we elicited incentive-compatible measures of confidence and over-confidence in the lab and correlate them with measures of right hand 2D:4D, used as a marker for the strength of prenatal testosterone exposure. We provide evidence that men with higher prenatal testosterone exposure (i.e., low 2D:4D ratio) are less likely to set unrealistically high expectations about their own performance. This in turn helps them to gain higher monetary rewards. Men exposed to low prenatal testosterone levels, instead, set unrealistically high expectations which results in self-defeating behavior.

**Keywords:** 2D:4D, testosterone, neuroeconomics, expectations, overconfidence, self-confidence, goals  
**JEL Classification:** C91, D03, D87

## OPEN ACCESS

### Edited by:

Pablo Brañas-Garza,  
Middlesex University, United Kingdom

### Reviewed by:

Maria Cubel,  
Brunel University London,  
United Kingdom  
Jaromír Kovářík,  
University of the Basque Country  
(UPV/EHU), Spain

### \*Correspondence:

Patricio S. Dalton  
p.s.dalton@uvt.nl

**Received:** 31 August 2017

**Accepted:** 10 January 2018

**Published:** 30 January 2018

### Citation:

Dalton PS and Ghosal S (2018)  
Self-confidence, Overconfidence and  
Prenatal Testosterone Exposure:  
Evidence from the Lab.  
*Front. Behav. Neurosci.* 12:5.  
doi: 10.3389/fnbeh.2018.00005

## 1. INTRODUCTION

Self-confidence and overconfidence play a crucial role in people's decisions and welfare. While positive thinking can enhance motivation and improve performance, being overly confident—i.e., believing one is better than one actually is—can be self-defeating (Benabou and Tirole, 2002). Indeed, overconfidence bias has been used to explain phenomena such as business failures (Camerer and Lovallo, 1999), stock market bubbles and excessively frequent trading (Barber and Odean, 2001; Grinblatt and Keloharju, 2009). An important question that arises is what determines the level of self-confidence and overconfidence. It is known that nurture does play a role. Mastering own experiences and observing successful experiences of similar others can influence people's confidence (Bandura, 1977). Does nature play any role too?

We address this question by examining whether prenatal testosterone exposure determines people's confidence and overconfidence about their own ability to perform a rather unfamiliar and challenging task<sup>1</sup>. As a marker for the strength of prenatal testosterone exposure we used the ratio of the length of the index finger to the length of the ring finger (2D:4D) of the right hand. We followed the vast literature started by Manning et al. (1998) which shows that individuals with conditions associated with very high prenatal testosterone levels exhibit significantly smaller 2D:4D (Brown et al., 2002)<sup>2</sup>. To measure confidence and overconfidence, we implemented an incentive-compatible

<sup>1</sup>Prenatal testosterone exposure has been shown to have important organizing effects on brain development, several psychological traits and behavior (see Tobet and Baum, 1987).

<sup>2</sup>The most direct evidence for the link between 2D:4D and prenatal testosterone exposure comes from Lutchmaya et al. (2004) who measure foetal oestrogen and testosterone levels before birth and record digit lengths at age two. They find that the right-hand digit ratio is significantly correlated with prenatal testosterone levels and the ratio of testosterone to oestrogen levels. See also Zheng and Cohn (2011).



scheme. We introduced participants to an unfamiliar task, and we asked them to report the number of tasks they expected to solve during the experiment. Their total final earnings depended on the precision of their estimate, so subjects had incentives to truthfully report their expected performance (i.e., their confidence in their own ability)<sup>3</sup>. Our experimental design also allowed us to measure subjects' degree of overestimation of their actual performance (i.e., overconfidence) in an incentive-compatible way. We paid the subjects piece-wise during their performance task, so, when performing, they had enough monetary incentives to perform up to their maximal potential. The difference between these two incentive-compatible measures (i.e., expected minus actual performance) constituted our incentive-compatible measure of overconfidence.

We found that, *ceteris paribus*, male subjects exposed to low prenatal testosterone levels were more likely to overestimate their actual performance. Such overestimation, rather than being a rational strategy to increase motivation and hence performance, showed to be self-defeating. Overconfident participants gained significantly less earnings than participants who were rather conservative in their expectations. This is in line with Benabou and Tirole's (2002) seminal model which predicts that overconfidence can harm welfare but individuals may nevertheless display it. Our paper provides empirical evidence for this theoretical finding and it also suggests a biological origin for such systematic overconfidence.

This paper contributes to three different strands of literatures. First, it contributes to the literature of psychology. Overconfidence is "perhaps the most robust finding in the psychology of judgment" (De Bondt and Thaler, 1995, p. 389). Here we provide evidence that it is—at least partially—biologically determined.

Second, it contributes to the literature of behavioral finance. Inasmuch our experimental results can be extrapolated to the world outside the laboratory, they suggest a plausible link between two well-known empirical findings in finance, namely that overconfident traders earn lower returns than more conservative traders Barber and Odean (2001) and that male traders with lower 2D:4D earn higher long term returns and remain longer time on business (Coates et al., 2009). Our findings would suggest that the higher success of traders with lower 2D:4D might be due to less overconfidence bias. Of course, this is just a conjecture that could be directly tested in the future.

Third, the paper contributes to an emerging literature in economics which studies the relationship between 2D:4D and economic preferences, skills and economic behavior. 2D:4D has been shown to be correlated with social preferences (van den Bergh and Dewitte, 2006; Millet and Dewitte, 2009; Buser, 2012; Brañas-Garza et al., 2013; Galizzi and Nieboer, 2015), risk preferences (Brañas-Garza et al., in press), cooperation in prisoner's dilemma (Sanchez-Pages and Turiegano, 2010), contributions to public goods (Cecchi and Duchoslav, 2016), cognitive reflection (Bosch-Domènech et al., 2014), social integration (Kovárik et al., 2017) and effort provision (Friedl

et al., 2018). In the domain of finance, low digit ratio individuals achieve higher trading profits (Coates and Herbert, 2008; Coates et al., 2009), are more likely to self-select into the financial services profession (Sapienza et al., 2009), and are more active and risk-taking traders (Cronqvist et al., 2016)<sup>4</sup>. However, to the best of our knowledge, there is not much work investigating the link between 2D:4D, confidence and overconfidence. Neyse et al. (2016) study the relation between 2D:4D and participants' prediction accuracy of their performance in a cognitive reflection test. They found that when using incentivized predictions, males with low digit ratios, on average, are less overconfident about their performance.

The rest of the paper is organized as follows. Section 2 introduces the experimental method. Section 3 describes the data and section 4 introduces the results. Section 5 concludes.

## 2. METHODS

We designed an experiment to measure the three variables of interest: (ex-ante) self-confidence, ex-post overconfidence and the second to fourth digit ratio (2D:4D). Through emails and leaflets, we recruited 255 undergraduate and graduate students from the University of Warwick. We conducted twelve sessions with approximately twenty students each. Each session lasted 60 min. The average payment was £ 14 including a show up fee of £ 5. In each session, the sequence of the experiment was as follows. Once each subject read and signed the consent form, the experimenter would read out loud the experimental instructions, which included a description of the task and the monetary payments<sup>5</sup>. Participants were informed that they had 20 min to complete the same task and that they would be paid 100 points (equivalent to £ 1) per completed task. Subjects were given 1 min of practice time to get familiar with the task and after that, we elicited their self-confidence in the following way<sup>6</sup>. We asked them to predict the number of tasks they expected to successfully complete in the 20 min of performance time. The answer to that question constituted our measure of self-confidence. In section 2.1 below we describe the incentive-compatible mechanism of self-confidence elicitation. Once the subjects reported their prediction, they started performing the task for 20 min. When they finished, they were asked to fill in a questionnaire, they were paid and their right hands were scanned. Below we describe in more detail the manner in which self-confidence, overconfidence and the 2D:4D were measured.

<sup>4</sup>Outside of economics, 2D:4D has been found to be correlated with many traits including reproductive success (Manning et al., 2000), sexual orientation (Robinson and Manning, 2000) and competitiveness in sports (Manning and Taylor, 2001).

<sup>5</sup>See Appendix A in Supplementary Material for the instructions and appendices B and C in Supplementary Material for a snapshot of the screen the subjects saw.

<sup>6</sup>1 min was only enough to understand what the task was about, but was not enough to understand how to fully solve it, except for someone who had previous expertise with a similar task. Out of the 257 subjects, only 5 subjects managed to solve the task during the practice time and we excluded them from our analysis. We explain this in more detail in section 3.

<sup>3</sup>The incentive-compatible scheme of payments we used was also implemented by Mobius and Rosenblat's (2006) to measure self-confidence in a lab setting. Next section describes in detail the mechanism.

## 2.1. Confidence, Overconfidence, and Incentives Scheme

Self-confidence is broadly defined as a feeling of trust in one's ability, quality and judgment. The literature of social psychology has operationalized this broad concept using two related constructs: "perceived self-efficacy" and "outcome expectations." Perceived self-efficacy is a judgment of capability to execute given types of performances; outcome expectations are judgments about the anticipated outcomes that would arise from such performances (Bandura, 1977, 1986)<sup>7</sup>.

Both psychological concepts are usually measured with surveys compounded of several rather broad statements to which the respondents have to agree or disagree following a Likert scale. For example, perceived self-efficacy scales include items such as "I can solve most problems if I invest the necessary effort" or "I can usually handle whatever comes my way." Outcome expectancy scales contain statements of the type "If I quit smoking I will save money" or "If I quit smoking I will gain weight."

Although these scales have been proven to be useful in many settings, they were not appropriate for the purpose of this paper for the following reasons. First, we required a unidimensional and easily interpretable measure of how confident the person was about his/her capacity to perform an unfamiliar task in the lab. These scales are rather multidimensional and general. Second, this paper also aimed at measuring overconfidence, so we needed to be able to evaluate how far were expectations from actual performance. The existing psychological scales are simply not developed to measure this construct. Finally, we needed to capture the true expectations of own performance and at the same time, we wanted to ensure that subjects performed up to their maximum capacity during performance time. To achieve this, we provided subjects with the following monetary incentive scheme. Subjects were asked to solve a practice task for 1 min. Once the practice period was over, their self-confidence  $C$  was measured by asking them to report how many tasks they expected to solve during the 20-min period. The subject received a piece rate of 100 points per solved task,  $P$ , minus 40 points for each task that he mispredicted when estimating future performance:

$$100 \times P - 40 \times |C - P|$$

The misprediction penalty provided the subjects with an incentive to truthfully report their perceived performance distribution. Note that this scheme implies that the effective piece rate of performance was 140 points for each successfully completed task as long as they stay below their estimate and 60 points for each successfully completed task thereafter. Hence, truthful elicitation of self-confidence implied that the marginal incentive during the performance period decrease (though remain positive) once reaching the estimated number of tasks. For this reason, we chose a generous exchange rate from points to money (£ 0.01 per point) to ensure that even 60 points represented a salient reward and the subject had high enough

incentives to continue putting effort. Moreover, once the subject reached his estimate, it meant that he/she figured out the way to solve the task, so the marginal cost of effort put thereafter is close to zero. Note that even if the participants chose to stop before the 20 min, they would have had to wait doing nothing until the 20 min have passed. Hence, they had two options once they reached  $C$ : to stop and wait doing nothing, or continue implementing mechanically the algorithm that they had already figured out and earn money. Almost all students chose the second option, so by revealed preferences, the marginal benefit of solving the task was higher than the marginal cost. As already argued, once the task has been figured out, the marginal cost of an additional task is close to zero<sup>8</sup>.

Above and beyond confidence, we were interested in measuring the degree of overconfidence. Moore and Healy (2008) defines overconfidence as the overestimation of one's actual performance and we apply this definition for this paper<sup>9</sup>. Like self-confidence, the degree of overconfidence is usually measured with answers to survey questionnaires, in a non-incentivised way. For the same reasons exposed above, we used an incentive compatible measure of overconfidence. A person was considered to be overconfident when he/she expected to perform better than his/her actual performance. This measure pins down overconfidence in an incentive compatible way because subjects had monetary incentives to both, announce their expectations as accurately as possible and perform as good as possible.

## 2.2. 2D:4D and Other Measures

At the end of the experiment, we scanned the right hand of each subject, we measured the length of their second and fourth finger, and calculated their ratio (2D:4D ratio)<sup>10</sup>. Finger length was measured by two independent research assistants using a digital caliper. All data analysis was done using the average of the two independent measures of ratios<sup>11</sup>.

In addition to the variables of interest, we collected independent data in a post-experiment questionnaire to construct variables that were used as controls in our regressions. In particular, we elicit risk attitudes using the Eckel and Grossman (2002) method in a non-incentivized way. This method involves a single choice among six hypothetical gambles. The gambles differ in expected return and variance. Each gamble has two possible outcomes with fifty percent probabilities of each

<sup>7</sup> Perceived self-efficacy is a very different concept to self-esteem. While perceived self-efficacy is a judgment of capability, self-esteem is a judgment of worth (Bandura, 1977, p. 309).

<sup>8</sup> We are not the first using this elicitation scheme to measure a decision-maker's incentive-compatible absolute self-confidence and performance. We use exactly the same incentive scheme proposed Mobius and Rosenblat's (2006) influential paper. There is other literature eliciting measures of relative self-confidence, that is, estimates of how much individuals expect to be above of below some sample statistics (e.g., median). However, in this paper, we are interested in absolute rather than relative self-confidence.

<sup>9</sup> Overconfidence has also been defined in the literature as the overplacement of one's performance relative to others and as the overestimation of the precision in one's knowledge (Moore and Healy, 2008).

<sup>10</sup> 2D:4D was determined from right-hand measurements only, because right-hand digit ratios have been shown previously to display more robust sex differences and are thus thought to be more sensitive to prenatal androgens.

<sup>11</sup> Both independent measures displayed a high repeatability (intraclass correlation 0.875). The results if we used the two measurements separately are qualitatively the same.

occurring. The higher the gamble, the higher expected payoff but also the higher the risk involved<sup>12</sup>.

We also used the General Self-Efficacy Scale (Schwarzer and Jerusalem, 1995) to measure generalized perceived self-efficacy (see Appendix D in Supplementary Material). This Likert-type scale consists of 10 statements. Subjects were asked to indicate how true they think each statement was for them. The scale was validated in several studies and widely used internationally (Schwarzer and Born, 1997). It captures, in a general way, the belief that one can perform well in a novel or difficult tasks.

## 2.3. The Task

For our experiment, we chose a computerized puzzle which consisted of a modified version of the so-called “Tower of Hanoi” (ToH) puzzle. The standard ToH consists of three straight bars, and a number of disks of different sizes which can slide onto any bar<sup>13</sup>. The puzzle starts with the disks in a pile in ascending order of size on one bar, the biggest at the bottom, thus making a conical shape. The challenge of the puzzle is to move the entire pile of disks to another bar, respecting the following rules: (a) only one disk can be moved at a time, (b) each move consists of taking the upper disk from one of the bars and sliding it onto another bar, on top of the other disks that may already be present on that bar and (c) no disk may be placed on top of a smaller disk. We used a slightly modified version of the original ToH to increase difficulty. In our case, instead of having disks of different sizes, there were disks of different colors. The rule was to always preserve the original order of colors of the disks (pink, green, blue, turquoise, brown). For example, brown could be moved on top of any other disks, but green could only be moved on top of the pink, etc<sup>14</sup>.

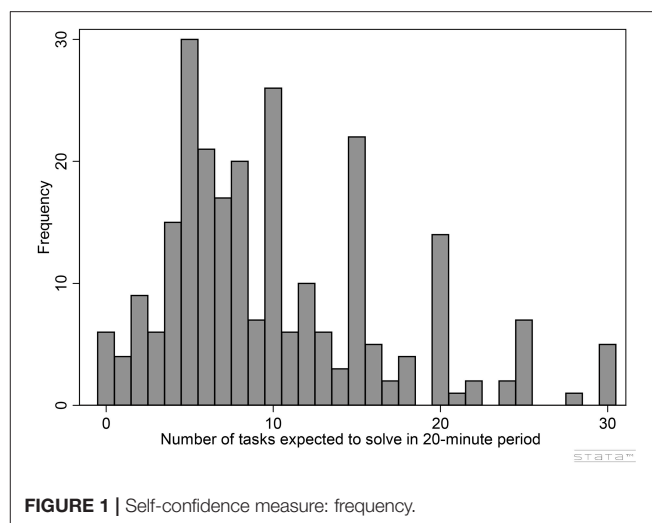
We chose this puzzle for several reasons. First, the rules of the task were easy to understand, which reduced the possibility of noise. Second, the task had a unique solution (involving thirty one moves), computed by backward induction. Third, it was quite unfamiliar to subjects and it constituted an Eureka-type of problem (Cooper and Kagel, 2005): it appeared to be challenging at first glance, but simple to solve once the algorithm is figured out. This is a desirable property for a self-confidence and overconfidence measure, since it allowed us to elicit expectations within a setting in which people had imperfect knowledge of their own abilities<sup>15</sup>. Indeed, in our experiment, only five subjects managed to solve the task in the practice time, but all eventually made it during the performance time.

## 3. DATA

Two hundred and fifty five students from Warwick University participated in the study. The sample was proportionally

**TABLE 1 |** Self-confidence: summary statistics.

	Mean	Std. Dev.	Min	Max	Obs
Whole sample	10.14343	6.629279	0	30	249
Female	8.48062	5.976648	0	30	129
Male	12.01667	6.848889	0	30	120



**FIGURE 1 |** Self-confidence measure: frequency.

balanced by gender<sup>16</sup>. Five subjects who solved the task in the practice time were excluded from all the analysis. We decided to exclude them because their prediction of expected performance would not involve any level of uncertainty about their capacity to perform. Further, we excluded one outlier with an overconfidence level forty times higher than the mean and two subjects who did not report their gender. Therefore, the final sample we analyze consisted of two hundred and forty nine subjects.

Table 1 shows the summary statistics of our experimental measure of self-confidence. On average, subjects expected to solve about ten ToHs in 20 min, with a standard deviation of about six. As Figure 1 shows, the frequency distribution of confidence in our data is quite disperse and rather skewed to the right, with a median at eight, a mode at five, a minimum at zero and a maximum at thirty. Finally, although this paper is not about gender differences, it is worth noticing that in average men expected to perform 40% better than women ( $P < 0.01$ )<sup>17</sup>.

We also looked at other variables that we expected to be positively correlated with our measure of self-confidence (see Table 2). As expected, we observed a significant positive correlation with Schwarzer and Jerusalem (1995) general measure of perceived self-efficacy ( $P < 0.01$ )<sup>18</sup>. Likewise, self-confidence was positively correlated with some proxies of the

<sup>12</sup>Since we did not provide material incentives to elicit risk preferences we label our proxy measure as risk attitude index.

<sup>13</sup>The standard ToH has been extensively studied by cognitive psychologists but very rarely used in economics (McDaniel and Rutström, 2001).

<sup>14</sup>A screenshot of the computerized puzzle can be seen in Appendix C (Supplementary Material).

<sup>15</sup>Imperfect knowledge of own ability is one of the key assumptions made by Benabou and Tirole (2002) to model self-confidence.

<sup>16</sup>32% of men and 21% of women reported to have played a similar game before, while 39% of men and 18% of women were enrolled in a maths related subject.

<sup>17</sup>This and all the tests reported hereafter are two sided.

<sup>18</sup>This correlation should be taken with caution though, since the measure of self-efficacy could be contaminated by the experience of each subject in the experiment.

ability to solve the task such as being enrolled in a mathematical oriented degree ( $P < 0.01$ ) and being familiar with the task ( $P < 0.10$ ). We also looked at its correlation with our risk attitude index, since one could expect that risk averse subjects set lower expectations. However we don't find evidence of a link between these two variables.

**Table 3** and **Figure 2** describe the data on overconfidence. Recall that those subjects whose expectations were higher (respectively lower) than their actual performance are classified as overconfident (respectively underconfident). As it can be seen in **Table 3**, the sample is equally divided between these two groups of subjects, with only 7% of

the subjects performing exactly the way they expected to perform. Interestingly, the number of overconfident (hence underconfident) subjects is equal for men and women.

Finally, **Table 4** summarizes the data on 2D:4D ratio. The average of 0.96 as well as the gender differences are in accordance with standard findings in the literature: male ratios are typically shorter than those of female.

## 4. RESULTS

### 4.1. Self-Confidence and Prenatal Testosterone Exposure

In **Table 5** we report the results of a linear regression analysis examining the relation between our measure of self-confidence and the digit ratio<sup>19</sup>. Self-confidence was significantly positively correlated with the digit ratio, suggesting that high self-confidence was associated with low prenatal testosterone exposure. When data were analyzed separately for men and women, we found that the effect was entirely driven by men. Also, as expected, men exhibited significantly higher self-confidence than women ( $P < 0.01$ ).

**TABLE 2 |** Self-confidence: Pair-wise correlations.

Construct	Variable	Self-confidence
Ability	Maths oriented degree	0.1759***
	Familiarity with the task	0.119*
Beliefs	Self-efficacy	0.1635***
Preferences	Risk-Attitude	0.0039

\*\*\*significant at 1%, \*significant at 10%.

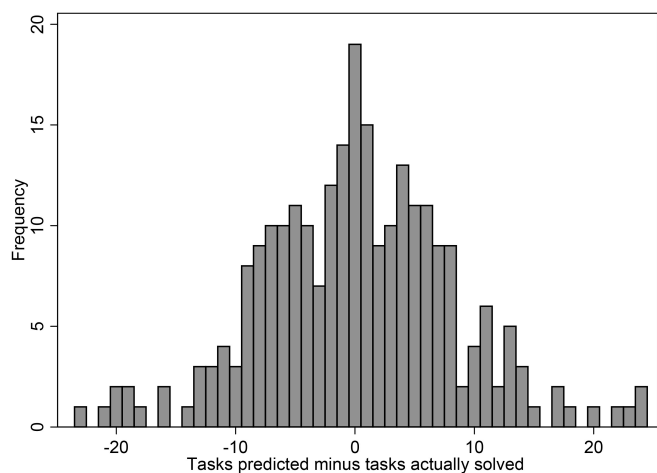
**TABLE 3 |** Predicted and actual performance.

Type	Predicted vs. actual performance	Total	Female	Male
Underconfident	Predicted < Actual Performance	114	58	56
Precise	Predicted = Actual Performance	19	7	12
Overconfident	Predicted > Actual Performance	116	59	57
		249	124	125

**TABLE 4 |** 2D/4D: Summary statistics.

	Mean	Std. Dev.	Min	Max	Obs
Whole sample	0.960252	0.03248	0.8467053	1.041442	249
Female***	0.968466	0.028316	0.8968218	1.041442	128
Male	0.951187	0.034542	0.8467053	1.028392	119

<sup>19</sup>Given that self-confidence is a count variable, we replicated our analysis using Negative Binomial Regressions and our results do not change. We chose Negative Binomial instead of Poisson regressions due to over dispersion in our data (variance greater than mean).



**FIGURE 2 |** Prediction minus actual performance.



**TABLE 5** | OLS regressions of 2D:4D on self-confidence.

	Both genders				Women				Men			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
Average digit ratio	26.749** (11.11)	30.298** (11.68)	30.513** (11.06)	29.594** (10.80)	12.32 (16.45)	13.01 (15.40)	13.07 (15.60)	13.10 (15.69)	36.734** (16.71)	39.409* (18.43)	39.201* (18.40)	38.099* (17.86)
Gender: Male = 1	3.931*** (0.96)	3.540*** (0.96)	3.572*** (0.94)	3.310*** (0.88)	—	—	—	—	—	—	—	—
Familiarity with task		0.994 (1.19)	1.006 (1.20)	0.918 (1.19)		1.74 (2.11)	1.737 (2.19)	1.511 (2.18)		−0.19 (1.06)	−0.17 (0.94)	−0.12 (0.95)
Math degree		1.676* (0.93)	1.573 (1.04)	1.612 (1.07)		−0.332 (1.95)	−0.335 (1.92)	−0.282 (1.97)		3.231** (1.19)	3.134** (1.24)	3.166** (1.22)
Risk attitude index			−0.16 (0.28)	−0.19 (0.29)			−0.005 (0.37)	−0.059 (0.39)			−0.38 (0.43)	−0.37 (0.42)
Self-efficacy				0.180* (0.09)				0.160* (0.089)				0.12 (0.24)
Observations	247	245	244	244	128	128	128	128	119	117	116	116

This table shows OLS regressions of number of repetitions of tasks expected to solve in 20 min after 1 min of practice time on the 2D:4D digit ratio. All regressions include sessions fixed effects and robust standard errors clustered by session are reported in brackets. \*\*\*significant at 1%, \*\*significant at 5%, \*significant at 10%.

The correlation between prenatal testosterone exposure and self-confidence may not reflect a causal relation between these variables but rather be due to a third variable, independently correlated with testosterone and self-confidence. For example, it may be that subjects enrolled in a mathematics oriented degree or who are familiar with the ToH, are also those who have been exposed to lower prenatal testosterone (i.e., high 2D:4D) and because of their better knowledge (and not directly because of the prenatal testosterone exposure) they expected to perform better than those with a low 2D:4D. However, when we control for these two factors, the estimated coefficient of self-confidence on 2D:4D remains substantially the same (Table 5, column II). The same happens with the risk attitude index and self-efficacy. When we include these variables in the regression, the association between prenatal testosterone exposure and self-confidence remains virtually unchanged (Table 5, columns III and IV). Interestingly, the degree of previous expertise with the task (measured with proxies such as being enrolled in a maths degree or familiarity with the task), has a significant positive correlation with male (rather than female) self-confidence, whereas perceived self-efficacy is significantly positively correlated with female (rather than male) self-confidence.

## 4.2. Overconfidence and Prenatal Testosterone Exposure

Table 6 reports results on the relation between our measure of overconfidence and digit ratio. Recall that overconfidence is defined as expectations minus actual performance, so this variable takes positive values when the person is overconfident, and is increasing in the degree of confidence. When we regressed this measure on digit ratio, we found that they were significantly positive correlated, suggesting that high overconfidence was associated with low prenatal testosterone exposure (Table 6). After controlling for possible confounding variables, like

previous experience with the task, risk attitude index and self-efficacy, the association between prenatal testosterone exposure and overconfidence became even stronger. (Table 6, columns III and IV). Again, we found this effect only in men. Also, as expected, we found that the higher the degree of previous expertise with the task and the higher the self-efficacy, the lower the overconfidence<sup>20</sup>.

## 4.3. Overconfidence and Experimental Earnings

So far we have shown that men who were exposed to higher prenatal testosterone in their mothers' womb were less likely to be overconfident. An important question that still remains unanswered regards the welfare effects of overconfidence. Was being overconfident good or bad for the subjects? Did overconfident subjects earn more money in the experiment than non-overconfident subjects?

As pointed out by Benabou and Tirole (2002), the answer is not straightforward. On the one hand, setting high expectations can improve earnings by motivating higher effort and hence improving performance. On the other hand, setting excessively high expectations can only increase the cost of not reaching them. Thus, whether overconfidence is in the end a good or a bad strategy is an empirical question. We examined this question by regressing an overconfidence dummy on the final experimental earnings (see Table 8). Our regressions confirm that being overconfident was on average a bad strategy in our experiment. Non-overconfident subjects who set their expectations below their actual potential ended up winning on average eight to

<sup>20</sup>In addition, we ran an Ordered Logit regression where the dependent variable took value zero if the predicted performance was lower than the actual performance, one if it was equal and two if it was higher. As shown in Table 7, the results remain qualitatively the same.

**TABLE 6 |** OLS regression of 2D:4D on expectations minus actual performance.

	Both genders				Women				Men			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
Average digit ratio	34.179** (14.50)	33.265** (13.05)	30.645** (13.18)	31.545** (14.14)	10.938 (23.33)	7.49 (23.32)	1.257 (23.15)	0.613 (22.93)	41.952** (17.91)	48.214*** (13.61)	48.417*** (14.96)	50.334*** (16.62)
Gender: Male = 1	0.585 (1.37)	1.2 (1.07)	1.659* (0.93)	1.946* (0.95)								
Familiarity with task		−5.897*** (1.38)	−5.537*** (1.42)	−5.417*** (1.43)	−3.888 (2.48)	−3.901 (2.43)	−3.6 (2.49)		−7.054*** (0.94)	−6.525*** (1.04)	−6.575*** (1.05)	
Math degree			−2.655** (1.06)	−2.746** (0.99)		−2.432 (2.16)	−2.471 (2.10)			−2.250* (1.22)	−2.416* (1.32)	
Risk attitude index				0.02 (0.35)			0.121 (0.44)					0.126 (0.50)
Self-efficacy				−0.210** (0.08)			−0.195 (0.14)					−0.238 (0.16)
Observations	247	245	244	244	128	128	128	128	119	117	117	116

This table shows OLS Regressions of a measure of expectations—actual performance on the 2D:4D digit ratio. All regressions include sessions fixed effects and robust standard errors are reported in brackets. \*\*\*significant at 1%, \*\*significant at 5%, \*significant at 10%.

**TABLE 7 |** Ordered logit regression of 2D:4D on under/over-confidence.

	Both genders				Women				Men			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
Average digit ratio	7.019** (3.58)	7.237* (3.95)	7.115* (4.00)	6.913* (4.19)	3.797 (6.68)	3.227 (6.92)	2.544 (6.97)	2.388 (7.18)	8.440* (4.80)	10.039** (4.42)	10.042** (4.40)	11.221** (4.98)
Gender: Male = 1	0.193 (0.22)	0.317* (0.19)	0.336* (0.18)	0.327* (0.17)	.	.	.	.	.	.	.	.
Familiarity with task		−1.280*** (0.31)	−1.265*** (0.31)	−1.264*** (0.31)	−1.093* (0.59)	−1.099* (0.59)	−1.081* (0.58)		−1.404*** (0.40)	−1.401*** (0.41)	−1.478*** (0.42)	
Math degree			−0.122 (0.30)	−0.133 (0.30)		−0.268 (0.40)	−0.26 (0.42)			−0.012 (0.43)	−0.07 (0.45)	
Risk attitude index				0.009 (0.08)			0.022 (0.09)					−0.061 (0.13)
Self-efficacy				−0.038 (0.03)			−0.013 (0.05)					−0.098* (0.06)
Observations	247	245	244	244	128	128	128	128	119	117	117	116

This table shows Ordered Logit Regressions of a variable that takes value 0 if Predicted < Actual Performance, 1 if Predicted = Actual Performance and 2 if Predicted > Actual Performance on the 2D:4D digit ratio. All regressions include sessions fixed effects and robust standard errors clustered by session are reported in brackets. \*\*\*significant at 1%, \*\*significant at 5%, \*significant at 10%.

nine British pounds more than overconfident subjects<sup>21</sup>. These results are true for both, men and women, and controlling for a series of possible confounders. The magnitude of the cost of overconfidence on earnings was very high: it more than doubled the cost of not having previous experience with the task. Interestingly, the 2D:4D ratio did not affect earnings directly, but through its effect on self-confidence.

The subjects who performed better in the lab seemed to have pursued a strategy that the psychologists know as “defensive

pessimism”: setting low expectations in uncertain situations to harness anxiety and thus perform better. This strategy was also discussed in the economic model of Benabou and Tirole (2002). In their theory, “defensive pessimism” comes as a result from assuming that ability is a substitute rather than a complement of effort in generating future pay-offs. This gives the person an incentive to discount or repress signals of high ability, as these would increase the temptation to “coast” or “slack off.” In other words, considering the possibility of failure may motivate higher effort to avoid that possibility, and it is a rational strategy to follow inasmuch it increases performance. This is, indeed, what we observe in our experimental data: overconfident

<sup>21</sup>Note that given that we created the dummies *Exceeded* and *Correct Expectations*, the benchmark variable for comparisons is *Unreached Expectations*.

**TABLE 8 |** OLS regression of under/over-confidence on actual earnings.

	Both Genders		Women		Men	
	I	II	I	II	I	II
Exceeded Expectations	8.750*** (0.76)	7.573*** (0.73)	7.774*** (1.01)	7.229*** (0.93)	8.851*** (1.21)	7.312*** (1.26)
Correct Expectations	4.461*** (0.52)	4.512*** (0.76)	3.869*** (1.26)	3.249*** (0.92)	5.549*** (1.56)	7.453*** (1.86)
Gender: Male = 1	3.282*** (1.08)	1.865** (0.91)	–	–	–	–
Familiarity with task		3.209** 1.06		2.416* (1.35)		3.550*** (1.26)
Math degree		3.560*** (0.98)		1.227 (1.70)		5.026*** 1.10
Risk attitude index		–0.2 (0.22)		–0.053 (0.21)		–0.577* (0.33)
Self-efficacy		0.285*** (0.06)		0.242** (0.09)		0.261 (0.18)
Average digit ratio		10.32 (11.21)		13.36 (15.95)		4.91 (14.67)
Observations	247	244	128	128	119	116

*Exceeded expectations is a dummy variable that takes value 1 if Expectations < Actual Performance and zero otherwise. Correct expectations is a dummy variable that takes value 1 if Expectations = Actual Performance and zero otherwise. The benchmark variable for comparison is unmet expectations or overconfidence (i.e., if Expectations > Actual Performance). The dependent variable is final experimental earnings measured in GBP. All regressions include sessions fixed effects and robust standard errors clustered by session are reported in brackets. \*\*\*significant at 1%, \*\*significant at 5%, \*significant at 10%.*

subjects gained substantially lower earnings than subjects who set more modestly their expectations. Overconfidence was self-defeating.

## 5. CONCLUSION

This paper examines the biological determinants of self-confidence and overconfidence. We provide evidence that men with higher prenatal testosterone exposure (i.e., low 2D:4D ratio) are less likely to set unrealistically high expectations about their own performance. Importantly, we also show that such bias has normative implications: overconfidence was detrimental for individuals' earnings. Our results are in line with the findings in Neyse et al. (2016) when they use incentive compatible measures of confidence and over-confidence. Both pieces of independent evidence using different tasks and samples confer further validation to our findings that men with low 2D:4D ratio are less overconfident.

The evidence in this paper can be understood as a plausible explanation of why male financial traders with higher prenatal testosterone exposure remain longer on business or have higher long term profits (Coates et al., 2009). According to our findings, these traders may be less likely to suffer from overconfidence bias,

and this helps them to be more successful in the long run. This interpretation is consistent with the empirical findings of Barber and Odean (2001), who show that overconfidence is negatively correlated with traders financial returns<sup>22</sup>.

Our paper also provides an alternative plausible channel through which prenatal testosterone exposure may affect behavior and outcomes in other settings. For instance, prenatal testosterone has been shown to be positively correlated with performance in a range of sports. The main explanation put forward is that it promotes the development of male fighting and competitiveness, which are useful traits to succeed in sports (Manning and Taylor, 2001). The evidence presented here suggests another alternative explanation: men with high prenatal testosterone exposure may succeed in sports because they may use “defensive pessimism” strategies. That is, they may set low expectations to harness anxiety and hence perform better.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of “the Ethics committee in the Faculty of Social Studies at the University of Warwick” with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the “Ethics committee in the Faculty of Social Studies at the University of Warwick.”

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

## ACKNOWLEDGMENTS

Both authors acknowledge support from ESRC-DFID grant RES-167-25-0364. We are grateful to Steven Lovelady and Alex Dobson for excellent research assistance in the lab and to Pablo Brañas-Garza, Burkhard Schipper, and Elena Cettolin for helpful comments. We also thank assistance of Mariela Dal Borgo and Dimitri Milgrom.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnbeh.2018.00005/full#supplementary-material>

<sup>22</sup>The other alternative explanations to Coates, Gurnell and Rustichini (2009) findings rely on risk preferences or preferences for competition. However, the evidence on the link between these two preferences and 2D:4D is mixed. While Sapienza et al. (2009), Apicella et al. (2008), and Schipper (2015) find no significant correlations between risk preferences and the digit ratio, Brañas-Garza et al. (in press) find that subjects with lower digit ratios tend to choose riskier lotteries. Furthermore, Pearson and Schipper (2012) finds no correlation between 2D:4D and competitive behavior in markets.

## REFERENCES

- 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
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychol. Rev.* 84, 191–215. doi: 10.1037/0033-295X.84.2.191
- Bandura, A. (1986). *Social Foundations of Thought and Action*. Englewood Cliffs, NJ: Albert Bandura.
- Barber, B. M., and Odean, T. (2001). Boys will be boys: gender, overconfidence, and common stock investment. *Q. J. Econ.* 116, 261–292. doi: 10.1162/003355301556400
- Benabou, R., and Tirole, J. (2002). Self-confidence and personal motivation. *Q. J. Econ.* 117, 871–915. doi: 10.1162/003355302760193913
- Bosch-Domènech, A., Brañas-Garza, P., and Espín, A. M. (2014). Can exposure to prenatal sex hormones (2D:4D) predict cognitive reflection? *Psychoneuroendocrinology* 43, 1–10. doi: 10.1016/j.psyneuen.2014.01.023
- Brañas-Garza, P., Kovárik, 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., 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.*
- Brown, W. M., Hines, M., Fane, B. A., and Breedlove, S. M. (2002). Masculinized finger length patterns in human males and females with congenital adrenal hyperplasia. *Horm. Behav.* 42, 380–386. doi: 10.1006/hbeh.2002.1830
- Buser, T. (2012). Digit ratios, the menstrual cycle and social preferences. *Games Econ. Behav.* 76, 457–470. doi: 10.1016/j.geb.2012.07.006
- Camerer, C., and Lovo, D. (1999). Overconfidence and excess entry: an experimental approach. *Amer. Econ. Rev.* 89, 306–318. doi: 10.1257/aer.89.1.306
- Cecchi, F., and Duchoslav, I. (2016). *Prenatal Trauma and Cooperation. Evidence from a Public Goods Game in Post-Conflict Uganda*. Mimeo. Available online at: [http://cega.berkeley.edu/assets/miscellaneous\\_files/Cecchi\\_Prenatal\\_Stress\\_and\\_Cooperation\\_WGAPE.pdf](http://cega.berkeley.edu/assets/miscellaneous_files/Cecchi_Prenatal_Stress_and_Cooperation_WGAPE.pdf)
- Coates J. M., and Herbert, J. (2008). Endogenous steroids and financial risk taking on a London trading floor. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6167–6172. doi: 10.1073/pnas.0704025105
- Coates, J. M., Gurnell, M., and Rustichini, A. (2009). Second-to-fourth digit ratio predicts success among high-frequency financial traders. *Proc. Natl. Acad. Sci. U.S.A.* 106, 623–628. doi: 10.1073/pnas.0810907106
- Cooper, D. J., and Kagel, J. H. (2005). Are two heads better than one? Team versus individual play in signaling games. *Amer. Econ. Rev.* 95, 477–509. doi: 10.1257/0002828054201431
- Cronqvist, H., Previtero, A., Siegel, S., and White, R. T. (2016). The fetal origins hypothesis in finance: prenatal environment and financial risk taking. *Rev. Finan. Stud.* 29, 739–786.
- De Bondt, W., and Thaler, R. H. (1995). “Financial decision-making in markets and firms: a behavioral perspective,” in *Handbooks in Operations Research and Management Science*, vol. 9, eds R. A. Jarrow, V. Maksimovic, and W. T. Ziemba (Amsterdam: North Holland), 385–410.
- Eckel, C., and Grossman, P. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evol. Hum. Behav.* 23, 281–295. doi: 10.1016/S1090-5138(02)00097-1
- Friedl, A., Neyse, L., and Schmidt, U. (2018). Payment scheme changes and effort adjustment: the role of 2D: 4D digit ratio. *J. Behav. Exp. Econ.* 72, 86–94. doi: 10.1016/j.soc.2017.11.007
- Galizzi, M. M., and Nieboer, J. (2015). Digit ratio and altruism: evidence from a large, multi-ethnic sample. *Front. Behav. Neurosci.* 9:41. doi: 10.3389/fnbeh.2015.00041
- Grinblatt, M., and Keloharju, M. (2009). Sensation seeking, overconfidence, and trading activity. *J. Finan.* 64, 549–578. doi: 10.1111/j.1540-6261.2009.01443.x
- Kovárik, J., Brañas-Garza, P., Davidson, M., Haim, D., Carcelli S., and Fowler J. (2017). Digit ratio (2D:4D) and social integration: an effect of prenatal sex hormones. *Netw. Sci.* 5, 476–489. doi: 10.1017/nws.2017.4
- 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
- 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. Reproduct.* 13, 3000–3004. doi: 10.1093/humrep/13.11.3000
- Manning, J. T., Barley, L., Walton, J., Lewis-Jones, D. I., Trivers, R. L., Singh, D., et al. (2000). The 2nd:4th digit ratio, sexual dimorphism, population differences, and reproductive success: evidence for sexually antagonistic genes? *Evol. Hum. Behav.* 21, 163–183. doi: 10.1016/S1090-5138(00)00029-5
- Manning, J. T., and Taylor, R. P. (2001). Second to fourth digit ratio and male ability in sport: implications for sexual selection in humans. *Evol. Hum. Behav.* 22, 61–69. doi: 10.1016/S1090-5138(00)00063-5
- Millet, K., and Dewitte, S. (2009). The presence of aggression cues inverts the relation between digit ratio (2D:4D) and prosocial behaviour in a dictator game. *Br. J. Psychol.* 100, 151–162. doi: 10.1348/000712608X324359
- McDaniel, T. M., and Rutström, E. E. (2001). Decision making costs and problem solving performance. *Exp. Econ.* 4, 145–161. doi: 10.1023/A:1011480621103
- Mobius M. M., and Rosenblat, T. S. (2006). Why beauty matters. *Am. Econ. Rev.* 96, 222–235. doi: 10.1257/000282806776157515
- Moore, D. A., and Healy, P. J. (2008). The trouble with overconfidence. *Psychol. Rev.* 115, 502. doi: 10.1037/0033-295X.115.2.502
- Neyse, L., Bosworth, S., Ring, P., and Schmidt, U. (2016). Overconfidence, incentives and digit ratio. *Sci. Rep.* 6:23294. doi: 10.1038/srep23294
- Pearson, M., and Schipper, B. C. (2012). The visible hand: finger ratio (2D: 4D) and competitive bidding. *Exp. Econ.* 15, 510–529. doi: 10.1007/s10683-011-9311-7
- Robinson, S. J., and Manning, J. T. (2000). The ratio of 2nd to 4th digit length and male homosexuality. *Evol. Hum. Behav.* 21, 333–345. doi: 10.1016/S1090-5138(00)00052-0
- 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
- Sapienza, P., Zingales, L., and Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15268–15273. doi: 10.1073/pnas.0907352106
- Schipper, B. C. (2015). *Sex Hormones and Choice under Risk*. Mimeo. Available online at: <http://faculty.econ.ucdavis.edu/faculty/schipper/hormrisk.pdf>
- Schwarzer, R., and Jerusalem, M. (1995). “Generalized self-efficacy scale,” in *Measures in Health Psychology: A users Portfolio. Causal and Control Beliefs*, eds J. Weinman, S. Wright, and M. Johnston (Windsor: NFER-NELSON), 35–37.
- Schwarzer, R., and Born, A. (1997). Optimistic self-beliefs: assessment of general perceived self-efficacy in thirteen cultures. *World Psychol.* 3, 177–190.
- Tobet, S. A., and Baum, M. J. (1987). Role for prenatal estrogen in the development of masculine sexual behavior in the male ferret. *Hormon. Behav.* 21, 419–429. doi: 10.1016/0018-506X(87)90001-8
- van den Bergh, B., and Dewitte, S. (2006). Digit ratio (2D:4D) moderates the impact of sexual cues on men’s decisions in ultimatum games. *Proc. R. Soc. Biol. Sci. B* 273, 2091–2095. doi: 10.1098/rspb.2006.3550
- Zheng, Z., and Cohn, M. J. (2011). Developmental basis of the sexually dimorphic digit ratio. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16289–16294. doi: 10.1073/pnas.1108312108

**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 © 2018 Dalton and Ghosal. 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) and the copyright owner 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.





# Discounting and Digit Ratio: Low 2D:4D Predicts Patience for a Sample of Females

Diego Aycinena<sup>1\*</sup> and Lucas Rentschler<sup>2</sup>

<sup>1</sup> Department of Economics, Universidad del Rosario, Bogotá, Colombia, <sup>2</sup> Department of Economics and Finance, Utah State University, Logan, UT, United States

Inter-temporal trade-offs are ubiquitous in human decision making. We study the relationship between preferences over such trade-offs and the ratio of the second digit to that of the forth (2D:4D), a marker for pre-natal exposure to sex hormones. Specifically, we study whether 2D:4D affects discounting. Our sample consists of 419 female participants of a Guatemalan conditional cash transfer program who take part in an experiment. Their choices in the convex time budget (CTB) experimental task allow us to make inferences regarding their patience (discounting), while controlling for present-biasedness and preference for smoothing consumption (utility curvature). We find that women with lower digit ratios tend to be more patient.

**Keywords:** 2D:4D, digit ratio, time preferences, discounting, convex time budget, testosterone, economic experiments, economic behavior

## OPEN ACCESS

### Edited by:

Pablo Brañas-Garza,  
Middlesex University, United Kingdom

### Reviewed by:

Balint Lenkei,  
Middlesex University, United Kingdom  
Antonio M. Espin,  
Middlesex University, United Kingdom  
Teresa García-Muñoz,  
University of Granada, Spain

### \*Correspondence:

Diego Aycinena  
diego.acyinena@urosario.edu.co

**Received:** 13 October 2017

**Accepted:** 18 December 2017

**Published:** 24 January 2018

### Citation:

Aycinena D and Rentschler L (2018)  
Discounting and Digit Ratio: Low  
2D:4D Predicts Patience for a Sample  
of Females.  
*Front. Behav. Neurosci.* 11:257.  
doi: 10.3389/fnbeh.2017.00257

## 1. INTRODUCTION

Human decisions involving inter-temporal outcomes are ubiquitous. For example, decisions involving savings and consumption, investments in physical and human capital, and career and health choices all involve trade-offs across time. Economists and other social scientists typically study inter-temporal choices using models which parameterize how an individual weights consumption at different points in time. In particular, discounted utility models assume that individuals place a higher weight on consumption that is sooner; that is, individuals discount the future. Richer models allow for other factors that may also affect inter-temporal choices, such as utility curvature (i.e., the preference to smooth consumption over time), and present biasedness (i.e., higher discounting of the future if choices involve present outcomes)<sup>1</sup>.

Time preferences are heterogeneous among individuals (Harrison et al., 2002; Andreoni et al., 2015). That is, individuals vary in the degree to which they discount the future (their patience), in their preference to smooth consumption, and in their degree of present-biasedness. Given this heterogeneity and that the domain of inter-temporal preferences includes choices over important human capital decisions, it is not surprising that measures of discounting correlate with smoking, alcohol consumption addiction, and drug abuse (Kirby et al., 1999; Mitchell, 1999; Petry, 2001; Chabris et al., 2008; Sutter et al., 2013). In addition, Cadena and Keys (2015) finds that impatient individuals are more likely to make investments that can be classified as dynamically inconsistent

<sup>1</sup>Discounting measures how much more a subject values consumption at an earlier date relative to a delayed later date. Present-biasedness refers to an increase in discounting when the earlier date under consideration is the present. See e.g., Laibson (1997); O'Donoghue and Rabin (1999).

and consequently end up with lower income on average. Golsteyn et al. (2014) finds that high discount rates have a negative relationship with school performance, labor supply, health and income. Kirby et al. (2002) also reports evidence of patience being positively correlated with literacy and schooling among the Tsimane' in Bolivia.

Thus, understanding the underlying determinants of inter-temporal preferences can help improve our understanding of human behavior over countless domains, as well as the welfare consequences thereof<sup>2</sup>. Indeed, we still know relatively little regarding the underlying determinants of inter-temporal preferences.

In this paper we examine whether a link exists between discounting and second-to-fourth digit length ratios (2D:4D)<sup>3</sup>. 2D:4D is a marker for pre-natal exposure to sex hormones (testosterone and estradiol) in males and females (Manning, 2002; Lutchmaya et al., 2004; Zheng and Cohn, 2011). Evidence suggests that exposure to sex hormones *in utero* has an organizational effect brain development (Goy and McEwen, 1980; Manning et al., 2001).

If exposure to sex hormones *in utero* has an effect on the brain, then examining a potential effect on time preferences seems warranted. Several studies find that higher cognitive ability is associated with more patience (Shamosh et al., 2008; Burks et al., 2009; Dohmen et al., 2010; Benjamin et al., 2013). Frederick (2005) introduced the cognitive reflection test (CRT), a simple test designed to capture the cognitive capacity to override an intuitive wrong answer and reflect upon the simple yet non-intuitive correct answer. High scores in this test correlate with higher cognitive abilities (as measured by the Wonderlic Personnel Test, the Need for Cognition Scale, etc.). Furthermore, Frederick finds that individuals with higher CRT scores are generally more patient (using hypothetical choices). In addition, Bosch-Domènech et al. (2014) reports that lower 2D:4D measures are associated with higher scores on the CRT. Collectively, these studies provide a rationale to examine the relationship between 2D:4D and discounting.

We use an experimental task, the convex time budget (CTB), to measure time preferences. This method has the advantage of allowing simultaneous structural estimation of discounting, utility curvature, and present-biasedness. The simultaneous estimation is important, as estimating them separately often results in estimates of discounting that are unrealistically high (Andersen et al., 2008).

External validity of time preferences measured via experimental tasks has been documented with different samples. Among school children, experimental measures of impatience are significant predictors of savings decisions, health behavior and school misconduct (Castillo et al., 2011;

Sutter et al., 2013). Experimentally elicited present-biasedness is correlated with credit card debt among a sample of adults in Massachusetts (Meier and Sprenger, 2010), and predicts payments for environmental services in a sample of Ugandan farmers (Clot and Stanton, 2014). With the experimental task and sample reported here, (Aycinena et al., 2017) shows that preferences for consumption smoothing predict choices among a menu of payment options with large stakes.

The main contribution of this paper is to the literature on hormones and economic behavior. Specifically, we contribute to the literature that examines economic behavior and 2D:4D as a proxy for prenatal exposure to hormones and economic behavior (e.g., Brañas-Garza and Rustichini, 2011; Millet, 2011; Apicella et al., 2015). This literature has examined economic parameters such as risk preferences (Garbarino et al., 2011; Aycinena et al., 2014; Branas-Garza et al., in press), altruism (Branas-Garza et al., 2013; Galizzi and Nieboer, 2015), overconfidence regarding cognitive abilities (Neyse et al., 2016), etcetera.

There has been limited attention paid to the relationship between 2D:4D and time preferences. Drichoutis and Nayga (2015) uses two experimental tasks involving multiple price list to separately measure risk and time preferences and relates them to 2D:4D. Their evidence is mixed, but suggests that there may be a negative relationship between 2D:4D and discounting. Our paper differs in several important ways: first, they have a final sample of 138 (77 female) university students, while we have a sample size of 419 females who are not students. Second, we use five independent measures of 2D:4D taken from scans of our subjects hands using software designed for this purpose. This is intended to minimize measurement error, and increase the reliability of our measurements. Drichoutis and Nayga (2015) use rulers to measure 2D:4D, and did not scan the hands of their subjects. Third, they used the Holt and Laury (2002) method to measure risk aversion (which is presumed to measure utility curvature). This method involves subjects choosing between lotteries. We employ the CTB task, which does not involve choices over lotteries. Lucas and Koff (2010) analyzes the relationship between 2D:4D and delay discounting, but does not consider other parameters involved in inter-temporal choices (consumption smoothing and present-biasedness). They only find a significant relationship for the right hand for women. They find that a lower 2D:4D ratio is associated with greater delay discounting. Our paper differs significantly from this study in that we use a large sample of non-students, use a different elicitation method and jointly estimate multiple parameters underlying intertemporal preferences.

In addition to contributing to the hormones and economic behavior literature, this study also contributes to the economics literature exploring time preferences on three fronts. First, a robust correlation between time preferences and 2D:4D would provide an exogenous determinant of individual time preferences which could serve as an exogenous instrument to examine causal relations between time preferences and other economic behavior. This could be an important tool to examine causal relationships; for instance, in the growing literature exploring the link between patience and social preferences (Curry et al., 2008; Espín et al., 2012, 2015). Second, most economic theories implicitly

<sup>2</sup>Several papers attempt to explore the covariates of time preferences (Lawrance, 1991; Pender, 1996; Harrison et al., 2002; Tanaka et al., 2010; Cassar et al., 2017). However, establishing a causal effect between the covariates and time preferences has proven to be challenging. For instance, Carvalho et al. (2016) attempts to explore the impact of poverty or lack of liquidity on discounting.

<sup>3</sup>The null hypothesis is that no correlation exists. As specified in our registered analysis plan, our alternative hypothesis is that 2D : 4D is negatively correlated with patience; that is, low digit ratio is related to a higher degree of patience.

or explicitly assume the stability of choice primitives (such as time and risk preferences) and there is empirical evidence of some stability in time preferences at the individual and aggregate levels (Kirby, 2009; Meier and Sprenger, 2015). The link between pre-natal exposure to hormones and time preferences suggests a (partial) mechanism through which time preferences can be heterogeneous across individuals and relatively stable over time. Finally, the third front links to the literature that shows that patience is correlated with higher cognitive ability (Shamosh et al., 2008; Burks et al., 2009; Dohmen et al., 2010; Benjamin et al., 2013). Given that cognitive ability seems to be correlated with 2D:4D (Brañas-Garza and Rustichini, 2011; Bosch-Domènech et al., 2014), our results may suggest a potential mechanism through which 2D:4D affects patience.

## 2. MATERIALS AND METHODS

*Acuerdo ministerial SP-M-466-2007* (regulating human clinical trials in Guatemala) did not apply to our study and no ethics committee has existed at our (former) institution in Guatemala. Nevertheless, we adhered to standard protocols involving studies that use experimental methods and measures of 2D:4D; specifically, no deception was used in the experiments, we obtained informed consent from participants, and we ensured privacy and security of data and decisions<sup>4</sup>.

### 2.1. Participants

Our sample consists of beneficiaries of Guatemala's Conditional Cash Transfer (CCT) program<sup>5</sup>. Due to CCT program requirements, our sample is 99.1% female and not representative for Guatemala<sup>6</sup>. As might be expected, relative to female respondents on a national representative survey, participants in our experiment are poorer, more likely to be or have been married, live in larger households and their living quarters are more precarious<sup>7</sup>.

<sup>4</sup>Given the anticipated low levels of schooling and literacy, assistants read the informed consent sheet to each individual, marked whether subjects gave informed oral consent, and signed the sheet.

<sup>5</sup>*Mi Bono Seguro* (My Security Bonus) is a targeted CCT program overseen by the *Ministerio de Desarrollo Social* (Ministry of Social Development) of Guatemala. It aims to improve human capital accumulation by promoting investments in health and education for poor households with pregnant women or children under the age of 16.

<sup>6</sup>As is conventional among CCT programs, females tend to be the recipients of the funds. This program uses geographic targeting and proxy means testing for eligibility. This program offers two types of conditional transfers: an education transfer and a health transfer. To obtain the health transfer all children under 15, and all pregnant or breastfeeding woman must attend regular medical check-ups. To obtain the education transfer all children between the ages of 6 and 15 must have a school attendance rate of at least 90%. Households may be eligible for both transfers.

<sup>7</sup>We compared our sample with the 2011 National Survey of Living Conditions (ENCOVI). ENCOVI is a national representative household survey focused on the measurement of living standards run by the National Institute of Statistics (INE) of Guatemala. To maximize comparability, we restricted attention to female ENCOVI respondents in a comparable age bracket. For detailed results of this comparison, see Aycinena et al. (2015). Not surprisingly, there are limitations with the comparison between our sample and the ENCOVI data. ENCOVI is a national representative survey that was implemented between March and August of 2011,

After dropping some observations, the final sample in our analysis consists of 419 individuals<sup>8</sup>. These subjects reside in seven different municipalities across three departments: (El Progreso, Escuintla, and Sacatepéquez) where we ran experimental sessions. Ages range from 20 to 76 (mean 35.9, median 35). All of these women, as a condition for eligibility in the CCT program either have children or were pregnant at the time of the experiment.

### 2.2. Experiment

Participants performed several independent experimental tasks. The first and main task elicits inter-temporal choices using a version of the CTB introduced by Andreoni and Sprenger (2012a,b). The other tasks (which are not used in the current analysis) involve choosing how to spread receipt of financial windfall gains over time when there is no cost associated with receiving funds earlier, eliciting a subject's willingness to forgo funds in order to maintain intra-household control of a financial windfall, and/or a hypothetical CTB which elicited how subjects believed they would behave if questions were asked at a future date.

Participants earn an initial amount of GTQ50 (approximately USD6.4 or PPP\$12.3) for taking part in the experiment<sup>9</sup>. In addition, they could earn between GTQ45 – GTQ100 (PPP\$11.1 – PPP\$24.7) based on their choices in the CTB. To put these amounts in context, CCT's entitled a household to receive GTQ150 (USD19.2 or PPP\$37) per month, provided all household members comply with the conditions. Median self-reported household monthly income for the sample was in the range from GTQ500 to GTQ1,000 (PPP\$123.5 to PPP\$246.9) and 90% of participants report monthly household income below GTQ2,000 (USD256 or PPP\$494).

#### 2.2.1. Convex Time Budget (CTB) Task

In the CTB, participants see a series of 24 questions, knowing in advance that one of them will be randomly selected to determine their earnings. Each question presents a choice among six options that involve a combination of money to be obtained at two different times:  $t$  and  $t + k$  days after the experiment<sup>10</sup>. Implicit in the options was a trade-off between receiving money earlier (at time  $t$ ) vs. delayed (time  $t + k$ ): each of these 24 questions allowed subjects to eliminate the delay of partial amounts of money, by “transforming” delayed money (at time  $t + k$ ) into early money

2 years before our field work began. This was, however, the closest LSM household data set available from INE.

<sup>8</sup>We dropped 4 men, 29 participants who showed no variation across all 24 choices, 36 potentially questionable observations (based on inconsistencies between the metadata in the image files and the session data), 1 individual for whom there is no consent form, and 2 individuals who refused to have their hands scanned.

<sup>9</sup>Guatemala's local currency is the Quetzal (GTQ). According to Guatemala's Central Bank, the average market exchange rate for the relevant period was GTQ7.8177 per USD. For 2013, World Development Indicators PPP conversion factor for private consumption was GTQ4.0499 per international dollar at purchasing power parity (PPPs).

<sup>10</sup>In the parlance of economics, each question presents six points uniformly distributed along an inter-temporal budget constraint regarding money at time  $t$  and at time  $t + k$ .

(at time  $t$ ) at a constant rate (marginal rate of transformation or MRT) that was weakly greater than one.

More specifically, in each question, one option is GTQ100 at time  $t + k$ , and GTQ0 at time  $t$  (not including the split payments participation fee). Each of the remaining five options involve shifting GTQ20 from time  $t + k$  to time  $t$  at a constant marginal transformation rate (MRT) or relative price, until only GTQ0 remains at time  $t + k$ . **Figure 1** illustrates the six options for a question (using  $MRT = 1.18$ ,  $t = 0$ , and  $k = 35$ ) as presented to participants<sup>11</sup>.

We used two values of  $t$ :  $t = \{0, 35\}$ . Each of these, were combined with two different delays:  $k = \{35, 63\}$ . The variation in the delay ( $k$ ) allows inference regarding discounting of future utility, and the variation in the early period ( $t = 0$  or  $t > 0$ ) allows inference regarding present-biasedness. For each of the four combinations of  $t$  and  $t + k$ , participants are presented with six questions, each with a different MRT. As previously mentioned, each question presented six options to choose from. These include two options “at the corners” (all the money delayed or all early) and four options of “interior choices” (involving combinations of both, delayed and early money). The availability of interior choices allows inference regarding preferences for consumption smoothing (Aycinena et al., 2017). **Table 1** summarizes the parameters used.

Payments were implemented via post-dated checks made out to the participant. As in Andreoni et al. (2015), to guarantee that the transaction costs associated with obtaining the two associated payments are the same, the GTQ50 participation payment is evenly divided between the payment at time  $t$  and the payment at time  $t + k$ <sup>12</sup>.

We vary three things between experimental sessions to control for order effects. First, for each pair of  $t$  and  $t + k$ , we varied the order in which participants see the associated six questions. In some sessions the relative price of money at time  $t$  is decreasing over the six questions, and in other sessions it is increasing. We refer to this as the decreasing opportunity cost (DOC) treatment. Second, in some sessions the options within a given question are ordered such that the amount at time  $t$  is monotonically decreasing, and in other sessions it is increasing. We refer to this as the decreasing soon amount (DSA) treatment. Third,

in some sessions, the GTQ25 payments for taking part in the experiment which was added to both the payment at time  $t$  and time  $t + k$  was explicitly shown in each question, and in others it was not. Note that this information was provided to participants prior to the CTB. This treatment simply varies the salience of the participation fee. We refer to this treatment as the included participation fee (IPF) treatment.

## 2.2.2. Sessions and Protocols

Experimental sessions took place in multipurpose rooms in the municipalities where subjects reside. We ran a total of 23 sessions with 16–24 subjects per session. Each session lasted between 3 and 4 h. All sessions were conducted by a session leader and a team of assistants.

Participants were asked to give informed consent upon arrival. After welcoming participants and giving a general introduction, the session leader projected at the front of the room and read aloud instructions for the CTB<sup>13</sup>. Afterwards, assistants ask each participant to answer several questions to ensure understanding. Then, assistants individually elicit answers for the first six questions (for  $t = 0$  and  $k = 35$ , with MRT varying across questions). As noted above, since many participants are illiterate it was important for assistants to provide individual support and show decision sheets (illustrating the available options with pictures of the relevant monetary amounts) for each question. Once all participants have answered the first six questions the session leader explains the changes for the following six questions and assistants individually elicit participant responses. This process continues until all 24 questions of the CTB have been answered.

Once the CTB task is complete, the session leader reads instructions for the remaining tasks and the experiment continues until all experimental tasks are completed. Participants then got a short break where beverages and snacks were provided. A bingo cage was used to determine the question from the CTB task that would be paid. Assistants individually interviewed each participant for a socioeconomic survey. Participants were then called individually to receive their checks and sign receipts. At this time they were asked if we could scan their hands. If they consented to this, their hands were then scanned.

## 2.3. Digit Ratio (2D:4D) Measures

We collected scanned images of the participants' hands<sup>14</sup>. After all images were collected, a research assistant randomly divided the images into five batches<sup>15</sup>. Each batch contained a total of 108 images, including 10 re-inserted images from other batches (so that each rater measured the 2D:4D ratio for a total of 50 subjects twice). These repeated measures serve as the basis for assessing the consistency of measurement for each rater.

<sup>11</sup>Since participants have low levels of literacy and numeracy, we presented all choices in the CTB using both numbers, and pictures of the associated quantities of money. Notice that each option specified the amount at time  $t$  and the amount at time  $t + k$ ; as well as the total amount. To further ensure that participants understood the task, assistants asked each participant the questions individually, resolved any questions as they arose and recorded the participant's decision.

<sup>12</sup>During the implementation there was a problem with the post-dated check payment mechanism, as some participants were able to cash checks earlier than the dates indicated on them. This would be problematic for our parameter estimates if participants anticipated that this was a possibility, as their effective MRT would then be equal to one in all cases. More specifically, if participants anticipated this, then we would expect that they would choose the option that would allow them to maximize the total amount of money over early and delayed payments. As long as the experimental MRT was greater than one, they would choose the option with the minimum early payment and maximum delayed payment. However, this is not what we observe. Reduced form regressions on early check cashing find no statistically significant correlation between cashing checks early and choosing options that concentrate amounts on delayed payments. Results are available upon request.

<sup>13</sup>The supplementary material shows the text of the instructions for both experimental tasks, translated from the original Spanish.




<sup>14</sup>Using a digital scanner is a common method for taking digit ratio measures that has been shown to be reliable (Kemper and Schwerdtfeger, 2009). An example of a scan can be seen in **Figure 2**.

<sup>15</sup>We split the measurement of images into batches to break the task into smaller sub-tasks, in an attempt to reduce the effects of fatigue or boredom for research assistants measuring the digit ratios.






# Pregunta #1




## Comprensión de instrucciones

1 Recibo hoy Q85  y además...  Dentro de 5 semanas: Q0   
TOTAL: Q85




---

2 Recibo hoy Q68  y además...  Dentro de 5 semanas: Q20   
TOTAL: Q88




---

3 Recibo hoy Q51  y además...  Dentro de 5 semanas: Q40   
TOTAL: Q91

---

4 Recibo hoy Q34  y además...  Dentro de 5 semanas: Q60   
TOTAL: Q94

---

5 Recibo hoy Q17  y además...  Dentro de 5 semanas: Q80   
TOTAL: Q97

---




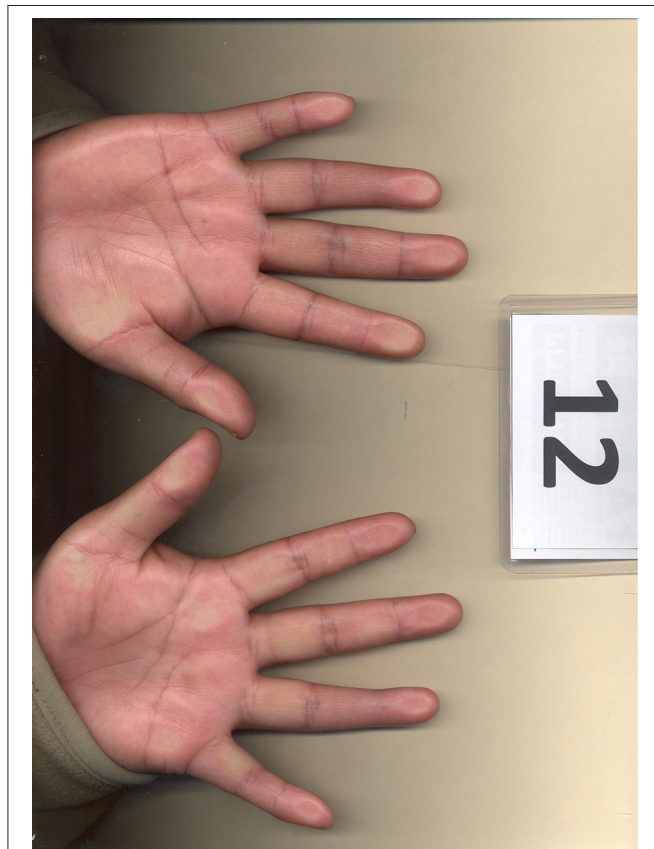
6 Recibo hoy Q0  y además...  Dentro de 5 semanas: Q100   
TOTAL: Q100

FIGURE 1 | Example CTB question, as presented to participants.

**TABLE 1** | Parameter summary for CTB task.

	1	2	3	4
<b>t</b>	<b>0</b>	<b>0</b>	<b>35</b>	<b>35</b>
<b>k</b>	<b>35</b>	<b>63</b>	<b>35</b>	<b>63</b>
MRT <sub>1</sub>	1.05	1.00	1.05	1.00
MRT <sub>2</sub>	1.11	1.05	1.11	1.05
MRT <sub>3</sub>	1.18	1.11	1.18	1.11
MRT <sub>4</sub>	1.25	1.33	1.25	1.33
MRT <sub>5</sub>	1.43	1.67	1.43	1.67
MRT <sub>6</sub>	1.82	2.22	1.82	2.22

**FIGURE 2** | Example of hand scan image used to measure 2D:4D.

Eight raters were instructed and received guidance on using the Autometric software (DeBruine, 2004) designed to measure digit ratios. They then independently measured both hands for each image in all five batches. The order in which each rater received the five batches was randomized.

Thus, we collected 8 independent 2D:4D measures for each hand of all participants. In addition, we had 50 randomly selected images measured twice by each rater. The repeated measures for the 50 randomly selected images allowed us to measure intra-rater consistency of 2D:4D measures. We drop the measures for three raters with an intraclass correlation coefficient (ICC)

**TABLE 2** | Within-rater consistency using repeated measures (for both hands).

	Left hand			Right hand		
	ICC	Rho	p-value	ICC	Rho	p-value
Rater 1	0.977	0.975	0.213	0.945	0.937	0.792
Rater 2	0.925	0.913	0.398	0.962	0.946	0.301
Rater 3	0.940	0.945	0.181	0.945	0.921	0.952
Rater 4	0.908	0.904	0.506	0.915	0.903	0.417
Rater 5	0.876	0.855	0.569	0.863	0.886	0.334

*Within-rater analysis of repeated measures. Table contains intra-class correlation coefficient (ICC), Spearman's rho correlation coefficients (Rho), and p-value for two-sided paired t-test for equality of means between raters measures for left and right hands, correspondingly.*

< 0.85. This leaves us with five high quality measures for each hand of each participant. The ICC for the repeated measures of the remaining raters range from 0.8625 to 0.9772 and the Spearman  $\rho$  range from 0.8548 to 0.9754. In no case are there statistically significant differences in the means of the repeated measures. **Table 2** shows measures of intra-rater consistency.

**Table 3** displays the between-rater correlation coefficients. Between rater correlation coefficients range from 0.8663 to 0.9392 for the right hand measures, and from 0.7546 to 0.9668 for the left hand.

We take the average across the five measures<sup>16</sup>. **Table 4** shows the summary statistics for the 2D:4D measures. The digit ratios for our sample are lower than those typically found in the literature. For the right hand, mean 2D:4D is 0.9322 (with a standard deviation of 0.0315); for the left hand the mean is 0.9337 (with a standard deviation of 0.0321)<sup>17</sup>. No statistical significant difference is found in variance or mean between hands. **Figure 3** illustrates the distribution of the average of all five measures for both hands.

Thus, our final 2D:4D data consists of the average of five (high quality) independent measures for the 419 final sample subjects.

### 3. RESULTS

#### 3.1. Plan of Analysis

Given the so called “replicability crisis” in scientific findings (see e.g., Ioannidis, 2005; Button et al., 2013; Aarts et al., 2015; Camerer et al., 2016), we attempted to limit the degrees of freedom available to us as researchers<sup>18</sup>.

<sup>16</sup>Voracek et al. (2007) suggests using the average of multiple independent measures by different raters.

<sup>17</sup>Dropping the highest and lowest (to mitigate the potential impact of outliers) and taking the average of three intermediate measures, we would have for the right hand a mean of 0.9323 (with a standard deviation of 0.0316), and for the left hand a mean of 0.9330 (with a standard deviation of 0.0322). Other samples tend to report higher 2D:4D measures; for instance Branas-Garza et al. (in press) reports mean of 0.9734 and 0.9775 for female left and right hands. Aycinena et al. (2014) reports a mean of 0.957 and 0.954 for female left and right hands. This difference might be due to the different ethnic compositions of the different samples.

<sup>18</sup>Studies may give researchers many degrees of freedom, even without explicit fishing (Gelman and Loken, 2014). In 2D:4D research this problem is not absent; if anything it may be exacerbated as there is no consensus regarding which hand

**TABLE 3** | Correlation coefficients for between-rater measures for left-hand and right-hand measures.

	Left hand					Right hand				
	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5
Rater 1	1.000					1.000				
Rater 2	0.897	1.000				0.880	1.000			
Rater 3	0.906	0.928	1.000			0.967	0.923	1.000		
Rater 4	0.939	0.858	0.899	1.000		0.956	0.872	0.930	1.000	
Rater 5	0.866	0.872	0.880	0.882	1.000	0.803	0.755	0.820	0.773	1.000

Table contains Spearman rho correlation coefficients between raters measures for left and right hands, respectively.

**TABLE 4** | Summary statistics of the 2D:4D ratio.

	Left hand	Right hand
Mean	0.933	0.931
Median	0.931	0.930
Standard deviation	0.032	0.032
Min	0.8492	0.8508
Max	1.1396	1.1006

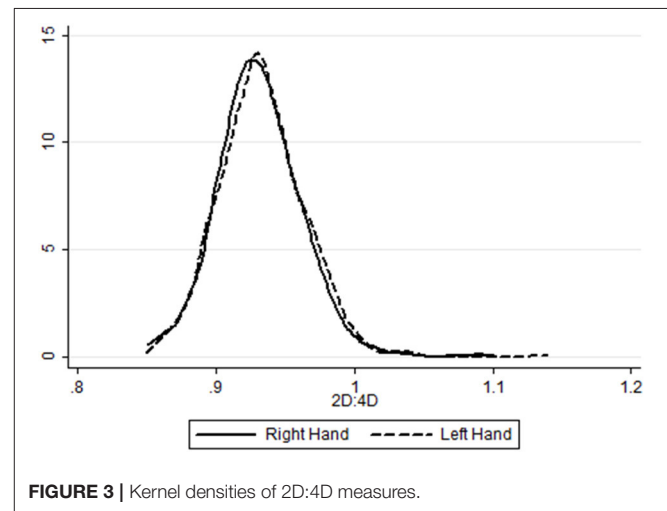
To limit the degrees available to us, we partnered with Anna Dreber to prepare an analysis plan<sup>19</sup>. In the plan, we specify that our main method of analysis will rely on the interval censored Tobit model to structurally estimate time-preference primitives, which allow discounting to vary with 2D:4D. Specifically, we estimate discounting ( $\delta$ ) as a linear function of 2D:4D (among other parameters).

In the analysis plan we also specify three robustness tests. First, we test robustness to changes in the background parameters, since (Andreoni et al., 2015) and (Aycinena et al., 2017) show that the structural estimates may be sensitive to whether or not the participation fee (among other background parameters) is included in the analysis. Thus we perform two robustness checks which modify assumptions about the background parameters.

Second, we examine whether the results are robust at the individual level. To do so, we structurally estimate time-preference primitives at the individual level, and test whether the individual level estimates for  $\delta$  are correlated with the individual 2D:4D measures. Finally, our third robustness check tests whether our results depend on the method of structural estimation. To do so, we drop the structural estimation approach and test whether 2D:4D measures predict choices of more delayed money using reduced form analysis.

to use, which measures (mean, median, etc.) to use, or the correct specification (linear, quadratic, etc.) to employ.

<sup>19</sup>We thank Anna Dreber for her time helping us prepare the analysis plan while she was blind to the data. The plan is posted at the Open Science Framework web platform: <https://osf.io/ey67f/register/564d31db8c5e4a7c9694b2be>. It should be noted that, technically, this is not a pre-analysis plan, since we developed it after data collection was finished. Nevertheless, we feel that by developing it jointly with a credible third party, it helps to reduce the degrees of freedom of our analysis.

**FIGURE 3** | Kernel densities of 2D:4D measures.

### 3.2. Theoretical and Econometric Framework

To analyze choices, we rely on a model inter-temporal preferences that assumes a time-separable quasi-hyperbolic utility function with constant relative risk aversion. Specifically, denoting the amount of money received by subject  $i$  at time  $t$  ( $t + k$ ) as  $x_{it}$  ( $x_{it+k}$ ), we assume that the following utility function underlies observed choices:

$$U(x_{it}, x_{it+k}) = \begin{cases} x_{it}^\alpha + \beta \delta^k x_{it+k}^\alpha & \text{if } t = 0 \\ x_{it}^\alpha + \delta^k x_{it+k}^\alpha & \text{if } t > 0. \end{cases} \quad (1)$$

Our framework includes three parameters that affect time-preferences: discounting ( $\delta$ ), present biasedness ( $\beta$ ) and utility curvature ( $\alpha$ ). The discount factor,  $\delta$ , captures the degree to which an individual discounts delays in consumption. A  $\delta = 1$  implies that individuals are so patient, that all else equal, they are indifferent to delays in consumption. The lower the value of  $\delta$  ( $\delta < 1$ ) implies higher discounting of delaying consumption, that is, less patience. Present biasedness,  $\beta < 1$ , captures how much (more) an individual discounts delaying consumption relative to immediate consumption. Note that  $\beta = 1$  implies a standard discounting model with no present biasedness. Finally  $\alpha$ , utility curvature, underlies preferences to inter-temporally smooth consumption. An  $\alpha = 1$  implies that consumption is

perfectly substitutable across time, thus no preference to smooth consumption in time. The lower the value of  $\alpha$  ( $\alpha < 1$ ) the higher the preference to smooth consumption. That is, all else equal, the lower  $\alpha$ , the more an individual is willing to sacrifice in order to attain a consumption profile that is smoother across time.

Notice that these three parameters are interrelated for time-preferences. That is, it is possible to observe the same choice by two individuals with very different levels of patience (different  $\delta$ 's) if there utility curvature ( $\alpha$ ) and/or present-biasedness ( $\beta$ ) also differ. Given this, it is important to estimate these three parameters jointly (see e.g., Andersen et al., 2008; Andreoni and Sprenger, 2012a).

### 3.3. Main Analysis: Structural Estimation

In our main analysis we employ interval censored tobit regressions<sup>20</sup>. This procedure jointly estimates three parameters:  $\alpha$ ,  $\beta$ , and  $\delta$ .

The parameter  $\delta$  is the aggregate measure of the time preferences in the population (see Andreoni et al., 2015 for a detailed description of the model and the estimation techniques). To test our hypothesis, we allow  $\delta$  to be a function of the 2D:4D ratio. As specified in our analysis plan, the functional form we assume is as follow:

$$\delta_i = \rho_0 + \rho_1 \cdot 2D:4D_i + \rho_2 \cdot DSA_i + \rho_3 \cdot DOC_i + \rho_4 \cdot IPF_i \quad (2)$$

<sup>20</sup>For the structural estimation, the covariance matrix was estimated using sandwich estimator for robust standard errors. See Aycinena et al. (2014) for a detailed description of the estimation method.

where experimental treatments [included participation fee (IPF) explicitly treatment, decreasing opportunity cost (DOC) treatment, decreasing soon amount (DSA) treatment] are included to control for differences in how the CTB task was presented to subjects.

The first two columns of **Table 5**, estimated separately, present results of the parameter estimates for the left and right hands of participants. The value for the parameter  $\alpha$  shows a strong preference for smoothing consumption over time. The  $\beta$  parameter is higher than one, thus it shows no evidence of present-biasedness<sup>21</sup>. Next we present results in which the parameter of interest,  $\delta$ , is a function of 2D:4D and treatment controls.

For the parametrization of the discount factor ( $\delta$ ), we see that the coefficient on 2D:4D is negative ( $-11.899$  for the left hand and  $-15.959$  for the right hand) and statistically significant for both hands at the 0.001 level. This implies that lower 2D:4D is correlated with a higher discount factor. That is, individuals with lower 2D:4D (a marker for higher exposure to testosterone *in utero*) make more patient choices.

Following our analysis plan, we also explore whether there is evidence of a non-linear effect of 2D:4D on discounting. Specifically, we examine whether there is a quadratic relationship by adding  $2D:4D^2$  as an explanatory variable. Under this specification (not reported but available from the authors upon request), we find that both the linear and squared coefficients

<sup>21</sup>Balakrishnan et al. (2017) suggests that present biasedness is only existent when payments are "truly immediate."

**TABLE 5 |** Parameter estimates.

	Main estimates		Robustness check 1.1		Robustness check 1.2	
	Left hand	Right hand	Left hand	Right hand	Left hand	Right hand
$\alpha$	0.540*** (0.017)	0.540*** (0.017)	0.727*** (0.010)	0.727*** (0.010)	0.877*** (0.005)	0.877*** (0.005)
$\beta$	1.105*** (0.017)	1.105*** (0.017)	1.096*** (0.016)	1.096*** (0.016)	1.111*** (0.020)	1.111*** (0.020)
$\rho_0$ (Constant)	9.778*** (1.630)	13.524*** (1.687)	9.150*** (1.530)	12.645*** (1.582)	11.665*** (2.039)	16.021*** (2.084)
$\rho_1$ (2D : 4D <sub>i</sub> )	-11.899*** (1.738)	-15.959*** (1.800)	-12.688*** (1.632)	-14.974*** (1.690)	-14.684*** (2.181)	-19.404*** (2.235)
$\rho_2$ (DSA <sub>i</sub> )	0.391*** (0.111)	0.361** (0.110)	0.359*** (0.104)	0.331** (0.104)	0.411*** (0.131)	0.375*** (0.130)
$\rho_3$ (DOC <sub>i</sub> )	0.159 (0.111)	0.186+ (0.111)	0.146 (0.105)	0.171 (0.104)	0.333** (0.132)	0.361*** (0.132)
$\rho_4$ (IPF <sub>i</sub> )	1.073*** (0.114)	1.094*** (0.114)	1.006*** (0.107)	1.026*** (0.107)	1.468*** (0.133)	1.495*** (0.134)
$\sigma$	1.521*** (0.040)	1.518*** (0.040)	2.411*** (0.065)	2.406*** (0.064)	6.639*** (0.117)	6.621*** (0.117)
Log-likelihood	16,235.6	16,226.1	16,204.9	16,186.7	18,480.4	18,461.6
BIC	-32,351.4	-32,332.5	-32,290.1	-32,253.7	-36,841.1	-36,803.5

Robust standard errors are reported in parenthesis.

+ $p < 0.1$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .



are negative, but none are statistically significant at conventional levels.

## 4. ROBUSTNESS CHECKS

### 4.1. Robustness to Changes in Background Parameters

It should be noted that the previous parameter estimates may be sensitive to whether or not the participation fee, among other background parameters, is included (e.g., Andreoni et al., 2015; Aycinena et al., 2017). Since all subjects received the participation fee, we included it (Q50, split evenly across two time periods) as a background parameter in the estimates reported in the previous section. For our first set of robustness checks, we test how sensitive our results are to modifying the background parameters.

We examine two alternative specifications of the background parameters. Our first examination involves dropping the participation fee from our analysis, so that  $x_{it}$  and  $x_{it+k}$  do not include the participation fee in our econometric analysis. We report the results for left and right hand in columns 3 and 4 of **Table 5** (under the heading “Robustness check 1.1”). For the second, we estimate the parameters with the explicit option displayed to participants, according to the IPF treatment<sup>22</sup>. The last two columns of **Table 5** (under the heading “Robustness check 1.2”) report the results of such estimates.

As the table shows, estimates of  $\alpha$  seem to be quite sensitive to the background parameters used. The estimate of  $\beta$  on the other hand, seems quite robust. Regarding our coefficient of interest, although not quite as sensitive as  $\alpha$ ,  $\delta$  does vary with the background parameters employed. Although the impact is not obvious due to the five parameters involved in the estimation of  $\delta$ , the mean value of  $\delta$  ranges from 0.6 to 0.85.

<sup>22</sup>Recall that in this treatment, some subjects were shown amounts in the CTB that explicitly included the participation fee, while others were shown amounts that did not include the participation fee.

Nevertheless, the point to note is that the coefficient on 2D:4D is negative and statistically significant ( $p < 0.001$ ) for both hands across all specifications. Thus, the relationship between 2D:4D and patience reported in the previous section seems robust to the specification of the background parameters.

### 4.2. Individual Level Estimates

The second robustness check involves attempting to estimate time preference primitives at the individual level. We use the interval censored Tobit model with 24 observations per individual (one observation for each of the 24 questions of the CTB) and attempt to jointly estimate  $\alpha$ ,  $\beta$ , and  $\delta$ .

Unfortunately, our individual estimates are very imprecise. For our parameter of interest,  $\delta$ , values range from 0 to  $1.4e^{191}$ , and the distribution is very skewed with a mean of  $3.4e^{188}$ , and for over half of the observations the estimate of  $\delta < 0.0001$ .<sup>23</sup> This lack of precision is not surprising given that for each individual, we have 24 observations to estimate eight parameters<sup>24</sup>. To try to overcome this problem, we restrict our analysis to individuals with an (arbitrarily defined) sensible  $\delta$  parameter: individuals with  $0 < \delta < 2$ . This reduces drastically our subsample to 168 individuals.

We use the parameter estimates for the 168 individuals of our restricted sub-sample as a dependent variable and estimate the following reduced form model (separately for left and right hands) using OLS:

$$\delta_i = \rho_0 + \rho_1 2D:4D_i + \rho_2 \cdot DSA_i + \rho_3 \cdot DOC_i + \rho_4 \cdot IPF_i + \epsilon_i \quad (3)$$

We present results in the first two columns of **Table 6**. For the sake of brevity, we only present the results for 2D:4D (point estimate of  $\rho_1$  and its standard error) and the adjusted  $R^2$ . The top row presents the 2D:4D coefficient for the left hand and the

<sup>23</sup>The 25th percentile is zero, with a mean of  $3.4e^{188}$  and median of .00001.

<sup>24</sup>The three parameters which measure preference primitives ( $\alpha$ ,  $\beta$ , and  $\delta$ ), in addition to the auxiliary parameters ( $\sigma$ , and the five cut-offs  $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$ ).

**TABLE 6 |** Reduced form analysis robustness checks.

	Robustness check 2		Robustness check 3.1		Robustness check 3.2	
	(1)	(2)	(1)	(2)	(1)	(2)
Left hand	−0.039 (0.332)	0.390 (0.332)	−2.608** (1.515)	−1.767+ (1.358)	69.401** (25.154)	44.125+ (22.902)
Adjusted $R^2$ /Pseudo $R^2$	−0.019	−0.005	0.021	0.041	0.217	0.265
Right hand	−0.305 (0.338)	−0.140 (0.382)	−3.396** (1.609)	−2.173* (1.423)	94.875** (27.502)	57.113* (24.925)
Adjusted $R^2$ /Pseudo $R^2$	−0.017	−0.008	0.022	0.042	0.221	0.266
Observations	168	168	10,053	10,053	10,053	10,053
Session fixed effects?	No	Yes	No	Yes	No	Yes
Surveyor fixed effects?	No	Yes	No	Yes	No	Yes

Point estimates for 2D:4D coefficient of the robustness checks. Robust standard errors are reported in parenthesis (clustered at the individual level for robustness checks 3.1 and 3.2). Robustness checks 2 and 3.2 are estimated using OLS; adjusted  $R^2$  for each hand is reported below standard errors. Robustness check 3.1 is estimated using ordered probits; Pseudo  $R^2$  is reported below the standard errors.

+ $p < 0.10$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

bottom row for the right hand, each estimated independently. None of the coefficients are statistically significant. The signs of the coefficients are consistent with our main analysis, except for the left hand when we include session and surveyor fixed effects. The adjusted  $R^2$  is negative for all four specifications of robustness check two, which indicates that the model is a very poor fit for the data<sup>25</sup>. Overall, this suggests that this approach was not successful in allowing us to test the robustness of the results<sup>26</sup>.

### 4.3. Reduced form Analysis

In our third robustness check, we bypass the structural estimation and directly examine choices with a reduced form approach. The independent variables we employ include our variable of interest (2D:4D), the marginal rate of transformation for the question ( $MRT_j$ ), the time when the early amount is to be received ( $t_j$ ), the delay ( $k_j$ ), and controls for our three treatment variables ( $DSA$ ,  $DOC$ ,  $IPF$ ). Since we have multiple observations per individual, we cluster standard errors at the individual level. In all of our reduced form analysis, we estimate the model for both right and left hand 2D:4D.

Since participants could choose among six discrete ordered options ( $Y_{ij} \in [1, 2, \dots, 6]$ ), we first examine this using an ordered probit model. Choosing option 1 maximizes the amount received in the early payment; choosing option 6 maximizes the amount received in the delayed payment. Thus, all else equal, a more impatient individual (i.e., with a lower  $\delta$ ) will tend to select lower options than a more patient individual (someone with a higher  $\delta$ ). If our results are robust, we would again expect a negative coefficient for 2D:4D.

We present the results (of our coefficients of interest) in the middle columns (Robustness check 3.1) of **Table 6**. Column (1) presents the coefficients for the model described above. We find that for both hands, coefficients are negative and statistically significant ( $p < 0.01$ ). Again, this supports the findings from the main estimates that lower 2D:4D individuals make more patient choices. Column (2) adds session and surveyor fixed effects. Under this specification, the coefficient for the left hand is no longer statistically significant at conventional levels ( $p < 0.1$ ).

For our second reduced form approach, we use ordinary least squares and the dependent variable is the early amount chosen ( $x_{ijt}$ ) by individual  $i$  in question  $j$ . We use the same independent variables, with our focus again being on the coefficient of the 2D:4D<sup>27</sup>. Notice the the higher the early amount chosen, the more impatient the individual (given the tradeoffs between early

and delayed amounts). Thus, in this approach, we expect a positive correlation between 2D:4D and our dependent variable.

Results for our coefficients of interest are reported in the last two columns (Robustness check 3.2) of **Table 6**. For the first specification (Column 1), the coefficients for both hands are positive and statistically significant ( $p < 0.01$ ). In column (2) we add session and surveyor fixed effects. In this case, the coefficient for the left hand is no longer statistically significant at conventional levels ( $p < 0.1$ ).

Again following our analysis plan, we perform an exploratory analysis of whether the relationship between 2D:4D and discounting is non-linear by adding 2D:4D<sup>2</sup> as an explanatory variable. We do not find any robust evidence for a non-linear relationship between 2D:4D and discounting. Coefficients are not statistically significant either in the ordered probit or the OLS model.

To summarize this last robustness test, we find that results do not depend crucially on the assumption and methods of the structural estimation. Using reduced form analysis, we find evidence that 2D:4D is negatively related to patience for both hands in the first specification, and for the right hand in the second.

## 5. DISCUSSION

In this study we investigate the impact of 2D:4D, as a proxy for pre-natal exposure to testosterone, on discounting. We use a large sample ( $N = 419$ ) of low income females from a wide age range. We rely on 24 choices per individual using the convex-time budget task with large stakes, and the average of five independent measures of 2D:4D.

We follow an analysis plan and jointly estimate time preference parameters and the curvature of the utility function, and allow the discount parameter ( $\delta$ ) to vary with 2D:4D. We find that, for both hands, 2D:4D is negatively correlated with discount factor ( $p < 0.001$ ). That is, we find that lower 2D:4D generates more patient choices.

We stick to our analysis plan and perform three robustness tests. First, we examine robustness of our results to varying background parameters; and find that our results are robust. Next, we attempt to estimate time-preference parameters at the individual level and correlate them with 2D:4D using reduced form models. Results of this second robustness check are mixed, since our individual level parameter estimates are very noisy. Our third robustness test involves replacing the parametric estimation method with a direct reduced form analysis. For each hand we run two tests using ordered probits and two using OLS. Given the criteria pre-specified in our analysis plan, our results are mixed. We pre-defined that we would consider a result to be significant if  $p - \text{value} < 0.05$  for both hands<sup>28</sup>. Specification (1) of robustness checks 3.1 and 3.2 satisfies this criteria. However, for specification (2), only the result for the right hand is significant at  $p < 0.05$ .

<sup>25</sup>It should be noted that this is not driven by the 2D:4D measure, as a model that excludes 2D:4D as an explanatory variable also has negative adjusted  $R^2$  of similar magnitude. More importantly, the partial  $R^2$  (or coefficient of partial determination) of the 2D:4D coefficient is always positive, suggesting that if anything, it helps the model fit of the data (although clearly not enough).

<sup>26</sup>Although our attempt to estimate parameters at the individual level failed, we believe important to stick to our analysis plan and report the attempt despite its failure.

<sup>27</sup>It should be noted that the analysis plan specified that the dependent variable for this approach would be the delayed amount chosen. That is a mistake, since the delayed amount is a linear transformation of the dependent variable used in the first approach (Robustness check 3.1). Results are qualitatively and statistically the same if we use delayed amount as our dependent variable.

<sup>28</sup>The analysis plan states: "Since we will look at the correlation between discounting and 2D:4D for both hands, there is concern about multiple testing. We will consider a result to be significant if the  $p$ -values corresponding to the coefficients of 2D:4D for both hands are  $< 0.05$ ."

Our result are in contrast to those of Lucas and Koff (2010), which reports that lower digit ratios are correlated with greater discounting among women. Our findings also differ from those of Drichoutis and Nayga (2015), which report no effect of digit ratio on (risk or) time preferences. These differences might stem from different samples, methods or protocols used.

However, our finding that lower 2D:4D leads to more patience is consistent with the combined results from other studies that relate 2D:4D, cognitive ability and patience. Bosch-Domènech et al. (2014) find that lower 2D:4D is associated with higher scores in the cognitive reflection test (CRT), and Frederick (2005) finds that higher CRT scores correlate with more patience (in hypothetical choices) and with higher cognitive abilities<sup>29</sup>. These results are also consistent with other studies which also find that higher cognitive ability is associated with more patience (Shamosh et al., 2008; Burks et al., 2009; Dohmen et al., 2010; Benjamin et al., 2013).

Why should we care about the relationship between 2D:4D and discounting? Time preferences, and discounting in particular, play an important role in human decision making over countless domains (health, human capital accumulation, labor supply, income, etc.) with important welfare consequences. Our results are thus important, as they point to a potential biological underpinning of time preferences.

On a more methodological note, this finding suggests an exogenous determinant of individual time preferences. This may have broad implications for economic studies on the causal effect of time preferences on different economic behavior. That is, our results could be an important advance in identification strategies for researchers seeking to identify causal relationships between time preferences and other economic behavior, by using 2D:4D as an exogenous instrument.

This study has several peculiarities. First, our sample also differs from typical 2D:4D samples, as we do not rely on a WEIRD (Western Educated Industrialized Rich Democratic) population sample (Henrich et al., 2010a,b). Rather, our sample is particular on different margins: low income non-Caucasian females enrolled in a conditional cash transfer program. In addition, the 2D:4D measures of our sample are lower than those typically found in the literature. As with most findings, our results should be replicated to improve our confidence in the findings (Maniadiis et al., 2017). In particular, this work should

be replicated with samples of men. One limitation of this study is that our sample is exclusively female. As Frederick (2005) noted, there is a higher correlation of time preferences with CRT for females than males.

## AUTHOR CONTRIBUTIONS

DA coordinated the study, designed the experiment, coordinated 2D:4D measurements, conducted statistical analysis and drafted the manuscript. LR coordinated the study, designed the experiment, conducted statistical analysis and drafted the manuscript. All authors gave final approval for publication.

## FUNDING

DA gratefully acknowledges financial support from Fundación Capital.

## ACKNOWLEDGMENTS

Special thanks to Anna Dreber for the encouragement to complete this project and for working with us (blind to the data) on the statistical analysis plan. Special thanks also to Szabolcs Blazsek for sharing code and helping with the estimating procedures. Betzy Sandoval provided excellent research assistance, including field supervision, data handling and project involvement. Jorge Chang, Betzy Sandoval, Ivone Gadala-María, Mario Sandari Gomez, Fernando Chang, Max Pfeifer, Josue Perez, and Rodrigo Gonzalez assisted taking measuring 2D:4D's. We are also grateful to Pablo Pastor, Alvaro Garcia, Raul Zurita, Raul E. Rueda, Arturo Melville, and Amy Benítez. This project would not have been possible without the collaboration of Fundación Capital and the Ministerio de Desarrollo Social.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnbeh.2017.00257/full#supplementary-material>

## REFERENCES

- Aarts, A., Anderson, J., Anderson, C., Attridge, P., Attwood, A., Fedor, A., et al. (2015). Estimating the reproducibility of psychological science. *Science* 349, 1–8. doi: 10.1126/science.aac471
- Andersen, S., Harrison, G., Lau, M., and Rutström, E. (2008). Eliciting risk and time preferences. *Econometrica* 76, 583–618. doi: 10.1111/j.1468-0262.2008.00848.x
- Andreoni, J., Kuhn, M., and Sprenger, C. (2015). Measuring time preferences: a comparison of experimental methods. *J. Econ. Behav. Organ.* 116, 451–464. doi: 10.1016/j.jebo.2015.05.018

<sup>29</sup>It should be noted that in Bosch-Domènech et al. (2014), 2D:4D still predicts CRT scores even after controlling for (hypothetical) patience).

- Andreoni, J., and Sprenger, C. (2012a). Estimating time preferences from convex budgets. *Am. Econ. Rev.* 102, 3333–3356. doi: 10.1257/aer.102.7.3333
- Andreoni, J., and Sprenger, C. (2012b). Risk preferences are not time preferences. *Am. Econ. Rev.* 102, 3357–3376. doi: 10.1257/aer.102.7.3357
- Apicella, C. L., Carré, J. M., and Dreber, A. (2015). Testosterone and economic risk taking: A review. *Adapt. Hum. Behav. Physiol.* 1, 358–385. doi: 10.1007/s40750-014-0020-2
- Aycinena, D., Baltaduonis, R., and Rentschler, L. (2014). Risk preferences and prenatal exposure to sex hormones for ladinos. *PLoS ONE* 9:e103332. doi: 10.1371/journal.pone.0103332
- Aycinena, D., Blazsek, S., Rentschler, L., and Sprenger, C. (2017). Intertemporal choice experiments and large stakes behavior. *Working Paper*.
- Aycinena, D., Sandoval, B., Blazsek, S., and Rentschler, L. (2015). Smoothing, discounting and demand for intra-household control for recipients of conditional cash transfer. *Working Paper*.

- Balakrishnan, U., Haushofer, J., and Jakiela, P. (2017). How soon is now? Evidence of present bias from convex time budget experiments. *NBER Working Paper*.
- Benjamin, D., Brown, S., and Shapiro, J. (2013). Who is “behavioral”? Cognitive ability and anomalous preferences. *J. Eur. Econ. Assoc.* 11, 1231–1255. doi: 10.2139/ssrn.675264
- Bosch-Domènech, A., Brañas-Garza, P., and Espín, A. (2014). Can exposure to prenatal sex hormones (2D: 4D) predict cognitive reflection? *Psychoneuroendocrinology* 43, 1–10. doi: 10.1016/j.psyneuen.2014.01.023
- Branas-Garza, P., Galizzi, 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.*
- Branas-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
- Burks, S., Carpenter, J., Goette, L., and Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7745–7750. doi: 10.1073/pnas.0812360106
- Button, K., Ioannidis, J., Mokrysz, C., Nosek, B., Flint, J., Robinson, E., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376. doi: 10.1038/nrn3475
- Cadena, B., and Keys, B. (2015). Human capital and the lifetime costs of impatience. *Am. Econ. J.* 7, 126–153. doi: 10.1257/pol.20130081
- Camerer, C., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science* 351, 1433–1436. doi: 10.1126/science.aaf0918
- Carvalho, L., Meier, S., and Wang, S. (2016). Poverty and economic decision-making: evidence from changes in financial resources at payday. *Am. Econ. Rev.* 106, 260–284. doi: 10.1257/aer.20140481
- Cassar, A., Healy, A., and Von Kessler, C. (2017). Trust, risk, and time preferences after a natural disaster: experimental evidence from thailand. *World Dev.* 94, 90–105. doi: 10.1016/j.worlddev.2016.12.042
- Castillo, M., Ferraro, P., Jordan, J., and Petrie, R. (2011). The today and tomorrow of kids: time preferences and educational outcomes of children. *J. Publ. Econ.* 95, 1377–1385. doi: 10.1016/j.jpubeco.2011.07.009
- Chabris, C., Laibson, D., Morris, C., Schuldt, J., and Taubinsky, D. (2008). Individual laboratory-measured discount rates predict field behavior. *J. Risk Uncertainty* 37:237. doi: 10.3386/w14270
- Clot, S., and Stanton, C. (2014). Present bias predicts participation in payments for environmental services: evidence from a behavioral experiment in Uganda. *Ecol. Econ.* 108, 162–170. doi: 10.1016/j.ecolecon.2014.07.026
- Curry, O., Price, M., and Price, J. (2008). Patience is a virtue: cooperative people have lower discount rates. *Pers. Individ. Diff.* 44, 780–785. doi: 10.1016/j.paid.2007.09.023
- DeBruine, L. M. (2004). *Autometric Software for Measurement of 2D: 4D Ratios*. Available online at: <http://www.facelab.org/debruine/Programs/autometric.php> (Accessed August 13, 2011).
- Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *Am. Econ. Rev.* 100, 1238–1260. doi: 10.1257/aer.100.3.1238
- Drichoutis, A., and Nayga, R. (2015). Do risk and time preferences have biological roots? *Southern Econ. J.* 82, 235–256. doi: 10.4284/0038-4038-2013.246
- Espín, A., Brañas-Garza, P., Herrmann, B., and Gamella, J. (2012). Patient and impatient punishers of free-riders. *Proc. R. Soc. B Biol. Sci.* 279:4923. doi: 10.1098/rspb.2012.2043
- Espín, A., Exadaktylos, F., Herrmann, B., and Brañas-Garza, P. (2015). Short- and long-run goals in ultimatum bargaining: impatience predicts spite-based behavior. *Front. Behav. Neurosci.* 9:214. doi: 10.3389/fnbeh.2015.00214
- Frederick, S. (2005). Cognitive reflection and decision making. *J. Econ. Perspect.* 19, 25–42. doi: 10.1257/089533005775196732
- Galizzi, 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
- Garbarino, E., Slonim, E., and Sydnor, J. (2011). Digit ratios (2D: 4D) as predictors of risky decision making for both sexes. *J. Risk Uncertainty* 42, 1–26. doi: 10.1007/s11166-010-9109-6
- Gelman, A., and Loken, E. (2014). The statistical crisis in science. *Am. Sci.* 102:460. doi: 10.1511/2014.111.460
- Golsteyn, B., Grönqvist, H., and Lindahl, L. (2014). Adolescent time preferences predict lifetime outcomes. *Econ. J.* 124, F739–F761. doi: 10.1111/ecoj.12095
- Goy, R., and McEwen, B. (1980). *Sexual Differentiation of the Brain*. Cambridge, MA: MIT Press.
- Harrison, G., Lau, M., and Williams, M. (2002). Estimating individual discount rates in denmark: a field experiment. *Am. Econ. Rev.* 92, 1606–1617. doi: 10.1257/000282802762024674
- Henrich, J., Heine, S., and Norenzayan, A. (2010a). The weirdest people in the world? *Behav. Brain Sci.* 33, 61–83. doi: 10.1017/S0140525X0999152X.
- Henrich, J., Heine, S. J., and Norenzayan, A. (2010b). Most people are not weird. *Nature* 466:29. doi: 10.1038/466029a
- Holt, C., and Laury, S. (2002). Risk aversion and incentive effects. *Am. Econ. Rev.* 92, 1644–1655. doi: 10.1257/000282802762024700
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med.* 2:e124. doi: 10.1371/journal.pmed.0020124
- Kemper, C., 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
- Kirby, K. (2009). One-year temporal stability of delay-discount rates. *Psychon. Bull. Rev.* 16, 457–462. doi: 10.3758/PBR.16.3.457
- Kirby, K., Godoy, R., Reyes-García, V., Byron, E., Apaza, L., Leonard, W., et al. (2002). Correlates of delay-discount rates: evidence from tsimane’amerindians of the bolivian rain forest. *J. Econ. Psychol.* 23, 291–316. doi: 10.1016/S0167-4870(02)00078-8
- Kirby, K., Petry, N., and Bickel, W. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *J. Exp. Psychol.* 128, 78–87.
- Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Q. J. Econ.* 112, 443–477.
- Lawrance, E. (1991). Poverty and the rate of time preference: evidence from panel data. *J. Polit. Econ.* 99, 54–77.
- Lucas, M., and Koff, E. (2010). Delay discounting is associated with the 2D: 4D ratio in women but not men. *Pers. Individ. Diff.* 48, 182–186. doi: 10.1016/j.paid.2009.10.002
- Lutchmaya, S., Baron-Cohen, S., Raggatt, P., Knickmeyer, R., and Manning, J. (2004). 2nd to 4th digit ratios, fetal testosterone and estradiol. *Early Hum. Dev.* 77, 23–28. doi: 10.1016/j.earlhumdev.2003.12.002
- Maniadis, Z., Tufano, F., and List, J. (2017). To replicate or not to replicate? Exploring reproducibility in economics through the lens of a model and a pilot study. *Econ. J.* 127, F209–F235. doi: 10.1111/ecoj.12527
- Manning, J. (2002). *Digit Ratio: A Pointer to Fertility, Behavior, and Health*. New Brunswick, NJ: Rutgers University Press.
- Manning, J., Baron-Cohen, S., Wheelwright, S., and Sanders, G. (2001). The 2nd to 4th digit ratio and autism. *Dev. Med. Child Neurol.* 43, 160–164. doi: 10.1111/j.1469-8749.2001.tb00181.x
- Meier, S., and Sprenger, C. (2010). Present-biased preferences and credit card borrowing. *Am. Econ. J.* 2, 193–210. doi: 10.1257/app.2.1.193
- Meier, S., and Sprenger, C. (2015). Temporal stability of time preferences. *Rev. Econ. Stat.* 97, 273–286. doi: 10.1162/REST\_a\_00433
- Millet, K. (2011). An interactionist perspective on the relation between 2D: 4D and behavior: an overview of (moderated) relationships between 2D: 4D and economic decision making. *Pers. Individ. Diff.* 51, 397–401. doi: 10.1016/j.paid.2010.04.005
- Mitchell, S. (1999). Measures of impulsivity in cigarette smokers and non-smokers. *Psychopharmacology* 146, 455–464.
- Neyse, L., Bosworth, S., Ring, P., and Schmidt, U. (2016). Overconfidence, incentives and digit ratio. *Sci. Rep.* 6:23294. doi: 10.1038/srep23294
- O’Donoghue, T., and Rabin, M. (1999). Doing it now or later. *Am. Econ. Rev.* 89, 103–124.
- Pender, J. (1996). Discount rates and credit markets: theory and evidence from rural india. *J. Dev. Econ.* 50, 257–296.
- Petry, N. M. (2001). Delay discounting of money and alcohol in actively using alcoholics, currently abstinent alcoholics, and controls. *Psychopharmacology* 154, 243–250. doi: 10.1007/s002130000638
- Shamosh, N., DeYoung, C., Green, A., Reis, D., Johnson, M., Conway, A., et al. (2008). Individual differences in delay discounting: relation to intelligence,



- working memory, and anterior prefrontal cortex. *Psychol. Sci.* 19, 904–911. doi: 10.1111/j.1467-9280.2008.02175.x
- Sutter, M., Kocher, M., Glätzle-Rützler, D., and Trautmann, S. (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *Am. Econ. Rev.* 103, 510–531. doi: 10.1257/aer.103.1.510
- Tanaka, T., Camerer, C., and Nguyen, Q. (2010). Risk and time preferences: linking experimental and household survey data from vietnam. *Am. Econ. Rev.* 100, 557–571. doi: 10.1257/aer.100.1.557
- Voracek, M., Manning, J., and Dressler, S. (2007). Repeatability and inter-observer error of digit ratio (2D: 4D) measurements made by experts. *Am. J. Hum. Biol.* 19, 142–146. doi: 10.1002/ajhb.20581
- Zheng, Z., and Cohn, M. (2011). Developmental basis of sexually dimorphic digit ratios. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16289–16294. doi: 10.1073/pnas.1108312108

**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.

The reviewers AME and BL and the handling Editor declared their shared affiliation.

Copyright © 2018 Aycinena and Rentschler. 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.



# A Context Dependent Interpretation of Inconsistencies in 2D:4D Findings: The Moderating Role of Status Relevance

Kobe Millet\* and Florian Buehler

Department of Marketing, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

## OPEN ACCESS

### Edited by:

Levent Neyse,  
Institut für Weltwirtschaft, Germany

### Reviewed by:

Giuseppe Danese,  
Universidade Católica Portuguesa,  
Portugal

Laura Kaltwasser,  
Berlin School of Mind and Brain,  
Humboldt University of Berlin,  
Germany

### \*Correspondence:

Kobe Millet  
kobe.millet@vu.nl

**Received:** 10 October 2017

**Accepted:** 14 December 2017

**Published:** 17 January 2018

### Citation:

Millet K and Buehler F (2018) A  
Context Dependent Interpretation of  
Inconsistencies in 2D:4D Findings:  
The Moderating Role of Status  
Relevance.  
Front. Behav. Neurosci. 11:254.  
doi: 10.3389/fnbeh.2017.00254

Whereas direct relationships between 2D:4D and dominance related attitudes or behavior often turn out to be weak, some literature suggests that the relation between 2D:4D and dominance is context-specific. That is, especially in status-challenging situations 2D:4D may be related to dominant behavior and its correlates. Based on this perspective, we interpret inconsistencies in the literature on the relation between 2D:4D and risk taking, aggression and dominance related outcomes and investigate in our empirical study how attitudes in low 2D:4D men may change as a function of the status relevance of the context. We provide evidence for the idea that status relevance of the particular situation at hand influences the attitude towards performance-enhancing means for low 2D:4D men, but not for high 2D:4D men. We argue that 2D:4D may be related to any behavior that is functional to attain status in a specific context. Implications for (economic) decision making are discussed.

**Keywords:** 2D:4D, digit ratio, social status, economic decision making, performance, dominance, context

## INTRODUCTION

*“Apart from economic payoffs, social status (social rank) seems to be the most important incentive and motivating force of social behavior.”*

Harsanyi (1976), p. 204.

In the present article, we focus on the potential importance of the drive for social status when studying relationships between 2D:4D and risk taking, performance, overconfidence, aggression or any other behavior that may be functional to attain status. We will argue that John Harsanyi's proposition of social status as one of the most important drivers of social decision making may especially hold for low 2D:4D men. Accordingly, we aim to illustrate how our status striving perspective may shed light on some puzzling inconsistencies in previous findings and provide some empirical evidence in support of our reasoning. Finally, we will discuss how these insights may be of relevance for the study of 2D:4D as a biological driver of economic decisions people make.

The second to fourth digit ratio or shortly 2D:4D is a biological marker referring to the relative length of the index (2nd digit) to the ring (4th digit) finger of someone's hand. A lower 2D:4D is assumed to be the result of prenatal exposure to increased levels of testosterone (Manning, 2002) and some direct evidence is provided in non-human mammals, for instance it has been shown that the enhancement of prenatal testosterone reduces 2D:4D in rats (Talarovicová et al., 2009) as well as in mice (Zheng and Cohn, 2011). Moreover, a lot of indirect evidence

in humans speaks towards this assumption, for instance ADHD (McFadden et al., 2005; de Bruin et al., 2006; Stevenson et al., 2007; Martel et al., 2008; Martel, 2009) and autism spectrum disorders (Manning et al., 2001; Milne et al., 2006; de Bruin et al., 2006; De Bruin et al., 2009), both thought to be influenced by prenatal testosterone, are related to 2D:4D as well. One of the most robust findings is the observation that 2D:4D is sexually dimorphic (Hönekopp and Watson, 2010). In general, males have a lower 2D:4D than females, not only in humans, but also in other mammals such as mice (Brown et al., 2002; Manning et al., 2003), rats (Talarovicová et al., 2009), bonobos (McIntyre et al., 2009) and baboons (McFadden and Bracht, 2003; Roney et al., 2004). Despite more evidence needed to validate 2D:4D as an indicator of prenatal testosterone, hundreds of publications in the last decade at least illustrate that 2D:4D is commonly accepted as an indirect biomarker of prenatal testosterone (Voracek, 2014).

Interestingly, 2D:4D has been related to sexually dimorphic behavior, such as aggression (Turanovic et al., 2017), risk taking (Brañas-Garza et al., in press), athletic achievement (Tester and Campbell, 2007), dominance (Manning and Fink, 2008) and according personality traits. Remarkably, surveying the existing literature it seems that the evidence for direct relationships between 2D:4D and personality measures is mixed (effects seem to be difficult to replicate at least). However, some relationships between 2D:4D and behavioral measures that are closely related to the same personality measures seem to be more robust in particular settings. We will conjecture below why these inconsistencies may arise.

Consider the mixed evidence for the relation between risk taking and 2D:4D. Whereas some find a negative relationship between 2D:4D and risk taking measures in both sexes (Dreber and Hoffman, 2007; Garbarino et al., 2011; Chicaiza-Becerra and Garcia-Molina, 2017) others do observe this effect among only men (Brañas-Garza and Rustichini, 2011; Stenstrom et al., 2011) or only women (Hönekopp, 2011) and even a larger amount of published studies did not find any significant association (Apicella et al., 2008; Sapienza et al., 2009; Aycinena et al., 2014; Kim et al., 2014; Drichoutis and Nayga, 2015; Schipper, 2015). Interestingly, some studies provide evidence for the idea that particular characteristics in the environment (Ronay and Von Hippel, 2010) or in the risk taking measure (Brañas-Garza et al., in press) may play a crucial role. Moreover, it is important to be aware that empirical evidence for the relation between 2D:4D and “real-world” risk taking looks much more convincing. For instance, it has been shown that low 2D:4D predicts risky driving behavior in traffic (as measured by the penalty point entries recorded on the driving license; Schwerdtfeger et al., 2010) as well as the likelihood to start a risky finance career (Sapienza et al., 2009). The relation between low 2D:4D and increased profitability of high-frequency financial traders (Coates et al., 2009) has also been explained by an increased tolerance for financial risk (Coates and Page, 2009). 2D:4D seems to be related to criminal risk taking actions too: Some evidence shows that imprisoned criminal offenders have a lower 2D:4D than nonoffenders (Hanoch et al., 2012) and low 2D:4D is related to increased criminal involvement (Ellis and Hoskin, 2015).

If we focus on the relation between 2D:4D and aggression, some recent meta-analyses have shown that the overall effect size of the relationship between 2D:4D and aggression measures is weak (Hönekopp, 2011; Turanovic et al., 2017). However, it is important to take into account that in the majority of the studies adopted in these meta-analyses aggression is measured in artificial settings or by self reports in questionnaires. Again, “real-world” aggressive behavior during sport contests seems to be more consistently related to 2D:4D (Perciavalle et al., 2013; Mailhos et al., 2016). Furthermore, typical studies focus on linear relationships between 2D:4D and aggression without taking the context into account. However, specific characteristics of the context may be crucial to observe any relationship with the dependent measure. At least, some data suggest that cues that point to challenges in the environment (such as aggression or provocation) are essential to observe a relationship between a lower 2D:4D and increased aggression levels (Millet and Dewitte, 2007; Kilduff et al., 2013) or decreased prosociality (Millet and Dewitte, 2009; Ronay and Galinsky, 2011). Accordingly, the relation between unprovoked aggression and 2D:4D in a simulated war game (McIntyre et al., 2007) may have emerged exactly because of the specific context in which the behavior took place.

What may be the reason for these seemingly inconsistent patterns of results? To find an answer, it may be instructive to look at the perspective presented in Millet (2011) and Ryckmans et al. (2015) to understand how specific characteristics of the particular study context and/or dependent measures may be crucial to observe effects between 2D:4D and the variable at hand. Ryckmans et al. (2015) remarked that the effect size of a direct linear relationship between 2D:4D and personality measures of dispositional dominance (see e.g., Manning and Fink, 2008) is weak at best despite more consistent evidence for the negative relationship between 2D:4D and performance in many different sports (Tester and Campbell, 2007; Hönekopp and Schuster, 2010), on the financial markets (Coates et al., 2009) and in cognitive tasks or academic assessments (Brosnan et al., 2011; Hopp et al., 2012; Bosch-Domènech et al., 2014). Moreover, strong relationships between 2D:4D and dominance related behavior or outcomes have been observed in non-human species such as macaques (Nelson et al., 2010) and baboons (Howlett et al., 2012, 2015). Ryckmans et al. (2015) propose that the activation of the dominance system is crucial to observe relations between 2D:4D and dominance and provide experimental evidence showing that male 2D:4D is indeed only associated with a dominant personality trait measure when the dominance system is likely to be activated (that is, after fictitious male-male interaction with another dominant man). This is in line with the perspective of Millet (2011), who argued that 2D:4D would only predict dominant-related behavior in those situations where status is at stake.

This perspective is consistent with empirical findings on circulating testosterone levels. Whereas a growing body of evidence points to the absence of a relationship between 2D:4D and circulating testosterone levels (Muller et al., 2011), low

2D:4D may reflect increased sensitivity to circulating levels of testosterone: Some recent studies show that testosterone administration only influences behavior for men and women with low 2D:4D (Carré et al., 2015; Buskens et al., 2016; Chen et al., 2016). In line with the biosocial model of status, testosterone seems to encourage behavior that is instrumental to dominate others (Mazur and Booth, 1998) and testosterone has especially high predictive validity in those situations when status is at stake (Newman and Josephs, 2009). Therefore, the reasoning that 2D:4D especially predicts dominant-related behavior in status challenging situations is consistent with this account.

In line with the perspective that testosterone is especially predictive when status is at stake we argue that 2D:4D is more likely to be related to status striving in specific, predictable situations than to unspecified measures of general risk taking, dominance, aggression or any other behavior *per se*. Based on this approach we would predict that only when status is at stake relationships between 2D:4D and context-specific goal-directed behavior would emerge (be it risk taking, aggression or even pro-social behavior). Following this reasoning, it is likely that for instance the relation between low 2D:4D and higher levels of aggressive behavior in soccer (Perciavalle et al., 2013; Mailhos et al., 2016) may be driven by the increased chance to win the particular game, but that 2D:4D and general personality measures of aggression are not related when measured in a controlled lab setting. First, status striving motivations are typically not activated when personality measures of aggression are assessed. Second, aggression is only one specific path towards status: whereas it may be functional to attain status in competitive and violent environments aggressive responses may also lead to the opposite effect (or be not effective at all) in other settings. At least some evidence is consistent with this idea as it has been shown that personality measures of aggression are not related to 2D:4D when people are exposed to a non-violent video, but that the relationship between 2D:4D and aggression emerges after exposure to a violent video (Millet and Dewitte, 2007; Kilduff et al., 2013).

Furthermore, the relation between 2D:4D and financial risk taking may predominantly emerge in experimental settings when the behavior is financially incentivized (Brañas-Garza et al., in press) as only higher actual payoffs in the experimental session are able to enhance relative status compared to other participants in the same experimental session. Similarly, relations between 2D:4D and “real-world” risk taking behavior may only arise when the risk one takes may lead to an increased status position. Whereas it has been claimed that increased tolerance for financial risk (Coates and Page, 2009) explains increased profitability of high-frequency financial traders (Coates et al., 2009) we would suggest otherwise: As profitability is status enhancing in this financial context, taking more risk can be considered the only viable option to potentially make the most profits. Thus, the urge to attain status is possibly a more important driving force than the proposed increased risk tolerance (Millet, 2009).

Given our interpretation of inconsistencies in the literature, we set up a study to investigate whether status relevance of the specific context is indeed important in the study of the

relation between 2D:4D and any goal-directed (i.e., potentially status-enhancing) behavior. Based on our reasoning we would predict a relationship between any behavior as long as it qualifies as a mean to enhance status in that specific context, but not if status is not relevant in the particular context at hand (and thus the same behavior is not functional anymore to attain status). We decided to focus on decisions without any financial outcome as merely the financial aspect by itself could already change the meaning of the decision. We simply manipulated one aspect of the context so that the same decision is considered functional to attain status or not. Based on our reasoning, we only expect a relationship between 2D:4D and the decision at hand when the decision is functional to attain status. More concretely, we provided a fictitious sports competition scenario in which winning either increased status (an important competition) or was status irrelevant (an unimportant competition). Interestingly, chances to win typical sports competitions can not only be increased by exercise, motivation, aggression, risk taking or physical superiority but also by the use of a wide spectrum of performance enhancing products, going from (legal) supplements to (illegal) doping. Therefore, we asked our participants about their evaluation of different products (both legal and illegal) that could potentially enhance performance. We included both legal and illegal products to create a realistic scenario (both types of products are generally perceived to be common practice in cycling competitions given the anecdotal evidence in popular media that professional cyclists make use of these). As these products are only functional to attain status in the status relevant condition, we expect that a relation between 2D:4D and attitude towards these performance enhancing means only emerges in the status relevant condition. More specifically, we predict that low 2D:4D men will generally be more positive about such performance enhancing means in the status relevant than in the status irrelevant situation. We do not make any *a priori* prediction with regard to the nature of the means (i.e., illegal vs. legal). By adopting this factor it may also provide insights into how far-reaching low 2D:4D men's ambitions may go. Albeit we make use of an imagination exercise and the attainment of status is therefore purely fictitious (i.e., a construct of participant's mind) in our experiment. We consider the design a rather conservative test of our hypothesis. If we observe a result that is consistent with this hypothesis despite the “imagination” part and lack of monetary incentivization, then *a fortiori* we would expect our hypothesis to hold in a framework with real, financially incentivized decisions.

## MATERIALS AND METHODS

One hundred and nine male students received partial course credits for their participation in the study. This study was carried out in accordance with the recommendations of the ethical guidelines of the faculty of Economics and Business Administration of the Vrije Universiteit Amsterdam with informed consent from all subjects. All subjects gave informed consent in accordance with the Declaration of Helsinki. The



protocol was approved by the FEWEB Research Ethics Review Board.

Upon arrival in the laboratory, each participant was assigned to a computer in a partially enclosed carrel. Participants did not see one another and could not talk. A maximum of 14 students participated at the same time. Participants were randomly assigned to one of two between-subjects conditions: a status-relevant vs. status-irrelevant condition. In the status-relevant condition, we asked participants to imagine that they are a professional cyclist and participate in the *most* important cycling race of their season. In the status-irrelevant condition on the other hand, we asked them to imagine to participate in the *least* important cycling race of their season. We chose to change only one word in the introduction to keep everything else constant. The meaning of performance changes depending on the specific context (least vs. most important): The striving to attain status (i.e., winning the race) is only activated in the context of an important race. After this introduction, we asked to what extent (on a 7-point Likert scale; 1: definitely not; 7: definitely yes) they would make use of different means to enhance their performance in the race: nutritional supplements (e.g., a protein shake), prohibited substances (e.g., EPO) and technological fraud (e.g., a hidden engine in the racing bike). Further, we asked them to rate on 7-point Likert scales how bad (=1) vs. good (=7) as well as how unethical (=1) vs. ethical (=7) each of these means are to enhance performance (see for descriptives **Table 1**). First, we composed a “legal means attitude” vs. “illegal means attitude” by averaging the three items related to nutritional supplements ( $\alpha = 0.85$ ) and averaging the six items related to the prohibited means ( $\alpha = 0.69$ ).

*A priori*, we determined to focus on right hand 2D:4D as androgenization is suggested to have a stronger impact on the right than on the left hand (e.g., Williams et al., 2000; McFadden and Shubel, 2002), gender differences are larger for right-hand 2D:4D (Hönekopp and Watson, 2010) and the right hand is more commonly used in previous research (Brañas-Garza and Rustichini, 2011). Hand scans were taken at the end of the session with a high-resolution scanner (Canon Lide 120) and afterwards two independent raters measured (by means of Photoshop CC 2015) the length of index (2nd) and ring (4th) finger. Finger lengths were measured from the bottom crease when there was

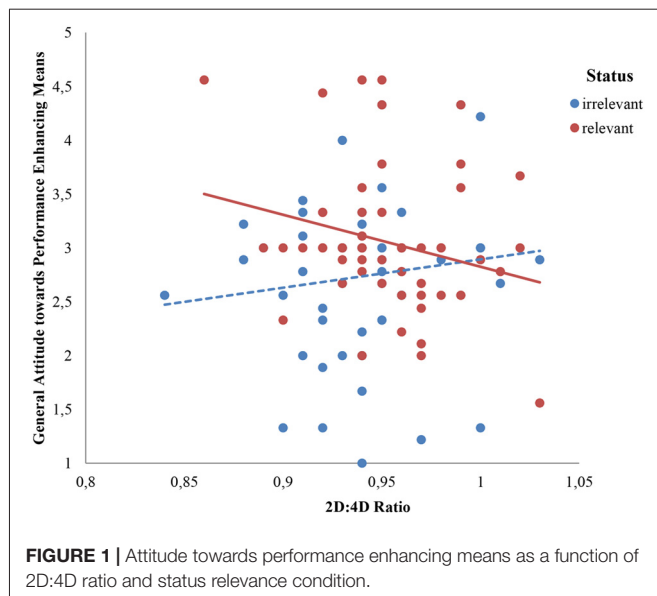
a band of creases at the base of the digit. Ratios of both raters were highly correlated ( $r = 0.87$ ), speaking towards the accuracy of the measurement. We averaged both ratios to obtain one single measure for 2D:4D and make use of this averaged 2D:4D in our analyses.

## RESULTS

We used both attitude measures as dependent variables (within: legal vs. illegal) in a mixed design with 2D:4D (mean-centered) and status relevance (between subjects: status relevant vs. irrelevant) as independent variables. A mixed-design analysis of variance assessed effects of the status relevance manipulation and 2D:4D on the attitude towards legal and illegal means to improve performance, which were included as repeated measures. We observed a more positive attitude towards legal ( $M = 5.75$ ,  $SD = 1.46$ ) than illegal means ( $M = 1.48$ ,  $SD = 0.72$ ,  $F_{(1,105)} = 818.90$ ,  $p = 0.000$ , partial  $\eta^2 = 0.87$ ). Further, a main effect of status relevance on general attitude towards performance enhancing means emerged ( $F_{(1,105)} = 7.97$ ,  $p = 0.006$ , partial  $\eta^2 = 0.07$ ), which was moderated by the nature of the means ( $F_{(1,105)} = 5.62$ ,  $p = 0.02$ , partial  $\eta^2 = 0.05$ ). Whereas status relevance influenced the attitude towards legal means ( $M_{\text{status relevant}} = 6.10$ ,  $SD = 1.12$  vs.  $M_{\text{status irrelevant}} = 5.37$ ,  $SD = 1.68$ ;  $F_{(1,105)} = 8.56$ ,  $p = 0.004$ , partial  $\eta^2 = 0.08$ ), it did not change attitude towards illegal means ( $M_{\text{status relevant}} = 1.52$  vs.  $M_{\text{status irrelevant}} = 1.42$ ,  $p = 0.53$ , partial  $\eta^2 = 0.004$ ). More interestingly and in line with our predictions, we also observed a marginally significant interaction effect between status relevance and 2D:4D ( $F_{(1,105)} = 3.90$ ,  $p = 0.05$ , partial  $\eta^2 = 0.04$ ). No other effects turned out to be significant (neither within or between; all  $ps > 0.14$ ). To be able to study the interaction between status relevance and 2D:4D in more detail we first calculated a general “attitude towards performance enhancing means” score by averaging the 9 item scores on the three performance enhancing means ( $\alpha = 0.68$ ) and used this measure in the remaining analyses. We aimed to provide insight into this interaction between 2D:4D and status relevance by: (1) calculating Spearman correlation coefficients between 2D:4D and general attitude scores within both conditions to examine in which condition 2D:4D and attitude scores are related; and (2) performing a spotlight analysis

**TABLE 1 |** Mean (*M*) and Standard Deviation (*SD*) of dependent measures (variable and single items) of the total sample and each condition.

Variable/Measure	Total sample ( <i>N</i> = 109)		Status irrelevant ( <i>N</i> = 53)		Status relevant ( <i>N</i> = 56)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<b>Nutritional Supplements</b>	<b>5.75</b>	<b>1.46</b>	<b>5.37</b>	<b>1.68</b>	<b>6.10</b>	<b>1.12</b>
Bad/Good	5.82	1.49	5.60	1.61	6.13	1.34
Unethical/Ethical	5.83	1.60	5.51	1.79	6.16	1.34
Intention	5.60	1.86	5.00	2.10	6.16	1.40
<b>Prohibited Substances</b>	<b>1.51</b>	<b>0.89</b>	<b>1.47</b>	<b>0.87</b>	<b>1.54</b>	<b>0.92</b>
Bad/Good	1.85	1.62	1.77	1.58	1.93	1.66
Unethical/Ethical	1.32	0.92	1.43	1.20	1.21	0.53
Intention	1.35	0.94	1.21	0.79	1.48	1.04
<b>Technological Fraud</b>	<b>1.44</b>	<b>0.70</b>	<b>1.38</b>	<b>0.64</b>	<b>1.51</b>	<b>0.75</b>
Bad/Good	1.67	1.46	1.51	1.31	1.82	1.59
Unethical/Ethical	1.33	0.84	1.40	0.99	1.27	0.67
Intention	1.33	0.84	1.23	0.67	1.43	0.97



(Irwin and McClelland, 2001; Spiller et al., 2013) as such analysis allows us to examine the effect of status relevance at different levels of 2D:4D. This analysis provides insights whether this effect of status relevance is especially driven by low 2D:4D men, high 2D:4D men or both. In accordance with our hypothesis, 2D:4D and the general attitude score were not related in the status irrelevant condition (Spearman's correlation coefficient  $r = 0.10$ ,  $p = 0.50$ ), but a negative relationship emerged when the situation described was status relevant (Spearman's correlation coefficient  $r = -0.27$ ,  $p = 0.04$ ; see **Figure 1**). The results from the spotlight analysis were also consistent with our prediction: For low 2D:4D men (one standard deviation below the mean), the attitude towards performance enhancing means in the status relevant condition was higher than the attitude towards these means in the status irrelevant condition ( $M_{\text{status relevant}} = 3.25$  vs.  $M_{\text{status irrelevant}} = 2.66$ ,  $\beta = 0.30$ ,  $SE = 0.095$ ,  $t_{(105)} = 3.10$ ,  $p = 0.002$ ). For high 2D:4D men (one standard deviation above the mean), the status relevance of the situation did not influence the attitude towards performance enhancing means ( $M_{\text{status relevant}} = 2.91$  vs.  $M_{\text{status irrelevant}} = 2.85$ ,  $\beta = 0.03$ ,  $SE = 0.10$ ,  $t_{(105)} = 0.31$ ,  $p = 0.76$ ).

## DISCUSSION

In line with a status drive perspective on 2D:4D, our findings indicate that low 2D:4D men are generally more positive towards performance-enhancing means to win a cycling competition when they believe that the competition at hand is important, but not so when the competition is not important. If low 2D:4D men would take legalness of means into account in their need to achieve status, a three-way interaction should have been observed. However, we did not find any evidence for a differentiation between legal (nutrition supplements) and illegal (EPO, a hidden engine in the bike) means thereby suggesting that low 2D:4D men may be more inclined "to do whatever it

takes to win" when stakes are high, but not when the outcome is irrelevant to attain personal status.

If this conclusion is correct, relationships between 2D:4D and any attitude, trait or behavior (be it greedy, impulsive, unethical, altruistic, selfish, ...) may emerge as long as these particular attitudes, traits and behaviors help to attain status in that specific situation. However, if the focal behavior is related to an outcome irrelevant to one's own status position, we do not expect any relationship between 2D:4D and the specific behavior at hand. Some recent evidence in a business context speaks towards this idea: lower use of prohibitive voice (i.e., expressing concerns about practices, behavior, incidents that may be harmful for the organization) is related to a low 2D:4D among low-ranked, but not among high-ranked employees (Bijleveld and Baalbergen, 2017). Albeit speculative, they argue (in line with our reasoning) that prohibitive voice in this particular setting can be considered status relevant for low-ranked, but not for high-ranked employees as it is important for low-ranked employees not to express prohibitive voice to attain or at least maintain status, whereas the use of prohibitive choice does not have any consequence for high-ranked employees (Bijleveld and Baalbergen, 2017).

We believe that a status striving perspective on 2D:4D may shed light on how 2D:4D may drive (economic) decisions. The failure of some studies to find a relationship between 2D:4D and attitudes or behavior may be due to an omitted variable problem, i.e., context: Depending on the particular context, the same behavior or attitudes may be functional in terms of possibilities to increase status or not. Only when considered functional in a specific setting, we would predict a relationship with 2D:4D. For instance, in a recent study it has been shown that 2D:4D only predicts risk taking with real monetary incentives (Brañas-Garza et al., in press). This observation is consistent with our reasoning considering that especially the context with incentivized choices is status-relevant: larger payoffs may directly lead to a higher perceived relative status among the sample of participants in the study. On the other hand, risk attitudes are by itself not directly related to any status-relevant outcome, which may explain why more often no association has been observed between 2D:4D and attitudinal risk taking measures. Our reasoning at least suggests that low 2D:4D men may be especially prone to take (financial) risk when they know that the potential outcome of the risk they take is status-enhancing, even when it is illegal or criminal (consistent with Hanoch et al., 2012; Ellis and Hoskin, 2015).

Following a similar reasoning, monetary incentives may not only change the meaning of financial risk responses but also of other behavioral measures. For instance, Neyse et al. (2016) found that low male 2D:4D is related to higher overconfidence levels when men are asked to predict own performance on a cognitive reflection test (as measured by overestimation, i.e., the individual estimate of the number of correct answers on a cognitive reflection test minus the actual number of correct answers on this test). Still, their effect only held when performance prediction accuracy is not monetarily incentivized: when more accurate predictions

are financially rewarded the relationship between 2D:4D and overconfidence actually reverses (Neyse et al., 2016). Following our rationale, we predict that overconfidence will increase or decrease among low 2D:4D men depending on its functionality to attain higher status. Overconfidence has been considered as a way to obtain status (Anderson et al., 2012), and the observed relationship between lower 2D:4D and higher overconfidence levels is thus consistent with our reasoning. However, incentivization of accuracy may actually change the meaning of the measurement. Under the assumption that larger pay-offs in the study at hand may directly lead to a higher perceived relative status among study participants, increased accuracy—and thus lower overconfidence—is actually functional to attain status in this particular setting.

Whereas our perspective may shed light on some inconsistencies in the 2D:4D literature, there is a need to further improve the theoretical perspective to provide insight into other inconsistencies. For instance, when taking a look at the relationship between 2D:4D and prosocial behavior, there have been observed both positive (Buser, 2012), negative (Millet and Dewitte, 2009) and curvilinear (Millet and Dewitte, 2006; Brañas-Garza et al., 2013; Galizzi and Nieboer, 2015) relationships in seemingly neutral situations as well as positive relationships in specific potentially “challenging” situations (Millet and Dewitte, 2009; Ronay and Galinsky, 2011). Whereas both prosocial and antisocial behavior have been considered as ways to attain status (Millet and Dewitte, 2009), it remains difficult to understand the inconsistency between findings in this stream of literature from the perspective we provide in the current manuscript. Some findings show how contextual characteristics are able to shift the relationship between 2D:4D and choices in ultimatum and dictator games (Van den Bergh and Dewitte, 2006; Millet and Dewitte, 2009; Ronay and Galinsky, 2011) and incentivization has been considered to be important as well (Brañas-Garza et al., 2013). Therefore, it seems to be crucial to consider the context in which the behavior took place as well as the specific nature of the measurement (e.g., incentivized or not, type of economic game, etc.). Though, the overall pattern of results in this domain remain difficult to explain from our perspective. For instance, our perspective does not allow to make any inference on how people with a “medium” 2D:4D may react differently compared to low and high 2D:4D people (Brañas-Garza et al., 2013; Galizzi and Nieboer, 2015). Still, we believe that the relations between 2D:4D and prosocial/antisocial choices are at least influenced by the perceived functionality to attain status in the specific context in which the study took place albeit other aspects seem to be crucial as well.

At least, our pattern of results corroborates the viewpoint that male 2D:4D is negatively related to performance in many domains because of the need for status. We suggest in line with Millet (2009) and Millet and Dewitte (2008) that low 2D:4D men may also self-select into those domains in which they excel (be it sports, music, cognitive performance or even performance on financial markets) as long as their superiority in that specific domain provides them with a feeling of higher relative standing. Remarkably, this self-selection perspective would predict a

relationship between 2D:4D and level of competition (e.g., lower 2D:4D in national vs. professional and/or recreational teams; Frick et al., 2017; Manning and Taylor, 2001), but not necessarily within competition. For instance, consider low 2D:4D men without the necessary skills to be part of a professional soccer team but still remain playing soccer at low level. Given the absence of superior performance they probably do so because of intrinsic motivation (i.e., the pleasure of the game) and not for the sake of status.

Our context-dependent perspective can be considered in line with the recent hypothesis that many of the relations between low 2D:4D and improved performance in sports (as well as in other domains) may be driven by the association between low 2D:4D and pronounced spikes of testosterone in challenge situations (Manning et al., 2014). Therefore one avenue for further research could focus on the interplay between circulating testosterone and 2D:4D by: (a) measuring circulating testosterone in different settings and investigate whether the relationship between 2D:4D and status-driven behavior is induced by enhanced circulating testosterone levels in these settings and thus increased testosterone sensitivity of low 2D:4D individuals; or (b) testing whether low 2D:4D predicts the production of testosterone levels in challenge situations. Such studies could at least provide further insight into the biological basis for the presumed relation between 2D:4D and status striving. We also would like to point out that our imbalanced sex ratio in the lab (only men participated) may have induced a more competitive setting by itself (see Griskevicius et al., 2012) and thereby increased circulating testosterone levels in general. Still, this assumption remains open for future research as well as the plausible hypothesis that a male biased sex ratio may have led to our specific pattern of results.

Finally, it is important to realize as well that it remains difficult to *ex ante* identify those contexts in which a particular behavior is considered functional to attain status or not. Whereas we are able to integrate many inconsistent findings in the literature based on this status striving perspective, further elaboration of the theoretical perspective is needed to get a better understanding of under what specific circumstances we may expect relationships between 2D:4D and other variables of interest. Therefore, another interesting avenue for further research is the study of the relation between 2D:4D and performance indices or specific decisions that may be considered functional or not to attain status in different contexts to provide insights into the generalizability of our findings. Further validation of our hypothesis would be especially desirable in incentivized laboratory or field studies in which (real) decisions need to be taken that are either functional or not to attain status in that particular context.

To conclude, in the present article we presented a theoretical perspective that provides an interpretation of inconsistencies in current 2D:4D literature. Further, we provided some empirical evidence for our reasoning that low 2D:4D men may do whatever it takes to attain status, thereby stressing the functionality of specific behavior towards this status goal in the particular context at hand. We hope that our interpretations, propositions and discussion are helpful in the formation and/or further development of a highly needed theoretical perspective to

understand how 2D:4D influences behavior and that the present analysis is at least helpful to identify interesting paths for future research.

## AUTHOR CONTRIBUTIONS

KM designed the study. FB carried out the experiment and collected data. KM and FB analyzed the data. KM wrote the

manuscript with support from FB. Both authors agree to be accountable for the content of the work.

## ACKNOWLEDGMENTS

We thank the Behavioral Lab of the School of Business and Economics (Vrije Universiteit Amsterdam) for data collection support.

## REFERENCES

- Anderson, C., Brion, S., Moore, D. A., and Kennedy, J. A. (2012). A status-enhancement account of overconfidence. *J. Pers. Soc. Psychol.* 103, 718–735. doi: 10.1037/a0029395
- 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
- Aycinena, D., Baltaduonis, R., and Rentschler, L. (2014). Risk preferences and prenatal exposure to sex hormones for latinos. *PLoS One* 9:e103332. doi: 10.1371/journal.pone.0103332
- Bijleveld, E., and Baalbergen, J. (2017). Prenatal exposure to testosterone (2D:4D) and social hierarchy together predict voice behavior in bankers. *PLoS One* 12:e0180008. doi: 10.1371/journal.pone.0180008
- Bosch-Domènech, A., Brañas-Garza, P., and Espín, A. M. (2014). Can exposure to prenatal sex hormones (2D:4D) predict cognitive reflection? *Psychoneuroendocrinology* 43, 1–10. doi: 10.1016/j.psyneuen.2014.01.023
- Brañas-Garza, P., Galizzi, 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
- Brosnan, M., Gallop, V., Iftikhar, N., and Keogh, E. (2011). Digit ratio (2D:4D), academic performance in computer science and computer-related anxiety. *Pers. Individ. Dif.* 51, 371–375. doi: 10.1016/j.paid.2010.07.009
- Brown, W. M., Finn, C. J., and Breedlove, S. M. (2002). Sexual dimorphism in digit-length ratios of laboratory mice. *Anat. Rec.* 267, 231–234. doi: 10.1002/ar.10108
- Buser, T. (2012). Digit ratios, the menstrual cycle and social preferences. *Games Econ. Behav.* 76, 457–470. doi: 10.1016/j.geb.2012.07.006
- Buskens, V., Raub, W., van Miltenburg, N., Montoya, E. R., and van Honk, J. (2016). Testosterone administration moderates effect of social environment on trust in women depending on second-to-fourth digit ratio. *Sci. Rep.* 6:e27655. doi: 10.1038/srep27655
- Carré, J. M., Ortiz, T. L., Labine, B., Moreau, B. J., Viding, E., Neumann, C. S., et al. (2015). Digit ratio (2D:4D) and psychopathic traits moderate the effect of exogenous testosterone on socio-cognitive processes in men. *Psychoneuroendocrinology* 62, 319–326. doi: 10.1016/j.psyneuen.2015.08.023
- Chen, C., Decety, J., Huang, P. C., Chen, C. Y., and Cheng, Y. (2016). Testosterone administration in females modulates moral judgment and patterns of brain activation and functional connectivity. *Hum. Brain Mapp.* 37, 3417–3430. doi: 10.1002/hbm.23249
- Chicaiza-Becerra, L. A., and Garcia-Molina, M. (2017). Prenatal testosterone predicts financial risk taking: evidence from Latin America. *Pers. Individ. Dif.* 116, 32–37. doi: 10.1016/j.paid.2017.04.021
- Coates, J. M., Gurnell, M., and Rustichini, A. (2009). Second-to-fourth digit ratio predicts success among high-frequency financial traders. *Proc. Natl. Acad. Sci. U S A* 106, 623–628. doi: 10.1073/pnas.0810907106
- Coates, J. M., and Page, L. (2009). A note on trader Sharpe Ratios. *PLoS One* 4:e8036. doi: 10.1371/journal.pone.0008036
- De Bruin, E. I., De Nijs, P. F., Verheij, F., Verhagen, D. H., and Ferdinand, R. F. (2009). Autistic features in girls from a psychiatric sample are strongly associated with a low 2D:4D ratio. *Autism* 13, 511–521. doi: 10.1177/1362361309335720
- de Bruin, E. I., Verheij, F., Wiegman, T., and Ferdinand, R. F. (2006). Differences in finger length ratio between males with autism, pervasive developmental disorder-not otherwise specified, ADHD, and anxiety disorders. *Dev. Med. Child Neurol.* 48, 962–965. doi: 10.1017/s001216220602118
- Dreber, A., and Hoffman, M. (2007). *Portfolio Selection in Utero*. Mimeo: Stockholm School of Economics.
- Drichoutis, A. C., and Nayga, R. M. (2015). Do risk and time preferences have biological roots? *South. Econ. J.* 82, 235–256. doi: 10.1002/soej.104284/0038-4038-2013.246
- Ellis, L., and Hoskin, A. W. (2015). Criminality and the 2D:4D ratio: testing the prenatal androgen hypothesis. *Int. J. Offender Ther. Comp. Criminol.* 59, 295–312. doi: 10.1177/0306624x13503813
- Frick, N. A., Hull, M. J., Manning, J. T., and Tomkinson, G. R. (2017). Relationships between digit ratio (2D:4D) and basketball performance in Australian men. *Am. J. Hum. Biol.* 29:e22937. doi: 10.1002/ajhb.22937
- 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
- Garbarino, E., Slonim, R., and Sydnor, J. (2011). Digit ratios (2D:4D) as predictors of risky decision making for both sexes. *J. Risk Uncertain.* 42, 1–26. doi: 10.1007/s11166-010-9109-6
- Griskevicius, V., Tybur, J. M., Ackerman, J. M., Delton, A. W., Robertson, T. E., and White, A. E. (2012). The financial consequences of too many men: sex ratio effects on saving, borrowing, and spending. *J. Pers. Soc. Psychol.* 102, 69–80. doi: 10.1037/a0024761
- Hanoch, Y., Gummerum, M., and Rolison, J. (2012). Second-to-fourth digit ratio and impulsivity: a comparison between offenders and nonoffenders. *PLoS One* 7:e47140. doi: 10.1371/journal.pone.0047140
- Harsanyi, J. C. (1976). *Essays on Ethics, Social Behaviour, and Scientific Explanation*. Dordrecht: Holland Boston.
- Hönekopp, J. (2011). Relationships between digit ratio 2D:4D and self-reported aggression and risk taking in an online study. *Pers. Individ. Dif.* 51, 77–80. doi: 10.1016/j.paid.2011.03.010
- 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
- Hopp, R. N., de Moraes, J. P., and Jorge, J. (2012). Digit ratio and academic performance in dentistry students. *Pers. Individ. Dif.* 52, 643–646. doi: 10.1016/j.paid.2011.12.013
- Howlett, C., Marshall, A. R., and Hughes, W. O. (2012). Digit ratios and dominance in female baboons (*Papio hamadryas* and *Papio ursinus*). *Int. J. Primatol.* 33, 1439–1452. doi: 10.1007/s10764-012-9636-x
- Howlett, C., Setchell, J. M., Hill, R. A., and Barton, R. A. (2015). The 2D:4D digit ratio and social behaviour in wild female chacma baboons (*Papio ursinus*) in relation to dominance, aggression, interest in infants, affiliation and heritability. *Behav. Ecol. Sociobiol.* 69, 61–74. doi: 10.1007/s00265-014-1817-5



- Irwin, J. R., and McClelland, G. H. (2001). Misleading heuristics and moderated multiple regression models. *J. Mark. Res.* 38, 100–109. doi: 10.1509/jmkr.38.1.100.18835
- Kilduff, L. P., Hopp, R. N., Cook, C. J., Crewther, B. T., and Manning, J. T. (2013). Digit ratio (2D:4D), aggression, and testosterone in men exposed to an aggressive video stimulus. *Evol. Psychol.* 11, 953–964. doi: 10.1177/147470491301100502
- Kim, Y., Kim, K., and Kim, T. H. (2014). Domain specific relationships of 2D:4D digit ratio in risk perception and risk behavior. *J. Gen. Psychol.* 141, 373–392. doi: 10.1080/00221309.2014.950188
- Mailhos, A., Buunk, A. P., Del Arca, D., and Tutte, V. (2016). Soccer players awarded one or more red cards exhibit lower 2D:4D ratios. *Aggress. Behav.* 42, 417–426. doi: 10.1002/ab.21638
- Manning, J. T. (2002). *Digit Ratio: A Pointer to Fertility, Behavior, and Health*. New Brunswick: Rutgers University Press.
- Manning, J. T., Baron-Cohen, S., Wheelwright, S., and Sanders, G. (2001). The 2nd to 4th digit ratio and autism. *Dev. Med. Child Neurol.* 43, 160–164. doi: 10.1017/s0012162201000317
- Manning, J. T., Callow, M., and Bundred, P. E. (2003). Finger and toe ratios in humans and mice: implications for the aetiology of diseases influenced by HOX genes. *Med. Hypotheses* 60, 340–343. doi: 10.1016/s0306-9877(02)00400-0
- Manning, J., Kilduff, L., Cook, C., Crewther, B., and Fink, B. (2014). Digit ratio (2D:4D): a biomarker for prenatal sex steroids and adult sex steroids in challenge situations. *Front. Endocrinol. Lausanne* 5:9. doi: 10.3389/fendo.2014.00009
- Manning, J. T., and Fink, B. (2008). Digit ratio (2D:4D), dominance, reproductive success, asymmetry, and sociosexuality in the BBC internet study. *Am. J. Hum. Biol.* 20, 451–461. doi: 10.1002/ajhb.20767
- Manning, J. T., and Taylor, R. P. (2001). Second to fourth digit ratio and male ability in sport: implications for sexual selection in humans. *Evol. Hum. Behav.* 22, 61–69. doi: 10.1016/s1090-5138(00)00063-5
- Martel, M. M. (2009). Conscientiousness as a mediator of the association between masculinized finger-length ratios and attention-deficit/hyperactivity disorder (ADHD). *J. Child Psychol. Psychiatry* 50, 790–798. doi: 10.1111/j.1469-7610.2009.02065.x
- Martel, M. M., Gobrogge, K. L., Breedlove, S. M., and Nigg, J. T. (2008). Masculinized finger-length ratios of boys, but not girls, are associated with attention-deficit/hyperactivity disorder. *Behav. Neurosci.* 122, 273–281. doi: 10.1037/0735-7044.122.2.273
- Mazur, A., and Booth, A. (1998). Testosterone and dominance in men. *Behav. Brain Sci.* 21, 353–363. doi: 10.1017/s0140525x98001228
- McFadden, D., and Bracht, M. S. (2003). The relative lengths and weights of metacarpals and metatarsals in baboons (*Papio hamadryas*). *Horm. Behav.* 43, 347–355. doi: 10.1016/s0018-506x(02)00048-x
- McFadden, D., and Shubel, E. (2002). Relative lengths of fingers and toes in human males and females. *Horm. Behav.* 42, 492–500. doi: 10.1006/hbeh.2002.1833
- McFadden, D., Westhafer, J. G., Pasanen, E. G., Carlson, C. L., and Tucker, D. M. (2005). Physiological evidence of hypermasculinization in boys with the inattentive type of attention-deficit/hyperactivity disorder (ADHD). *Clin. Neurosci. Res.* 5, 233–245. doi: 10.1016/j.cnr.2005.09.004
- McIntyre, M. H., Barrett, E. S., McDermott, R., Johnson, D. D., Cowden, J., and Rosen, S. P. (2007). Finger length ratio (2D:4D) and sex differences in aggression during a simulated war game. *Pers. Individ. Dif.* 42, 755–764. doi: 10.1016/j.paid.2006.08.009
- McIntyre, M. H., Herrmann, E., Wobber, V., Halbwax, M., Mohamba, C., de Sousa, N., et al. (2009). Bonobos have a more human-like second-to-fourth finger length ratio (2D:4D) than chimpanzees: a hypothesized indication of lower prenatal androgens. *J. Hum. Evol.* 56, 361–365. doi: 10.1016/j.jhevol.2008.12.004
- Millet, K. (2009). Low second-to-fourth-digit ratio might predict success among high-frequency financial traders because of a higher need for achievement. *Proc. Natl. Acad. Sci. U S A* 106:E30. doi: 10.1073/pnas.0900396106
- Millet, K. (2011). An interactionist perspective on the relation between 2D:4D and behavior: an overview of (moderated) relationships between 2D:4D and economic decision making. *Pers. Individ. Dif.* 51, 397–401. doi: 10.1016/j.paid.2010.04.005
- Millet, K., and Dewitte, S. (2006). Second to fourth digit ratio and cooperative behavior. *Biol. Psychol.* 71, 111–115. doi: 10.1016/j.biopsycho.2005.06.001
- Millet, K., and Dewitte, S. (2007). Digit ratio (2D:4D) moderates the impact of an aggressive music video on aggression. *Pers. Individ. Dif.* 43, 289–294. doi: 10.1016/j.paid.2006.11.024
- Millet, K., and Dewitte, S. (2008). A subordinate status position increases the present value of financial resources for low 2D:4D men. *Am. J. Hum. Biol.* 20, 110–115. doi: 10.1002/ajhb.20689
- Millet, K., and Dewitte, S. (2009). The presence of aggression cues inverts the relation between digit ratio (2D:4D) and prosocial behaviour in a dictator game. *Br. J. Psychol.* 100, 151–162. doi: 10.1348/000712608x324359
- Milne, E., White, S., Campbell, R., Swettenham, J., Hansen, P., and Ramus, F. (2006). Motion and form coherence detection in autistic spectrum disorder: relationship to motor control and 2:4 digit ratio. *J. Autism Dev. Disord.* 36, 225–237. doi: 10.1007/s10803-005-0052-3
- Muller, D. C., Giles, G. G., Bassett, J., Morris, H. A., Manning, J. T., Hopper, J. L., et al. (2011). Second to fourth digit ratio (2D:4D) and concentrations of circulating sex hormones in adulthood. *Reprod. Biol. Endocrinol.* 9:57. doi: 10.1186/1477-7827-9-57
- Nelson, E., Hoffman, C. L., Gerald, M. S., and Shultz, S. (2010). Digit ratio (2D:4D) and dominance rank in female rhesus macaques (*Macaca mulatta*). *Behav. Ecol. Soc.* 64, 1001–1009. doi: 10.1007/s00265-010-0915-2
- Newman, M. L., and Josephs, R. A. (2009). Testosterone as a personality variable. *J. Res. Personality* 43, 258–259. doi: 10.1016/j.jrp.2008.12.028
- Neyse, L., Bosworth, S., Ring, P., and Schmidt, U. (2016). Overconfidence, incentives and digit ratio. *Sci. Rep.* 6:23294. doi: 10.1038/srep23294
- Perciavalle, V., Di Corrado, D., Petralia, M. C., Gurrissi, L., Massimino, S., and Coco, M. (2013). The second-to-fourth digit ratio correlates with aggressive behavior in professional soccer players. *Mol. Med. Rep.* 7, 1733–1738. doi: 10.3892/mmr.2013.1426
- Ronay, R., and Galinsky, A. D. (2011). Lex talionis: testosterone and the law of retaliation. *J. Exp. Soc. Psychol.* 47, 702–705. doi: 10.1016/j.jesp.2010.11.009
- Ronay, R., and Von Hippel, W. (2010). Power, testosterone, and risk-taking. *J. Behav. Decis. Mak.* 23, 473–482. doi: 10.1002/bdm.671
- Roney, J. R., Whitham, J. C., Leoni, M., Bellem, A., Wieblenowski, N., and Maestripieri, D. (2004). Relative digit lengths and testosterone levels in Guinea baboons. *Horm. Behav.* 45, 285–290. doi: 10.1016/j.yhbeh.2003.12.008
- Ryckmans, J., Millet, K., and Warlop, L. (2015). The influence of facial characteristics on the relation between male 2D:4D and dominance. *PLoS One* 10:e0143307. doi: 10.1371/journal.pone.0143307
- Sapienza, P., Zingales, L., and Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proc. Natl. Acad. Sci. U S A* 106, 15268–15273. doi: 10.1073/pnas.0907352106
- Schipper, B. C. (2015). *Sex Hormones and Choice Under Risk. Working Papers No. 127*. Riverside, CA: University of California, Department of Economics.
- Schwerdtfeger, A., Heims, R., and Heer, J. (2010). Digit ratio (2D:4D) is associated with traffic violations for male frequent car drivers. *Accid. Anal. Prev.* 42, 269–274. doi: 10.1016/j.aap.2009.08.001
- Spiller, S. A., Fitzsimons, G. J., Lynch Jr, J. G., and McClelland, G. H. (2013). Spotlights, floodlights, and the magic number zero: simple effects tests in moderated regression. *J. Mark. Res.* 50, 277–288. doi: 10.1509/jmr.12.0420
- Stenstrom, E., Saad, G., Nepomuceno, M. V., and Mendenhall, Z. (2011). Testosterone and domain-specific risk: digit ratios (2D:4D and rel2) as predictors of recreational, financial and social risk-taking behaviors. *Pers. Individ. Dif.* 51, 412–416. doi: 10.1016/j.paid.2010.07.003
- Stevenson, J. C., Everson, P. M., Williams, D. C., Hipskind, G., Grimes, M., and Mahoney, E. R. (2007). Attention deficit/hyperactivity disorder (ADHD) symptoms and digit ratios in a college sample. *Am. J. Hum. Biol.* 19, 41–50. doi: 10.1002/ajhb.20571
- Talarovicová, A., Kršková, L., and Blažeková, J. (2009). Testosterone enhancement during pregnancy influences the 2D:4D ratio and open field motor activity of rat siblings in adulthood. *Horm. Behav.* 55, 235–239. doi: 10.1016/j.yhbeh.2008.10.010
- Tester, N., and Campbell, A. (2007). Sporting achievement: what is the contribution of digit ratio? *J. Pers.* 75, 663–678. doi: 10.1111/j.1467-6494.2007.00452.x

- Turanovic, J. J., Pratt, T. C., and Piquero, A. R. (2017). Exposure to fetal testosterone, aggression and violent behavior: a meta-analysis of the 2D:4D digit ratio. *Aggress. Violent Behav.* 33, 51–61. doi: 10.1016/j.avb.2017.01.008
- Van den Bergh, B., and Dewitte, S. (2006). Digit ratio (2D:4D) moderates the impact of sexual cues on men's decisions in ultimatum games. *Proc. Biol. Sci.* 273, 2091–2095. doi: 10.1098/rspb.2006.3550
- Voracek, M. (2014). No effects of androgen receptor gene CAG and GGC repeat polymorphisms on digit ratio (2D:4D): a comprehensive meta-analysis and critical evaluation of research. *Evol. Hum. Behav.* 35, 430–437. doi: 10.1016/j.evolhumbehav.2014.05.009
- Williams, T. J., Pepitone, M. E., Christensen, S. E., Cooke, B. M., Huberman, A. D., Breedlove, N. J., et al. (2000). Finger-length ratios and sexual orientation. *Nature* 404, 455–456. doi: 10.1038/35006555
- Zheng, Z., and Cohn, M. J. (2011). Developmental basis of sexually dimorphic digit ratios. *Proc. Natl. Acad. Sci. U S A* 108, 16289–16294. doi: 10.1073/pnas.1108312108
- 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 © 2018 Millet and Buehler. 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.



# Prenatal Temperature Shocks Reduce Cooperation: Evidence from Public Goods Games in Uganda

Jan Duchoslav\*

Development Economics Group, Wageningen University, Wageningen, Netherlands

Climate change has not only led to a sustained rise in mean global temperature over the past decades, but also increased the frequency of extreme weather events. This paper explores the effect of temperature shocks *in utero* on later-life taste for cooperation. Using historical climate data combined with data on child and adult behavior in public goods games, I show that abnormally high ambient temperatures during gestation are associated with decreased individual contributions to the public good in a statistically and economically significant way. A 1 standard deviation rise in mean ambient temperature during gestation is associated with a 10% point decrease in children's cooperation rate in a dichotomous public goods game, and the reduced taste for cooperation lasts into adulthood.

**Keywords:** climate change, temperature shocks, public goods game, cooperation, fetal origins, Africa

## 1. INTRODUCTION

Climate scientists have reached solid consensus that global climate change is occurring over a decade ago (Oreskes, 2005). There has been a sustained rise of mean global temperature, and extreme temperatures have become increasingly common (see **Figure 1**, adapted from Coumou and Rahmstorf, 2012). The focus of scientific discourse on the topic has therefore shifted toward estimating the economic implications of future climate change as well as finding feasible, effective countermeasures and mitigation strategies (Dell et al., 2009). The severity of the former justifies the costs of the latter. Careful assessment of the damage function is thus of utmost importance.

Recent contributions to this literature have investigated the effects of immediate temperature on outcomes ranging from economic production (Dell et al., 2009; Burke et al., 2015) through the onset of conflict (Hsiang et al., 2013), to mortality rates (Barreca et al., 2016) and human reproductive behavior—with consequences for physical health and educational outcomes of the offspring (Wilde et al., 2017), and potentially for overall population growth (Barreca et al., 2015). Inspired by another growing body of literature—that on fetal origins, i.e., the impact of intrauterine conditions during gestation on later-life outcomes—I take a step back and consider behavioral implications of temperature shocks *in utero*. Using historical variation in ambient temperature as a natural experiment, and behavior in an incentivized public goods game as an outcome measure, I assess the impact of unusually high temperatures during gestation on later-life taste for cooperation—a preference essential to much economic production. I find that abnormally high ambient temperatures during gestation significantly reduce cooperativeness in children, and that this effect lasts into adulthood.

Stemming from an observation by the epidemiologist Barker (1990) that low birth-weight and premature birth are associated with coronary heart disease in later life, the fetal origins literature has grown considerably beyond the medical field into other domains including economics, psychology

## OPEN ACCESS

### Edited by:

Ulrich Schmidt,  
Institut für Weltwirtschaft, Germany

### Reviewed by:

Rosemarie Perry,  
New York University, United States  
Sheng Miao,  
Salk Institute for Biological Studies,  
United States

### \*Correspondence:

Jan Duchoslav  
jan.duchoslav@wur.nl

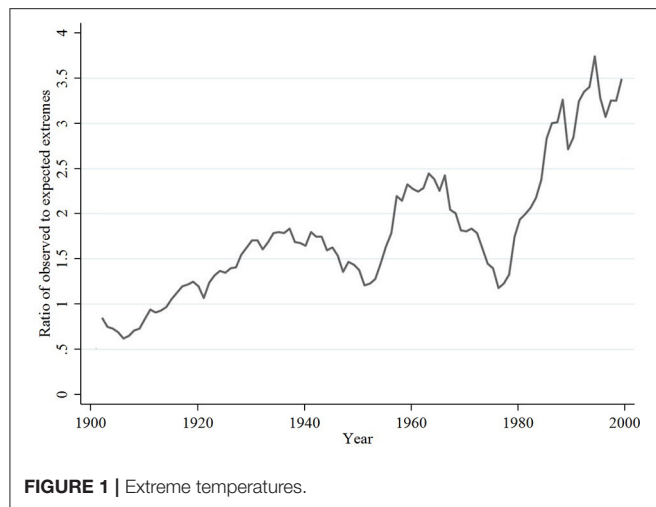
**Received:** 31 August 2017

**Accepted:** 08 December 2017

**Published:** 21 December 2017

### Citation:

Duchoslav J (2017) Prenatal Temperature Shocks Reduce Cooperation: Evidence from Public Goods Games in Uganda. *Front. Behav. Neurosci.* 11:249. doi: 10.3389/fnbeh.2017.00249



or management science. Conditions *in utero* and their proxies have now been linked to later life outcomes ranging from educational achievement (Bhutta et al., 2002; Almond, 2006) through trading ability (Coates et al., 2009) to sexual identity (Csathó et al., 2003). Using the 1918 influenza pandemic as a natural experiment, Almond (2006) finds that mother's illness during pregnancy reduces the educational attainment and income of the offspring. Other natural experiments make use of the Ramadan (Almond and Mazumder, 2011) and the Nazi invasion of the Netherlands (van Os and Selten, 1998) to show that fasting and stress (respectively) during pregnancy increase the chance of mental disability in the offspring. In a similar fashion, Banerjee et al. (2010) use the case of the advancing phylloxera infestation of French vineyards to show that negative income shocks during gestation reduce adult height—a marker of overall health.

In another strand of the fetal origins literature, various markers of conditions *in utero* such as preterm birth, the ratio between the lengths of the index and ring fingers (2D:4D), and especially weight at birth are linked to later-life outcomes. As a direct consequence of intrauterine growth retardation, preterm birth, or both, low birth weight is a telltale sign of adverse conditions *in utero*. The exact nature of the physiological processes that lead to low weight at birth (often collectively referred to as intrauterine programming) are still subject to vigorous scientific debate. There is however growing consensus that they may involve hormonal imbalances in early pregnancy, decreased fetal nutritional intake in late pregnancy (whether due directly to low maternal nutritional intake or to suboptimal placental size, blood flow or function), and low fetal oxygen supply throughout gestation. These can in turn be triggered by conditions as diverse as maternal malnutrition, stress, disease, substance abuse, and environmental exposure (such as to high altitude or ambient temperature) (Fowden et al., 2006). As a general marker of unfavorable intrauterine conditions, low birth weight has been linked to various later-life outcomes ranging from cardiovascular disease (Barker, 1990) to lower income (Black et al., 2007; Bharadwaj et al., 2017), behavioral problems (Hille et al., 2007), and reduced cognitive abilities (Hack et al.,

2005; Figlio et al., 2014), which in turn reduce the taste for cooperation (Moore et al., 1998; Zhang et al., 2015)<sup>1</sup>.

Considering the abundant evidence that ambient temperature during gestation is one of the factors affecting birth weight (Wells and Cole, 2002; Lawlor et al., 2005; Deschênes et al., 2009) and preterm birth (Lajinian et al., 1997; Yackerson et al., 2008; Flouris et al., 2009), its effects on later-life outcomes in general and social preferences in particular have received surprisingly little attention. To be sure, much of this non-experimental strand of fetal origins literature consists of comparative cohort studies without sufficient controls for socioeconomic and behavioral confounders (Black et al., 2007; Dell et al., 2009; Deschênes et al., 2009; Zhang et al., 2015 being noteworthy exceptions), and is therefore prone to suffer from omitted variable bias. Taken as a whole, this body of literature nonetheless points toward a link between ambient temperature during gestation and later-life outcomes. The methodologically well-executed study by Deschênes et al. (2009) (as well as that by Lawlor et al., 2005) further suggests that it is relative—rather than absolute—temperature shocks that matter in this respect<sup>2</sup>.

To my knowledge, the hitherto only study to look at the effect of *in utero* temperature shocks on later-life outcomes links temperature during gestation to depression in adulthood (Adhvaryu et al., 2015). The present paper fills in part of the remaining gap by studying the effects of ambient temperature during gestation on the taste for cooperation. I describe the experimental design and my empirical strategy in section 2, present the results in section 3, and conclude in section 4.

## 2. EXPERIMENTAL DESIGN AND DATA

Employing new data from behavioral games, anthropometric measurements and an extensive socioeconomic survey conducted in Uganda, I exploit the quasi-experimental variation in weather to gauge the impact of prenatal temperature shocks on later-life cooperation.

I use several distinct datasets in my analysis<sup>3</sup>. Temperature data come from Willmott and Matsuura's (2015) gridded monthly time series interpolated from weather station observations. I combine the temperature values with my main and secondary self-collected datasets. The main set contains data from a survey of primary school pupils from Northern Uganda, and also includes their choices in a one-shot dichotomous public

<sup>1</sup>Additionally, preterm birth—another marker of adverse intrauterine conditions—predicts poor educational attainment (Bhutta et al., 2002). The 2D:4D ratio—a marker of prenatal stress—predicts cognitive abilities (Bosch-Domènech et al., 2014), risk preferences (Sapienza et al., 2009; Cronqvist et al., 2016), and prosocial preferences (Buser, 2012; Brañas-Garza et al., 2013; Galizzi and Nieboer, 2015; Cecchi and Duchoslav, 2018).

<sup>2</sup>Wells and Cole (2002) come to a different conclusion, providing between-population evidence that absolute—rather than relative—temperature shocks during gestation drive changes in birth-outcomes. Although they control for various confounding factors such as income and nutritional intake, these are aggregated at the national and yearly levels. Due to this design, the study cannot distinguish between regular seasonal variation in temperature and relative temperature shocks.

<sup>3</sup>See Appendix B in Supplementary Materials for an overview of variable definitions.



goods game, as well as their anthropometric measurements. The secondary set contains data from a household survey from Southern Uganda, and records of the behavior of the representatives of these households in a standard public goods game.

## 2.1. Temperature

I construct my measure of ambient air temperature during gestation using Willmott and Matsuura's (2015) historical time series—one of two publicly available datasets with values spatially interpolated from terrestrial weather station measurements. Although the alternative dataset produced by Harris et al. (2014) is generally more popular, that of Willmott and Matsuura is better suited for my purposes as it uses a much denser set of weather stations in East Africa (as well as globally)<sup>4</sup>. The dataset contains a single temperature value for each historical month and spatial grid cell of  $0.5 \times 0.5^\circ$  (roughly  $55 \times 55$  km in Uganda).

My behavioral and survey data come from two clusters of locations—one in Northern Uganda, spanning four neighboring grid cells ( $2.5\text{--}3.5^\circ\text{N}$ ,  $32.5\text{--}33.5^\circ\text{E}$ ), and one from a single grid cell in Southern Uganda ( $0.5\text{--}1.0^\circ\text{S}$ ,  $30.0\text{--}30.5^\circ\text{E}$ ). Since the differences between the values in the four neighboring northern cells in any given month are minimal, I use their mean values for all observations in the northern cluster, obtaining a single monthly temperature value for each of the two location clusters. Temperature variation within each cluster of location thus stems from temporal, rather than geographical differences. To obtain the value of ambient temperature during the gestation of a respondent in the northern cluster, I average the temperature in the northern location during the month of his or her birth and in the preceding 8 months. The values for respondents from the southern locations are constructed analogously. For an individual born in November, for example, I average the monthly values from March until November.

The historical monthly means are plotted in **Figure 2**, where the gray curves represent the monthly temperature means in the two clusters, the red curve represents the temperatures in

the months in which the respondents in the northern cluster (main sample) were gestating, and the blue curve denotes the temperatures in the months in which the respondents in the southern cluster (secondary sample) were gestating. The mean values of ambient air temperature during the gestation (9 months) of individual respondents are denoted by black circles.

Mean temperatures of 9-month-long gestational periods have, by construction, a much smaller variance than monthly mean temperatures (as **Figure 2** illustrates). Similarly, the variance of monthly means is smaller than that of daily means. Basing my analysis on overall mean values therefore somewhat reduces its sensitivity. Using more detailed temperature data such as a set of 9 monthly values for each individual would not, however, correspond to the level of precision with which I can determine the dates of conception—and thus the periods of gestation—of the respondents. By its nature, I can only infer an individual's probable date of conception from their date of birth. The possibility of premature and late births introduces in such inference a level of uncertainty which is only aggravated by the fact that I only know the month (rather than the exact date) of birth of my respondents. The margin of error associated with these imprecisions can easily be more than a month. In extreme cases, there would thus be no overlap between actual and assumed values of temperatures in any given month of gestation. Using instead the mean value over the whole assumed period of gestation largely reduces the effect of such inaccuracies.

## 2.2. Main sample

My main sample consists of 531 children and their caregivers from Pader district in Northern Uganda. The children come from 42 primary schools visited in June and July 2014. In each school, 16 pupils were randomly selected from a list of those enrolled at the beginning of the year<sup>5</sup>.

I measure children's and caregivers' willingness to cooperate by involving them in a one-shot dichotomous public goods game similar to those in Cárdenas et al. (2009) and Barr et al. (2014). In each school, children were randomly assigned to groups of 8, but were not told which other 7 children (of the 15 participating in that school) belonged to their group. Each child then anonymously selected either a "private card" or a "group card"<sup>6</sup>. By choosing the private card, the respondent allotted 4 candies to himself, but none to the other unknown members of the group. By selecting the group card, the respondent instead ensured 1 candy for each of the 8 group members, including himself (see Appendix C in Supplementary Materials for a reproduction of the two cards). In this set up, total welfare is maximized when all 8 game participants opt for the group card, such that they each receive 8 candies. A sole free rider selecting the private card would receive 11 candies, but in the

<sup>4</sup>The dataset produced by Harris et al. (2014) at the Climate Research Unit at the University of East Anglia is clearly more popular than that assembled by Willmott and Matsuura (2015) at the University of Delaware. While Willmott and Matsuura's dataset has been cited by 149 studies since its publication in 2009, the CRU dataset has been cited 751 times since its publication 5 years later according to Google Scholar. In 2015 alone, the CRU dataset was cited in 226 publications on Africa, compared to only 20 citations of Willmott and Matsuura (including citations of previous versions of their dataset). I find the preference for Harris et al. odd, given that their data is based on a much sparser set of weather stations than Willmott and Matsuura's. Willmott and Matsuura use measurements from between 3 and 19 (on average 8) weather stations within a  $5^\circ$  (about 555 km) radius of my main research site in Northern Uganda, and between 3 and 19 (with an average of 6) stations within a  $5^\circ$  radius from my secondary research site in Southern Uganda, depending on the month and year of measurement. Harris et al.'s use only 0 to 4 (on average 3) and 0 to 2 (on average 1) stations respectively in the same regions, and until 1941, the nearest weather station used in their dataset was Harare— $21^\circ$  ( $2,300$  km) from my main research site and  $17^\circ$  ( $1,900$  km) from my secondary research site. Willmott and Matsuura's data predict over 25% of the variation in actual temperature data from Entebbe, Uganda, while Harris et al.'s predict less than 7%. I therefore consider Willmott and Matsuura (2015) superior to Harris et al. (2014) in the East African context, despite the overwhelming popularity of the latter.

<sup>5</sup>Out of a total of 672, the caregivers of 141 pupils did not know their children's birth date. These children were excluded from my analysis.

<sup>6</sup>Contrarily to many public goods games in which participants can choose their preferred contribution level, I opted for a dichotomous choice, effectively reducing the game to a prisoner's dilemma: respondents could either cooperate or not. While this reduced my ability to pick up the nuances present in the experimental sample, I believe that it facilitated the decision making process, especially for the youngest.

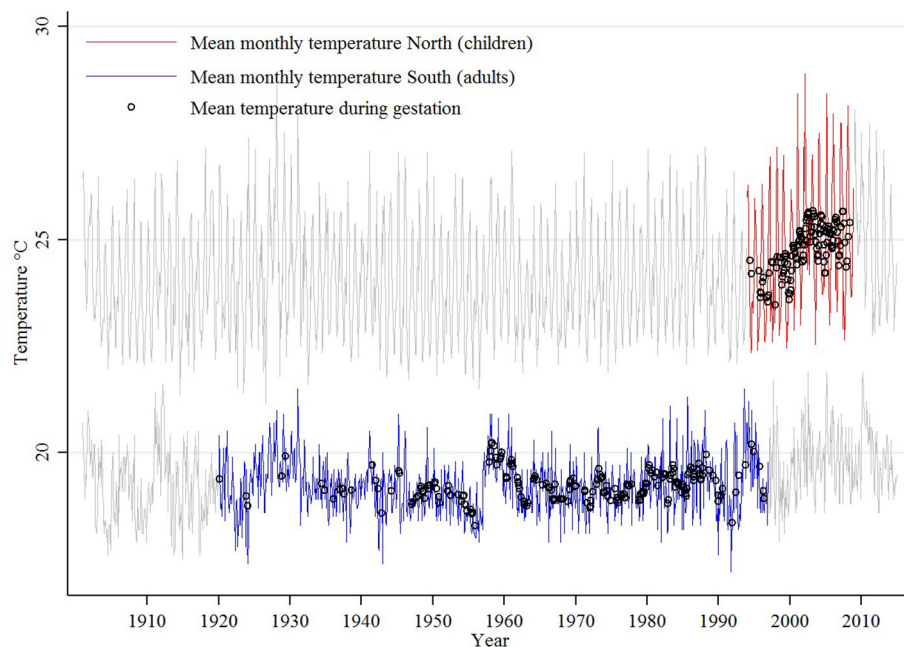


FIGURE 2 | Temperature trends.

Nash equilibrium, everyone selects the private card and ends up with only 4 candies each.

Caregivers played a similar public goods game, but made their decisions in the isolated environment of their home, unaware of the identity of the other 8 participants with whom they were grouped. If they chose the private card, they received 4,000 UGX (roughly 1.5 USD at the time). Choosing the group card instead meant an allocation of 1,000 UGX to each anonymous member of the group, including themselves. In the Nash equilibrium, each participant thus received 4,000 UGX, total welfare was maximized at a return of 8,000 UGX for each group member, and a sole free rider would earn 11,000 UGX. 26% of the children chose the cooperative option, while the cooperation rate among their caregivers was 34%.

The descriptive statistics for the children are presented in **Table 1A**. The mean ambient temperature faced by the mothers of the children in my sample during their pregnancy was 24.9°C. The children are on average 10 years old, and girls and boys are equally represented. The height of the children in the sample is practically identical to the mean for their age, but their body mass is 1.39 standard deviation below the mean for their age (de Onis et al., 2007)<sup>7</sup>. This suggests that some may have been nutritionally deprived in their early life, which could confound my results (I address this issue below). Children's cognitive ability was measured through standard Raven's progressive matrices (Kaplan and Saccuzzo, 2012). It is an intelligence quotient (IQ)

adjusted for age and scaled relative to the sample (with a mean at 100 and a standard deviation of 15). I further proxy for child prenatal stress by the second-to-fourth (2D:4D) digit ratio—a marker of hormonal exposure *in utero*<sup>8</sup>.

Child postnatal conflict exposure—a potentially important confounding factor considering that most children in the sample were born during a period of civil war in Northern Uganda—is a composite measure derived from the exposure of the caregiver and the child's year of birth. Given their young age at the time of the conflict, children were not asked any war-related questions. Instead, I use caregiver responses to an adapted version of the War Trauma Questionnaire (Macksoud, 1992; Papageorgiou et al., 2000)<sup>9</sup>. It consists of 23 yes–no questions about various violent events witnessed by the caregiver, from which I construct a conflict exposure index using the number of positive responses as a measure of exposure (Bellows and Miguel, 2009) and normalizing it for the sample. To proxy the child's postnatal

<sup>7</sup>Based on WHO recommendations for treating outliers (de Onis et al., 2007), I truncate the anthropometric data at 6 standard deviations from the mean. This results in 15 and 14 dropped observations for height-for-age and BMI-for-age respectively.

<sup>8</sup>The lengths of the index and ring fingers were measured on the palmar surface of the right hand, from the midpoint of the palmar digital crease to the tip of the finger. The state of the art in measuring finger lengths is to use an office scanner to take a perfectly flat image of the palmar surface of the hand, and computer software to measure the exact lengths. Given the constraints due to the remoteness of the field location, I instead used clipboards and tape measures, allowing only for precision to the nearest 1 mm. This resulted in measurement error of  $\pm 33.3\%$  at the mean of the estimates. A pilot in which 30 raters each separately measured the digit lengths of 35 individuals revealed comparable margins of error. While the precision of this measurement is still well below that obtained in laboratory settings (see Voracek et al., 2007), my measurements should be at least as accurate as those in other field studies which sometimes only report whether the index finger is longer, shorter, or of the same length as the ring finger (Buser, 2012).

<sup>9</sup>Any questions about shelling and bombardment are irrelevant in the Ugandan setting, and were therefore omitted from the questionnaire.

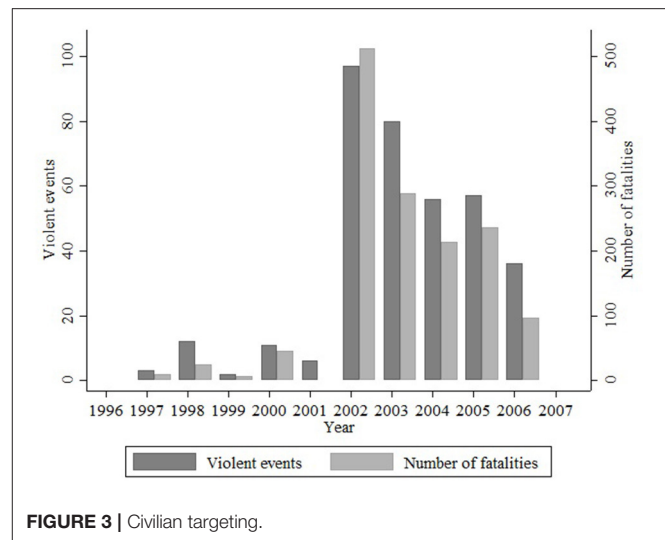
**TABLE 1** | Descriptive statistics (main sample).

	Mean	Std. dev.	Min.	Max.	N
<b>(A) CHILDREN</b>					
Temperature	24.915	0.505	23.464	25.681	531
Cooperation	0.256	0.437	0	1	531
Female	0.493	0.5	0	1	531
Age in years	10.478	2.731	6	20	531
Height-for-age	−0.025	1.005	−3.391	3.571	516
BMI-for-age	−1.387	0.935	−3.99	2.26	517
IQ-for-age	100.619	14.815	73.993	148.458	488
Prenatal stress	0	1	−3.273	5.028	215
Postnatal conflict exposure	0	1	−4.971	2.885	477
Precipitation	3.507	0.559	2.401	5.124	531
Consumer prices	−2.113	2.163	−5.954	4.937	531
<b>(B) CAREGIVERS</b>					
Age in years	40.588	9.738	20	94	522
Age at birth of child	30.207	9.545	10	85	522
Female	0.58	0.494	0	1	531
Acholi	0.976	0.155	0	1	531
Christian	0.998	0.043	0	1	531
Years of education	3.84	3.431	0	18	531
Married	0.772	0.42	0	1	531
Functional literacy	0.234	0.423	0	1	531
Conflict exposure	0	1	−3.035	1.688	528
Household size	7.945	2.491	2	20	531
Wealth	0	1	−1.54	3.068	531
Risk aversion	0.516	0.5	0	1	531
Cooperation	0.335	0.473	0	1	531

conflict exposure I weight the caregiver's conflict exposure index by the portion of violence their child could have potentially witnessed after birth. To obtain the weights, I divide the number of civilian fatalities that occurred in Pader district following the child's birth by the total number of civilian fatalities recorded in the district throughout the length of the conflict (**Figure 3**)<sup>10</sup>. For example, a child born in December 2003—by which time 62% of reported fatalities took place—whose caregiver's conflict exposure is 87% is likely to have witnessed 38% of the violence that the caregiver was exposed to. For my purposes, the child's conflict exposure index would therefore be 33% ( $((1 - 0.62) \times 0.87 = 0.33)$ ).

Finally, precipitation and consumer prices during gestation are constructed analogously to the temperature variable.

To account for further environmental and genetic effects on preferences, I also interviewed each child's main caregiver—the adult household member with whom the child spends most time. The descriptive statistics for the caregivers are presented in **Table 1B**. About half of the caregivers in my sample chose to cooperate in the public goods game. Caregivers are on average 41 years of age<sup>11</sup>, 58% are female. Additionally, I collected

**FIGURE 3** | Civilian targeting.

information about their education level and risk preferences. All caregivers were exposed to at least some kind of conflict-related violence, though the level of exposure varies greatly<sup>12</sup>. Almost all respondents are Christian and belong to the Acholi ethnic group. A typical household is composed of 8 people. I also collected information about their relative asset wealth (Sahn and Stifel, 2003).

In my setting, information about the current main caregivers can only serve as a proxy for environmental and genetic influences to which the children have been subjected throughout their lives. Of the 531 caregivers in my sample, only 265 are biological mothers of the children, while 206 are their biological fathers. The remaining 60 were grandparents, uncles or aunts, other relatives, and siblings (in descending order of prevalence). One caregiver was not related to the child at all. Nonetheless, the average caregiver in my sample had been taking care of the child for 82% of the child's life, making the information about the caregivers a strong proxy for the environment surrounding the children.

## 2.3. Secondary Sample

My main dataset contains rich information about the children and their environment, but suffers from two important shortcomings. The first is its conflict setting. If temperature shocks invite conflict (O'Loughlin et al., 2012; Hsiang et al., 2013), then the physiological effects of ambient temperature during gestation would be hard to disentangle from the effects of temperature-induced conflict. Second, it does not allow me to repeat the analysis using the caregiver's behavior and temperature during their gestation, because I only know the caregivers' year of birth. This means that I cannot tell whether the behavioral effects that temperature shocks *in utero* have on children last into

<sup>10</sup>Source: ACLED Version 5, 1997–2013 (Raleigh et al., 2010).

<sup>11</sup>Nine caregivers did not know their age, reducing the number of observations to 522.

<sup>12</sup>Three caregivers refused to complete the conflict exposure module of the survey, reducing the number of observations to 528. This reduction carries over to the measure of child postnatal conflict exposure, which is derived from that of their caregiver.

adulthood. To address these concerns, I turn to a second sample of 257 adults from Sheema district in Southern Uganda, which was untouched by the conflict in the north.

In July and August 2014 I visited 45 villages in the district, and surveyed a random sample of 10 households per village selected from a census. A randomly selected adult representative of each surveyed household was invited to participate in an incentivized public goods game<sup>13</sup>.

The game was played in groups of 5 participants who could anonymously decide to contribute between 0 and 5 tokens (worth 1,000 UGX or 0.38 USD each) to a common pot, keeping the rest for themselves. Shared funds were doubled and redistributed equally (after rounding). After an initial practice round, 5 rounds of the game were played with each group, though the participants did not know beforehand how many rounds the game would last<sup>14</sup>. One round was selected at random for payment.

In this design, total welfare is maximized when all participants contribute their entire endowment of 5 tokens to the common pot, receiving 10 each in return. Nevertheless, free riders could receive up to 13 tokens, and the Nash equilibrium is reached with all players keeping their 5 tokens. On average, participants contributed 3.44 tokens to the public good. The descriptive statistics for the game participants are presented in **Table 2**. The mean ambient temperature faced by the mothers of the adults in my sample during their gestation was 19.2°C. The participants are on average 42 years old, a third are female, and 84% are married. On average, they fell just short of completing primary education, and half are functionally literate. Nearly the whole sample is ethnically Ankole and Christian by religion.

### 3. ANALYSIS AND RESULTS

#### 3.1. Main Finding

I hypothesize that exposure to high ambient temperatures during an individual's gestation may impact his or her later-life preference for cooperation. Combining the findings of Wells and Cole (2002), Lawlor et al. (2005), and Deschênes et al. (2009) with those of Hack et al. (2005) and Zhang et al. (2015), I expect prenatal exposure to high ambient temperatures to reduce cooperative behavior. I analyze this relationship by fitting the following linear probability model (LPM):

$$\Pr(\text{Cooperation}_{iym} = 1 | \text{Temperature}_{ym}, \mathbf{x}_{iym}, \mathbf{z}_{iym}) = \alpha + \beta \text{Temperature}_{ym} + \gamma' \mathbf{x}_{iym} + \delta' \mathbf{z}_{iym} + \zeta_m + \eta_s + \varepsilon_{iym} \quad (1)$$

where  $\text{Cooperation}_{iym}$  equals 1 if child  $i$  born in month  $m$  of year  $y$  and attending school  $s$  selects the cooperative option,  $\text{Temperature}_{ym}$  is the mean ambient temperature during the likely gestation of children born in month  $m$  of year  $y$ ,  $\mathbf{x}_{iym}$  is a vector of individual child characteristics (female, age,

**TABLE 2 |** Descriptive statistics (secondary sample).

	Mean	Std. dev.	Min.	Max.	N
Temperature	19.222	0.347	18.278	20.233	257
Contribution	3.437	1.174	0	5	257
Age in years	42.004	15.711	16	92	257
Female	0.335	0.473	0	1	257
Years of education	6.743	3.866	0	17	257
Married	0.844	0.363	0	1	257
Munyankole	0.977	0.151	0	1	257
Christian	0.996	0.062	0	1	257
Functional literacy	0.498	0.501	0	1	257
Wealth	0	1	-1.212	5.515	257

age  $\times$  female),  $\mathbf{z}_{iym}$  is a vector of caregiver characteristics (female, age, age  $\times$  female, Acholi, years of education),  $\zeta_m$  are month-of-birth fixed effects,  $\eta_s$  are school fixed effects, and  $\varepsilon_{iym}$  is a stochastic error term. Standard errors are clustered at the level of running month of birth.

Estimating the model without controls, I find that exposure to high ambient temperature during gestation is negatively correlated with the child's probability of contribution to the public good. Parametrically, a 1°C increase in mean ambient temperature during gestation reduces the child's probability of contribution by 7.6% points (**Table 3**, column 1). At mean prevalence of 25.6%, this is equivalent to a 30% reduction in the likelihood of cooperation.

To account for non-temperature seasonal confounds and unobserved background characteristics potentially related to season of birth similar to those described by Buckles and Hungerman (2013) in the United States, I include calendar month fixed effects, which only increases the magnitude of the detected effect of temperature (**Table 3**, column 2). The relationship could potentially also be driven by other child characteristics. Prosocial preferences develop throughout childhood and adolescence, and become increasingly gender-dependent with approaching adulthood (Eisenberg et al., 2006). Controlling for age, gender and their interaction, however, does not change the interpretation of the result (**Table 3**, column 3), nor does controlling for caregiver characteristics and school fixed effects to account for family and peer demographics (**Table 3**, columns 4 and 5), both of which have been linked to children's prosocial behavior (Eisenberg et al., 2006).

#### Result 1

Exposure to abnormally high ambient temperature during gestation decreases later-life taste for cooperation. A 1°C (1 s.d.) increase in mean ambient temperature during gestation decreases the probability of cooperation in a public goods game by up to 20% points (10% points), leading to a 16% (8%) drop in total welfare.

The result holds when subjected to a battery of robustness checks. It remains practically unchanged when estimated by probit and logit models (see Table A.1 in the Supplementary

<sup>13</sup>Out of a total of 450 randomly selected household representatives, 193 either did not know their month and year of birth, or did not show up to play the public goods game. These people were excluded from my analysis.

<sup>14</sup>Withholding the information about the exact length of the game helps ensure that all rounds, including the final one, are played in a strategic way.



**TABLE 3 |** High temperature decreases taste for cooperation.

	Child cooperation				
	(1)	(2)	(3)	(4)	(5)
Temperature	−0.076** (0.038)	−0.134** (0.055)	−0.140** (0.069)	−0.138* (0.073)	−0.204*** (0.069)
Month of birth FE	N	Y	Y	Y	Y
Child characteristics	N	N	Y	Y	Y
Caregiver characteristics	N	N	N	Y	Y
School FE	N	N	N	N	Y
N	531	531	531	522	522
Area under ROC curve	0.556	0.612	0.618	0.624	0.756

Notes: SE clustered at the level of running month of birth in parentheses. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . Child characteristics: Female, Age (in months), Age  $\times$  Female. Caregiver characteristics: Female, Age (in years), Age  $\times$  Female, Acholi, Years of education.

Materials). It is not driven by outliers—excluding observations with high-leverage (the most extreme values of independent variables) and high-influence observations (observations whose deletion from the dataset would most change the magnitude of the estimated coefficients) does not significantly affect the result (see Table A.2, columns 2 and 3 in Supplementary Materials). Limiting the analysis to children born in the same area where they were interviewed also does not affect the result (see Table A.2, column 4 in Supplementary Materials). Including the mean air temperature during a 9-month period 1 year prior to the assumed period of gestation as a placebo treatment leaves the result unaffected, as does assuming other periods of gestation<sup>15</sup>, and using an alternative source of temperature data (see Table A.3 in Supplementary Materials).

To better understand the main result, I estimate temperature effects on later life cooperation for each pregnancy trimester. This not only serves to better pinpoint the critical period of exposure, but also provides an indication of the potential mechanisms at play. The first trimester is crucial to brain development, and it is the time when epigenetic programming of the endocrine system takes place. The third trimester, when the fetus gains the most weight, is crucial for general health. It is clear from Figure 4, which shows the effects of mean ambient temperature in each trimester (and their 90% confidence intervals), that the result is driven mainly by exposure in the first trimester. The magnitude of the effect of temperature shocks in the first gestational trimester is about twice as large as those of temperature shocks in the second and third trimesters. The effect in the first trimester is also the only statistically significant one in my estimation ( $p = 0.05$ ), but that could well be due to my underpowered

<sup>15</sup>In my analysis, I assume the period of gestation to correspond with the calendar month of the respondent's birth and the previous 8 months. To obtain the mean temperature during the gestational period of a respondent born in November, for example, I average the mean temperatures from March until November. As discussed in section 2.1, this is quite a simplification. A child born full term on the 1st of November would have gestated between February and October, while a child born 1 month prematurely on the 30th of November would have gestated between April and November. As a robustness check, I re-estimate model (1) using these two extremes as alternative individual regressors.

estimation<sup>16</sup>. Additively, they make up the overall temperature effect throughout gestation. The fact that most of the effect seems to be concentrated in the first trimester suggests that the observed behavioral effects may be linked directly to altered brain development, changes in endocrine regulation, or both, rather than indirectly to general health.

### 3.2. Indirect Temperature Effects and Other Factors

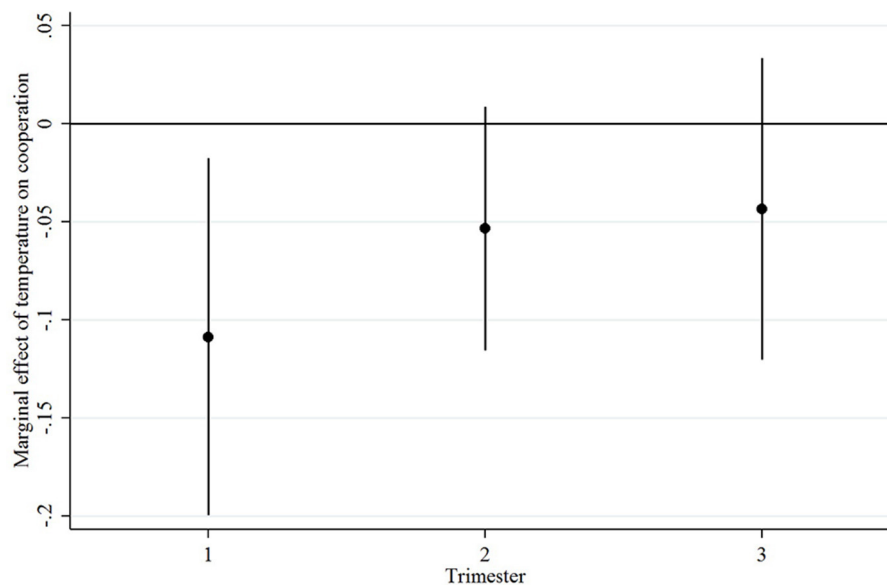
Both in theory and in my data, temperature is strongly negatively correlated with precipitation, which in turn affects agricultural yields and—by extension—food prices. The combination of high temperatures and low precipitation during gestation could thus lead to malnutrition in infancy, whose negative consequences for the child's cognitive abilities can persist for years (Beckett et al., 2006). On the other hand, low precipitation levels decrease the likelihood of malaria contraction (Craig et al., 1999), and could thus also have a positive effect on later-life outcomes (Barreca, 2010).

Controlling for the environmental covariates and the indicators of early-life deprivation, I find that high precipitation during gestation decreases children's taste for cooperation, while high consumer prices increase it (Table 4, columns 1 and 3). This suggests that—at least in the context of Northern Uganda—the effects of precipitation during gestation on later-life prosocial preferences via exposure to malaria dominate those via agricultural yields, and that—unsurprisingly—the mean sampled household is likely to be a net food producer. Importantly, however, precipitation and consumer price effects do not wash away the effect of temperature itself (Table 4, columns 4 and 6).

High cognitive abilities proxied by the age-adjusted IQ predict higher probability of contributing to the public good in accordance with Zhang et al. (2015). From Beckett et al. (2006), I would expect height-for-age and BMI-for-age—both markers of early-life nutritional deprivation—to also be positively correlated with child cooperation. Instead, I estimate their effects to be statistically insignificant and significantly negative respectively (see Table 4, columns 2 and 3). Their inclusion in the model does not however alter my main result (Table 4, columns 5 and 6).

There is increasingly conclusive evidence that high temperatures may trigger or intensify violent conflict

<sup>16</sup>Since I estimate effect sizes as beta coefficients in a multivariate regression in one sample, I cannot directly analyze the power of the estimation in the sense of the probability of detecting a difference in the proportions or means of two samples, nor the related minimal detectable effect given the sample size. To get a rough idea of the power of the estimation, I split the sample at the median value of *Temperature*, and consider the half with high values of *Temperature* as shocked and the rest as not shocked (effectively recoding *Temperature* as a binary variable). Assuming these to be a treatment and a control group in an experiment, and requiring power of 0.80 and significance of 0.05, I could only detect a difference between the cooperation rates in the two groups 1.7 times larger than the observed one (or 1.5 times larger if I could use the full sample of 672 observations). Allowing for full variation in *Temperature* and using additional controls in the multivariate regression setting should improve the power of the estimation, but still likely leaves it far below ideal. To be sure, this does not mean that the estimates which I find to be statistically significant are not so. Rather, it means that I cannot rule out with sufficient certainty that the coefficients which seem to be statistically insignificant in my estimations are not in fact different from zero.



**FIGURE 4** | Temperature effects are strongest in the first trimester.

(O'Loughlin et al., 2012; Hsiang et al., 2013). Pre- and post-natal exposure to conflict have in turn been found to influence social preferences: Conflict-induced prenatal stress reduces contributions to the public good in later life (Cecchi and Duchoslav, 2018), while post-natal exposure leads to more prosocial behavior within close networks (Voors et al., 2012; Bauer et al., 2014; Gilligan et al., 2014). Many of the children in my sample were born during a period of civil war in Northern Uganda. Using a sub-sample for which information on war exposure and prenatal stress is available,<sup>17</sup> I find that prenatal stress (proxied by a z-score of the reverse 2D:4D ratio—a marker of prenatal stress) indeed reduces the taste for cooperation (Table 5, column 1). Unlike other studies (Voors et al., 2012; Bauer et al., 2014; Gilligan et al., 2014), I find no statistically significant relationship between postnatal conflict exposure and cooperation, though this could be due to the crudeness of my measure of conflict exposure (see section 2.2 for details). Importantly, the inclusion of these war-related controls does not wash away the effect of ambient temperature during gestation; it rather makes it stronger (Table 5, column 2).

The preferences of children may be influenced by those of their caregivers through both environmental and—when the two are blood related—genetic mechanisms (Dohmen et al., 2012). Controlling for caregiver preferences, I find that a child's social preferences are strongly correlated with the social preferences of

their main caregiver, but not with the caregiver's risk preferences. Children are about 10% points more likely to contribute to the public good if their main caregiver contributes to it as well in a separate game (Table 6, column 1). The effect of ambient temperature during gestation is however not affected by these controls (Table 6, column 2), and the results hold when analysis is restricted to caregivers who are biological parents of their children (Table 6, columns 3 and 4).

Finally, it is conceivable that different types of parents are more likely to conceive at times with different weather and climate patterns. If the different types of parents would also have different social preferences, such self-selection could bias my results. In my setting, much of any such bias should be absorbed by the month of birth fixed effects. To further verify that no self-selection bias is present, I regress a battery of caregiver characteristics on mean temperature during the child's gestation according to the following model:

$$y_{iym} = \alpha + \beta Temp_{ym} + \gamma' x_{iym} + \delta_m + \zeta_s + \varepsilon_{iym} \quad (2)$$

where  $y_{iym}$  refers to one of the following characteristics of the caregiver of child  $i$  born in month  $m$  of year  $y$  in village  $s$ : gender, marital status, functional literacy, risk aversion, public goods game choice, age at birth of child, years of education, conflict exposure, wealth, and household size. All other notation is the same as above.

If parents did not self-select into conceiving at the onset of a particularly hot (or cold) 9-months period based on these characteristics, the estimated  $\beta$  coefficients should be statistically insignificant. I summarize the estimated  $\beta$  coefficients and their 95% confidence intervals in Figure 5. As expected, none is statistically different from zero, indicating no detectable parent self-selection bias.

<sup>17</sup>Part of the children in this study were also interviewed in 2012, at which time I measured the lengths of their fingers to calculate the 2D:4D ratio. I made the same measurements for this study, but after explaining to the research assistants that the digit ratio is "usually around 1", the frequency of precisely that value being reported increased dramatically. While I do not believe that this was a result of intentional misenumeration, it does constitute a heavy bias, forcing me to discard the 2014 2D:4D values. This reduced the available sample to those children interviewed in 2012.

**TABLE 4 |** Other environmental factors and early life deprivation.

	Child cooperation					
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature				−0.205** (0.087)	−0.197*** (0.070)	−0.221*** (0.082)
Precipitation	−0.079 (0.048)		−0.102** (0.050)	−0.090* (0.050)		−0.116** (0.050)
Consumer prices	0.020** (0.009)		0.019** (0.009)	0.004 (0.011)		0.001 (0.010)
Height-for-age		0.023 (0.021)	0.026 (0.021)		0.020 (0.020)	0.024 (0.020)
BMI-for-age		−0.039* (0.022)	−0.048** (0.023)		−0.044* (0.022)	−0.048** (0.023)
IQ-for-age		0.004** (0.001)	0.004*** (0.001)		0.004*** (0.001)	0.004*** (0.001)
Child and caregiver characteristics	Y	Y	Y	Y	Y	Y
Month of birth & School FE	Y	Y	Y	Y	Y	Y
N	522	468	468	522	468	468
Area under ROC curve	0.751	0.782	0.787	0.760	0.792	0.798

Notes: Probit marginal effects. SE clustered at the level of running month of birth in parentheses. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . Child characteristics: Female, Age (in months), Age  $\times$  Female. Caregiver characteristics: Female, Age (in years), Age  $\times$  Female, Acholi, Years of education.

**TABLE 5 |** Conflict exposure.

	Child cooperation	
	(1)	(2)
Temperature		−0.250** (0.109)
Prenatal stress	−0.050* (0.027)	−0.052* (0.027)
Postnatal conflict exposure	−0.013 (0.023)	−0.015 (0.022)
Child characteristics	Y	Y
Month of birth and School FE	Y	Y
N	211	211
Area under ROC curve	0.869	0.883

Notes: \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . SE clustered at the level of running month of birth in parentheses. Child characteristics: Female, Age, Age<sup>2</sup>, Age  $\times$  Female, Age<sup>2</sup>  $\times$  Female.

## Result 2

The relationship between ambient temperature during gestation and cooperation is stable and robust to controlling for other environmental factors, early life deprivation markers, pre- and post-natal conflict exposure and caregiver preferences.

### 3.3. Long-Term effects

To gauge the long-term effects of ambient temperature shocks during gestation on the taste for cooperation and to test the external validity of my main finding, I apply a similar analytical approach to a sample of adults from a different part of the country playing a different type of public goods game. I first fit

**TABLE 6 |** Caregiver preferences.

	Child cooperation			
	All caregivers (1)	(2)	Biological parents only (3)	(4)
Temperature		−0.200*** (0.070)		−0.191** (0.075)
Caregiver cooperation	0.104** (0.045)	0.103** (0.044)	0.110** (0.047)	0.113** (0.046)
Caregiver risk aversion	−0.028 (0.037)	−0.024 (0.037)	−0.004 (0.042)	0.002 (0.041)
Child characteristics	Y	Y	Y	Y
Month of birth & School FE	Y	Y	Y	Y
N	522	522	462	462
Area under ROC curve	0.752	0.762	0.740	0.752

Notes: SE clustered at the level of running month of birth in parentheses. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . Child characteristics: Female, Age (in months), Age  $\times$  Female. Caregiver characteristics: Female, Age (in years), Age  $\times$  Female, Acholi, Years of education.

the following OLS model:

$$\text{Contribution}_{iym} = \alpha + \beta \text{Temperature}_{ym} + \gamma' \mathbf{x}_{iym} + \delta_m + \zeta_s + \varepsilon_{iym} \quad (3)$$

where  $\text{Contribution}_{iym}$  represents the average amount of tokens contributed to the public good by participant  $i$  born in month  $m$  of year  $y$  and living in village  $s$ .  $\text{Temperature}$  has the same meaning as above.  $\mathbf{x}_{iym}$  is a vector of personal characteristics of participant  $i$  born in month  $m$  of year  $y$  and living in village  $s$ , which is comprised of Age (age in months), Female (a dummy equal to 1 if the participant is female), and their interaction.  $\delta_m$  are month of year fixed effects,  $\zeta_s$  village fixed effects, and  $\varepsilon_{iym}$  is a stochastic error term. Standard errors are clustered at the levels of running month of birth and game group.

Estimating the model both with and without controls, I find a negative and statistically significant effect of ambient temperature during gestation on contribution to the public good, with every 1°C increase in temperature lowering contributions to the public good by nearly 0.5 token or some 13% (Table 7). The result is robust to outlier exclusion (see Table A.4 in the Supplementary Materials) as well as to a placebo test by older temperatures (see Table A.5 in the Supplementary Materials).

## Result 3

The negative effects of exposure to unusually high ambient temperature during gestation on later-life taste for cooperation last into adulthood. A 1°C (1 s.d.) increase in mean ambient temperature during gestation decreases contributions to the public good by about 13% (5%), thus decreasing total welfare by 6% (2%).

## 4. DISCUSSION AND CONCLUSION

When Montesquieu (1748) wrote that excess heat makes people “slothful and dispirited,” he pointed out that the fact is often

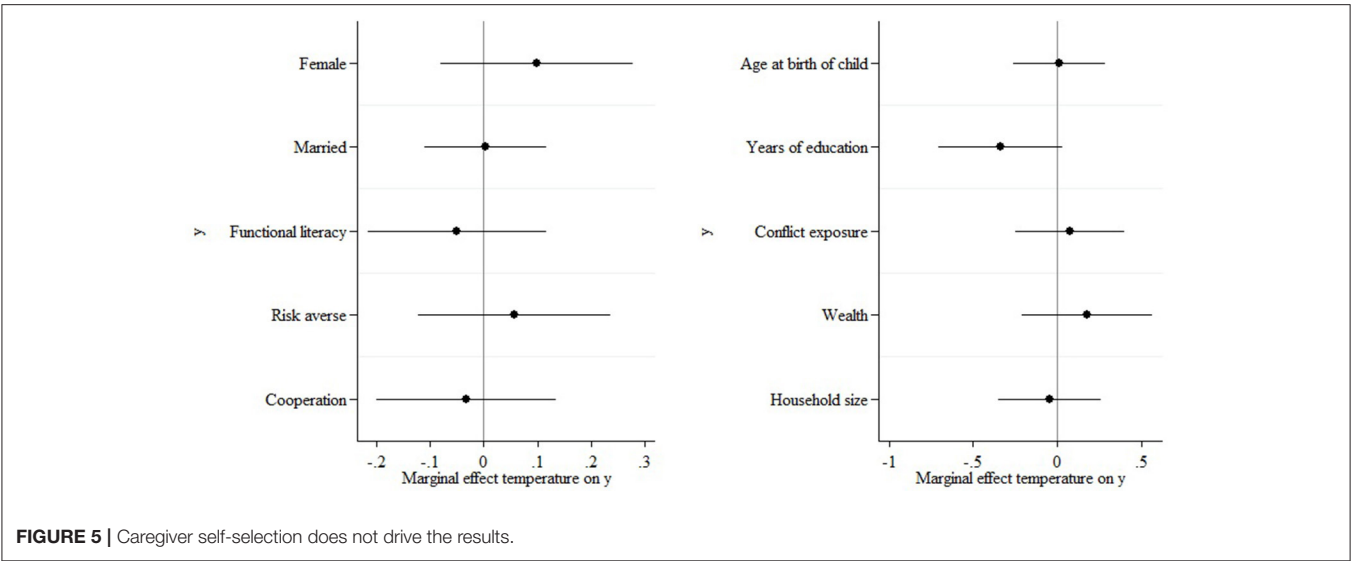


TABLE 7 | Long-term effects.

	Contribution	
	(1)	(2)
Temperature	−0.341* (0.193)	−0.468** (0.204)
Personal characteristics	N	Y
Month of birth & Village FE	N	Y
N	257	257
adj. R <sup>2</sup>	0.006	0.140

Notes: SE clustered at the level of running month of birth and game group in parentheses. \**p* < 0.10, \*\**p* < 0.05, \*\*\**p* < 0.01. Personal characteristics: Female, Age, Age × Female.

used as a justification for slavery. It is perhaps due to the negative connotations of this argument that few social scientists studied the effects of heat on human behavior until quite recently. With global climate change driving temperatures to historically unprecedented levels, this attitude has drastically shifted.

There is now some cross-country evidence suggesting that prevailing extreme temperatures negatively affect health outcomes (Wells and Cole, 2002), and hamper economic production (Burke et al., 2015). Looking exclusively at such cross-country studies, one could be tempted to conclude that it is absolute temperature that drives health and behavioral changes, and that geographical location largely predetermines health and economic outcomes. In such a world, children in tropical countries would be born underweight (Wells and Cole, 2002), suffer from the various negative consequences of poor birth outcomes (Black et al., 2007), and grow up in inefficient economies (Burke et al., 2015). In the context of this study, they would become less cooperative than their luckier counterparts from more temperate climates.

Within-country analyses, however, paint a more complex picture. Due to their longitudinal nature, they have to control for any trends and seasonal patterns not associated with temperature (typically by including time fixed effects in their models), effectively netting out seasonal and long-term temperature patterns as well. Their findings suggest that unexpected deviations from normal temperatures—rather than absolute temperatures—are responsible for observed health and behavioral changes (Dell et al., 2009; Deschênes et al., 2009; Hsiang et al., 2013). In the context of this paper, one would thus expect a person born in an unusually warm year in Northern Uganda to be less cooperative than their neighbor born in an unusually cold year. One would, however, not know whether they should be more or less cooperative than somebody born on the same day in North Holland, for example.

Relying on longitudinal data from two locations in Uganda, I follow Dell et al.’s (2009) recommendation to include time fixed effects in this paper. I find that exposure to higher than normal ambient temperatures during gestation reduces the probability that a child contributes to the public good. The estimated effect is large, and lasts into adulthood. It is most pronounced in the first gestational trimester, which is consistent with the hypothesis that the mechanism through which temperature shocks during gestation alter later-life behavior is linked directly to altered brain development, changes in endocrine regulation, or both, rather than indirectly to general health. Due to the reduced form of this study, I cannot unfortunately make any conclusive claims in this regard, and I leave the establishment of precise causal links to future research into the physiological mechanisms of intrauterine programming—a topic of vigorous scientific debate. I do, however, show a clear correlation between abnormally high ambient temperatures during gestation and reduced cooperation in later life. The relationship is robust to controlling for potential confounders including other environmental factors, markers of early-life deprivation, prenatal stress, postnatal conflict exposure and



caregiver preferences, and is therefore unlikely to be of spurious nature.

Thus, people's willingness to cooperate—a prerequisite for much of economic production—may decline as the likelihood of extreme temperatures increases. The welfare implications of this are substantial in my stylized behavioral games. Their estimation in practice is, however, beyond the scope of this paper, and should instead be the focus of future research. Similarly, it will be important to study the extent to which adaptation to new climatic realities may mitigate the behavioral effects of higher temperatures. Until these questions are answered, at least the possibility of such effects should be taken into account when constructing the damage function of climate change and assessing the benefits of climate policies.

## ETHICS STATEMENT

The study was carried out in accordance with the recommendations of the Social Sciences Ethics Committee at Wageningen University. All subjects or their caregivers gave written informed consent in accordance with the declaration

of Helsinki. The protocol was approved by the Social Sciences Ethics Committee at Wageningen University.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## ACKNOWLEDGMENTS

This research received funding from The Netherlands Organisation for Scientific Research (NWO) as part of project number W 07.72.108, and under grant number 453.10.001. I thank James Fenske, Francesco Cecchi, Erwin Bulte, the two referees, and seminar participants at the University of Oxford and at Wageningen University for their insightful comments.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnbeh.2017.00249/full#supplementary-material>

## REFERENCES

- Adhvaryu, A., Fenske, J., Kala, N., and Nyshadham, A. (2015). *Fetal Origins of Mental Health: Evidence from Africa*. CSAE Working Paper WPS/2015-15.
- Almond, D. (2006). Is the 1918 influenza pandemic over? Long-term effects of *in utero* influenza exposure in the post-1940 U.S. population. *J. Polit. Econ.* 114, 672–712. doi: 10.1086/507154
- Almond, D., and Mazumder, B. (2011). Health capital and the prenatal environment: the effect of Ramadan observance during pregnancy. *Am. Econ. J.* 3, 56–85. doi: 10.1257/app.3.4.56
- Banerjee, A., Duflo, E., Postel-Vinay, G., and Watts, T. (2010). Long-run health impacts of income shocks: wine and phylloxera in nineteenth-century France. *Rev. Econ. Stat.* 92, 714–728. doi: 10.3386/w12895
- Barker, D. J. P. (1990). The fetal and infant origins of adult disease: the womb may be more important than the home. *Br. Med. J.* 301, 1111.
- Barr, A., Packard, T., and Serra, D. (2014). Participatory accountability and collective action: experimental evidence from Albania. *Eur. Econ. Rev.* 68, 250–269. doi: 10.1016/j.euroecorev.2014.01.010
- Barreca, A., Clay, K., Deschênes, O., Greenstone, M., and Shapiro, J. S. (2016). Adapting to climate change: the remarkable decline in the US temperature-mortality relationship over the 20th century. *J. Polit. Econ.* 124, 105–159. doi: 10.1086/684582
- Barreca, A., Deschênes, O., and Guldi, M. (2015). Maybe Next Month? Temperature Shocks, Climate Change, and Dynamic Adjustments in Birth Rates.
- Barreca, A. I. (2010). The long-term economic impact of *in utero* and postnatal exposure to malaria. *J. Hum. Resour.* 45, 865–892. doi: 10.1353/jhr.2010.0027
- Bauer, M., Cassar, A., Chytilová, J., and Henrich, J. (2014). War's enduring effects on the development of egalitarian motivations and in-group biases. *Psychol. Sci.* 25, 47–57. doi: 10.1177/0956797613493444
- Beckett, C., Maughan, B., Rutter, M., Castle, J., Colvert, E., Groothues, C., et al. (2006). Do the effects of early severe deprivation on cognition persist into early adolescence? Findings from the English and Romanian adoptees study. *Child Dev.* 77, 696–711. doi: 10.1111/j.1467-8624.2006.00898.x
- Bellows, J., and Miguel, E. (2009). War and local collective action in Sierra Leone. *J. Publ. Econ.* 93, 1144–1157. doi: 10.1016/j.jpubeco.2009.07.012
- Bharadwaj, P., Lundborg, P., and Rooth, D.-O. (2017). Birth weight in the long run. *J. Hum. Resour.* doi: 10.3368/jhr.53.1.0715-7235R. [Epub ahead of print].
- Bhutta, A. T., Cleves, M. A., Casey, P. H., Cradock, M. M., and Anand, K. J. S. (2002). Cognitive and behavioral outcomes of school-aged children who were born preterm. *JAMA* 288, 728–737. doi: 10.1001/jama.288.6.728
- Black, S. E., Devereux, P. J., and Salvanes, K. G. (2007). From the cradle to the labor market? The effect of birth weight on adult outcomes. *Q. J. Econ.* 122, 409–439. doi: 10.3386/w11796
- Bosch-Domènech, A., Brañas-Garza, P., and Espín, A. M. (2014). Can exposure to prenatal sex hormones (2D:4D) predict cognitive reflection? *Psychoneuroendocrinology* 43, 1–10. doi: 10.1016/j.psyneuen.2014.01.023
- 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
- Buckles, K. S., and Hungerman, D. M. (2013). Season of birth and later outcomes: old questions, new answers. *Rev. Econ. Stat.* 95, 711–724. doi: 10.1162/REST\_a\_00314
- Burke, M., Hsiang, S. M., and Miguel, E. (2015). Global non-linear effect of temperature on economic production. *Nature* 527, 235–239. doi: 10.1038/nature15725
- Buser, T. (2012). Digit ratios, the menstrual cycle and social preferences. *Games Econ. Behav.* 76, 457–470. doi: 10.1016/j.geb.2012.07.006
- Cárdenas, J. C., Chong, A., and Nopo, H. (2009). To what extent do Latin Americans trust, reciprocate, and cooperate?: evidence from experiments in six Latin American countries. *Economía* 9, 45–88. doi: 10.1353/eco.0.0027
- Cecchi, F., and Duchoslav, J. (2018). The effect of prenatal stress on cooperation: evidence from a violent conflict in Uganda. *Eur. Econ. Rev.* 101, 35–56. doi: 10.1016/j.euroecorev.2017.09.015
- Coates, J. M., Gurnell, M., and Rustichini, A. (2009). Second-to-fourth digit ratio predicts success among high-frequency financial traders. *Proc. Natl. Acad. Sci. U.S.A.* 106, 623–628. doi: 10.1073/pnas.0810907106
- Coumou, D., and Rahmstorf, S. (2012). A decade of weather extremes. *Nat. Clim. Change* 2, 491–496. doi: 10.1038/nclimate1452
- Craig, M., Le Sueur, D., and Snow, B. (1999). A climate-based distribution model of malaria transmission in Sub-Saharan Africa. *Parasitol. Today* 15, 105–111. doi: 10.1016/S0169-4758(99)01396-4
- Cronqvist, H., Previtero, A., Siegel, S., and White, R. E. (2016). The fetal origins hypothesis in finance: prenatal environment, the gender gap, and investor behavior. *Rev. Finan. Stud.* 29, 739–786. doi: 10.1093/rfs/hhv065

- Csathó, Á., Osváth, A., Bicsák, É., Karádi, K., Manning, J., and Kállai, J. (2003). Sex role identity related to the ratio of second to fourth digit length in women. *Biol. Psychol.* 62, 147–156. doi: 10.1016/S0301-0511(02)00127-8
- de Onis, M., Onyango, A. W., Borghi, E., Siyam, A., Nishida, C., and Siekmann, J. (2007). Development of a WHO growth reference for school-aged children and adolescents. *Bull. World Health Organ.* 85, 660–667. doi: 10.2471/blt.07.043497
- Dell, M., Jones, B. F., and Olken, B. A. (2009). Temperature and income: reconciling new cross-sectional and panel estimates. *Am. Econ. Rev.* 99, 198–204. doi: 10.1257/aer.99.2.198
- Deschênes, O., Greenstone, M., and Guryan, J. (2009). Climate change and birth weight. *Am. Econ. Rev.* 99, 211–217. doi: 10.1257/aer.99.2.211
- Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2012). The intergenerational transmission of risk and trust attitudes. *Rev. Econ. Stud.* 79, 645–677. doi: 10.1093/restud/rdr027
- Eisenberg, N., Fabes, R. A., and Spinrad, T. L. (2006). “Prosocial development,” in *Handbook of Child Psychology*, eds W. Damon, R. M. Lerner, and N. Eisenberg (Hoboken, NJ: Wiley), 646–719.
- Figlio, D., Guryan, J., Karbownik, K., and Roth, J. (2014). The effects of poor neonatal health on children’s cognitive development. *Am. Econ. Rev.* 104, 3921–3955. doi: 10.1257/aer.104.12.3921
- Flouris, A. D., Spiropoulos, Y., Sakellariou, G. J., and Koutedakis, Y. (2009). Effect of seasonal programming on fetal development and longevity: links with environmental temperature. *Am. J. Hum. Biol.* 21, 214–216. doi: 10.1002/ajhb.20818
- Fowden, A. L., Giussani, D. A., and Forhead, A. J. (2006). Intrauterine programming of physiological systems: causes and consequences. *Physiology* 21, 29–37. doi: 10.1152/physiol.00050.2005
- 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
- Gilligan, M. J., Pasquale, B. J., and Samii, C. (2014). Civil war and social cohesion: lab-in-the-field evidence from Nepal. *Am. J. Polit. Sci.* 58, 604–619. doi: 10.1111/ajps.12067
- Hack, M., Taylor, H. G., Drotar, D., Schluchter, M., Cartar, L., Andreias, L., et al. (2005). Chronic conditions, functional limitations, and special health care needs of school-aged children born with extremely low-birth-weight in the 1990s. *JAMA* 294, 318–325. doi: 10.1001/jama.294.3.318
- Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H. (2014). Updated high resolution grids of monthly climatic observations—the CRU TS3.10 Dataset. *Int. J. Climatol.* 34, 623–642. doi: 10.1002/joc.3711
- Hille, E. T. M., Weisglas-Kuperus, N., van Goudoever, J. B., Jacobusse, G. W., Ens-Dokkum, M. H., de Groot, L., et al. (2007). Functional outcomes and participation in young adulthood for very preterm and very low birth weight infants: the dutch project on preterm and small for gestational age infants at 19 years of age. *Pediatrics* 120, e587–e595. doi: 10.1542/peds.2006-2407
- Hsiang, S. M., Burke, M., and Miguel, E. (2013). Quantifying the influence of climate on human conflict. *Science* 341:1235367. doi: 10.1126/science.1235367
- Kaplan, R., and Saccuzzo, D. (2012). *Psychological Testing: Principles, Applications, and Issues*. Boston, MA: Cengage Learning.
- Lajinian, S., Hudson, S., Applewhite, L., Feldman, J., and Minkoff, H. L. (1997). An association between the heat-humidity index and preterm labor and delivery: a preliminary analysis. *Am. J. Public Health* 87, 1205–1207. doi: 10.2105/AJPH.87.7.1205
- Lawlor, D. A., Leon, D. A., and Davey Smith, G. (2005). The association of ambient outdoor temperature throughout pregnancy and offspring birthweight: findings from the aberdeen children of the 1950s cohort. *BJOG Int. J. Obstetr. Gynaecol.* 112, 647–657. doi: 10.1111/j.1471-0528.2004.00488.x
- Macksoud, M. (1992). Assessing war trauma in children: a case study of lebanese children. *J. Refugee Stud.* 5, 1–15. doi: 10.1093/jrs/5.1.1
- Montesquieu (1748). *The Spirit of Laws*. Montesquieu, La Brède.
- Moore, C., Barresi, J., and Thompson, C. (1998). The cognitive basis of future-oriented prosocial behavior. *Soc. Dev.* 7, 198–218. doi: 10.1111/1467-9507.00062
- O’Loughlin, J., Witmer, F. D. W., Linke, A. M., Laing, A., Gettelman, A., and Dudhia, J. (2012). Climate variability and conflict risk in east africa. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18344–18349. doi: 10.1073/pnas.1205130109
- Oreskes, N. (2005). The scientific consensus on climate change. *Science* 306:1686. doi: 10.1126/science.1103618
- Papageorgiou, V., Frangou-Garunovic, A., Iordanidou, R., Yule, W., Smith, P., and Vostanis, P. (2000). War trauma and psychopathology in bosnian refugee children. *Eur. Child Adolesc. Psychiatr.* 9, 84–90. doi: 10.1007/s007870050002
- Raleigh, C., Linke, A., Hegre, H., and Karlsen, J. (2010). Introducing ACLED: an armed conflict location and event dataset. *J. Peace Res.* 47, 651–660. doi: 10.1177/0022343310378914
- Sahn, D., and Stifel, D. (2003). Exploring alternative measures of welfare in the absence of expenditure data. *Rev. Income Wealth* 49, 463–489. doi: 10.1111/j.0034-6586.2003.00100.x
- Sapienza, P., Zingales, L., and Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15268–15273. doi: 10.1073/pnas.0907352106
- van Os, J., and Selten, J.-P. (1998). Prenatal exposure to maternal stress and subsequent schizophrenia: the May 1940 invasion of The Netherlands. *Br. J. Psychiatry* 172, 324–326. doi: 10.1192/bjp.172.4.324
- Voors, M. J., Nillesen, E. E., Verwimp, P., Bulte, E. H., Lensink, R., and Van Soest, D. P. (2012). Violent conflict and behavior: a field experiment in burundi. *Am. Econ. Rev.* 102, 941–964. doi: 10.1257/aer.102.2.941
- Voracek, M., Manning, J. T., and Dressler, S. G. (2007). Short report 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
- Wells, J. C. K., and Cole, T. J. (2002). Birth weight and environmental heat load: a between-population analysis. *Am. J. Phys. Anthropol.* 119, 276–282. doi: 10.1002/ajpa.10137
- Wilde, J., Apouey, B. H., and Jung, T. (2017). The effect of ambient temperature shocks during conception and early pregnancy on later life outcomes. *Eur. Econ. Rev.* 97, 87–107. doi: 10.1016/j.euroecorev.2017.05.003
- Willmott, C. J., and Matsuura, K. (2015). Terrestrial air temperature: 1900–2014 gridded monthly time series. Available online at: [http://climate.geog.udel.edu/~climate/html\\_pages/Global2014/README.GlobalTsT2014.html](http://climate.geog.udel.edu/~climate/html_pages/Global2014/README.GlobalTsT2014.html)
- Yackerson, N., Piura, B., and Sheiner, E. (2008). The influence of meteorological factors on the emergence of preterm delivery and preterm premature rupture of membrane. *J. Perinatol.* 28, 707–711. doi: 10.1038/jp.2008.69
- Zhang, S., Gao, W., and Fan, B. (2015). Cognitive ability and cooperation: evidence from the public goods experiments. *Ann. Econ. Fin.* 16, 43–68.

**Conflict of Interest Statement:** The author declares 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 Duchoslav. 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.



# 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 (Stirratt and Perrett, 2010) and cheating more in a lottery (Haselhuhn and Wong, 2011; Geniole et al., 2014). However, Stirratt 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 one's 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

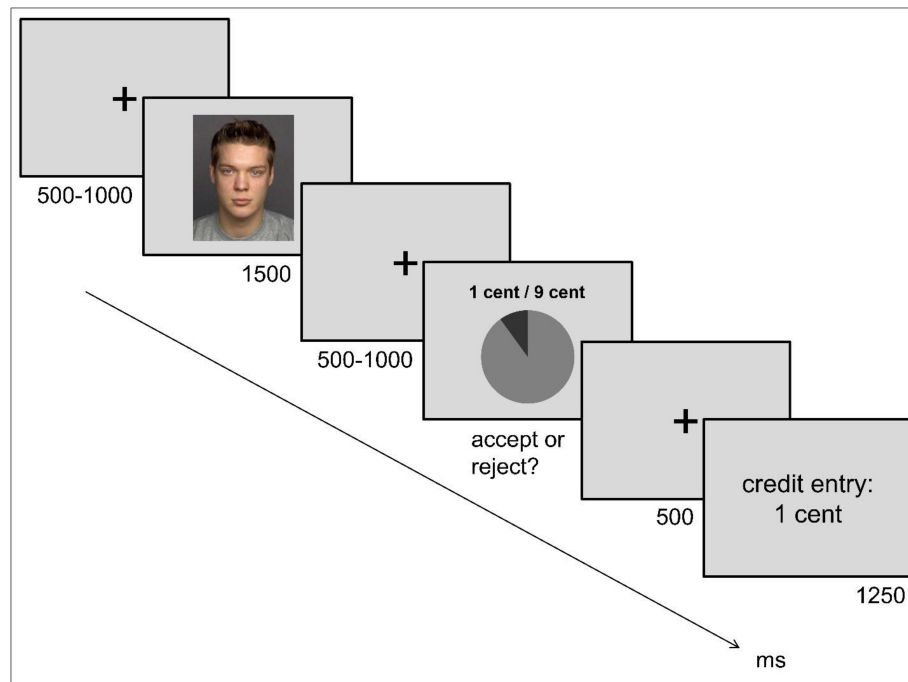
is assigned a continuous value of social orientation, which can be categorized to competitive, individualistic, prosocial and altruistic. Previous research indicates that SVO is a valid predictor of the cooperative tendency in social dilemmas (Bogaert et al., 2008; Balliet et al., 2009). In the analyses, 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



**FIGURE 1 |** Trial Scheme of the Ultimatum Game. Each trial started with a fixation cross shown for a variable time of 500–1,000 ms, followed by a photograph of a proposer for 1,500 ms, and another fixation cross presented for 500–1,000 ms; then, participants received an offer about splitting 10 cent which they had to accept or reject via button press. Afterwards, a fixation cross was presented again for 500 ms. Participants received feedback about the sum booked to their account before the next trial started after 1,250 ms. Portrait taken from Ebner et al. (2010) for which the depicted individual gave consent to be displayed in research-related publications.

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

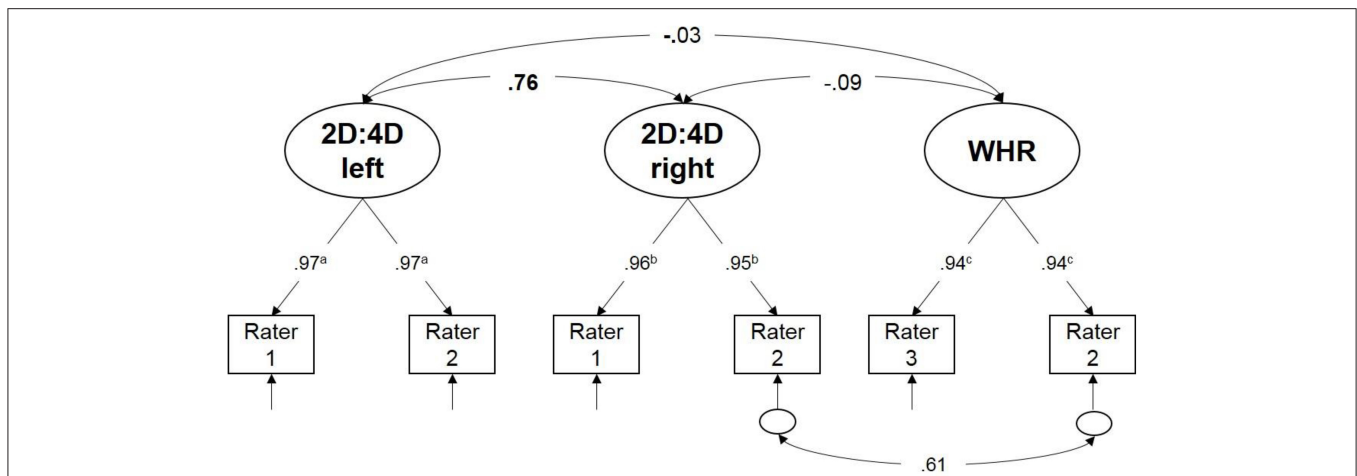
prosociality by means of three parcels of SVO responses (see above) and a further indicator of total offers in the proposer part of the Ultimatum Game. Negative reciprocity as a latent variable is measured by rejection rates of unfair offers in three independent experimental blocks and assertiveness 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



**FIGURE 2 |** Schematic representation of the measurement model of prenatal and pubertal organizational effects of hormones. Rectangles represent measured variables and circles are used to depict latent variables. 2D:4D at the left and the right hand was measured by two different raters. Rater specific values are used as measured variables to estimate latent variables that represent prenatal and pubertal organizational effects accounted for measurement error due to the rater. Unidirectional path represent factor loadings and bidirectional path are used for depicting correlations. Short arrows (with a small circle) represent residual variance (non-reliability of a measured variables). For simplicity, we use only arrows to indicate error variance if there is no residual correlations between residuals. 2D:4D left—left hand 2D:4D estimation taking rater induced measurement error into account; 2D:4D right—right hand 2D:4D estimation; WHR—facial width-to-height ratio estimate taking rater induced measurement error into account; a, b, c indices on factor loadings are used to indicate that non-standardized loadings were fixed to equality within factors; standardized loadings are depicted in the figure. Significant relationships between latent factors at  $p < 0.05$  are written bold.

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.

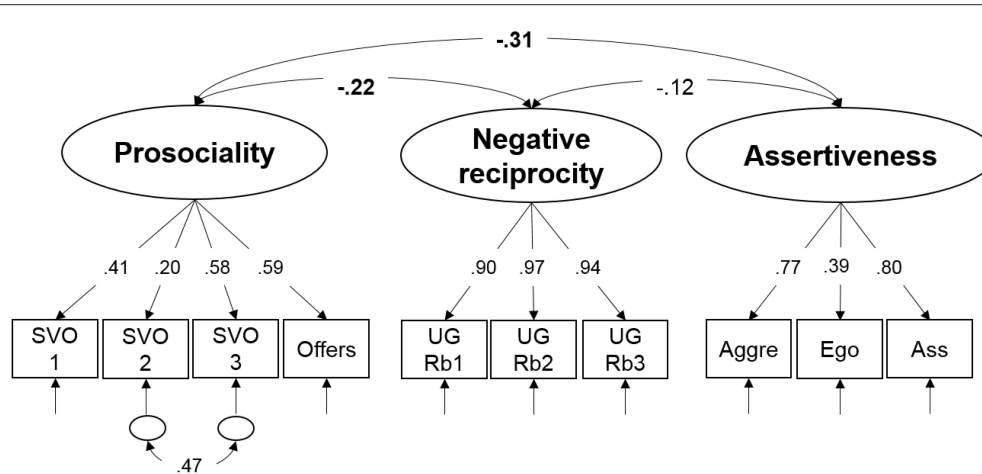
### Structural Model of Organizational Hormonal Effects and Behavior

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,



**FIGURE 3 |** Schematic representation of the measurement model of prosociality, negative reciprocity and assertiveness. See the caption of **Figure 2** explaining the general conventions of the graphical language visualizing latent variable models. SVO, Social value orientation; UG, Ultimatum Game; Rb1, Rejection of unfair offers in bloc 1; Rb2, Rejection of unfair offers in bloc 2; Rb3, Rejection of unfair offers in bloc 3; Aggre, Aggressive; Ego, Egoistic; Ass, Assertive. Significant relationships between latent factors at  $p < 0.05$  are depicted in bold.

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**.

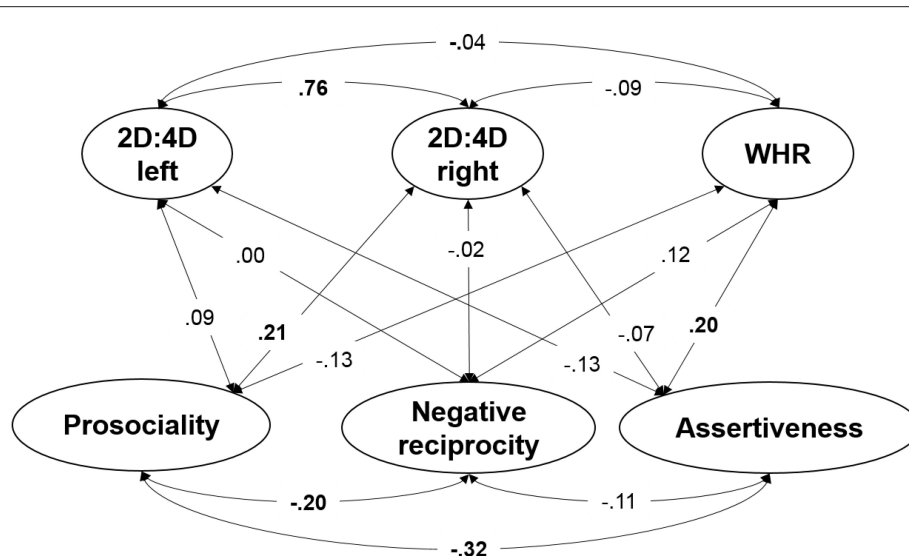
### Curvilinear Relations between Organizational Effects of Prenatal and Pubertal Hormones, Prosocial Behavior, and Related Traits

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 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 (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.

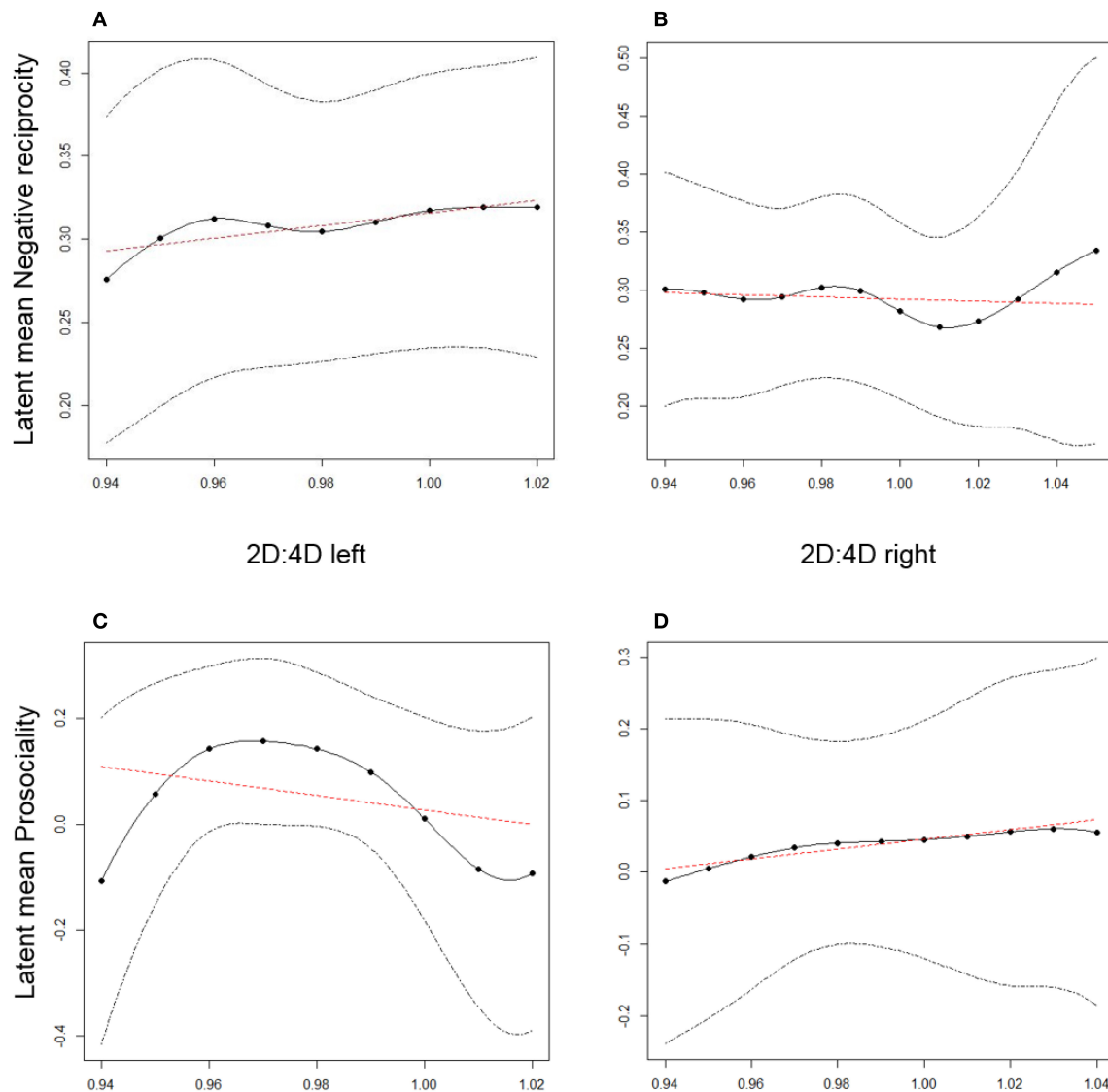
Model	$\chi^2$	df	CFI	RMSEA	$\Delta\chi^2$	$\Delta df$
Configural invariance	203.54	177	0.986	0.042	–	–
Weak (metric) invariance	230.27	193	0.981	0.048	26.72	16
Strong (scale) invariance	284.59	209	0.958	0.069	54.31*	16

\* $p < 0.01$ ; 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.





**FIGURE 5 |** Parameter gradients for latent factor means as estimated by Local Structural Equations to test non-linear associations between 2D:4D and prosociality. In (A,C) the x-axis depicts average values of the measured 2D:4D across the two raters at the left hand side. In (B,D) the x-axis depicts average values of the measured 2D:4D across the two raters at the right hand side.

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 (Burris 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 Flidner, Alf Mante, Karsten Manske, Astrid Kiy, Friederike Rüffer, 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., Maestriperi, 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., Diken, 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., Snoddy, 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.21083
- 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 Desdicioglu, 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.jpubeco.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.



# The (Null) Effect of Affective Touch on Betrayal Aversion, Altruism, and Risk Taking

Lina Koppel<sup>1,2</sup>, David Andersson<sup>1,2</sup>, India Morrison<sup>1</sup>, Daniel Västfjäll<sup>1,2,3,4</sup> and Gustav Tinghög<sup>1,2,5\*</sup>

<sup>1</sup> Center for Social and Affective Neuroscience, Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden, <sup>2</sup> JEDI Lab, Division of Economics, Department of Management and Engineering, Linköping University, Linköping, Sweden, <sup>3</sup> Division of Psychology, Department of Behavioral Sciences and Learning, Linköping University, Linköping, Sweden, <sup>4</sup> Decision Research, Eugene, OR, United States, <sup>5</sup> The National Center for Priority Setting in Health Care, Department of Medical and Health Sciences, Linköping University, Linköping, Sweden

Pleasant touch is thought to increase the release of oxytocin. Oxytocin, in turn, has been extensively studied with regards to its effects on trust and prosocial behavior, but results remain inconsistent. The purpose of this study was to investigate the effect of touch on economic decision making. Participants ( $n = 120$ ) were stroked on their left arm using a soft brush (touch condition) or not at all (control condition; varied within subjects), while they performed a series of decision tasks assessing betrayal aversion (the Betrayal Aversion Elicitation Task), altruism (donating money to a charitable organization), and risk taking (the Balloon Analog Risk Task). We found no significant effect of touch on any of the outcome measures, neither within nor between subjects. Furthermore, effects were not moderated by gender or attachment. However, attachment avoidance had a significant effect on altruism in that those who were high in avoidance donated less money. Our findings contribute to the understanding of affective touch—and, by extension, oxytocin—in social behavior, and decision making by showing that touch does not directly influence performance in tasks involving risk and prosocial decisions. Specifically, our work casts further doubt on the validity of oxytocin research in humans.

**Keywords:** touch, oxytocin, betrayal aversion, altruism, risk taking, trust

## OPEN ACCESS

### Edited by:

Pablo Brañas-Garza,  
Middlesex University, United Kingdom

### Reviewed by:

Gideon Nave,  
California Institute of Technology,  
United States  
Sandra Racionero-Plaza,  
Universidad Loyola Andalucía, Spain

### \*Correspondence:

Gustav Tinghög  
gustav.tinghog@liu.se

**Received:** 28 August 2017

**Accepted:** 11 December 2017

**Published:** 19 December 2017

### Citation:

Koppel L, Andersson D, Morrison I, Västfjäll D and Tinghög G (2017) The (Null) Effect of Affective Touch on Betrayal Aversion, Altruism, and Risk Taking.  
*Front. Behav. Neurosci.* 11:251.  
doi: 10.3389/fnbeh.2017.00251

## INTRODUCTION

Touch plays a vital role for social and psychological well-being and is said to have a “Midas effect” on judgments and decisions, promoting prosocial behavior (Crusco and Wetzel, 1984; Schirmer et al., 2016). Pleasant touch is also thought to increase the release of oxytocin (Walker et al., 2017). Oxytocin, in turn, has been extensively studied with regards to its effects on trust and prosocial behavior, but results remain inconsistent. In this study, we indirectly investigated the presumed effect of endogenously released oxytocin by gently stroking participants on their forearm while they performed a series of decision tasks assessing betrayal aversion, altruism, and risk taking. Our findings contribute to the understanding of touch—and, by extension, oxytocin—in social behavior and decision making.

The first evidence for a causal link between oxytocin and trust was provided by Kosfeld et al. (2005), who found that intranasally administered oxytocin increased investments in a trust game. However, this finding has been difficult to replicate. Some researchers have found that intranasal

oxytocin has no effect on initial investments in the trust game, but that it influences investments following trust betrayal (Baumgartner et al., 2008); others have found that the effect of oxytocin on trust and responses to trust betrayal is moderated by gender (Yao et al., 2014) or that it only applies to individuals high in attachment avoidance (De Dreu, 2012). A recent review and meta-analysis found no consistent effect of intranasal oxytocin on trust (Nave et al., 2015). One potential caveat of these studies is the controversial assumption that intranasal oxytocin passes the blood–brain barrier and reaches target brain areas (Leng and Ludwig, 2016). However, studies correlating plasma levels of endogenously released oxytocin with trust have also yielded mixed results. Some researchers have found that oxytocin has no effect on investments in the trust game, but that it influences the amount returned by trustees (Morhenn et al., 2008) or that the level of oxytocin is higher following the receipt of an intentional monetary transfer compared to an equivalent transfer that is determined by a random lottery (Zak et al., 2005). Others have found a U-shaped pattern such that individuals who are either high or low in plasma oxytocin are both more trusting and more trustworthy than participants with moderate levels of oxytocin (Zhong et al., 2012). A drawback of several of these studies that could help explain the inconsistent findings is that they have used unextracted samples of plasma oxytocin that have been shown to be unreliable (McCullough et al., 2013; Christensen et al., 2014). In addition, the oxytocin literature suffers from issues such as publication bias (Lane et al., 2016) and low statistical power (Walum et al., 2016). In sum, the evidence that oxytocin directly influences behavior remains sparse. If there is an effect, it is likely moderated by a variety of factors.

In the present study, we aimed to experimentally manipulate the levels of endogenously released oxytocin by gently stroking participants' forearm with a soft brush. Slow, gentle touch is perceived as pleasant and activates areas of the brain that are associated with interoception and reward, such as the insula, caudate, and dorsolateral prefrontal cortex (Perini et al., 2015). Gentle stroking of the skin at a speed of 1–10 cm/s also activates a specific type of nerve fibers, C-tactile (CT) afferents, that respond optimally to the type of touch that is perceived as most pleasant (Löken et al., 2009). The pleasant and relaxing effects of CT-optimal touch mirror those of exogenously administered or endogenously released oxytocin, suggesting that activation of CT fibers increases the release of oxytocin (Walker et al., 2017, see also Uvnäs-Moberg et al., 2015), although this link has yet to be established empirically. Furthermore, it has been shown that people spontaneously stroke other humans, but not objects, at CT-optimal speeds (Croy et al., 2016), which supports the idea that touch, in particular the kind of touch that activates CT fibers, plays a vital role in the formation and maintenance of social bonds (Olausson et al., 2010). Thus, it seems reasonable to hypothesize that affective touch influences economic behavior; however, to the best of our knowledge, no such studies exist.

We investigate the effect of touch on betrayal aversion, altruism, and risk taking. Betrayal aversion refers to the reluctance to take risk when the outcome depends on a human counterpart rather than when it is determined by nature (i.e., chance; Bohnet and Zeckhauser, 2004). The first evidence for this

tendency was provided by Bohnet and Zeckhauser (2004), who elicited participants' minimum acceptable probability (MAP) of getting an even split (the good outcome) for which they were willing to take a risk in a standard trust game compared to an equivalent risk-only trust game. They found that participants' MAPs were greater in the trust game than in the risk-only trust game, indicating that people infer a cost from the possibility of being betrayed by another person, above and beyond the monetary cost. This finding has been replicated across several cultures (Bohnet et al., 2008). More recent neuroimaging studies have shown that playing a trust game with a human counterpart rather than a computer activates areas of the brain that are associated with emotion regulation and negative affect, including the right anterior insula, medial frontal cortex, and right dorsolateral prefrontal cortex (Aimone et al., 2014). Furthermore, betrayal averse participants show less amygdala activity before choosing a risky compared to certain option in a non-social risk task but not in an equivalent social risk task and show greater activity in the striatum, which is involved in reward, after receiving a social than a non-social outcome (Lauharatanahirun et al., 2012).

Betrayal aversion has been suggested as one of the mechanisms by which oxytocin increases trust (Engelmann and Fehr, 2017). The prediction that follows is that touch, because it presumably increases oxytocin levels, reduces betrayal aversion. For instance, Baumgartner et al. (2008) gave male participants intranasal oxytocin or placebo and compared decisions made by investors in a trust game both before and after they received feedback that the trustees did not reciprocate in 50% of cases. Oxytocin had no significant effect on investments before feedback, but after feedback participants who had received oxytocin invested more than those who had received placebo. This suggests that oxytocin reduces the sensitivity to betrayal of trust. However, note that this study relied on intranasal oxytocin despite controversial underlying assumptions. More recent research has failed to replicate the findings (Klackl et al., 2013). Another study that is of particular relevance to the present study was conducted by Morhenn et al. (2008), who compared participants' behavior in a one-shot trust game following either a 15-min massage or a 15-min rest. They found no difference in investors' behavior, but trustees who had received a massage returned more money than trustees who had rested. Most importantly, for participants who had rested, both oxytocin levels and the amount received from the investor predicted the amount returned by the trustees, but for participants who had received a massage, only oxytocin predicted the amount returned. These findings suggest that touch—and oxytocin—promotes prosocial behaviors, although note that these researchers used unextracted samples of plasma oxytocin that may be unreliable (see McCullough et al., 2013; Christensen et al., 2014).

The effect of touch on altruism is more difficult to predict. Previous research suggests that touch increases positive valuations and makes people more prosocial overall, an effect known as the Midas effect (Crusco and Wetzel, 1984; Schirmer et al., 2016). For instance, restaurant guests give larger tips after having been touched on the shoulder by the waitress (Crusco and Wetzel, 1984) and people who have been touched are more likely

to help a stranger (Kleinke, 1977; Guéguen and Fischer-Ikoku, 2003). The prediction that follows from this line of research is that touch increases altruism. However, the oxytocin literature gives a more complex picture. Some researchers have found that oxytocin increases donations to charitable organizations (Barraza et al., 2011; Marsh et al., 2015) and that it increases monetary contributions in a social dilemma (Israel et al., 2012). Others have found that the effect of oxytocin on altruism depends on contextual factors, such as whether the target is a stranger or a close other (Pornpattananangkul et al., 2017) and whether they belong to the in-group or to the outgroup (De Dreu et al., 2010). An alternative explanation of these inconsistencies is that oxytocin promotes mentalizing, i.e., the ability to take someone else's perspective (Domes et al., 2007; but see also Radke and de Bruijn, 2015, and Leppanen et al., 2017, who found no support of this suggestion). Zak et al. (2007) found that intranasal oxytocin increased monetary offers in an ultimatum game but not in an equivalent dictator game. The difference between these two tasks is that in the ultimatum game, the investor has to take the recipient's reaction into account because the recipient can reject the investor's offer, resulting in zero earnings for both players. In contrast, in the dictator game, the recipient simply obtains whatever amount the investor offers, which does not require perspective taking to the same extent. Furthermore, a recent fMRI study showed that oxytocin had no effect on the frequency of altruistic decisions, but that it increased activity in the left temporo-parietal junction, a region that has been implicated in theory of mind, when participants observed others being helped (Hu et al., 2016). Following this line of research, touch should have no direct effect on altruism. However, note again that these studies used intranasal oxytocin.

Oxytocin has mostly been studied in terms of its role in social relationships and behavior, so its effect on risk taking in the non-social domain is unclear. Physical contact has been shown to increase financial risk taking, especially if the toucher is female and if the touch involves a tap on the shoulder rather than a handshake (Levav and Argo, 2010). Somatosensory stimulation in the form of thermal pain also increases risk seeking (Koppel et al., 2017). On the other hand, individuals who have received oxytocin are not more risk seeking than participants who have received a placebo, as shown in studies comparing the effect of intranasal oxytocin on decisions in a trust game to decisions in an equivalent risk game (e.g., Kosfeld et al., 2005). To our knowledge, only one published study has investigated the effect of intranasal oxytocin using a risk-taking task that does not involve another person, and it found no main effect of oxytocin on risk taking (Patel et al., 2015). However, a three-way interaction appeared such that men (but not women) who had received oxytocin were less risk taking if they were told that others were watching them perform the task (which resulted in social stress). Thus, if touch influences risk taking, it may do so via some mechanism other than increased oxytocin, such as increased positive affect.

To the best of our knowledge, our study is the first to investigate the effect of CT-optimal touch on economic decision making. We implemented a crossover design in which all participants completed the decision tasks both with and without

touch (in counterbalanced order), which allowed us to explore the effects both within and between subjects. Furthermore, we investigated betrayal aversion, altruism, and risk taking using three standard economic decision-making tasks: the Betrayal Aversion Elicitation Task (BAET), a dictator game, and the Balloon Analog Risk Task (BART).

## MATERIALS AND METHODS

### Participants

One hundred and twenty participants (43% female) were recruited from a subject pool at Linköping University, Sweden. Participants signed up using ORSEE (Greiner, 2015). Participants were Swedish-speaking students from a variety of disciplines. Ages ranged from 19 to 54 years ( $M = 24.8$ ,  $SD = 6.0$ ). A power calculation indicated that 101 participants were needed to detect a 0.25 effect size with 70% power within subjects. All participants gave written informed consent in accordance with the Declaration of Helsinki and were compensated with the amount earned on one randomly selected task. The procedures were approved by the regional ethics committee.

### Materials

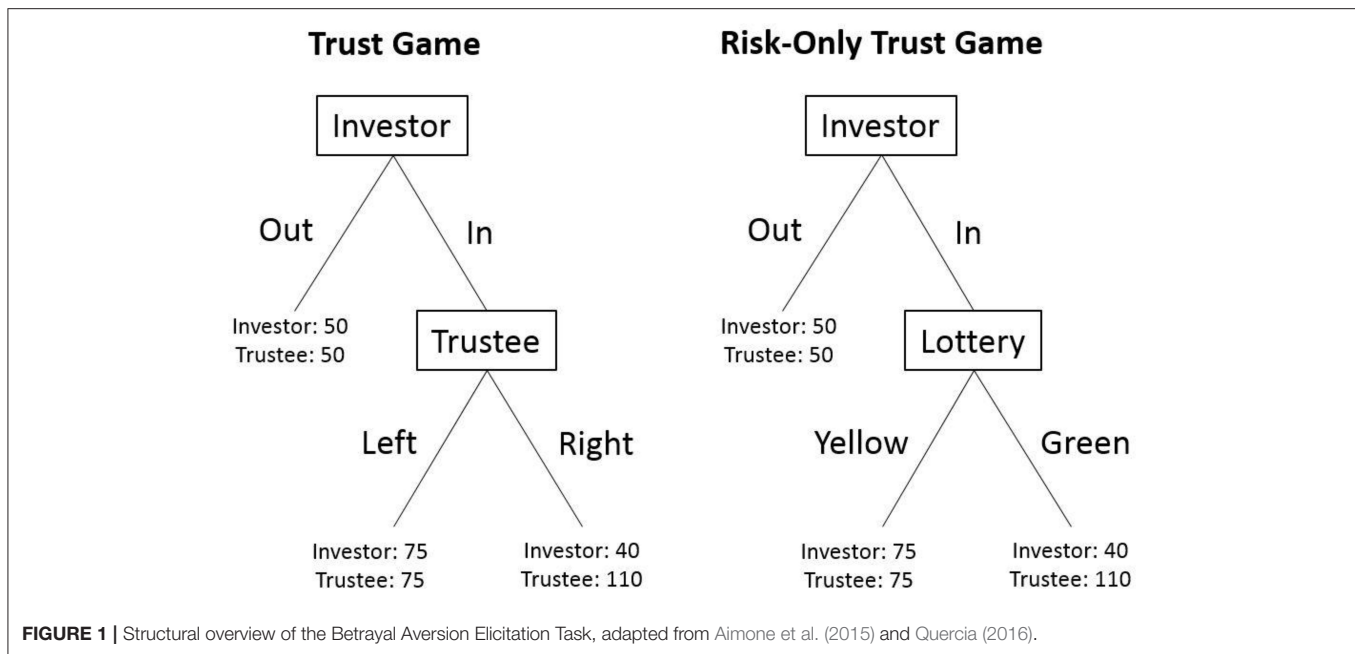
#### Betrayal Aversion Elicitation Task (BAET)

Betrayal aversion was assessed using the Betrayal Aversion Elicitation Task (BAET; Aimone et al., 2015), which consists of two games: a trust game and a risk-only trust game (illustrated in **Figure 1**). In the trust game, the participant plays in the role of investor and is randomly paired with one other participant that plays in the role of trustee. The investor's task is to choose between *in* (trust) and *out* (don't trust). If they choose *out*, both the investor and the trustee receive 50 SEK (~6 USD). If they choose *in*, the amount they receive depends on the trustee's choice. The trustee chooses between *left* (reciprocate) and *right* (betray). If they choose *left*, both the trustee and the investor receive 75 SEK. If they choose *right*, the investor receives 40 SEK and the trustee receives 110 SEK.

All participants played in the role of investor. Prior to the study, a group of 20 participants completed the same trust game but in the role of trustee. That is, they indicated whether they would choose *left* or *right* if the investor chose *in*. The results from this part of the experiment determined investors' payoff. The investors' task was to indicate whether they chose *in* or *out*, for each possible value of the number of trustees that chose *left*. They made their decisions by filling out a choice list table consisting of 21 rows reporting all possible proportions of trustees choosing *left*, starting with "20 out of 20" in the first row and ending with "0 out of 20" in the last row. This elicitation method has been shown to increase participants' understanding of the task and to result in less noisy valuations, compared to an open-ended elicitation method (Quercia, 2016).

The risk-only trust game is identical to the trust game, except payoffs depend on a random lottery rather than on the trustee's decision. The lottery was described as an urn containing 20 colored balls that each can be either yellow or green. If a yellow ball is drawn, both the investor and the trustee receive 75 SEK. If a green ball is drawn, the investor receives 40 SEK and the trustee





receives 110 SEK. The actual number of yellow and green balls was predetermined by the number of the 20 previous participants in the trust game who had chosen *left* and *right*, respectively. Thus, the probability of drawing a yellow ball in the risk-only trust game is the same as the probability of being paired with a trustee who chose *left* in the trust game.

The variable of interest in the BAET is the MAP of being paired with a trustee who chose *left* (in the trust game) or drawing a yellow ball (in the risk-only trust game) for which a participant is willing to choose *in*. We inferred each participant's MAP by calculating the mean between the last proportion for which they chose *in* and the first proportion for which they chose *out*, going from the top to the bottom of the choice list table<sup>1</sup>. Participants' betrayal aversion (BA) was then calculated as  $BA = MAP_{TG} - MAP_{ROTG}$ . If  $MAP_{TG} > MAP_{ROTG}$ , participants are said to be betrayal averse. If  $MAP_{TG} < MAP_{ROTG}$ , participants are said to be betrayal seeking. If  $MAP_{TG} = MAP_{ROTG}$ , participants are said to be betrayal neutral.

### Dictator Game

Altruism was assessed using a dictator game in which participants distributed 100 SEK (~12 USD) between themselves and UNICEF. Participants indicated how much they wanted to keep for themselves and how much they wanted to give to UNICEF,

using two sliding scales that ranged from 0 to 100 SEK, in 1 SEK increments. The sum of the scales had to equal 100 SEK.

### Balloon Analog Risk Task (BART)

Risk taking was assessed using the Balloon Analog Risk Task (BART; Lejuez et al., 2002). On each of 30 trials, participants were presented with a picture of a balloon and were instructed that they could pump up the balloon to earn money. Each pump earned them 0.10 SEK. However, if they pumped up a balloon so much that it exploded, they earned 0 SEK on that trial. Risk taking is operationalized as the average number of pumps per trial, excluding trials on which the balloon exploded. We refer to this variable as the *adjusted average pumps*.

### Self-report Measures: Touch Pleasantness, Game Understanding, and Attachment

Participants rated how pleasant and relaxing the touch was using two visual analog scales ranging from -10 (*very unpleasant/not relaxing at all*) to 10 (*very pleasant/very relaxing*). Game understanding was assessed following Quercia (2016; see Supplementary Materials)<sup>2</sup>. Attachment was assessed using the Revised Adult Attachment Scale (Collins, 1996), which consists of 18 items measuring how participants generally feel in important close relationships. The scale assesses both attachment avoidance and attachment anxiety. Participants indicated how

<sup>1</sup>We did not force participants to switch between *in* and *out* only once in the choice list table. As a result, 6–9% of participants in each task had multiple switching points in their responses. In two of these cases, the switching points occurred in the middle of the table, allowing us to infer the participant's MAP by taking the average between the first and the last switching point. The rest of the participants with multiple switching points were excluded from the analysis. Participants were also excluded if they selected *out* in the first row and switched to *in* at some point in the table. If a participant chose *in* for all rows of the table, their MAP was set to 0; if they chose *out* for all rows, it was set to 1.

<sup>2</sup>Contrary to Quercia (2016) and Aimone et al. (2015), we did not assess game understanding immediately following the instructions and before participants filled in the choice list table, because we wanted to avoid priming analytical, "system 2" thinking. Leaving the comprehension questions to the end of the experiment also ensured that the two rounds of the task were identical to the greatest extent possible. We assume that participants understood the task if we were able to infer a MAP from their responses. Results from the comprehension questions are reported in the Supplementary Materials and are similar to those reported in Quercia (2016).

characteristic each item was of them on a Likert-type scale from 1 (*not characteristic at all*) to 5 (*very characteristic*). Cronbach's alpha in our study was 0.84 for attachment avoidance and 0.86 for attachment anxiety. Finally, participants were asked to guess the purpose and hypotheses of the study and to report their suspicion of deception in the Betrayal Aversion Elicitation Task.

Complete instructions for all tasks are provided in the Supplementary Materials. All tasks except the BART were administered in Qualtrics. The BART was administered in Inquisit 5.

## Procedure

We implemented a crossover design in which participants performed the decision tasks twice: once in a touch condition and once in a no-touch control condition. The order of the tasks was the same for all participants—i.e., (1) Betrayal Aversion Elicitation Task, (2) dictator game, (3) Balloon Analog Risk Task—but the order of the touch and control conditions was counterbalanced between participants. Thus, participants served as their own controls.

Participants were seated at a desk equipped with a computer and were instructed to rest their left arm behind a curtain, palm facing down. The experimenter sat on the other side of the curtain. In the touch condition, the experimenter gently stroked the participant on the dorsal part of left forearm at a speed of 3 cm/s using a goat hair brush. This stroking procedure and velocity is optimal for activating CT fibers (Löken et al., 2009). The self-report measures confirmed that participants indeed perceived the touch as pleasant ( $M = 5.58$ ,  $SD = 4.26$ ) and relaxing ( $M = 3.92$ ,  $SD = 5.11$ ). The brushing began 60 s before the instructions for the first task were displayed and continued until completion of the last task. Thus, participants received touch both while reading the instructions for each task and while performing that task. In the control condition, participants received no touch, but the experimenter remained seated behind the curtain. Participants read the instructions for each task at their own pace immediately before completing that task. After completing all decision tasks twice (once in the touch condition and once in the control condition), participants filled out the self-report measures and were compensated for participating.

## Data Analysis

We first investigated whether the proportion of participants who were classified as betrayal averse, betrayal neutral, and betrayal seeking differed between the touch and control condition, using a McNemar-Bowker test of symmetry. We then performed a paired samples *t*-test to investigate whether participants were on average less betrayal averse in the touch condition compared to the control condition. We also performed regression analyses in order to confirm the results from the *t*-test while controlling for factors such as age and gender. Our regression model was specified as follows:

$$y_{ik} = \beta_0 + \beta_1 \text{Touch} + \beta_2 \text{Round} + \beta_4 \mathbf{X}_i + \epsilon_{ik}$$

where the dependent variable  $y_{ik}$  indicates the betrayal aversion ( $\text{MAP}_{\text{TG}} - \text{MAP}_{\text{ROTG}}$ ) for participant  $i$  on round  $k$ . *Touch* is a

dummy for the touch condition and *Round* is a dummy for the second round of the tasks, i.e., the second time participants performed the Betrayal Aversion Elicitation Task.  $\mathbf{X}_i$  is the control variables age and gender. Alternative model (2) also included the interaction terms *Touch*  $\times$  *Round*, which allows the effect of touch to differ across the two task rounds, and *Touch*  $\times$  *Gender*, which allows the effect of touch to differ across genders. Alternative model (3) added the control variables attachment anxiety and attachment avoidance and alternative model (4) also included the interaction terms *Touch*  $\times$  *Anxiety* and *Touch*  $\times$  *Avoidance*, which allow the effect of touch to vary with attachment styles. The models were estimated using OLS and standard errors were corrected for clustering on the individual level.

Paired samples *t*-tests and regressions as specified above were also performed for altruism and risk taking, with mean amount donated to UNICEF and adjusted average pumps as dependent variables. We also investigated whether touch pleasantness correlated with betrayal aversion, altruism, and risk taking in the touch condition. Finally, we repeated all analyses using the corresponding between-subjects tests, to investigate the effect of touch in the first round of each task. The between-subjects analyses were performed because participants' responses are likely to be relatively consistent between the first and second round of the tasks and because the manipulation may be fairly obvious to participants, thus potentially influencing the results.

## RESULTS

### The Effect of Touch on Betrayal Aversion

Figure 2 displays the percentage of participants in each condition who were classified as betrayal averse ( $\text{MAP}_{\text{TG}} > \text{MAP}_{\text{ROTG}}$ ), betrayal neutral ( $\text{MAP}_{\text{TG}} = \text{MAP}_{\text{ROTG}}$ ), and betrayal seeking ( $\text{MAP}_{\text{TG}} < \text{MAP}_{\text{ROTG}}$ ). In the touch condition, 26% of participants were betrayal averse, 46% were betrayal neutral, and 29% were betrayal seeking. In the control condition, 35% of participants were betrayal averse, 43% were betrayal neutral, and 22% were betrayal seeking. Thus, participants were less betrayal averse in the touch condition. However, a McNemar-Bowker test of symmetry indicated that there was not a significant difference in the proportions of betrayal averse, betrayal neutral, and betrayal seeking participants between the touch and control conditions,  $p = 0.475^3$ .

Figure 3A displays the average betrayal aversion ( $\text{MAP}_{\text{TG}} - \text{MAP}_{\text{ROTG}}$ ) in the touch and control conditions (see also Supplementary Table 1). A paired samples *t*-test indicated that there was no significant difference in betrayal aversion between the two conditions,  $M_{\text{touch}} = -0.005$  (95% CI  $[-0.039, 0.030]$ ),  $M_{\text{control}} = 0.017$  (95% CI  $[-0.016, 0.048]$ ),  $t_{(99)} = -0.48$ ,  $p = 0.633$ . The regression analyses found no significant effect either (see Table 1). That is, participants were not significantly less betrayal averse in the touch condition compared to the control condition,  $\beta = -0.021$ ,  $p = 0.320$ . Touch pleasantness

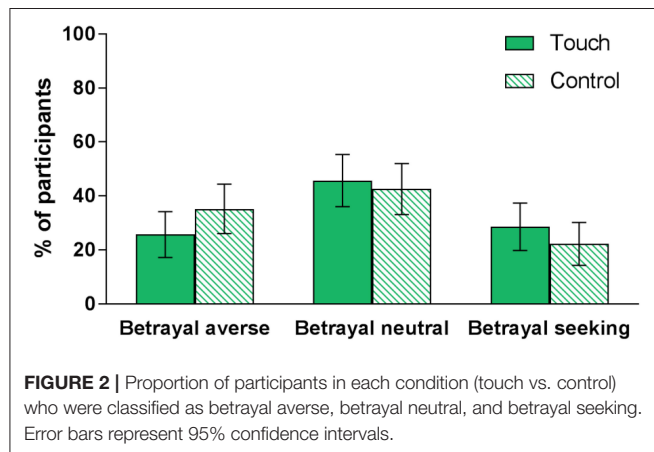
<sup>3</sup>The difference in proportions of betrayal averse, betrayal neutral, and betrayal seeking participants in the first round of the Betrayal Aversion Elicitation Task was also non-significant, Chi-Square test,  $p = 0.727$  (see Supplementary Figure 1).

did not correlate with betrayal aversion in the touch condition, Spearman's  $\rho = -0.09$ ,  $p = 0.348$ .

Because participants' responses are likely to be relatively consistent between the first and second round of the task, we also performed a between-subjects analysis to investigate the effect of touch in the first round, i.e., the first time participants performed the task. **Figure 3B** displays the results from this analysis (see also Supplementary Table 2). An independent samples  $t$ -test indicated that there was no significant difference in betrayal aversion between the two conditions,  $M_{\text{touch}} = -0.030$  (95% CI  $[-0.090, 0.030]$ ),  $M_{\text{control}} = 0.008$  (95% CI  $[-0.047, 0.064]$ ),  $t_{(102)} = -0.95$ ,  $p = 0.344$ . Regression analyses found no significant effects either (see Supplementary Table 2). There was a weak, negative correlation between betrayal aversion and touch pleasantness ratings, Spearman's  $\rho = -0.28$ ,  $p = 0.047$ . However, this correlation seemed to be driven by an outlier. When the outlier was excluded, the correlation was no longer significant, Spearman's  $\rho = -0.24$ ,  $p = 0.099$ .

## The Effect of Touch on Altruism

**Figure 4A** displays the mean amount donated to UNICEF in the dictator game, separated by condition (touch vs. control). There was no significant difference in donations between the two conditions,  $M_{\text{touch}} = 35.24\%$  (95% CI  $[26.88, 39.62]$ ),



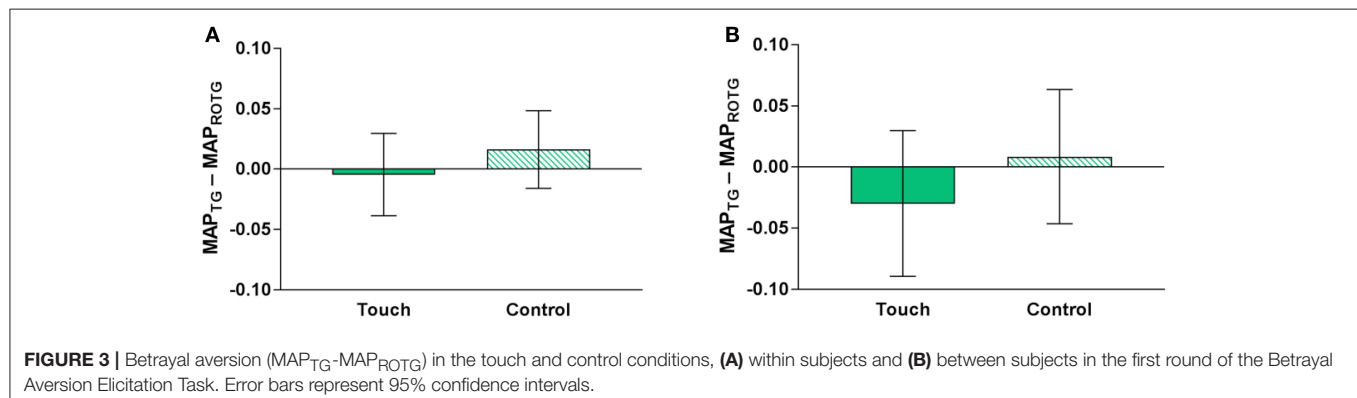
**FIGURE 2 |** Proportion of participants in each condition (touch vs. control) who were classified as betrayal averse, betrayal neutral, and betrayal seeking. Error bars represent 95% confidence intervals.

$M_{\text{control}} = 32.70\%$  (95% CI  $[26.34, 39.06]$ ), paired samples  $t_{(119)} = 0.46$ ,  $p = 0.649$ . The regression analyses found no significant effect either (see **Table 2**). That is, participants did

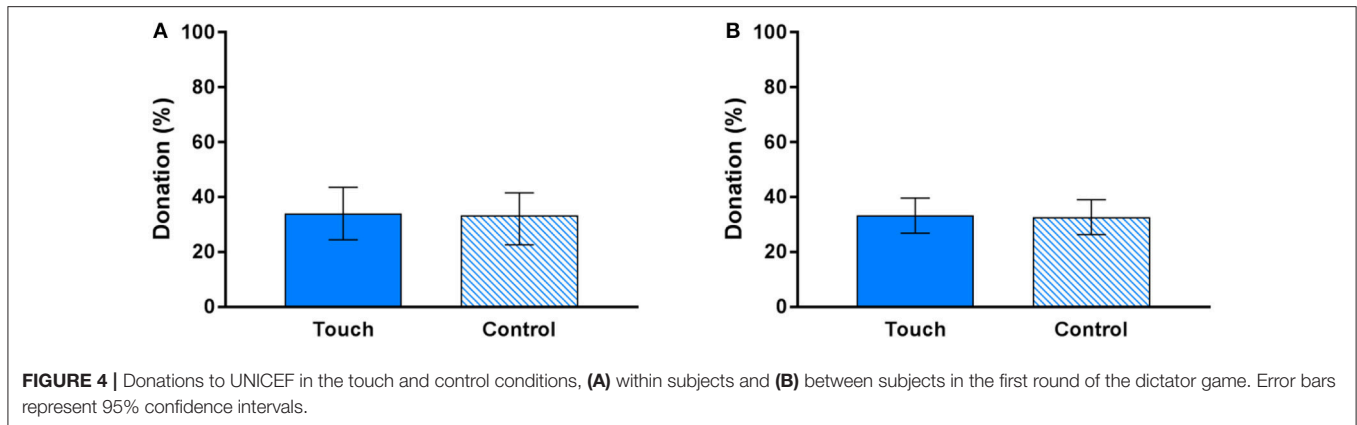
**TABLE 1 |** Regression analyses of betrayal aversion.

	(1)	(2)	(3)	(4)
Touch	-0.021 (0.021)	-0.017 (0.045)	-0.021 (0.021)	0.042 (0.099)
Round	0.032 (0.021)	0.013 (0.034)	0.032 (0.021)	0.014 (0.034)
Touch × Round		0.035 (0.055)		0.034 (0.055)
Female	-0.012 (0.026)	0.013 (0.036)	-0.009 (0.027)	-0.009 (0.027)
Touch × Female		-0.050 (0.041)		0.018 (0.040)
Age	-0.001 (0.003)	-0.001 (0.003)	-0.001 (0.003)	-0.001 (0.003)
Anxiety			-0.005 (0.013)	-0.004 (0.014)
Touch × Anxiety				-0.002 (0.017)
Avoidance			0.015 (0.023)	0.024 (0.039)
Touch × Avoidance				-0.020 (0.044)
Constant	0.022 (0.071)	0.023 (0.075)	-0.001 (0.099)	-0.028 (0.134)

This table reports OLS coefficient estimates (robust standard errors corrected for clustering on the individual level in parentheses). The dependent variable is participants' betrayal aversion ( $MAP_{TG} - MAP_{ROTG}$ ). "Touch" is a dummy for the touch condition. "Round" is a dummy for the second round of the tasks, i.e., the second time the participants performed the tasks. "Touch × Round" is the interaction between the touch condition and the task round, allowing the effect of touch to differ across the two task rounds. "Female" is a gender dummy. "Touch × Female" is the interaction between the touch condition and gender, allowing the effect of touch to differ between men and women. "Age" is the participant's age in years. "Anxiety" is the participant's score on the attachment anxiety subscale. "Touch × Anxiety" is the interaction between the touch condition and attachment anxiety, allowing the effect of touch to vary with the level of attachment anxiety. "Avoidance" is the participant's score on the attachment avoidance subscale. "Touch × Avoidance" is the interaction between the touch condition and attachment avoidance, allowing the effect of touch to vary with the level of attachment avoidance. All  $ps > 0.10$ .



**FIGURE 3 |** Betrayal aversion ( $MAP_{TG} - MAP_{ROTG}$ ) in the touch and control conditions, **(A)** within subjects and **(B)** between subjects in the first round of the Betrayal Aversion Elicitation Task. Error bars represent 95% confidence intervals.



not donate more money in the touch compared to the control condition,  $\beta = 0.550$ ,  $p = 0.650$ . There was an interaction between touch and gender such that women donated more money to UNICEF in the touch than in the control condition; however, this interaction was only significant at the 10% level,  $\beta = 4.472$ ,  $p = 0.072$ . Furthermore, there was a significant effect of attachment avoidance such that those high in attachment avoidance donated less money,  $\beta = -11.874$ ,  $p = 0.014$ . However, this finding should be interpreted with caution since it is uncorrected for multiple hypothesis testing. Attachment anxiety had no significant effect and there were no interactions between touch and attachment anxiety or attachment avoidance. Touch pleasantness did not correlate with amount donated in the touch condition, Spearman's  $\rho = 0.10$ ,  $p = 0.256$ .

As with betrayal aversion, we also conducted between-subjects analyses to investigate the effect of touch in the first round. **Figure 4B** displays the mean amount donated to UNICEF in the first round of the dictator game, separated by condition. There was no significant difference between the two conditions,  $M_{\text{touch}} = 34.00\%$  (95% CI [24.43, 43.57]),  $M_{\text{control}} = 33.33\%$  (95% CI [24.55, 42.12]), independent samples  $t_{(118)} = 0.10$ ,  $p = 0.918$ . Regression analyses found no significant effects either, apart from the effect of attachment avoidance mentioned above (see Supplementary Table 4). Touch pleasantness did not correlate with amount donated in the touch condition, Spearman's  $\rho = -0.12$ ,  $p = 0.343$ .

## The Effect of Touch on Risk Taking

**Figure 5A** displays the adjusted average number of pumps per trial in the BART, separated by condition (touch vs. control). There was no significant difference in the number of pumps between the two conditions,  $M_{\text{touch}} = 36.14$  (95% CI [33.58, 38.69]),  $M_{\text{control}} = 36.40$  (95% CI [33.83, 38.97]), paired samples  $t_{(118)} = -0.38$ ,  $p = 0.708$ , thus indicating that affective touch does not influence risk taking<sup>4</sup>. The regression analyses found no significant effect either (see **Table 3**). That is, participants were not more risk taking in the touch condition compared to the

control condition,  $\beta = -0.237$ ,  $p = 0.714$ . However, there was a significant effect of *Round*, such that participants were more risk taking in the second compared to the first round of the tasks,  $\beta = 3.128$ ,  $p < 0.0001$ . This is expected given that the number of pumps increases toward the end of the task (Lejuez et al., 2002). There was also a significant effect of gender, indicating that women were less risk taking than men,  $\beta = -7.487$ ,  $p = 0.003$ . This is in line with previous findings from the BART (Lejuez et al., 2002) and from other measures of risk taking (Byrnes et al., 1999; Charness and Gneezy, 2012). Again, note that these  $p$ -values are uncorrected and should be interpreted with caution. Touch pleasantness did not correlate with risk taking in the touch condition, Spearman's  $\rho = 0.08$ ,  $p = 0.400$ .

**Figure 5B** displays the adjusted average number of pumps per trial in the first round of the BART, separated by condition. There was no significant difference in the number of pumps between the two conditions,  $M_{\text{touch}} = 35.55$  (95% CI [31.77, 39.33]),  $M_{\text{control}} = 33.85$  (95% CI [30.34, 37.35]), paired samples  $t_{(117)} = 0.66$ ,  $p = 0.510$ . Regression analyses found no significant effect either (see Supplementary Table 7). Touch pleasantness did not correlate with risk taking, Spearman's  $\rho = 0.17$ ,  $p = 0.203$ .

## DISCUSSION

We investigated the effect of pleasant touch on betrayal aversion, altruism, and risk taking. Pleasant touch activates CT fibers in the skin, which are thought to mediate the oxytocin-enhancing effects of touch (Walker et al., 2017). Our results indicate no effect of touch on any of the outcome variables, neither within subjects nor between subjects. Furthermore, there were no significant interactions between touch and gender or attachment styles.

Given the lack of consistency in previous studies investigating the effect of oxytocin on trust (Nave et al., 2015), it is perhaps unsurprising that we find no effect of touch on betrayal aversion. Several issues have been pointed out in the oxytocin literature, including publication bias (Lane et al., 2016), low statistical power (Walum et al., 2016), lack of evidence that intranasal oxytocin reaches target brain areas (Leng and Ludwig, 2016), and unreliable measures of plasma oxytocin (McCullough et al., 2013; Christensen et al., 2014). This suggests that what we

<sup>4</sup>One participant's data was lost due to technical issues. Separating the task into the first, middle, and last 10 trials yielded no significant results (see Supplementary Table 5, 6).



**TABLE 2 |** Regression analyses of altruism.

	(1)	(2)	(3)	(4)
Touch	0.550 (1.208)	−0.501 (6.618)	0.550 (1.213)	−4.348 (9.756)
Round	−1.383 (1.208)	0.422 (6.517)	−1.383 (1.213)	−0.794 (6.386)
Touch × Round		−1.773 (12.822)		−1.086 (12.571)
Female	2.103 (6.347)	−0.160 (6.469)	0.490 (6.637)	−1.795 (6.772)
Touch × Female		4.472* (2.462)		4.535* (2.596)
Age	0.780 (0.490)	0.787 (0.499)	0.877 (0.463)	0.881* (0.474)
Anxiety			−1.021 (3.587)	−1.219 (3.642)
Touch × Anxiety				0.413 (1.278)
Avoidance			−11.409 (4.679)	−11.874** (4.783)
Touch × Avoidance				0.936 (1.627)
Constant	13.155 (13.443)	13.496 (14.442)	43.781** (17.287)	46.086** (18.711)

This table reports OLS coefficient estimates (robust standard errors corrected for clustering on the individual level in parentheses). The dependent variable is the amount donated to UNICEF. "Touch" is a dummy for the touch condition. "Round" is a dummy for the second round of the tasks, i.e., the second time the participants performed the tasks. "Touch × Round" is the interaction between the touch condition and the task round, allowing the effect of touch to differ across the two task rounds. "Female" is a gender dummy. "Touch × Female" is the interaction between the touch condition and gender, allowing the effect of touch to differ between men and women. "Age" is the participant's age in years. "Anxiety" is the participant's score on the attachment anxiety subscale. "Touch × Anxiety" is the interaction between the touch condition and attachment anxiety, allowing the effect of touch to vary with the level of attachment anxiety. "Avoidance" is the participant's score on the attachment avoidance subscale. "Touch × Avoidance" is the interaction between the touch condition and attachment avoidance, allowing the effect of touch to vary with the level of attachment avoidance. \* $p < 0.10$ , \*\* $p < 0.05$ .

think we know about oxytocin in humans may not be true. Furthermore, as suggested by Bartz et al. (2011; see also Shamay-Tsoory and Abu-Akel, 2016), the effect of oxytocin on trust and prosocial behavior—if there is one—is likely constrained by both individual and contextual factors. For example, previous studies have suggested that oxytocin reduces investments in a trust game following betrayal in women but not in men (Yao et al., 2014) and that oxytocin increases trust and reduces betrayal aversion in individuals that are high, compared to low, in attachment avoidance (De Dreu, 2012). However, in our study, we found no significant interactions between touch and gender or attachment. Regarding contextual factors, previous studies have shown that oxytocin increases trust when trustees are described as trustworthy but not when they are described as untrustworthy (Mikolajczak et al., 2010) and that oxytocin increases trust and altruism toward the in-group but results in defensive behaviors toward the outgroup (De Dreu et al., 2010). Oxytocin also increased cooperation (which requires some

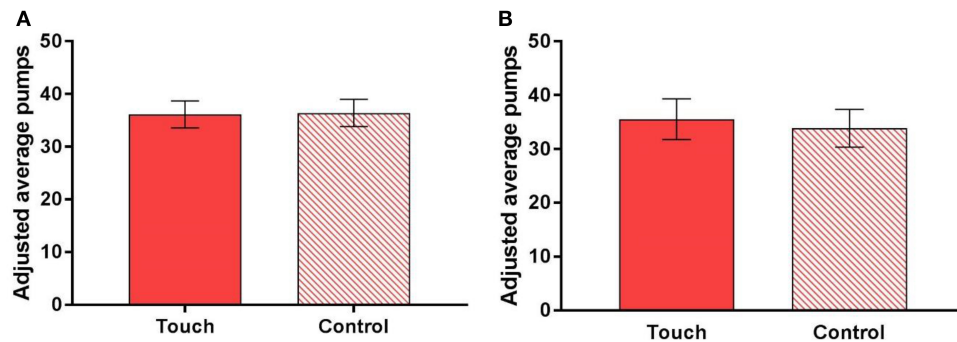
**TABLE 3 |** Regression analyses of risk taking.

	(1)	(2)	(3)	(4)
Touch	−0.237 (0.645)	1.814 (2.650)	−0.237 (0.647)	2.937 (3.829)
Round	3.128*** (0.645)	5.179** (2.489)	3.128*** (0.647)	5.268** (2.500)
Touch × Round		4.101 (4.848)		−4.289 (4.860)
Female	−7.487*** (2.421)	−7.532*** (2.519)	−7.882*** (2.473)	−7.897*** (2.539)
Touch × Female		−0.001 (1.297)		−0.102 (1.328)
Age	−0.097 (0.172)	−0.082 (0.167)	−0.090 (0.175)	−0.074 (0.171)
Anxiety			1.155 (1.478)	1.140 (1.453)
Touch × Anxiety				0.143 (0.815)
Avoidance			−1.171 (2.012)	−0.954 (2.016)
Touch × Avoidance				−0.518 (0.979)
Constant	40.488*** (4.542)	39.105*** (4.588)	40.622*** (6.864)	38.632*** (6.918)

This table reports OLS coefficient estimates (robust standard errors corrected for clustering on the individual level in parentheses). The dependent variable is adjusted average pumps, i.e., the average number of pumps per trial in the BART excluding trials on which the balloon exploded. "Touch" is a dummy for the touch condition. "Round" is a dummy for the second round of the tasks, i.e., the second time the participants performed the tasks. "Touch × Round" is the interaction between the touch condition and the task round, allowing the effect of touch to differ across the two task rounds. "Female" is a gender dummy. "Touch × Female" is the interaction between the touch condition and gender, allowing the effect of touch to differ between men and women. "Age" is the participant's age in years. "Anxiety" is the participant's score on the attachment anxiety subscale. "Touch × Anxiety" is the interaction between the touch condition and attachment anxiety, allowing the effect of touch to vary with the level of attachment anxiety. "Avoidance" is the participant's score on the attachment avoidance subscale. "Touch × Avoidance" is the interaction between the touch condition and attachment avoidance, allowing the effect of touch to vary with the level of attachment avoidance. \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

degree of trust) in a coordination game when there had been prior contact between participants but reduced cooperation when there had been no prior contact (Declerck et al., 2010). This finding is particularly noteworthy because in the study by Kosfeld et al. (2005), which provided the initial evidence for a causal link between oxytocin and trust, participants introduced themselves to each other before they played the trust game. In contrast, participants in our study played with anonymous counterparts. Therefore, given that oxytocin may enhance the salience of social cues (Shamay-Tsoory and Abu-Akel, 2016), the absence of an effect of touch on betrayal aversion in our study could, at least in part, be due to the lack of social information.

An alternative explanation for our null effect of touch on betrayal aversion is that the size of betrayal aversion was small to begin with, indicating that participants made little difference between the trust game and the risk-only trust game. Early studies reported betrayal aversion sizes ranging from 0.08 to



**FIGURE 5 |** Average number of pumps per trial in the Balloon Analog Risk Task, excluding trials on which the balloon exploded and separated by condition (touch vs. control), (A) within subjects and (B) between subjects in the first round of the task. Error bars represent 95% confidence intervals.

0.22 (Bohnet et al., 2008). Studies using the Betrayal Aversion Elicitation Task, which assesses betrayal aversion within subjects, have reported betrayal aversion sizes of 0.04 (Aimone et al., 2015) and 0.07 (Quercia, 2016). Betrayal aversion in our study was  $-0.005$  (indicating slightly betrayal seeking or betrayal neutral) in the touch condition and 0.017 (slightly betrayal averse) in the control condition. Furthermore, the proportion of participants that could be categorized as betrayal averse was lower than the proportion of participants that could be categorized as betrayal neutral, which contradicts previous findings that people generally are betrayal averse (Bohnet and Zeckhauser, 2004; Bohnet et al., 2008; Aimone et al., 2015). One possible explanation for these discrepancies is that participants in previous studies (e.g., Bohnet and Zeckhauser, 2004; Bohnet et al., 2008; Aimone et al., 2015; Quercia, 2016) were tested in groups, meaning that any betrayal occurred there and then as a result of the decision of another participant that was present in the same room. In contrast, participants in our study were tested individually and played with anonymous counterparts who had already made their decisions prior to the study. Therefore, the potential betrayal may have felt less personal, which, in turn, may have reduced the negative affective experience associated with the possibility of being betrayed (Lauharatanahirun et al., 2012; Aimone et al., 2014).

The reason we found no effect of touch on altruism could, again, be that there is no direct, causal effect and/or that it depends on individual and contextual factors. Some researchers have found that oxytocin increases donations to charitable organizations (Barraza et al., 2011; Marsh et al., 2015) and that it increases monetary contributions both to the in-group and to the outgroup in a social dilemma (Israel et al., 2012). The prediction that follows from this line of research is that touch increases altruism, which is not what we found in the present study. Instead, our findings are in line with studies showing no effect of oxytocin on altruism (Zak et al., 2007; Hu et al., 2016). Nevertheless, other researchers have found that the effect depends on the closeness of the relationship to the target (Pornpattananangkul et al., 2017) and whether the target belongs to the in-group or the outgroup (De Dreu et al., 2010). We did not take such contextual factors into account. We did find a trend

such that touch increased altruism in women more than in men, but this interaction was significant only at the 10% significance level and should be interpreted with caution.

The lack of an effect of touch on risk taking makes sense given that previous studies have found no effect of oxytocin on non-social risk taking (e.g., Kosfeld et al., 2005; Patel et al., 2015). However, it is at odds with studies showing that brief physical contact increases risk taking (Levav and Argo, 2010). A limitation of our study is that we did not measure actual hormone levels, so we cannot rule out the possibility that our lack of effects is due to a failure to increase oxytocin. An alternative possibility is that touch increases positive affect, which, in turn, reduces betrayal aversion and increases altruism and risk taking. Indeed, CT-optimal touch is perceived as pleasant and rewarding (Perini et al., 2015) and positive affect has been suggested as one of the mechanisms underlying the Midas effect (Schirmer et al., 2016). However, studies finding an effect of touch on altruism and risk taking have investigated touch in the form of brief physical contact, such as a tap on the shoulder (Kleinke, 1977; Crusco and Wetzel, 1984; Guéguen and Fischer-Iokou, 2003; Levav and Argo, 2010). Here, we investigated the effect of continuous, gentle stroking that lasted throughout the decision phase. This distinction is important for several reasons. First, it is possible that any effect in our study was reduced because the manipulation was obvious to participants. Second, it is possible that participants attributed any affective changes to the touch and that the influence on behavior diminished as a result. Third, incidental affect from the touch may not have been strong enough to override integral affect from the BART, which is rich in affective cues (for a discussion of the integration of incidental and integral affect in decision making, see Västfjäll et al., 2016). Moreover, in studies reporting an effect of brief touch on prosocial behavior, the prosocial behavior was directed toward the toucher, such as the waitress receiving a tip (Crusco and Wetzel, 1984). In contrast, participants in our study donated money to a charitable organization. It is possible that increases in positive affect are attributed to the person delivering the touch, and that touch therefore promotes prosocial behavior only toward the toucher.

In conclusion, we found no effect of touch on betrayal aversion, altruism, or risk taking. These results add to a growing body of research suggesting that oxytocin has no direct, causal effect on trust and prosocial behaviors. Nonetheless, we remain optimistic that touch plays a vital role for social and psychological well-being. It is possible that its effects on economic decision making and behavior are dependent on the social context in a way that may be difficult to study in a laboratory setting. Future research should continue to investigate the circumstances under which affective touch—and its hormonal correlates—influences social behaviors and economic decision making.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of human research guidelines, Regional Ethics Board for East Gothland with written informed consent from all subjects.

## REFERENCES

- Aimone, J. A., Houser, D., and Weber, B. (2014). Neural signatures of betrayal aversion: an fMRI study of trust. *Proc. Biol. Sci.* 281:20132127. doi: 10.1098/rspb.2013.2127
- Aimone, J., Ball, S., and King-Casas, B. (2015). The betrayal aversion elicitation task: an individual level betrayal aversion measure. *PLoS ONE* 10:e0137491. doi: 10.1371/journal.pone.0137491
- Barraza, J. A., McCullough, M. E., Ahmadi, S., and Zak, P. J. (2011). Oxytocin infusion increases charitable donations regardless of monetary resources. *Horm. Behav.* 60, 148–151. doi: 10.1016/j.yhbeh.2011.04.008
- Bartz, J. A., Zaki, J., Bolger, N., and Ochsner, K. N. (2011). Social effects of oxytocin in humans: context and person matter. *Trends Cogn. Sci.* 15, 301–309. doi: 10.1016/j.tics.2011.05.002
- Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., and Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron* 58, 639–650. doi: 10.1016/j.neuron.2008.04.009
- Bohnet, I., Greig, F., Herrmann, B., and Zeckhauser, R. (2008). Betrayal aversion: evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *Am. Econ. Rev.* 98, 294–310. doi: 10.1257/aer.98.1.294
- Bohnet, I., and Zeckhauser, R. (2004). Trust, risk and betrayal. *J. Econ. Behav. Organ.* 55, 467–484. doi: 10.1016/j.jebo.2003.11.004
- Byrnes, J. P., Miller, D. C., and Schafer, W. D. (1999). Gender differences in risk taking: a meta-analysis. *Psychol. Bull.* 125, 367–383. doi: 10.1037/0033-2909.125.3.367
- Charness, G., and Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *J. Econ. Behav. Organ.* 83, 50–58. doi: 10.1016/j.jebo.2011.06.007
- Christensen, J. C., Shiyanov, P. A., Estep, J. R., and Schlager, J. J. (2014). Lack of association between human plasma oxytocin and interpersonal trust in a prisoner's dilemma paradigm. *PLoS ONE* 9:e116172. doi: 10.1371/journal.pone.0116172
- Collins, N. L. (1996). Working models of attachment: implications for explanation, emotion, and behavior. *J. Pers. Soc. Psychol.* 71, 810–832. doi: 10.1037/0022-3514.71.4.810
- Croy, I., Luong, A., Tricoli, C., Hofmann, E., Olausson, H., and Sailer, U. (2016). Interpersonal stroking touch is targeted to C tactile afferent activation. *Behav. Brain Res.* 297, 37–40. doi: 10.1016/j.bbr.2015.09.038
- Crusco, A. H., and Wetzel, C. G. (1984). The midas touch: the effects of interpersonal touch on restaurant tipping. *Pers. Soc. Psychol. Bull.* 10, 512–517. doi: 10.1177/0146167284104003

## AUTHOR CONTRIBUTIONS

LK, IM, DV, and GT developed the study concept and design. LK collected the data and LK and DA analyzed the results. All authors contributed to the interpretation of the data. LK wrote the first draft of the paper and DA, IM, DV, and GT provided revisions. All authors approved the final version.

## FUNDING

This research was funded by Marianne and Marcus Wallenberg Foundation grant No. 2014.0187 and Ragnar Söderberg Foundation.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnbeh.2017.00251/full#supplementary-material>

- Declerck, C. H., Boone, C., and Kiyonari, T. (2010). Oxytocin and cooperation under conditions of uncertainty: the modulating role of incentives and social information. *Horm. Behav.* 57, 368–374. doi: 10.1016/j.yhbeh.2010.01.006
- De Dreu, C. K. W. (2012). Oxytocin modulates the link between adult attachment and cooperation through reduced betrayal aversion. *Psychoneuroendocrinology* 37, 871–880. doi: 10.1016/j.psyneuen.2011.10.003
- De Dreu, C. K. W., Greer, L. L., Handgraaf, M. J. J., Shalvi, S., Van Kleef, G. A., Baas, M., et al. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science* 328, 1408–1411. doi: 10.1126/science.1189047
- Domes, G., Heinrichs, M., Michel, A., Berger, C., and Herpertz, S. C. (2007). Oxytocin improves “mind-reading” in humans. *Biol. Psychiatry* 61, 731–733. doi: 10.1016/j.biopsych.2006.07.015
- Engelmann, J., and Fehr, E. (2017). “The neurobiology of trust and social decision-making: the important role of emotions,” in *Social Dilemmas: New Perspectives on Reward and Punishment*, eds P. A. M. van Lange, B. Rockenbach, and T. Yamagishi (New York, NY: Oxford University Press). doi: 10.1093/oso/9780190630782.001.0001
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *J. Econ. Sci. Assoc.* 1, 114–125. doi: 10.1007/s40881-015-0004-4
- Guéguen, N., and Fischer-lokou, J. (2003). Tactile contact and spontaneous help: an evaluation in a natural setting. *J. Soc. Psychol.* 143, 785–787. doi: 10.1080/00224540309600431
- Hu, Y., Scheele, D., Becker, B., Voos, G., David, B., Hurlmann, R., et al. (2016). The effect of oxytocin on third-party altruistic decisions in unfair situations: an fMRI study. *Sci. Rep.* 6:20236. doi: 10.1038/srep20236
- Israel, S., Weisel, O., Ebstein, R. P., and Bornstein, G. (2012). Oxytocin, but not vasopressin, increases both parochial and universal altruism. *Psychoneuroendocrinology* 37, 1341–1344. doi: 10.1016/j.psyneuen.2012.02.001
- Klackl, J., Pfundmair, M., Agroskin, D., and Jonas, E. (2013). Who is to blame? Oxytocin promotes nonpersonalistic attributions in response to a trust betrayal. *Biol. Psychol.* 92, 387–394. doi: 10.1016/j.biopsycho.2012.11.010
- Kleinke, C. L. (1977). Compliance to requests made by gazing and touching experimenters in field settings. *J. Exp. Soc. Psychol.* 13, 218–223. doi: 10.1016/0022-1031(77)90044-0
- Koppel, L., Andersson, D., Morrison, I., Posadzy, K., Västfjäll, D., and Tinghög, G. (2017). The effect of acute pain on risky and intertemporal choice. *Exp. Econ.* 20, 878–893. doi: 10.1007/s10683-017-9515-6
- Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., and Fehr, E. (2005). Oxytocin increases trust in humans. *Nature* 435, 673–676. doi: 10.1038/nature03701

- Lane, A., Luminet, O., Nave, G., and Mikolajczak, M. (2016). Is there a publication bias in behavioural intranasal oxytocin research on humans? opening the file drawer of one laboratory. *J. Neuroendocrinol.* 28. doi: 10.1111/jne.12384
- Lauharatanahirun, N., Christopoulos, G. I., and King-Casas, B. (2012). Neural computations underlying social risk sensitivity. *Front. Hum. Neurosci.* 6:213. doi: 10.3389/fnhum.2012.00213
- Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., et al. (2002). Evaluation of a behavioral measure of risk taking: the BALLOON Analogue Risk Task (BART). *J. Exp. Psychol. Appl.* 8, 75–84. doi: 10.1037/1076-898X.8.2.75
- Leng, G., and Ludwig, M. (2016). Intranasal oxytocin: myths and delusions. *Biol. Psychiatry* 79, 243–250. doi: 10.1016/j.biopsych.2015.05.003
- Leppanen, J., Ng, K. W., Tchanturia, K., and Treasure, J. (2017). Meta-analysis of the effects of intranasal oxytocin on interpretation and expression of emotions. *Neurosci. Biobehav. Rev.* 78, 125–144. doi: 10.1016/j.neubiorev.2017.04.010
- Levay, J., and Argo, J. J. (2010). Physical contact and financial risk taking. *Psychol. Sci.* 21, 804–810. doi: 10.1177/0956797610369493
- Löken, L. S., Wessberg, J., Morrison, I., McGlone, F., and Olausson, H. (2009). Coding of pleasant touch by unmyelinated afferents in humans. *Nat. Neurosci.* 12, 547–548. doi: 10.1038/nn.2312
- Marsh, N., Scheele, D., Gerhardt, H., Strang, S., Enax, L., Weber, B., et al. (2015). The neuropeptide oxytocin induces a social altruism bias. *J. Neurosci.* 35, 15696–15701. doi: 10.1523/JNEUROSCI.3199-15.2015
- McCullough, M. E., Churchland, P. S., and Mendez, A. J. (2013). Problems with measuring peripheral oxytocin: can the data on oxytocin and human behavior be trusted? *Neurosci. Biobehav. Rev.* 37, 1485–1492. doi: 10.1016/j.neubiorev.2013.04.018
- Mikolajczak, M., Gross, J. J., Lane, A., Corneille, O., de Timary, P., and Luminet, O. (2010). Oxytocin makes people trusting, not gullible. *Psychol. Sci.* 21, 1072–1074. doi: 10.1177/0956797610377343
- Morhenn, V. B., Park, J. W., Piper, E., and Zak, P. J. (2008). Monetary sacrifice among strangers is mediated by endogenous oxytocin release after physical contact. *Evol. Hum. Behav.* 29, 375–383. doi: 10.1016/j.evolhumbehav.2008.04.004
- Nave, G., Camerer, C., and McCullough, M. (2015). Does oxytocin increase trust in humans? a critical review of research. *Perspect. Psychol. Sci.* 10, 772–789. doi: 10.1177/1745691615600138
- Olausson, H., Wessberg, J., Morrison, I., McGlone, F., and Vallbo, Å. (2010). The neurophysiology of unmyelinated tactile afferents. *Neurosci. Biobehav. Rev.* 34, 185–191. doi: 10.1016/j.neubiorev.2008.09.011
- Patel, N., Grillon, C., Pavletic, N., Rosen, D., Pine, D. S., and Ernst, M. (2015). Oxytocin and vasopressin modulate risk-taking. *Physiol. Behav.* 139, 254–260. doi: 10.1016/j.physbeh.2014.11.018
- Perini, I., Olausson, H., and Morrison, I. (2015). Seeking pleasant touch: neural correlates of behavioral preferences for skin stroking. *Front. Behav. Neurosci.* 9:8. doi: 10.3389/fnbeh.2015.00008
- Pornpattananangkul, N., Zhang, J., Chen, Q., Kok, B. C., and Yu, R. (2017). Generous to whom? the influence of oxytocin on social discounting. *Psychoneuroendocrinology* 79, 93–97. doi: 10.1016/j.psychen.2017.02.016
- Quercia, S. (2016). Eliciting and measuring betrayal aversion using the BDM mechanism. *J. Econ. Sci. Assoc.* 2, 48–59. doi: 10.1007/s40881-015-0021-3
- Radke, S., and de Bruijn, E. R. A. (2015). Does oxytocin affect mind-reading? a replication study. *Psychoneuroendocrinology* 60, 75–81. doi: 10.1016/j.psychen.2015.06.006
- Schirmer, A., Wijaya, M. T., and Liu, S. (2016). “The midas effect: how somatosensory impressions shape affect and other-concern,” in *Affective Touch and the Neurophysiology of CT Afferents*, eds H. Olausson, J. Wessberg, I. Morrison, and F. McGlone (New York, NY: Springer), 283–299.
- Shamay-Tsoory, S. G., and Abu-Akel, A. (2016). The social salience hypothesis of oxytocin. *Biol. Psychiatry* 79, 194–202. doi: 10.1016/j.biopsych.2015.07.020
- Uvnäs-Moberg, K., Handlin, L., and Petersson, M. (2015). Self-soothing behaviors with particular reference to oxytocin release induced by non-noxious sensory stimulation. *Front. Psychol.* 6:1529. doi: 10.3389/fpsyg.2014.01529
- Västfjäll, D., Slovic, P., Burns, W. J., Erlandsson, A., Koppel, L., Asutay, E., et al. (2016). The arithmetic of emotion: integration of incidental and integral affect in judgments and decisions. *Front. Psychol.* 7:325. doi: 10.3389/fpsyg.2016.00325
- Walker, S. C., Trotter, P. D., Swaney, W. T., Marshall, A., and McGlone, F. P. (2017). C-tactile afferents: cutaneous mediators of oxytocin release during affiliative tactile interactions? *Neuropeptides* 64, 27–38. doi: 10.1016/j.npep.2017.01.001
- Walum, H., Waldman, I. D., and Young, L. J. (2016). Statistical and methodological considerations for the interpretation of intranasal oxytocin studies. *Biol. Psychiatry* 79, 251–257. doi: 10.1016/j.biopsych.2015.06.016
- Yao, S., Zhao, W., Cheng, R., Geng, Y., Luo, L., and Kendrick, K. M. (2014). Oxytocin makes females, but not males, less forgiving following betrayal of trust. *Int. J. Neuropsychopharmacol.* 17, 1785–1792. doi: 10.1017/S146114571400090X
- Zak, P. J., Kurzban, R., and Matzner, W. T. (2005). Oxytocin is associated with human trustworthiness. *Horm. Behav.* 48, 522–527. doi: 10.1016/j.yhbeh.2005.07.009
- Zak, P. J., Stanton, A. A., and Ahmadi, S. (2007). Oxytocin increases generosity in humans. *PLoS ONE* 2:e1128. doi: 10.1371/journal.pone.0001128
- Zhong, S., Monakhov, M., Mok, H. P., Tong, T., Lai, P. S., Chew, S. H., et al. (2012). U-shaped relation between plasma oxytocin levels and behavior in the trust game. *PLoS ONE* 7:e51095. doi: 10.1371/journal.pone.0051095

**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 Koppel, Andersson, Morrison, Västfjäll and Tinghög. 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.





# Investigating Gender Differences under Time Pressure in Financial Risk Taking

Zhixin Xie<sup>1</sup>, Lionel Page<sup>1\*</sup> and Ben Hardy<sup>2</sup>

<sup>1</sup> Queensland Behavioral Economics Group, School of Economics and Finance, Queensland University of Technology, Brisbane, QLD, Australia, <sup>2</sup> School of Finance and Management, SOAS University of London, London, United Kingdom

There is a significant gender imbalance on financial trading floors. This motivated us to investigate gender differences in financial risk taking under pressure. We used a well-established approach from behavior economics to analyze a series of risky monetary choices by male and female participants with and without time pressure. We also used second to fourth digit ratio (2D:4D) and face width-to-height ratio (fWHR) as correlates of pre-natal exposure to testosterone. We constructed a structural model and estimated the participants' risk attitudes and probability perceptions via maximum likelihood estimation under both expected utility (EU) and rank-dependent utility (RDU) models. In line with existing research, we found that male participants are less risk averse and that the gender gap in risk attitudes increases under moderate time pressure. We found that female participants with lower 2D:4D ratios and higher fWHR are less risk averse in RDU estimates. Males with lower 2D:4D ratios were less risk averse in EU estimations, but more risk averse using RDU estimates. We also observe that men whose ratios indicate a greater prenatal exposure to testosterone exhibit a greater optimism and overestimation of small probabilities of success.

**Keywords:** testosterone, 2D:4D, fWHR, time pressure, risk taking

## OPEN ACCESS

### Edited by:

Pablo Brañas-Garza,  
Middlesex University, United Kingdom

### Reviewed by:

Marcello Sartarelli,  
University of Alicante, Spain  
Valerio Capraro,  
Centrum Wiskunde & Informatica,  
Netherlands

### \*Correspondence:

Lionel Page  
lionel.page@qut.edu.au

**Received:** 31 August 2017

**Accepted:** 28 November 2017

**Published:** 15 December 2017

### Citation:

Xie Z, Page L and Hardy B (2017)  
Investigating Gender Differences  
under Time Pressure in Financial Risk  
Taking.  
Front. Behav. Neurosci. 11:246.  
doi: 10.3389/fnbeh.2017.00246

## INTRODUCTION

Why are there so few women trading in the markets? The last 50 years have seen more and more women participating in the workforce. In many professions, the percentage of women approaches or exceeds 50% (see for example, Chambers Partners, 2014; Kaiser Family Foundation, 2015; Catalyst, 2016). Yet some professions stay firmly outside of this evolution. Professional traders on financial trading floors are unambiguously one of these cases. Although women represent more than half the workforce in financial services (Sethi et al., 2013) they are typically in marketing, compliance or HR roles (Jäkel and Moynihan, 2016). What scant data there is suggests that women comprise 15% of junior investment and trading roles (Green et al., 2009; Lietz, 2012).

The causes of this gender imbalance are still not well understood. While in some professions it is argued that an invisible ceiling prevents the access of women, (Korzec, 2000; Williams and Richardson, 2010; Truss, 2016) this is unlikely to be the case in finance, where performance pressure pushes firms to look for the best talent at all costs. A number of explanations have been advanced in both the academic and practitioner literature for the relative absence of women. Some explanations suggest that there are fundamental differences in cognition between the sexes (e.g., Sapienza et al., 2009), some that there are psychological differences (see Charness and Rustichini, 2011) and some

that social factors account for differences in behavior (Byrnes et al., 1999; Saqib and Chan, 2015) and that this, in turn accounts for the differences in representation. This study investigates a potential factor driving gender imbalance on trading floors: differences between men's and women's risk preferences, particularly under time pressure.

Trading is a pressurized activity where stakes are high and time is short (Oberlechner and Nimgade, 2005; Kocher and Sutter, 2006). To examine the relationship between risk-taking, time pressure and gender, we use a standard risk elicitation experiment with substantial incentives, where biological markers of prenatal exposure to testosterone are measured for men and women and where choices are observed under different degrees of time pressure.

This paper contributes to three distinct bodies of research: the literature on gender differences in risk attitudes, the literature on gender differences in financial behavior and careers, and the literature on stability of preferences.

There is a substantial body of research on gender differences in risk attitudes. One of the most common and consistent findings in the risk preference literature has been that men take more risk than women (Powell and Ansic, 1997; Byrnes et al., 1999; Eckel and Grossman, 2002; Croson and Gneezy, 2009). Croson and Gneezy (2009) discussed some explanations of the gender difference in risk taking, which included emotions, overconfidence and risk as challenge or threats. The search for the roots of these gender differences has pointed to the role played by the androgen hormone testosterone. Testosterone (T) is an androgenic hormone which plays a pivotal role in sexual differentiation. This organizing role of testosterone is what alters the course of fetal development from the default female pattern—in effect, it is what makes men men. In addition to this organizing and differentiating role, testosterone, is also thought to modulate behavior in a number of ways. Testosterone levels have been positively associated with a number of behaviors in adult men, including aggression (Archer, 2006), sensation seeking (Roberti, 2004), hostility (Hartgens and Kuipers, 2004), mate-seeking (Roney et al., 2003), and dominance (Mazur and Booth, 1998). Research in economics has shown that markers of pre-natal exposure to testosterone—in effect, measures of testosterone's organizing effects—have an impact on risk attitude (Coates and Page, 2009; Brañas-Garza and Rustichini, 2011; Garbarino et al., 2011; Brañas-Garza et al., 2017). We complement this research by investigating how prenatal testosterone exposure affects risk attitude decomposed into outcome sensitivity and probability sensitivity (in a RDU model).

This paper also contributes to the substantial literature on gender differences in financial behavior, which have been observed in both real and experimental markets. In the real market, men believe they are more competent than women (Graham et al., 2009), are more overconfident (Grinblatt and Keloharju, 2009), and trade more often than women (Barber and Odean, 2001). Deaves et al. (2010) found no gender effect in trading but observed that women traded less than men. Experimental studies, such as Fellner and Maciejovsky (2007), find that women submitted fewer offers and engaged in fewer trades than men. Eckel and Füllbrunn (2015) showed that

all-male markets yield significant price bubbles while all-female markets produced prices that were below fundamental value. A variety of reasons have been suggested for these differences in observed behavior. Research has suggested that men are more competitive (Niederle and Vesterlund, 2007), so drive harder to beat others. Men are perceived as selfish (Aguar et al., 2009; Brañas-Garza et al., 2016) and actually are more selfish (Rand, 2016).

One of the differences between men and women is in levels of testosterone. Coates et al. (2010) proposed a hypothesis suggesting that the irrational exuberance observed during market bubbles is mediated by testosterone. They speculated that men and women traders are likely to behave differently with male traders' behavior driving market instability. In the present study, we compare men and women's financial risk taking under time pressure. Time pressure is a key aspect of financial decisions on the trading floor. Traders make decisions in financial markets within seconds after new information becomes available (Busse and Green, 2002). In the light of this we theorized that gender differences under time pressure may be one of the factors driving the gender imbalance observed in these environments. If men and women make different decisions under time pressure then it may be that the market favors one decision making profile over another, and so favors one gender over another. Kocher et al. (2013) found that risk aversion for gains was robust under time pressure, whereas risk-seeking for losses turned into risk aversion under time pressure. For mixed prospects, i.e., a mixture of gains and losses, subjects became more loss-averse and more gain-seeking under time pressure. Nursimulu and Bossaerts (2014) found that the time-varying sensitivities translated into decreased risk aversion and increased probability distortions for gains under extreme time pressure. Capraro et al. (2017) examined the effect of time pressure and degree of deliberation on decisions about the allocation of resources. They did not, however, examine gender effects. Although there has been work on social preferences and time pressure, there is less work on risk attitudes under time pressure and very little on gender difference in risk attitudes under time pressure.

Finally, by investigating variations in risk preferences under time pressure, the paper contributes to the literature on the stability of economic preferences. The stability of preferences has been a shibboleth of much economic theory since Stigler and Becker's seminal paper (Stigler and Becker, 1977). Recent research, however, has shown that preferences are not as stable as hitherto supposed. Both explicit factors, for example time pressure (Kocher and Sutter, 2006), and implicit ones, such as levels of the hormone cortisol (Kandasamy et al., 2014), mean that people make different choices. Research in a number of fields has shown that time pressure affects the nature of interpersonal interaction, such as the levels of cooperation (Rand et al., 2012, 2014; Capraro and Cococcioni, 2015, 2016; Rand, 2016). Despite this, the impact of time pressure has been largely ignored by economics (Kocher and Sutter, 2006; De Paola and Gioia, 2016) and, what work there has been, has not clearly delineated the influence of time pressure on decision-making. Work rooted in experimental psychology has examined the speed vs. accuracy trade-off. Speedy decisions are thought to be of poorer quality,

as time pressure prevents effective information processing. This, in turn, leads individuals to fall back on heuristics rather than the information presented (see Kocher and Sutter, 2006). Where risk appetite is evaluated, most research has suggested that risk-taking increases with time pressure (Huber and Kunz, 2007; Young et al., 2012; Kocher et al., 2013; Hu et al., 2015). Only Young et al. (2012) examined gender differences, but found none.

Our research finds that, in line with previous research, male participants took more risk. In addition, we identified three patterns which shed new light on gender differences in risk attitudes. First, the degree of testosterone that men are exposed to *in utero* correlates with riskier decisions in later life. Second, testosterone exposure was associated with more optimism and overweighting of small probabilities of chances under time pressure for male participants, relative to female participants.

## MOTIVATIONS AND HYPOTHESES

There are two broad classes of explanation for why women are underrepresented in front office roles. The first is that women behave differently to men, and in ways which are not valued in financial services. The second group is that the front office provides an environment that neither welcomes women, nor is attractive to them. These two positions poles of the argument could be stylized as nature and nurture.

This paper focusses on the nature element of the debate. The differences between men and women begin at the moment of fertilization where the fusion of genetic material from each parent determines whether the fetus develops as a male or female. How do these biological differences play out so that, years later, men and women make, on average, very different decisions?

Biological sex is determined at conception and many of its effects are cemented *in utero*. The default pattern for developing embryos is female, but the Y chromosome contains the *SRY* gene which transforms the indifferent gonad into male testes. These testes then produce testicular hormones (e.g., testosterone) which confers the male primary and secondary sex characteristics. Between 12 and 18 weeks of gestation male fetal plasma testosterone levels reach nine times that of females causing the formation of male external genitalia and conformational alterations in the brain and spinal cord (Breedlove and Hampson, 2002). This testosterone peak also affects the length of the digits. Intra-uterine testosterone levels have been found negatively correlated with the ratio between the second and fourth digits (index and ring fingers, known as the 2D:4D ratio) (Lutchmaya et al., 2004). Higher concentrations of fetal testosterone produce lower 2D:4D ratios and men typically have lower 2D:4D ratios than women (Manning et al., 1998; McIntyre, 2006). Interestingly, no relationship between testosterone and 2D:4D ratio is observed (Hollier et al., 2015) when testosterone levels in umbilical blood are measured at birth. This may be a timing issue, as the *in utero* testosterone peak (see above) has passed and the post-partum peak (Swerdlow et al., 2002) has yet to occur.

During puberty, another androgen peak results in the development of male secondary sex characteristics and has

further effects on cerebral architecture. Again, this pubertal peak affects bodily conformation, notably in the ratio between facial width and height, or fWHR (Verdonck et al., 1999; Weston et al., 2007), with males having larger ratios than females.

These markers of testosterone exposure can be readily measured and impact on risk-taking and decision-making. Coates et al. (2009) found that male traders with lower 2D:4D had higher profitability and Coates and Page (2009) found that this result was entirely driven by greater risk-taking. Garbarino et al. (2011) designed a financially motivated decision-making experiment and found that: men had lower 2D:4D ratios than women and the difference was significant; women made more risk-averse choices compared with men, and both men and women with smaller digit ratios made riskier financial choices with effect being identical for men and women. Barel (2017) found that only women exhibited more financial risk taking with lower 2D:4D but higher optimism levels. However, no significant correlation between the 2D:4D and risk preferences were observed by Schipper (2014). Drichoutis and Nayga (2015) found no effect of digit ratio on either risk or time preferences. Studies using 2D:4D ratios are potentially confounded by a number of factors such as ethnic groups (Manning et al., 2007). Consequently, the relationship between 2D:4D and risk-taking is not conclusive. Brañas-Garza et al. (2017) provide a detailed review of this research. Little is known about the associations with fWHR. The differential impact of testosterone exposure on risk preferences for both genders remains inconclusive.

The 2D:4D ratio has been shown, in men, to be negatively correlated with good visual and spatial performance (Manning and Taylor, 2001; Kempel et al., 2005), dominance and masculinity (Fink et al., 2007), sensation-seeking (Fink et al., 2006), and overconfidence (Dalton and Ghosal, 2014; Neyse et al., 2016). Overconfident investors and those investors most prone to sensation seeking were found trading more frequently (Grinblatt and Keloharju, 2009). Tester and Campbell (2007) found that the significant relationship between the 2D:4D ratio and sporting achievement was nearly identical in both men and women. However, several traits were only found in women, for instance, sensation-seeking, psychoticism, neuroticism (Austin et al., 2002), verbal fluency (Manning, 2002) social cognition (Williams et al., 2003), and cognitive reflection (Bosch-Domènech et al., 2014). The predictions of the face width-to-height ratio (fWHR) were mostly found in men. Carré and McCormick (2008) found that male undergraduate students had a larger fWHR, higher scores of trait dominance, and more reactive aggression than female students. However, the individual differences in fWHR predict reactive aggression in men but not in women. Valentine et al. (2014) supported the finding that fWHR is a physical marker of dominance and men with higher ratios are more attractive to women. Lefevre et al. (2014) suggested links between fWHR and self-reported aggression in both men and women, as well as dominance in men, but not in women.

This study examines the relation between gender and risk-taking in situations with and without time pressure. We summarize our investigation in three questions:

Question 1: Does time pressure increase an appetite for risk?

Question 2: Is higher testosterone exposure associated with higher risk-taking?

Question 3: Is there heterogeneity by gender?

## METHODS

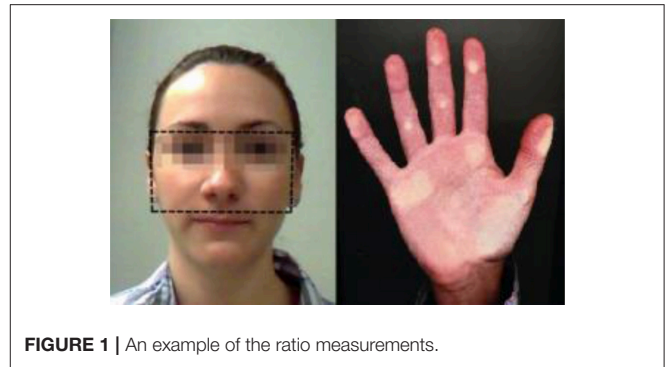
### Experimental Design

The experiment was programmed in zTree (Fischbacher, 2007) and conducted at Queensland University of Technology (QUT). Participants were recruited via the Queensland Behavioral Economics Group (QuBE) website, powered by Online Recruitment System for Economics and Experiments (ORSEE) (Greiner, 2004).<sup>1</sup> 154 students (74 females and 80 males) in total participated in 9 experimental sessions in this study and each experimental session lasted around 30–40 min. Upon entry to a laboratory at QuBE, participants were randomly assigned to a computer terminal. They were asked to complete the task individually and independently.

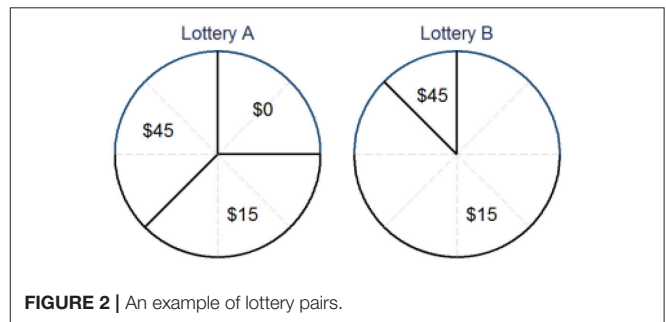
To measure the markers of participants' testosterone exposure, photographs of their faces were taken and right hands were scanned (see **Figure 1**). Then, the facial width was measured by the distance between the left and the right zygion (bizygomatic width) and the facial height was measured by the distance between the upper lip and brow (upper facial height Carré and McCormick, 2008, see photograph in **Figure 1**). The lengths of the second and fourth digits were measured from the basal crease (i.e., the crease closest to the base of the finger) to the central point of the fingertip (Garbarino et al., 2011; Neyse and Brañas-Garza, 2014).

Participants then engaged in a standard risk preference elicitation task using Random Lottery Pair design (Hey and Orme, 1994). This task consists of three phases and 30 decisions between pairs of lotteries per phase (90 decisions in total). Further, to investigate the role of time pressure, there are different time constraints imposed in each phase: no constraint, 8 and 4 s to make a decision in one lottery pair. These are the time constraints chosen by Kocher et al. (2013) in their study of risky decisions under time pressure. An 8 s constraint represents a moderate time pressure, while 4 s is a situation of extreme time pressure where participants have very little time to make a decision after discovering the different outcomes and their probability. We adopt a within-subject approach, which allows us to gain statistical power by controlling for unobservable characteristics. However, it also runs the risk of creating ordering effects. Therefore, to mitigate this risk, we randomized the order of the phases across experimental sessions.

Participants were presented with a pair of pie charts describing the probabilities of four fixed monetary prizes of 0, 15, 30, and \$45 (Australian Dollars).<sup>2</sup> An example of lottery pairs is shown



**FIGURE 1** | An example of the ratio measurements.



**FIGURE 2** | An example of lottery pairs.

in **Figure 2**. In this example, Lottery A offers a \$0 prize with a probability of 25%, \$15 with a probability of 37.5% and \$45 with a probability of 37.5%, whilst Lottery B offers a \$15 prize with a probability of 87.5 and \$45 with a probability of 12.5%.<sup>3</sup> Hence, the expected payoff is \$22.5 for Lottery A and \$18.75 for Lottery B. There were no numerical references to the probabilities and expected payoffs displayed; participants had to judge them from the pie chart within the given time constraint. No indifference choice was allowed between the two lotteries.

At the end of the 90 decisions, one lottery pair was randomly chosen and the participant's decision in this particular lottery pair was chosen. The "roulette wheel" of this lottery was then spun on their computer screen to determine their final payments. The instructions were explained in the form of a PowerPoint presentation before the start of the experiment, and they are shown in the Appendix.

### Estimation Procedure

To study risk preferences, we fit a rank dependent utility model (RDU). We use this model because of its general form. It contains

with a point in the Marschak-Machina triangle (space representing all possible lotteries with three outcomes). In this triangle, EU decision makers have linear indifference curves. The selection of lotteries creates pairs of points in the triangle, by varying the location of these pairs of points, the choice among lotteries reveals the slopes of the indifference curves in the triangle and whether these slopes are not constant (revealing that decision makers violate EU, for instance because they weight probabilities).

<sup>3</sup>The lotteries used in each phase had different probabilities, however, they were drawn randomly from the same pool. They did not differ in characteristics on average. The number of lotteries in each phase (30 decisions) limits the risk of substantial differences across phase due to the random selection of lotteries.

<sup>1</sup>The research ethics require participants being anonymous and unidentifiable during and after the experiment, therefore participants' personal information, such as age, faculty and ethnic groups were not collected. The ethics committee at QUT Business School approved this research and participants gave written informed consents before partaking the experiment.

<sup>2</sup>We used the set of lotteries from Conte et al. (2011) and Hey (2001) also used by many other studies including Moffatt (2005) and Conte et al. (2011). Hey (2001) explained the logic of the choice of lotteries. Each lottery can be associated



expected utility (EU) as a special case, allows us to disentangle risk preferences between a sensitivity to *payoffs* via the curvature of a *utility* function and a sensitivity to *probabilities* via the curvature of a *probability weighting* function (Wakker, 2010).

The utility of each lottery can be determined by the function:

$$V = \sum_{k=1}^K w_k \times U_k \quad (1)$$

Where

$$\begin{cases} w_i = \omega(p_i + \dots + p_n) & i = 1, \dots, n-1, \\ -\omega(p_{i+1} + \dots + p_n), & \\ w_i = \omega(p_i), & i = n \end{cases}$$

In the equations above,  $k = 1, \dots, K$  and  $K$  is the number of possible prizes in a lottery. The subscript of  $w_i$  indicates that the prizes were ranked from the smallest to the biggest. The probability weighting function  $\omega(p)$  is then applied to the aggregated probabilities, so the decision weights  $w_i$  are derived by the differences in these transformed aggregated probabilities.

We use the power constant relative risk aversion (CRRA) utility function:

$$U(x) = \frac{x^{1-\alpha}}{1-\alpha} \quad (2)$$

where  $x$  is each prize in a lottery and  $\alpha (\neq 1)$  is the coefficient and yet to be estimated. If  $\alpha > 0$ , it corresponded to a risk-averse attitude toward the actual payoff;  $\alpha < 0$  reflects a risk-loving attitude;  $\alpha = 0$  means risk-neutral.

Furthermore, we use the two-parameter weighting function in (Lattimore et al., 1992):

$$\omega(p) = \frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma} \quad (3)$$

where  $\delta, \gamma > 0$ . The parameter  $\gamma$  determines the curvature (concavity or convexity) of the probability weighting. If  $\gamma > 1$ , the function has an “S-shape.” It means that a small probability is underweighted by the agent. For example, while in **Figure 2**, the probability to win \$45 is 12.5% in Lottery B, an agent would act as if he/she believed this probability is lower than 12.5%. If  $\gamma < 1$  the function has an “inverse S-shape.” It means that a small probability is over weighted. Then an agent thinks that his or her chance receiving \$45 is  $>12.5\%$ .

The parameter  $\delta$  provides an additional weight on the probability weighting function. If  $\delta < 1$ , the probabilities are down weighted, indicating a pessimistic view of the payoffs. For example, an agent thinks that the chance of receiving \$45 is  $<12.5\%$  and that the chance of receiving \$15 is  $<87.5\%$ . On the contrary, if  $\delta > 1$ , the probabilities are over weighted, indicating that an agent holds an optimistic view toward the overall chances. Additionally, the EU is a special case when both  $\gamma = \delta = 1$ .

We estimate these parameters using a random utility approach whereby the decision maker sometimes does not select the option with the highest utility due to cognitive errors. We use a “context utility” specification, making the variance of these

cognitive errors depend on the magnitude of the payoffs being considered in the decision situation. This specification has been found to be better than alternatives which assume that errors are the same between different context of choice (Wilcox, 2011). The difference in utility between the two lotteries in a pair is modeled as:

$$\nabla V = \frac{\lambda(V_A - V_B)}{U(z_{\max}) - U(z_{\min})} \quad (4)$$

where  $\lambda$  represented the overall scale of the errors and the denominator is the influence of the specific context on the error in one lottery pair. The subscript of “A” and “B” represent the two lotteries and  $z_{\max}$  and  $z_{\min}$  denote the maximum and minimum possible payoffs in one pair.

The parameters  $\alpha, \gamma, \delta$  as the reflection of participants’ risk preference, and their perception of probabilities, are estimated by maximum likelihood method by using pooled data and clustering standard errors at each participant level. Therefore, the likelihood function is written as:

$$\ln L(\alpha, \gamma, \delta; y) = \sum_m ((\ln \Phi(\nabla V) | y_m = 1) + (\ln(1 - \Phi(\nabla V)) | y_m = 0)) \quad (5)$$

where  $y_m = 1(0)$  denotes the choice of lottery A (B) chosen in each pair  $m$ .

For ease of interpretation by the reader (and the authors), the 2D:4D ratios were reversed as R2D:4D, so that a higher ratio suggests higher testosterone exposure—just as higher fWHR suggests higher testosterone exposure. Both ratios are standardized. We also introduce two variables: “Male” as a gender dummy variable and “Time” as a categorical variable to measure the phases under three different time constraints. The parameters  $\alpha, \gamma$  and  $\delta$  are written as linear combination of variables, as written by the below equations, jointly in the maximum likelihood estimation:

$$\begin{aligned} \alpha &= \beta_0 + \beta_1 \text{Male} + \beta_2 \text{Ratio} + \beta_3 \text{Male} \times \text{Ratio} + \beta_4 \text{Time} \\ &\quad + \beta_5 \text{Male} \times \text{Time} + \beta_6 \text{Ratio} \times \text{Time} \\ &\quad + \beta_7 \text{Male} \times \text{Time} \times \text{Ratio} \\ \gamma &= \mu_0 + \mu_1 \text{Male} + \mu_2 \text{Ratio} + \mu_3 \text{Male} \times \text{Ratio} + \mu_4 \text{Time} \\ &\quad + \mu_5 \text{Male} \times \text{Time} + \mu_6 \text{Ratio} \times \text{Time} \\ &\quad + \mu_7 \text{Male} \times \text{Time} \times \text{Ratio} \\ \delta &= \varphi_0 + \varphi_1 \text{Male} + \varphi_2 \text{Ratio} + \varphi_3 \text{Male} \times \text{Ratio} + \varphi_4 \text{Time} \\ &\quad + \varphi_5 \text{Male} \times \text{Time} + \varphi_6 \text{Ratio} \times \text{Time} \\ &\quad + \varphi_7 \text{Male} \times \text{Time} \times \text{Ratio}. \end{aligned} \quad (6)$$

Therefore, our estimates are the parameters leading to the highest likelihood. After estimating our structural models from (1) to (6) jointly, we obtain two sets of estimations separately by using fWHR as “Ratio” (Estimation 1 in **Table 2**) and by using R2D:4D as “Ratio” (Estimation 2 in **Table 2**) in model (6). We can then investigate the interrelation between parameters and the effects of variables, by interpreting the coefficients for the sub-groups, for example, if  $\beta_3$  is significantly not equal to 0, it means the fWHR or

**TABLE 1** | Summary statistics.

Subsample	Observations ( <i>N</i> )	fWHR	2D:4D	Expected return of chosen lotteries	Variance of chosen lotteries
Males	80	1.842 (0.140)	0.963 (0.032)	22.581 (7.46)	141.899 (120.68)
Females	74	1.875 (0.108)	0.967 (0.042)	22.579 (7.50)	134.125 (115.20)
Mann-Whitney Test ( $H_0$ : Females = Males)		$z = 2.305$ $p = 0.021$	$z = -1.548$ $p = 0.122$	$z = -0.236$ $p = 0.814$	$z = -3.263$ $p = 0.021$

Standard deviations are in the parentheses.

R2D:4D has significantly different effects on males and females' risk attitude in our experiment.

## RESULTS

The summary statistics are presented in **Table 1**. The average fWHR for male participants in our experiment is 1.842 ( $SD = 0.140$ ), and the average for females is 1.875 ( $SD = 0.108$ ). The fWHR is not normally distributed in our sample. Therefore, we use a nonparametric test, the Mann-Whitney test to examine the differences between two gender groups. We find that male participants have lower fWHR than females (test statistic: 2.305 and  $p = 0.021$ ). The average 2D:4D ratio for males is 0.963 ( $SD = 0.032$ ) and for females is 0.967 ( $SD = 0.042$ ). However, the differences in 2D:4D ratio between male and female participants in our experiment are not significant (test statistic:  $-1.548$  and  $p = 0.122$ ).

The expected return of chosen lotteries for males is 22.581 ( $SD = 7.46$ ), showing no significant difference ( $p = 0.814$ ) from females of 22.579 ( $SD = 7.50$ ). However, females chose the lotteries with significantly ( $p = 0.021$ ) lower variance (134.125,  $SD = 115.20$ ) than males (141.899,  $SD = 120.68$ ). This suggests that female participants in our experiment have less appetite for risk. We have also used the Brown and Forsythe (1974) to examine the equality of the variances of chosen lotteries. The test result suggests that male participants have higher variances, as the Levene's robust test statistic ( $W_0$ ) is 10.498 with  $p = 0.001$ .

The CRRA function parameter  $\alpha$  is separately estimated under EU and RDU. The EU model is simply estimated like the RDU model with the parameters  $\gamma$ ,  $\delta$  each set to 1. Results for EU and RDU parameters are presented in **Table 2**.

We find that participants tend to have a concave utility function reflecting risk aversion ( $\alpha > 0$ ), both in the EU and RDEU estimation. We also find that the probability weighting function displays the typical "inverse S-shape" with the parameter  $\gamma$  being below 1 for men and women. These results are consistent with previous findings (Harrison and Rutström, 2008; Bruhin et al., 2010).

Q3: We find that males are less risk averse both in the EU and RDU estimations ( $\beta_1 < 0$  in Estimation 1 and 2). However, we do not find baseline gender differences in probability perception

(the coefficients  $\mu_1$  and  $\varphi_1$  are not significantly different from 0 in **Table 2**).<sup>4</sup>

Q2 and Q3: There is some indication of a link between exposure and risk aversion. We find that R2D:4D has a negative effect on the risk-attitude parameter  $\alpha$  only for males, but not for females ( $\beta_3$  in Estimation 2 is  $-0.164$  and significant with  $p < 0.05$ ) in the EU estimations. It shows that males with higher R2D:4D have more appetite for risk (less risk-averse). We do not find an association between fWHR and any changes of risk taking<sup>5</sup> in the EU estimations ( $\beta_2$  in Estimation 1 and 2 are not significant).

In the RDU estimations, we find that fWHR ( $\beta_2$  is  $-0.037$  with  $p < 0.1$  in Estimation 1) and R2D:4D ( $\beta_2$  is  $-0.037$  with  $p < 0.05$  in Estimation 2) have a negative effect on females' risk-attitude, but positive effect on males' risk-attitude ( $\beta_3$  is  $0.084$  with  $p < 0.01$  and  $0.069$  with  $p < 0.1$  in Estimation 1 and 2). This suggests that females with higher ratios have more appetite for risk (less risk-averse), while the relationship is opposite for the males.

The differences in  $\alpha$  across the two models are to be expected. The reason is that the risk attitudes are only represented by  $\alpha$  in the EU model, while they are represented by  $\alpha$ ,  $\gamma$ , and  $\delta$  in the RDU model. In the case where EU is the best model, we should expect the RDU model to have a similar  $\alpha$  and  $\gamma = 1$ ,  $\delta = 1$ . Whenever people weight probabilities,  $\gamma$  and  $\delta$  are going to differ from 1. In such a case, there is no reason to expect the  $\alpha$  to be the same in the EU and RDU as the  $\alpha$  in the EU will partially adjust itself to explain part of the risk attitudes reflected in the  $\gamma$  and  $\delta$  in the RDEU model.

There is a clearer indication of a link with the attitudes to probabilities for males (but not for female participants). The inverse S-shape of the probability weighting function is more pronounced for males with higher ratios ( $\mu_3$  is  $-0.160$ ,  $p < 0.05$  in Estimation 1 and  $-0.148$ ,  $p < 0.05$  in Estimation 2). And male participants with higher ratios are more optimistic ( $\varphi_3$  is  $0.820$ ,  $p < 0.01$  in Estimation 1 and  $0.945$ ,  $p < 0.01$  in Estimation 2). It suggests that male participants with higher ratios overweight their chances of receiving bigger payoffs and are more optimistic

<sup>4</sup>Note that gender differences can still exist overall due to gender differences in other variables such as prenatal exposure to testosterone which can have an influence on risk preferences. We look into this below.

<sup>5</sup>The  $\beta_3$  in Estimation 1 is  $-0.115$ , but not significant as  $p > 0.1$  in the EU estimations.

**TABLE 2 |** Estimation Results on fWHR and R2D:4D.

Estimation 1	EU				Estimation 2	EU			
	$\alpha$	$\alpha$	$\gamma$	$\delta$		$\alpha$	$\alpha$	$\gamma$	$\delta$
Male	-0.135**	-0.071***	0.059	0.316	Male	-0.096**	-0.075**	0.054	0.108
( $\beta_1, \beta_1, \mu_1, \varphi_1$ )	(-2.37)	(-2.60)	(1.01)	(1.31)	( $\beta_1, \beta_1, \mu_1, \varphi_1$ )	(-2.12)	(-2.50)	(0.94)	(0.59)
fWHR	-0.026	-0.037*	0.067	-0.026	R2D:4D	-0.021	-0.037**	0.038	-0.068
( $\beta_2, \beta_2, \mu_2, \varphi_2$ )	(-0.79)	(-1.79)	(1.32)	(-0.14)	( $\beta_2, \beta_2, \mu_2, \varphi_2$ )	(-0.76)	(-2.09)	(1.52)	(-0.39)
Male $\times$ fWHR	-0.115	0.084***	-0.160**	0.820***	Male $\times$ R2D:4D	-0.164**	0.069*	-0.148**	0.945***
( $\beta_3, \beta_3, \mu_3, \varphi_3$ )	(-1.64)	(2.97)	(-2.41)	(3.35)	( $\beta_3, \beta_3, \mu_3, \varphi_3$ )	(-2.19)	(1.80)	(-2.38)	(3.24)
Under 8 s	-0.183***	-0.085***	-0.303***	-0.373***	Under 8 s	-0.167***	-0.088**	-0.297***	-0.364***
( $\beta_4, \beta_4, \mu_4, \varphi_4$ )	(-5.78)	(-2.82)	(-7.84)	(-3.00)	( $\beta_4, \beta_4, \mu_4, \varphi_4$ )	(-5.47)	(-2.22)	(-6.34)	(-2.90)
Under 4 s	-0.045	-0.020	-0.157***	-0.119	Under 4 s	-0.024	-0.025	-0.149***	-0.168
( $\beta_4, \beta_4, \mu_4, \varphi_4$ )	(-1.19)	(-0.64)	(-2.93)	(-0.78)	( $\beta_4, \beta_4, \mu_4, \varphi_4$ )	(-0.61)	(-0.66)	(-2.69)	(-1.03)
Male $\times$ Under 8 s	0.017	-0.043	-0.035	-0.328	Male $\times$ Under 8 s	-0.014	-0.010	-0.031	-0.112
( $\beta_5, \beta_5, \mu_5, \varphi_5$ )	(0.25)	(-0.70)	(-0.57)	(-1.58)	( $\beta_5, \beta_5, \mu_5, \varphi_5$ )	(-0.25)	(-0.03)	(-0.25)	(-0.23)
Male $\times$ Under 4 s	0.007	0.052	-0.062	0.049	Male $\times$ Under 4 s	0.008	0.064	-0.064	0.075
( $\beta_5, \beta_5, \mu_5, \varphi_5$ )	(0.10)	(1.10)	(-0.83)	(0.19)	( $\beta_5, \beta_5, \mu_5, \varphi_5$ )	(0.13)	(1.25)	(-0.88)	(0.32)
fWHR $\times$ Under 8 s	0.084**	-0.032	0.034	-0.051	R2D:4D $\times$ Under 8 s	0.064	-0.032	0.014	-0.073
( $\beta_6, \beta_6, \mu_6, \varphi_6$ )	(2.22)	(-1.17)	(0.75)	(-0.32)	( $\beta_6, \beta_6, \mu_6, \varphi_6$ )	(1.52)	(-1.29)	(0.57)	(-0.46)
fWHR $\times$ Under 4 s	0.108***	-0.008	0.052	-0.276	R2D:4D $\times$ Under 4 s	0.058*	-0.003	0.042	-0.147
( $\beta_6, \beta_6, \mu_6, \varphi_6$ )	(2.78)	(-0.25)	(0.72)	(-1.59)	( $\beta_6, \beta_6, \mu_6, \varphi_6$ )	(1.71)	(-0.06)	(1.23)	(-0.48)
Male $\times$ fWHR $\times$ Under 8 s	-0.158*	-0.067	-0.032	-0.650***	Male $\times$ R2D:4D $\times$ Under 8 s	-0.288***	-0.097	-0.031	-0.558
( $\beta_7, \beta_7, \mu_7, \varphi_7$ )	(-1.76)	(-0.97)	(-0.46)	(-2.97)	( $\beta_7, \beta_7, \mu_7, \varphi_7$ )	(-2.68)	(-0.11)	(-0.07)	(-0.59)
Male $\times$ fWHR $\times$ Under 4 s	-0.162**	0.057	-0.027	0.378*	Male $\times$ R2D:4D $\times$ Under 4 s	-0.169	0.029	-0.028	-0.011
( $\beta_7, \beta_7, \mu_7, \varphi_7$ )	(-1.99)	(1.49)	(-0.32)	(1.69)	( $\beta_7, \beta_7, \mu_7, \varphi_7$ )	(-1.37)	(0.44)	(-0.44)	(-0.03)
Constant	0.480***	0.558***	0.844***	1.394***	Constant	0.476***	0.549***	0.853***	1.381***
( $\beta_0, \beta_0, \mu_0, \varphi_0$ )	(15.43)	(33.68)	(20.54)	(10.31)	( $\beta_0, \beta_0, \mu_0, \varphi_0$ )	(15.70)	(34.80)	(18.18)	(11.19)

Z statistics in parentheses, \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

toward their chances of winning monetary outcomes. A similar association was not found for female participants.

Q1: We find some indication that time pressure increases risk aversion with  $\alpha$  being smaller in the 8 s time pressure condition ( $\beta_4$  is  $-0.183$  with  $p < 0.01$  and  $-0.167$  with  $p < 0.01$  in Estimation 1 and 2). This result is in line with previous findings (Kocher and Sutter, 2006). However, we do not find an overall significant effect in our extreme time pressure condition (4 s).

There is also a clear effect of time pressure on the probability weighting parameter. The “inverse S-shape” appears more pronounced in the time pressure conditions ( $\mu_4$  is  $-0.303$  with  $p < 0.01$  for 8 s and  $-0.157$  with  $p < 0.01$  for 4 s in Estimation 1;  $\mu_4$  is  $-0.297$  with  $p < 0.01$  for 8 s and  $-0.149$  with  $p < 0.01$  for 4 s in Estimation 2). We also find more optimism, but only in the 8 s time pressure condition ( $\varphi_4$  is  $-0.373$  with  $p < 0.01$  and  $-0.364$  with  $p < 0.01$  for 8 s in Estimations 1 and 2).

Q1 and Q3: However, we do not find notable baseline gender differences in risk attitude under time pressure (8 and 4 s conditions), as  $\beta_5$ ,  $\mu_5$  and  $\varphi_5$  are not significantly different from zero in Estimation 1 and 2.<sup>6</sup>

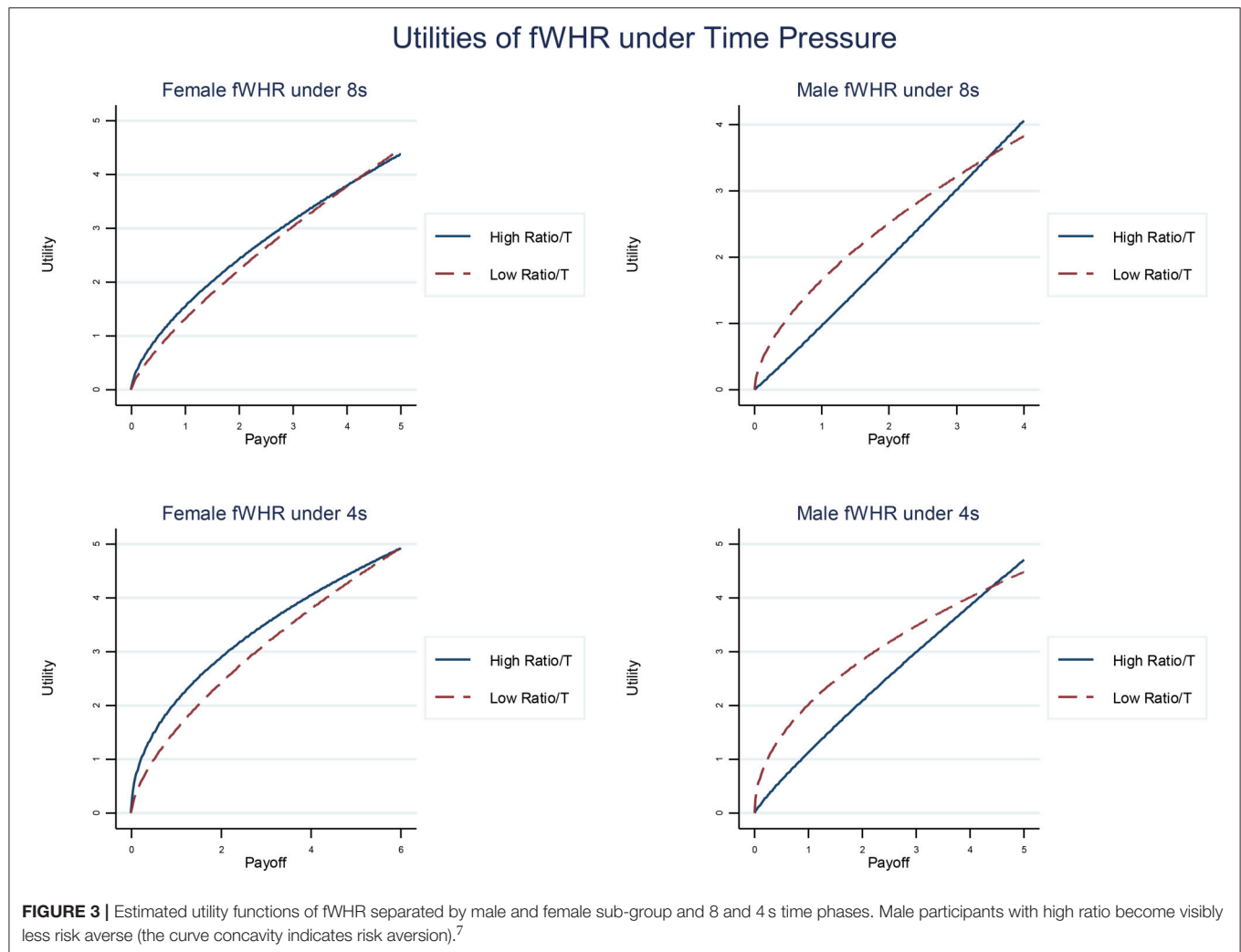
Q1, Q2, and Q3: When looking at the coefficient of risk aversion, there is a differential effect of time pressure by gender

as a function of fWHR in the 8 s time pressure condition ( $\beta_6$  is  $0.084$  with  $p < 0.05$  while  $\beta_7$  is  $-0.158$  and marginally significant with  $p < 0.1$  in Estimation 1). In the phase with extreme time pressure, we also find that female participants with higher fWHR have more risk-averse attitude, while males with higher ratios have more appetite for risk ( $\beta_6$  is  $0.108$  with  $p < 0.01$  while  $\beta_7$  is  $-0.162$  with  $p < 0.05$  in Estimation 1). Further, in the Estimation 2, we find that female participants with higher R2D:4D have more risk-averse attitude in the 4 s time pressure condition ( $\beta_6$  is  $0.058$  and marginally significant with  $p < 0.1$ ), whereas males with higher R2D:4D have more appetite for risk in the 8 s time pressure condition ( $\beta_7$  is  $-0.288$  with  $p < 0.01$ ).

The previous results decompose the effect on risk attitudes and probability perception of gender, prenatal exposure and time pressure. Once this decomposition is done, we can look into how different subgroups differ. We present here our estimation of the parameters  $\alpha$  (see **Figure 3**),  $\gamma$  and  $\delta$  (see **Figure 4**) at the aggregated level for meaningful subgroups for the fWHR ratios (overall, similar results are found for 2D:4D).

In terms of sensitivity to outcomes, the male subgroup with higher fWHR shows less curvature under time pressure in their utility function than that with lower ratios (see right column in **Figure 3**), but similar association is not found in female subgroup (see left column in **Figure 3**). The curvature in utility suggests that the risk attitude of an agent: concave as risk-averse  $\alpha > 0$ ;

<sup>6</sup>Gender differences in risk attitude under time pressure can still be present due to gender differences in prenatal exposure to testosterone. We look into this below.



convex as risk-loving  $\alpha < 0$ . Less curvature in utility function suggests more appetite for risk.

As the prizes in the lottery are rearranged from the biggest to the smallest in a rank-dependent manner, the left bottom region in a probability weighting function reveals if the probabilities of the prizes are over weighted or under weighted. For example, in any subfigure in **Figure 4**, the estimated functions are above the diagonal line in the left bottom region. It means that the actual probabilities are over weighted.

In terms of sensitivity to probability, under time pressure, males, with higher fWHR overestimate probabilities (**Figure 4**) of receiving bigger payoffs and have a more optimistic view

about probabilities than those with lower ratios. However, we observe the opposite effect in the female sub-group (see left column in **Figure 4**). These effects are more pronounced under extreme time pressure (by comparing the top and bottom rows in **Figure 4**).

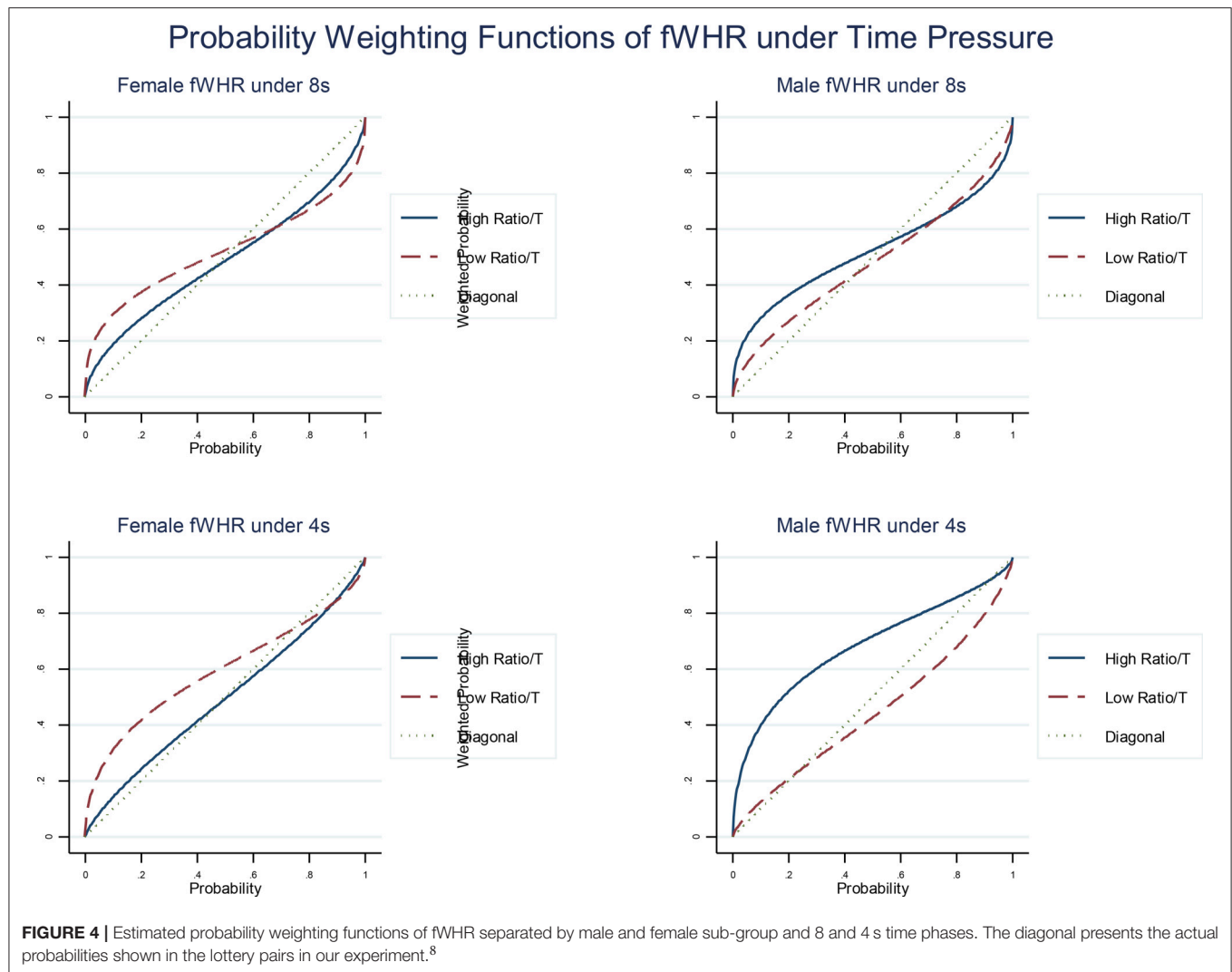
To answer our Questions 1–3 in section Motivations and Hypotheses directly, the equations (6) in our structural models are also estimated by using: (1) firstly, the “Time” variable, which is a categorical variable to measure the phases under three different time constraints, as covariates; (2) then adding “Male” and “Ratio” variables into the covariates; (3) finally, adding the interactions in to the covariates. The EU estimates are shown in Table 3 in Appendix (Supplementary Material) and the RDU estimates are shown in Tables 4, 5 in Appendix (Supplementary Material).

Based on our findings discussed above, we can now answer the three questions raised in section Motivations and Hypotheses and summarize our results:

**Result 1:** Time pressure increases an appetite for risk. Participants under time pressure become more optimistic.

<sup>7</sup>We calculate the utilities based on our estimations of Estimation 1 and 2 in **Table 2**. As fWHR and R2D:4D are standardized, we use value of 1 as high ratio and  $-1$  as low ratio. Therefore, for example, the utility for females with high ratios in the 8s time condition is calculated as:  $\alpha = 0.480 - 0.026 \times 1 - 0.183 + 0.084 \times 1 = 0.355$ , and the utility for low ratios is 0.239. The utilities for males in the 8s time condition are calculated as  $-0.036$  for high ratios and 0.395 for low ratios. The utilities for females in the 4s time condition are calculated are 0.517 for high ratios and 0.354 for low ratios, and that for males are 0.112 for high ratios and 0.503 for low ratios.





**Result 2:** We do not find enough evidence to support the hypothesis that higher testosterone exposure is associated with higher risk-taking. We observed mixed results in EU and RDU estimates.

**Result 3:** We find that male participants are less risk averse and that the gender gap in risk attitudes increases under moderate time pressure. We also observe that men with higher testosterone exposure exhibit a greater optimism and overestimation of small probabilities of success.

## DISCUSSION AND CONCLUSION

This study looked into gender difference in risk attitude under pressure and the potential role of prenatal exposure. We find that males are less risk averse than female participants, in line with existing research. We disentangled the different aspects of risk preferences, giving us new insights into these gender differences.

<sup>8</sup>The calculations of the probability weighting functions apply the same method of calculations in the utilities in Figure 3.

We found that gender differences were clearer in the sensitivity to probability than in the sensitivity to outcomes.

When looking at prenatal exposure to testosterone, we find that males with high fWHR and R2D:4D sought more risk and overweighted small probabilities of high gain. They also were more optimistic about outcomes than the females. Females with high fWHR and R2D:4D did the opposite, taking less risk. Time pressure also, on average, made males more optimistic.

In summary, men, and particularly those with high fWHR and R2D:4D took more risk and were more bullish about pursuing an elusive chance of winning, especially under time pressure.

These results show that prenatal testosterone exposure alters risk-taking in men; particularly under time pressure. Previous research has shown that a low 2D:4D (or high R2D:4D) ratio associated with high testosterone exposure predicted a longer survival of professional traders (Coates and Page, 2009). As a consequence, men with a low 2D:4D ratio were likely to be overrepresented in the population of traders. Our result

may help make sense of this finding given that the male participants with low 2D:4D ratios displayed a greater propensity to take risks under time pressure. The results of the present research did not find such an effect of time pressure on women. If women traders are seen as taking fewer risks than their male counterparts, particularly in response to time pressure, then, in a market which values activity, they may be seen as less appropriate candidates. Moreover, if they make it past the selection phase, they may well not be retained as they do not measure up to the accepted yardstick for performance.

As well as demonstrating marked differences between men and women in decision making, this research also clearly confirms that preferences are not stable and that time pressure affects choice. Because each participant was exposed to the same information in each case, there was no information difference. Rather time pressure was likely to have, interfered with information processing, thereby producing differing results. The nature of this instability was complex, being influenced by both time pressure and the long-term organizational effects of testosterone. Previous experimental and theoretical studies have argued that deliberation may have a non-linear effect on moral choices (Moore and Tenbrunsel, 2014) and cooperation (Capraro and Cococcioni, 2016). A non-linear relationship has also been observed by some authors between circulating testosterone and risk taking (Stanton et al., 2011). The consequence of all this is that the useful simplification of assuming that preferences are stable, may lead to forget the fact that preference instability is substantial, widespread and non-linear.

Our results suggest that if the market privileges risk taking and confidence under time pressure then a combination of physiological predisposition and preference instability may favor the employment of men. This, in turn, may explain the preponderance of men in the market. This is difficult to prove in any definitive sense as counterfactuals are not readily available. Care also has to be taken in extrapolating from a laboratory study to global markets as the requirements of controlling for factors except those under investigation inevitably means that a degree of verisimilitude is lost. The risk-taking task, for example is a stylized one with a limited number of parameters. The sample size, relative to financial markets, is small, and does not, necessarily, mirror the profile of those in financial markets. Moreover, the choices are single shot interactions, rather than the dynamic, ongoing and varied interactions observed in real markets. This study only looks at the organizational effects of testosterone manifest in 2D:4D ratio and fWHR, not at the activational effects of circulating testosterone. It also does not address other hormones, such as cortisol, which have been demonstrated to affect risk-taking (Kandasamy et al., 2014). Despite this, our findings on gender differences, the role of prenatal testosterone exposure and of time pressure provide some clues as to why women may be at a perceived disadvantage in a

pressurized trading environment. This, in turn, may mean that they are less likely to be recruited and retained.

To provide a fuller picture, there are a number of questions for further research to address. The first is to examine risk taking when the probability distribution is less clearly defined. This ambiguity may affect the results. The second question is whether the nature of risk taking changes when there is interaction between participants. These sorts of interaction studies have been undertaken in hormone research (e.g., Cueva et al., 2015). They improve external validity but sometimes at the expense of mechanistic clarity. Third, external validity could be improved by conducting the task with different groups of bank employees. It may be that different functions have different risk profiles, so traders may differ from asset managers, for example. Fourth, further research should sample circulating hormone levels to explore the interaction between activational (circulating) and organizational (i.e., those shaping development) hormones.

This research provides a piece of the puzzle as to why women are underrepresented in a number of areas of finance. But does it matter that these areas are male dominated? Markets are well-served by diversity as a means of tempering herd instincts. A market that is skewed in favor of employing men may, therefore, bring its own set of problems. Some researchers, for example, Coates and colleagues (Coates et al., 2010), have suggested that improving gender diversity may improve market stability. This is supported by experimental evidence which suggested both gender (Cueva and Rustichini, 2015) and hormonal diversity (Cueva et al., 2015) improve market stability. Although there are many explanations for aggregate behavior in financial markets, the effect of gender, preference stability and hormonal exposure may have significant repercussions.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the ethics committee from School of Economics and Finance at Queensland University of Technology with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the ethics committee from School of Economics and Finance at Queensland University of Technology.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnbeh.2017.00246/full#supplementary-material>

## REFERENCES

- Aguiar, F., Brañas-Garza, P., Cobo-Reyes, R., Jimenez, N., and Miller, L. M. (2009). Are women expected to be more generous? *Exp. Econ.* 12, 93–98. doi: 10.1007/s10683-008-9199-z
- Archer, J. (2006). Testosterone and human aggression: an evaluation of the challenge hypothesis. *Neurosci. Biobehav. Rev.* 30, 319–345. doi: 10.1016/j.neubiorev.2004.12.007
- Austin, E. J., Manning, J. T., Mcinroy, K., and Mathews, E. (2002). A preliminary investigation of the associations between personality, cognitive ability and digit ratio. *Pers. Individ. Diff.* 33, 1115–1124. doi: 10.1016/S0191-8869(02)00002-8
- Barber, B. M., and Odean, T. (2001). Boys will be boys: gender, overconfidence, and common stock investment. *Q. J. Econ.* 116, 261–292. doi: 10.1162/003355301556400
- Barel, E. (2017). 2D:4D, optimism, and risk taking. *Curr. Psychol.* 1–9. doi: 10.1007/s12144-017-9598-4
- Bosch-Domènech, A., Brañas-Garza, P., and Espín, A. M. (2014). Can exposure to prenatal sex hormones (2D:4D) predict cognitive reflection? *Psychoneuroendocrinology* 43, 1–10. doi: 10.1016/j.psyneuen.2014.01.023
- 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
- Brañas-Garza, P., Capraro, V., and Rascon, E. (2016). *Gender Differences in Altruism: Expectations, Actual Behaviour and Accuracy of Beliefs*. Available online at: <https://ssrn.com/abstract=2796221>
- Brañas-Garza, P., Galizzi, M., and Nieboer, J. (2017). Experimental and self-reported measures of risk taking and digit ratio (2D: 4D): evidence from a large, systematic study. *Int. Econ. Rev.*
- Breedlove, S. M., and Hampson, E. (2002). “Sexual differentiation of the brain and behavior,” in *Behavioral Endocrinology, 2nd Edn.*, eds J. B. Becker, S. M. Breedlove, D. Crews, and M. M. McCarthy (Cambridge, MA; London: MIT Press), 776.
- Brown, M. B., and Forsythe, A. B. (1974). Robust tests for the equality of variances. *J. Am. Stat. Assoc.* 69, 364–367. doi: 10.1080/01621459.1974.10482955
- Bruhin, A., Fehr-Duda, H., and Epper, T. (2010). Risk and rationality: uncovering heterogeneity in probability distortion. *Econometrica* 78, 1375–1412. doi: 10.3982/ECTA7139
- Busse, J. A., and Green, T. C. (2002). Market efficiency in real time. *J. Finan. Econ.* 65, 415–437. doi: 10.1016/S0304-405X(02)00148-4
- Byrnes, J. P., Miller, D. C., and Schafer, W. D. (1999). Gender differences in risk taking: a meta-analysis. *Psychol. Bull.* 125, 367–383. doi: 10.1037/0033-2909.125.3.367
- Capraro, V., and Cococcioni, G. (2016). Rethinking spontaneous giving: extreme time pressure and ego-depletion favor self-regarding reactions. *Sci. Rep.* 6:27219. doi: 10.1038/srep27219
- Capraro, V., and Cococcioni, G. (2015). Social setting, intuition and experience in laboratory experiments interact to shape cooperative decision-making. *Proc. R. Soc. B* 282:20150237. doi: 10.1098/rspb.2015.0237
- Capraro, V., Corgnet, B., Espín, A. M., and Hernán-González, R. (2017). Deliberation favours social efficiency by making people disregard their relative shares: evidence from USA and India. *R. Soc. Open Sci.* 4:160605. doi: 10.1098/rsos.160605
- 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. Lond. B Biol. Sci.* 275, 2651–2656. doi: 10.1098/rspb.2008.0873
- Catalyst, I. (2016). *Women in Accounting [Online]*. Catalyst, Inc., Available online at: [http://www.catalyst.org/knowledge/women-accounting-footnote4\\_zipkayp](http://www.catalyst.org/knowledge/women-accounting-footnote4_zipkayp) (Accessed Oct 18, 2017).
- Chambers Partners, L. (2014). *Law Firm Gender Diversity [Online]*. Chambers Partners, Ltd. Available online at: <http://www.chambersstudent.co.uk/where-to-start/newsletter/2014-gender-in-the-law-survey/law-firm-gender-diversity> (Accessed Oct 18, 2017).
- Charness, G., and Rustichini, A. (2011). Gender differences in cooperation with group membership. *Games Econ. Behav.* 72, 77–85. doi: 10.1016/j.geb.2010.07.006
- Coates, J. M., and Page, L. (2009). A note on trader sharpe ratios. *PLoS ONE* 4:e8036. doi: 10.1371/journal.pone.0008036
- Coates, J. M., Gurnell, M., and Rustichini, A. (2009). Second-to-fourth digit ratio predicts success among high-frequency financial traders. *Proc. Natl. Acad. Sci. U.S.A.* 106, 623–628. doi: 10.1073/pnas.0810907106
- Coates, J. M., Gurnell, M., and Sarayai, Z. (2010). From molecule to market: steroid hormones and financial risk-taking. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 331–343. doi: 10.1098/rstb.2009.0193
- Conte, A., Hey, J. D., and Moffatt, P. G. (2011). Mixture models of choice under risk. *J. Econometr.* 162, 79–88. doi: 10.1016/j.jeconom.2009.10.011
- Crosen, R., and Gneezy, U. (2009). Gender differences in preferences. *J. Econ. Lit.* 47, 448–474. doi: 10.1257/jel.47.2.448
- Cueva, C., and Rustichini, A. (2015). Is financial instability male-driven? gender and cognitive skills in experimental asset markets. *J. Econ. Behav. Organ.* 119, 330–344. doi: 10.1016/j.jebo.2015.08.014
- Cueva, C., Roberts, R. E., Spencer, T., Rani, N., Tempest, M., Tobler, P. N., et al. (2015). Cortisol and testosterone increase financial risk taking and may destabilize markets. *Sci. Rep.* 5:11206. doi: 10.1038/srep11206
- Dalton, P., and Ghosal, S. (2014). *Self-Confidence, Overconfidence and Prenatal Testosterone Exposure: Evidence from the Lab*. Scottish Institute for Research in Economics.
- De Paola, M., and Gioia, F. (2016). Who performs better under time pressure? results from a field experiment. *J. Econ. Psychol.* 53, 37–53. doi: 10.1016/j.joep.2015.12.002
- Deaves, R., Lüders, E., and Schröder, M. (2010). The dynamics of overconfidence: evidence from stock market forecasters. *J. Econ. Behav. Organ.* 75, 402–412. doi: 10.1016/j.jebo.2010.05.001
- Drichoutis, A. C., and Nayga, R. M. (2015). Do risk and time preferences have biological roots? *South. Econ. J.* 82, 235–256. doi: 10.4284/0038-4038-2013.246
- Eckel, C. C., and Füllbrunn, S. C. (2015). Thar she blows? Gender, competition, and bubbles in experimental asset markets. *Am. Econ. Rev.* 105, 906–920. doi: 10.1257/aer.20130683
- Eckel, C. C., and Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evol. Hum. Behav.* 23, 281–295. doi: 10.1016/S1090-5138(02)00097-1
- Fellner, G., and Maciejovsky, B. (2007). Risk attitude and market behavior: evidence from experimental asset markets. *J. Econ. Psychol.* 28, 338–350. doi: 10.1016/j.joep.2007.01.006
- Fink, B., Neave, N., and Seydel, H. (2007). Male facial appearance signals physical strength to women. *Am. J. Hum. Biol.* 19, 82–87. doi: 10.1002/ajhb.20583
- Fink, B., Neave, N., Laughton, K., and Manning, J. T. (2006). Second to fourth digit ratio and sensation seeking. *Pers. Individ. Diff.* 41, 1253–1262. doi: 10.1016/j.paid.2006.05.002
- Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* 10, 171–178. doi: 10.1007/s10683-006-9159-4
- Garbarino, E., Slonim, R., and Sydnor, J. (2011). Digit ratios (2D:4D) as predictors of risky decision making for both sexes. *J. Risk Uncertain.* 42, 1–26. doi: 10.1007/s11166-010-9109-6
- Graham, J. R., Harvey, C. R., and Huang, H. (2009). Investor competence, trading frequency, and home bias. *Manag. Sci.* 55, 1094–1106. doi: 10.1287/mnsc.1090.1009
- Green, C., Jegadeesh, N., and Tang, Y. (2009). Gender and job performance: evidence from wall street. *Financial Anal. J.* 65, 65–78. doi: 10.2469/faj.v65.n6.1
- Greiner, B. (2004). *The Online Recruitment System Orsee 2.0—a Guide for the Organization of Experiments in Economics*. University of Cologne, Working paper series in economics, Vol. 10, 63–104.
- Grinblatt, M., and Keloharju, M. (2009). Sensation seeking, overconfidence, and trading activity. *J. Finance* 64, 549–578. doi: 10.1111/j.1540-6261.2009.01443.x
- Harrison, G. W., and Rutström, E. E. (2008). “Risk aversion in the laboratory,” in *Risk Aversion in Experiments*, eds J. C. Cox and G. W. Harrison (Bingley: Emerald Group Publishing Limited), 41–196.
- Hartgens, F., and Kuipers, H. (2004). Effects of androgenic-anabolic steroids in athletes. *Sports Med.* 34, 513–554. doi: 10.2165/00007256-200434080-00003
- Hey, J. D. (2001). Does repetition improve consistency? *Exp. Econ.* 4, 5–54. doi: 10.1023/A:1011486405114
- Hey, J. D., and Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica* 62, 1291–1326. doi: 10.2307/2951750

- 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
- Hu, Y., Wang, D., Pang, K., Xu, G., and Guo, J. (2015). The effect of emotion and time pressure on risk decision-making. *J. Risk Res.* 18, 637–650. doi: 10.1080/13669877.2014.910688
- Huber, O., and Kunz, U. (2007). Time pressure in risky decision-making: effect on risk defusing. *Psychol. Sci.* 49:415.
- Jäkel, A., and Moynihan, T. (2016). *Women in Financial Service 2016*. London: Oliver Wyman Group.
- Kaiser Family Foundation (2015). *Employer Health Benefits [Online]*. Kaiser Family Foundation and Health Research & Educational Trust. Available online at: <http://files.kff.org/attachment/report-2015-employer-health-benefits-survey> (Accessed Oct 18, 2017).
- Kandasamy, N., Hardy, B., Page, L., Schaffner, M., Graggaber, J., Powlson, A. S., et al. (2014). Cortisol shifts financial risk preferences. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3608–3613. doi: 10.1073/pnas.1317908111
- Kempel, P., Gohlke, B., Klempau, J., Zinsberger, P., Reuter, M., and Hennig, J. (2005). Second-to-fourth digit length, testosterone and spatial ability. *Intelligence* 33, 215–230. doi: 10.1016/j.intell.2004.11.004
- Kocher, M. G., and Sutter, M. (2006). Time is money—Time pressure, incentives, and the quality of decision-making. *J. Econ. Behav. Organ.* 61, 375–392. doi: 10.1016/j.jebo.2004.11.013
- Kocher, M. G., Pahlke, J., and Trautmann, S. T. (2013). Tempus fugit: time pressure in risky decisions. *Manag. Sci.* 59, 2380–2391. doi: 10.1287/mnsc.2013.1711
- Korzec, R. (2000). The glass ceiling in law firms: a form of sex-based discrimination the glass ceiling law firms: a form of sex-based discrimination. *J. Employ. Discrimination Law* 251, 251–263.
- Lattimore, P. K., Baker, J. R., and Witte, A. D. (1992). The influence of probability on risky choice: a parametric examination. *J. Econ. Behav. Organ.* 17, 377–400. doi: 10.1016/S0167-2681(95)90015-2
- Lefevre, C. E., Wilson, V. A., Morton, F. B., Brosnan, S. F., Paukner, A., and Bates, T. C. (2014). Facial width-to-height ratio relates to alpha status and assertive personality in capuchin monkeys. *PLoS ONE* 9:e93369. doi: 10.1371/journal.pone.0093369
- Lietz, N. G. (2012). *In the Hot Finance Jobs, Women Are Still Shut Out*. Cambridge, MA: Harvard Business School Publishing Corporation.
- 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
- Manning, J. T. (2002). *Digit Ratio: A Pointer to Fertility, Behavior, and Health*. New Brunswick, NJ: Rutgers University Press.
- Manning, J. T., and Taylor, R. P. (2001). Second to fourth digit ratio and male ability in sport: implications for sexual selection in humans. *Evol. Hum. Behav.* 22, 61–69. doi: 10.1016/S1090-5138(00)00063-5
- Manning, J. T., Churchill, A. J., and Peters, M. (2007). The effects of sex, ethnicity, and sexual orientation on self-measured digit ratio (2D:4D). *Arch. Sex. Behav.* 36, 223–233. doi: 10.1007/s10508-007-9171-6
- 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
- Mazur, A., and Booth, A. (1998). Testosterone and dominance in men. *Behav. Brain Sci.* 21, 353–363. doi: 10.1017/S0140525X98001228
- McIntyre, M. (2006). The use of digit ratios as markers for perinatal androgen action. *Reprod. Biol. Endocrinol.* 4:10. doi: 10.1186/1477-7827-4-10
- Moffatt, P. G. (2005). Stochastic choice and the allocation of cognitive effort. *Exp. Econ.* 8, 369–388. doi: 10.1007/s10683-005-5375-6
- Moore, C., and Tenbrunsel, A. E. (2014). “Just think about it”? Cognitive complexity and moral choice. *Organ. Behav. Hum. Decis. Process.* 123, 138–149. doi: 10.1016/j.obhdp.2013.10.006
- Neyse, L., and Brañas-Garza, P. (2014). *Digit Ratio Measurement Guide*. Kiel: Institute for the World Economy (IfW).
- Neyse, L., Bosworth, S., Ring, P., and Schmidt, U. (2016). Overconfidence, incentives and digit ratio. *Sci. Rep.* 6:23294. doi: 10.1038/srep23294
- Niederle, M., and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much?. *Q. J. Econ.* 122, 1067–1101. doi: 10.1162/qjec.122.3.1067
- Nursimulu, A., and Bossaerts, P. (2014). Excessive volatility is also a feature of individual level forecasts. *J. Behav. Finance* 15, 16–29. doi: 10.1080/15427560.2014.877016
- Oberlechner, T., and Nimgade, A. (2005). Work stress and performance among financial traders. *Stress Health* 21, 285–293. doi: 10.1002/smi.1063
- Powell, M., and Ansic, D. (1997). Gender differences in risk behaviour in financial decision-making: an experimental analysis. *J. Econ. Psychol.* 18, 605–628. doi: 10.1016/S0167-4870(97)00026-3
- Rand, D. G. (2016). Cooperation, fast and slow: meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychol. Sci.* 27, 1192–1206. doi: 10.1177/0956797616654455
- Rand, D. G., Greene, J. D., and Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature* 489, 427–430. doi: 10.1038/nature11467
- Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., et al. (2014). Social heuristics shape intuitive cooperation. *Nat. Commun.* 5:577. doi: 10.1038/ncomms4677
- Roberti, J. W. (2004). A review of behavioral and biological correlates of sensation seeking. *J. Res. Pers.* 38, 256–279. doi: 10.1016/S0092-6566(03)00067-9
- Roney, J. R., Mahler, S. V., and Maestripieri, D. (2003). Behavioral and hormonal responses of men to brief interactions with women. *Evol. Hum. Behav.* 24, 365–375. doi: 10.1016/S1090-5138(03)00053-9
- Sapienza, P., Zingales, L., and Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15268–15273. doi: 10.1073/pnas.0907352106
- Saqib, N. U., and Chan, E. Y. (2015). Time pressure reverses risk preferences. *Organ. Behav. Hum. Decision Process.* 130, 58–68. doi: 10.1016/j.obhdp.2015.06.004
- Schipper, B. C. (2014). *Sex Hormones and Choice under Risk*. University of California, Davis, Department of Economics Working Papers 147.
- Sethi, B., Miller, E., and Philpott, R. (2013). *Mending the Gender Gap: Advancing Tomorrow's Women Leaders in Financial Services*. PricewaterhouseCoopers, Available online at: [www.pwc.com](http://www.pwc.com)
- Stanton, S. J., Liening, S. H., and Schultheiss, O. C. (2011). Testosterone is positively associated with risk taking in the Iowa gambling task. *Hormon. Behav.* 59, 252–256. doi: 10.1016/j.yhbeh.2010.12.003
- Stigler, G. J., and Becker, G. S. (1977). De gustibus non est disputandum. *Am. Econ. Rev.* 67, 76–90.
- Swerdlow, R. S., Wang, C., and Sinha Hikim, A. P. (2002). “The hypothalamic - pituitary-gonadal axis in men,” in *Hormones, Brain and Behavior*, eds D. W. Pfaff, A. P. Arnold, A. M. Etgen, S. E. Fahrbach, and R. T. Rubin (Amsterdam; London: Academic Press), 1–37.
- Tester, N., and Campbell, A. (2007). Sporting achievement: what is the contribution of digit ratio? *J. Pers.* 75, 663–678. doi: 10.1111/j.1467-6494.2007.00452.x
- Truss, L. (2016). “RE: Women in the legal industry - GOV.UK,” in *Type to Spark21 Conference* (London).
- Valentine, K. A., Li, N. P., Penke, L., and Perrett, D. I. (2014). Judging a man by the width of his face: the role of facial ratios and dominance in mate choice at speed-dating events. *Psychol. Sci.* 25, 806–811. doi: 10.1177/0956797613511823
- 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. Orthodontics* 21, 137–143. doi: 10.1093/ejo/21.2.137
- Wakker, P. P. (2010). *Prospect Theory: For Risk and Ambiguity*. Cambridge, UK: Cambridge University Press.
- Weston, E. M., Friday, A. E., and Liö, P. (2007). Biometric evidence that sexual selection has shaped the hominin face. *PLoS ONE* 2:e710. doi: 10.1371/journal.pone.0000710
- Wilcox, N. T. (2011). ‘Stochastically more risk averse’: a contextual theory of stochastic discrete choice under risk. *J. Econom.* 162, 89–104. doi: 10.1016/j.jeconom.2009.10.012



- Williams, J. C., and Richardson, V. T. (2010). *New Millennium, Same Glass Ceiling? The Impact of Law Firm Compensation Systems on Women*, Chicago, IL: The Project for Attorney Retention .
- Williams, J. H., Greenhalgh, K. D., and Manning, J. T. (2003). Second to fourth finger ratio and possible precursors of developmental psychopathology in preschool children. *Early Hum. Dev.* 72, 57–65. doi: 10.1016/S0378-3782(03)00012-4
- Young, D. L., Goodie, A. S., Hall, D. B., and Wu, E. (2012). Decision making under time pressure, modeled in a prospect theory framework. *Organ. Behav. Hum. Decision Process.* 118, 179–188. doi: 10.1016/j.obhdp.2012.03.005

**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 Xie, Page and Hardy. 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.



# Stress Induces Contextual Blindness in Lotteries and Coordination Games

Isabelle Brocas<sup>1</sup>, Juan D. Carrillo<sup>1</sup> and Ryan Kendall<sup>2\*</sup>

<sup>1</sup> LABEL and Department of Economics, University of Southern California, Los Angeles, CA, United States, <sup>2</sup> Department of Economics, University College London, London, United Kingdom

In this paper, we study how stress affects risk taking in three tasks: individual lotteries, Stag Hunt (coordination) games, and Hawk-Dove (anti-coordination) games. Both control and stressed subjects take more risks in all three tasks when the value of the safe option is decreased and in lotteries when the expected gain is increased. Also, subjects take longer to take decisions when stakes are high, when the safe option is less attractive and in the conceptually more difficult Hawk-Dove game. Stress (weakly) increases reaction times in those cases. Finally, our main result is that the behavior of stressed subjects in lotteries, Stag Hunt and Hawk-Dove are all highly predictive of each other ( $p$ -value  $< 0.001$  for all three pairwise correlations). Such strong relationship is not present in our control group. Our results illustrate a “contextual blindness” caused by stress. The mathematical and behavioral tensions of Stag Hunt and Hawk-Dove games are axiomatically different, and we should expect different behavior across these games, and also with respect to the individual task. A possible explanation for the highly significant connection across tasks in the stress condition is that stressed subjects habitually rely on one mechanism to make a decision in all contexts whereas unstressed subjects utilize a more cognitively flexible approach.

## OPEN ACCESS

### Edited by:

Monica Capra,  
Claremont Graduate University,  
United States

### Reviewed by:

Jan B. Engelmann,  
University of Amsterdam, Netherlands  
Salomon Israel,  
Hebrew University of Jerusalem, Israel

### \*Correspondence:

Ryan Kendall  
Ryan.Kendall@ucl.ac.uk

**Received:** 18 August 2017

**Accepted:** 13 November 2017

**Published:** 11 December 2017

### Citation:

Brocas I, Carrillo JD and Kendall R  
(2017) Stress Induces Contextual  
Blindness in Lotteries and  
Coordination Games.  
*Front. Behav. Neurosci.* 11:236.  
doi: 10.3389/fnbeh.2017.00236

**Keywords:** stress, contextual blindness, lotteries, coordination games, risk taking

## 1. INTRODUCTION

How does stress influence human behavior? While a significant amount of the work in this direction connects chronic stress with poor health outcomes, stress has also been shown to influence decision-making. The pioneering theory suggests that any stress above an optimal level unambiguously decreases performance (Yerkes-Dodson Law, Yerkes and Dodson, 1908). In spite of this Law's intuitive appeal, subsequent research has unveiled a far more subtle relationship between stress and choice, even in purely objective tasks<sup>1</sup>. In particular, the recent literature has shown a complex relationship between stress and an individual's preference to take risks (reviews in Mather and Lighthall, 2012; Starcke and Brand, 2012). Studies using incentivized lotteries find that stressed males choose more risky lotteries while stressed females choose less risky lotteries (Preston et al., 2007; Lighthall et al., 2009; Van Den Bos et al., 2009)<sup>2</sup>. In addition, compared to a one-time increase in stress, chronic stress experienced over the course of 8 days has been shown to more significantly

<sup>1</sup>For example, subjects under stress are less accurate at identifying visual cues located on the periphery of their vision, but these same subjects are actually *more accurate* than their non-stressed counterparts at identifying cues directly in front of them (Hockey, 1970).

<sup>2</sup>A differential effect across gender is not surprising since, in general, stress is theorized to affect men and women differently (Taylor et al., 2000).

increase risk-aversion (Kandasamy et al., 2014). Finally, cortisol has been shown to play a role in the preference of subjects to avoid ambiguity—a concept closely related to risk (Danese et al., 2017).

There is also a small literature studying the relationship between individual lotteries and two-player coordination (“Stag Hunt”) and anti-coordination (“Hawk-Dove”) strategic situations (or “games”). Results in this area are inconclusive. While some papers suggest a correlation between risk taking in individual lotteries and risk taking in Stag Hunt games (Heinemann et al., 2009; Chierchia and Coricelli, 2015), others do not find any significant relationship (Neumann and Vogt, 2009; Al-Ubaydli et al., 2013; Büyükboyacı, 2014). Imaging studies have found correlations in neural activity between choices in lotteries and Stag Hunt games but no correlation between choices in lotteries and Hawk-Dove games or between choices in the two games (Nagel et al., 2014). The authors conclude that Stag Hunt games engage brain networks associated to risk while Hawk-Dove games engage brain networks associated to strategic thinking.

Our paper lies at the intersection of these two literatures by studying the effect of stress on risk-taking in lotteries and multi-player games of strategy—Stag Hunt and Hawk-Dove<sup>3</sup>. Our laboratory experiment relies on a novel way to represent these three tasks in an identical context that differs in the minimal amount to uniquely distinguish each task (Figure 1). Using this method, differences in behavior across tasks can best be explained by cognitive flexibility in response to fundamental differences across tasks rather than spurious differences in presentations.

Our first result is to show that subjects in both the control and stress condition behave in line with our theoretical predictions. In particular, our participants take more risks in all three tasks as the value of the safe option is decreased. They also take more risks in the individual lottery choice as the probability of the high payoff is increased (Result 1). Our second and main result is that stress impairs cognitive flexibility. More precisely, the choices made by stressed subjects in lotteries, Stag Hunt and Hawk-Dove are all highly and positively correlated with each other. In contrast, control subjects show a (weak) correlation between lotteries and Stag Hunt and no significant correlation between the other pairs of tasks. A cluster analysis reveals that about one-half of the subjects under stress allocate a similar and significant fraction of their endowment to the safe option in all tasks. These subjects are responsible for strengthening the behavioral relationship between tasks (Result 2). Finally, we show that subjects take more time to respond when stakes are high, when the safe option is less attractive and in Hawk-Dove (arguably, the conceptually more difficult game). Stress also tends to increase reaction times in all tasks (Result 3).

The findings suggest that some subjects under stress are oblivious to the fundamental differences that distinguish the three tasks (objective probabilities of lotteries, strategic complementarity of risk-taking in Stag Hunt, and strategic

substitutability of risk-taking in Hawk-Dove). This *contextual blindness* fits in with recent findings which demonstrate that stress promotes habits in humans at the expense of goal-directed performance (Schwabe and Wolf, 2009). It has been shown that people under stress have an increased reliance on automatic over controlled cognitive processes (Schwabe et al., 2012) and are less likely to adjust their initial strategies (Kassam et al., 2009). One underlying mechanism that could lead to contextual blindness is the suppressed activation in the left temporoparietal junction (TPJ) caused by a stressful environment (Engelmann et al., 2017). Impairment of the TPJ has been shown to negatively impact a subject's ability to understand and predict the behavior of others (Samson et al., 2004) which is particularly important in games such as Hawk-Dove. Taken together, the results provide a framework for stress inducing intuitive, rather than deliberative, decision-making (Yu, 2016). Interestingly, previous research on decision-making under risk and stress has made it clear that “such habitual responses do not map neatly onto risk-aversion or risk-seeking” (Buchanan and Preston, 2014). Our paper shows that, rather than a story connecting stress and risk preferences, there is a more complex relationship between stress and risk evaluation across contexts.

A main implication of contextual blindness is that subjects under stress are generally more predictable. Knowing a subject's behavior in any one task is highly predictive of his behavior in the other two tasks. In addition, stress may affect the way we view the agency of our opponent. In our experiment, the behavior of stressed subjects was similar whether they were facing an objective probability or a strategic opponent. When facing an opponent, they expected the same behavior in games that are opposite in nature. One implication from this is that stress causes people to treat others as if they have less sophistication or less agency, which may have other ramifications in social settings.

The paper is organized as follows. Section 2 describes our experimental design and predictions, with particular emphasis on the methodological contributions. Section 3 analyzes the aggregate data in each task and treatment. Section 4 studies the effect of stress on decision-making both across and within tasks, which provides our main result pertaining to contextual blindness. Section 5 investigates how stress and task complexity affect reaction times. Section 6 concludes.

## 2. DESIGN AND PROCEDURES

### 2.1. Experimental Design

We first describe our experimental design. Further details regarding implementation, timing, and exclusion criteria are relegated to Appendix A1.

#### 2.1.1. Stress Inducement and Hormonal Analysis

To induce a stress response in our treatment group, we closely followed the protocol of the Socially Evaluated Cold Pressor Test (SECPT, Schwabe et al., 2008). This task requires subjects to place their hand in ice water while their face is video recorded. All 72 subjects in the stress group successfully passed our requirements for completing the SECPT. To measure hormonal changes, we followed the “passive drool” protocol provided by the laboratory

<sup>3</sup>There is also a literature relating stress to behavior in multi-person games. However, it is only tangentially related to our work as it focuses mainly on the effect of stress on prosocial or anti-social behavior (see Buchanan and Preston, 2014; Van Den Bos and Flik, 2015 for summaries).



FIGURE 1 | Screenshot of LO (Method 1), HD (Method 2), and SH (Method 3).

that ran our assay analysis (ZRT Labs). Each subject was required to submit 3 saliva samples in order to collect data on their baseline, peak, and end cortisol levels. All samples were viable and were used to measure the amount of circulating cortisol.

### 2.1.2. Timeline and Saliva Sample Collection

Since stress responses widely vary across individuals, we followed most of the literature on stress (Preston et al., 2007; Lighthall et al., 2009; Van Den Bos et al., 2009) and implemented a between-subjects design, with *Control* and *Stress* subjects (such method also avoids learning and endowment effects). The timeline of the experiment was the following. First, we provided detailed instructions of the tasks and performed a comprehension quiz. Subjects submitted their “Baseline” saliva sample. Subjects in the control treatment started the tasks immediately after the Baseline sample, whereas subjects in the stress treatment performed the SECPT before starting the tasks. Twenty five minutes after the Baseline saliva sample, all subjects were instructed to stop making choices in the task, and we collected the “Peak” saliva sample. Subjects completed the remaining tasks along with a brief demographic survey. They were shown all their choices and outcomes and provided the “End” saliva sample. One outcome was then randomly chosen by the computer to be used for payment. The average intra- and inter-assay coefficients of variation were no greater than 7 and 8%, respectively.

The procedure had a limitation. Indeed, due to the absence of the SECPT task, the experiment took less time in the control treatment than in the stress treatment. This is reflected in **Figure 2**, where the average time between the Baseline and End saliva sample is 47.6 and 56.6 min, respectively. Ideally, the control treatment should have included a “placebo” task to replace the SECPT (e.g., hand immersion in warm water during 3 min) both to equalize the length and attention demand of the

experiment and to have the saliva samples taken at approximately the same intervals.

### 2.1.3. Participants and Sessions

The study was reviewed by the University Park Institutional Review Board at the University of Southern California (UP-14-00663). Experiments were conducted at the Los Angeles Behavioral Economics Laboratory (LABEL) at the University of Southern California. To participate in the experiment, subjects could not eat, drink anything other than water, smoke, exercise, ingest caffeine, or chew gum within 1 h upon arriving at the laboratory. Subjects were also excluded if they had been asleep within 2 h prior to arriving at the lab or used any lip products at any time after 8 a.m. on the day of the experiment.

All sessions started at 3 p.m. and lasted no longer than 5:15 p.m. They had either 6 or 8 subjects with, at most, two more subjects of one gender in a session. We gathered data on a total of 144 subjects. One subject (stress group) was excluded due to a baseline cortisol 15 times the average of the sample, so our data is comprised of the choices of 143 subjects (71 stress, 66 female).

## 2.2. Tasks

Each subject made choices in three experimental tasks: individual lotteries (LO), Stag Hunt games (SH), and Hawk-Dove games (HD). All three tasks have a *Safe* option  $S$  and a two-state *Risky* option,  $R_H$  and  $R_L$ , so that  $R_L < S < R_H$ . The inherent nature of risk in each task differs. LO is an individual choice problem, where the (objective) probability of earning  $R_H$ ,  $p \equiv \Pr(R_H)$ , is known before the choice is made. SH and HD are two-person, simultaneous, non-cooperative games, where the probability of earning  $R_H$  depends on the choice of another subject in the room. In SH, the probability of earning  $R_H$  is *increasing* in the level of risk chosen by the other subject (a coordination game



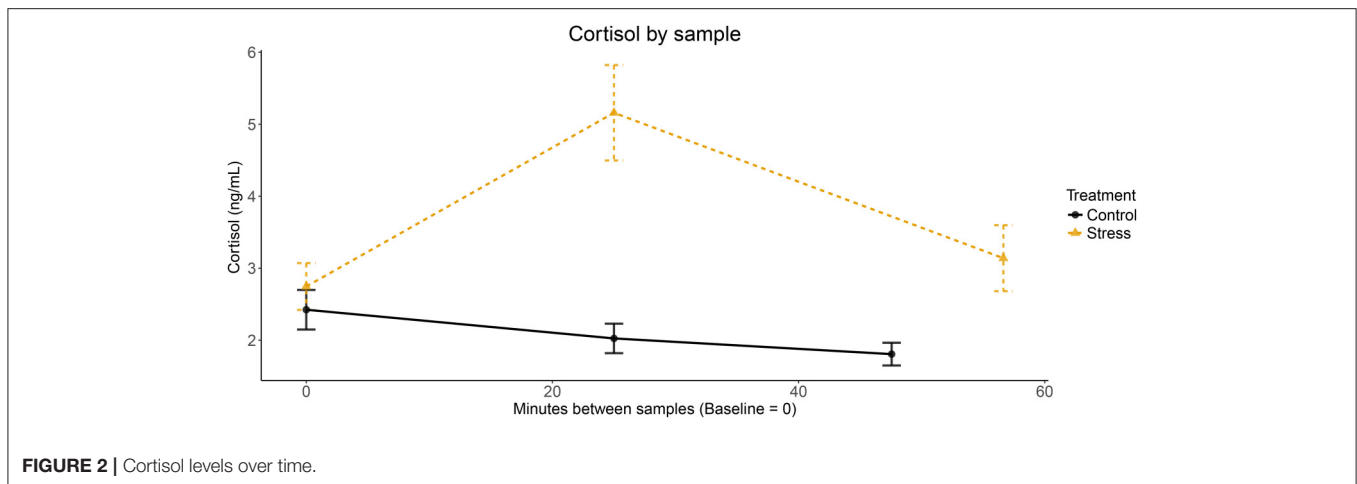


FIGURE 2 | Cortisol levels over time.

where risk-taking is a strategic complement), whereas in **SH** it is *decreasing* in the level of risk chosen by the other subject (an anti-coordination game where risk-taking is a strategic substitute). The basic structure of the tasks is summarized in **Table 1**<sup>4</sup>.

To implement these three tasks, we construct the following novel design. In each round, subjects are given 100 tokens, that they must allocate between the *Safe* and *Risky* options (neutrally labeled “Option A” and “Option B” in the experiment). The computer then randomly selects a ball from an urn with 100 green and orange balls (see below). For any token allocation  $x$  ( $\in \{0, \dots, 100\}$ ) to *Safe* and  $100 - x$  to *Risky*, the payoff obtained by the subject is:

$$\begin{aligned} & \frac{x}{100} S + \frac{100 - x}{100} R_H && \text{if the computer draws a green ball} \\ & \frac{x}{100} S + \frac{100 - x}{100} R_L && \text{if the computer draws an orange ball} \end{aligned}$$

In words, each token allocated to *Safe* yields  $\frac{S}{100}$  whereas each token allocated to *Risky* yields either  $\frac{R_H}{100}$  or  $\frac{R_L}{100}$ . As  $x$  decreases, the spread between the payoff if the computer draws a green and an orange ball increases. If the subject sets  $x = 100$ , she obtains  $S$  for sure. If the subjects sets  $x = 0$ , she obtains either  $R_H$  (green ball) or  $R_L$  (orange ball).

As described, for each token allocated to *Risky*, the probability of earning payoffs  $\frac{R_H}{100}$  and  $\frac{R_L}{100}$  are simply the proportion of green balls and orange balls in the computer’s urn, respectively. The only difference between our three tasks **LO**, **SH**, and **HD** is the way in which the number of green and orange balls is determined:

- In **LO**, the number of green and orange balls is fixed and known (given by  $p$ ).
- In **SH**, the number of green and orange balls is equal to the number of tokens that the participant with whom the subject is matched allocates to *Risky* and *Safe*, respectively.

<sup>4</sup>As it is well-known, **SH** is a coordination game with two pure-strategy equilibria (*Safe-Safe* and *Risky-Risky*) and one mixed-strategy equilibrium whereas **HD** is an anti-coordination game with two pure-strategy equilibria (*Safe-Risky* and *Risky-Safe*) and one mixed-strategy equilibrium.

TABLE 1 | Experimental tasks.

Lotteries - LO		Stag Hunt - SH		Hawk-Dove - HD			
Risky:	$R_H \text{w.p. } p$	Risky	Safe	Risky	Safe		
	$R_L \text{w.p. } 1 - p$	Risky	$R_H, R_H$	$R_L, S$	Risky	$R_L, R_L$	$R_H, S$
Safe:	S	Safe	$S, R_L$	$S, S$	Safe	$S, R_H$	$S, S$

- In **HD**, the number of green and orange balls is equal to the number of tokens that the participant with whom the subject is matched allocates to *Safe* and *Risky*, respectively.

In addition, in **SH** and **HD** subjects are told that their choice affects the number of green and orange balls in the urn of the participant with whom they are matched in the exact same way. That is, in **SH** (**HD**) the more tokens a subject allocates to *Risky*, the *more* (*less*) likely it is that the other participant earns  $R_H$ .

**Figure 1** provides screenshots of the **LO** (top), **HD** (bottom left) and **SH** (bottom right) tasks. At the top of the screen, the subject is told the current task (neutrally labeled as “Method 1,” “Method 2,” and “Method 3,” respectively). She is also reminded how the number of green and orange balls in her urn is determined. At the center of the screen, the subject can observe the parameters of the current round. In these three tasks,  $S = \$21$ ,  $R_H = \$53$  and  $R_L = \$13$ . At the bottom of the screen, there is a slider that the subject can use to allocate her 100 tokens across *Safe* and *Risky*. As the subject moves the slider to test different token allocations, the earnings for each ball color are calculated and presented in real-time on the screen. In all three screenshots, the subject has set  $x = 29$ . After the subject is satisfied with the allocation of tokens, she has to click the “CONFIRM” button to submit her choice.

Our experiment has two methodological contributions that we would like to emphasize. First, the contextual presentation of the three tasks is almost identical. Only the information concerning the determination of green and orange balls is changed. Capturing the inherently different natures of risk in such a symmetric way serves an important purpose: different behavior is likely to be only in response to the meaningful differences

between these tasks, rather than to superficial differences in presentation or comprehension. Second, endowing subjects with 100 tokens that can be allocated across *Safe* and *Risky* can be used to measure “interior” behavior. In lotteries, it is analogous to portfolio diversification. In games, it is analogous to allowing subjects to play mixed strategies. In both cases, it provides more information than the standard binary choice method.

### 2.3. Payoff-Variants, Stakes, and Equilibria

Subjects played a total of 48 rounds, 16 rounds of each task all with different payoffs. The experiment was broken up into blocks of 4 consecutive rounds of the same task, and all sessions started with a **LO** block, which was arguably simpler. Before each block, subjects were shown a screen reminding them that a new block was starting. This screen ensured that subjects would be aware of which task (**LO**, **SH**, or **HD**) they were playing next. For the games, subjects were randomly and anonymously rematched after each round. For the lotteries, they were playing an individual decision problem (the exact experimental instructions are in Appendix B). To avoid learning effects, subjects did not see the behavior of their partner nor the color of the ball drawn by the computer in each round. At the end of the 48 rounds, subjects observed all their choices and those of their partners. One round was randomly drawn by the computer and the outcome in that round was used for payment. Subjects earned an average of \$31, with a minimum of \$1 (twice) and a maximum of \$53 (three subjects). In addition to these earnings, all subjects were paid a \$5 show-up fee.

We chose the payoffs in order to provide substantial variation in monetary stakes and equilibrium predictions. First, define:

$$\Delta \equiv R_H - R_L \quad (1)$$

as a measure of the monetary *stakes*. For all tasks, we set  $\Delta \in \{10, 20, 30, 40\}$ . In the analysis, we will refer to “low stakes” as  $\Delta \in \{10, 20\}$  and “high stakes” as  $\Delta \in \{30, 40\}$ . Second, given a triplet  $(R_L, S, R_H)$ , the mixed-strategy Nash equilibrium of the **SH** game is:

$$\alpha \equiv \frac{S - R_L}{R_H - R_L} \quad (2)$$

where  $\alpha$  is the probability of choosing *Risky*. For each  $\Delta$ , we choose  $(R_L, S, R_H)$  so that  $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$ . This gives 16 combinations of stakes and mixed equilibrium predictions in **SH**.

Finally, notice that once we fix  $\Delta$ , then  $\alpha$  is proportional to  $S$  the payoff of the *Safe* option.

Notice that for a given triplet  $(R_L, S, R_H)$ , the mixed-strategy Nash equilibrium of **HD** is:

$$1 - \alpha \equiv \frac{R_H - S}{R_H - R_L} \quad (3)$$

where  $1 - \alpha$  is the probability of choosing *Risky*. Therefore, the same payoff-triplets as in **SH** provide also 16 combination of stakes ( $\Delta \in \{10, 20, 30, 40\}$ ) and mixed-strategy equilibria ( $1 - \alpha \in \{0.8, 0.6, 0.4, 0.2\}$ ) in **HD**. Last, we use the technique developed by Jessie and Kendall (2015) to select the payoffs in a way that the differences between games are only in the component that the Nash Equilibrium uses to make predictions. **Table 2** provides a sample of eight games used in the experiment and Appendix A2 provides the entire list.

Finally, to create the **LO** tasks, we choose the payoffs  $(R_L, S, R_H)$  of the **SH** and **HD** games corresponding to the extreme mixed-strategy Nash equilibria of the games:  $\alpha = 0.2$  and  $\alpha = 0.8$ . Using these payoffs, we set the lottery probability of the high payoff  $R_H$  to  $p = 0.2$  and  $p = 0.8$ . Creating four lotteries in this way for  $\Delta \in \{10, 20, 30, 40\}$  yields a total of 16 **LO** tasks. **Table 3** provides some examples of lotteries.

### 2.4. Predictions

Our model has three parameters  $(\Delta, \alpha, p)$  in the **LO** tasks and two parameters  $(\Delta, \alpha)$  in the **SH** and **HD** tasks.

Predictions in **LO** are standard. Fixing the other two parameters, *Risky* becomes more attractive as  $p$  increases (first-order stochastic increase in the risky option) and  $\alpha$  decreases ( $S$  closer to  $R_L$ ). The effect of  $\Delta$  is less clear. For example, increasing  $\Delta$  makes *Risky* more desirable when  $p = 0.8$  and  $\alpha = 0.2$  and less desirable when  $p = 0.2$  and  $\alpha = 0.8$ .

**TABLE 3 |** Examples of payoff-variants in **LO** tasks.

<b>LO</b> ( $\alpha = 0.2$ ; $\Delta = 40$ ; $p = 0.8$ )	<b>LO</b> ( $\alpha = 0.2$ ; $\Delta = 30$ ; $p = 0.2$ )	<b>LO</b> ( $\alpha = 0.8$ ; $\Delta = 20$ ; $p = 0.8$ )	<b>LO</b> ( $\alpha = 0.8$ ; $\Delta = 10$ ; $p = 0.2$ )
<i>Safe</i> : 21 w.p. 1 <i>Risky</i> : 53 w.p. 0.8 13 w.p. 0.2	<i>Safe</i> : 22 w.p. 1 <i>Risky</i> : 46 w.p. 0.2 16 w.p. 0.8	<i>Safe</i> : 30 w.p. 1 <i>Risky</i> : 34 w.p. 0.8 14 w.p. 0.2	<i>Safe</i> : 28 w.p. 1 <i>Risky</i> : 30 w.p. 0.2 20 w.p. 0.8

**TABLE 2 |** Examples of payoff-variants in **SH** and **HD** tasks.

<b>SH</b> ( $\alpha = 0.2, \Delta = 40$ )		<b>SH</b> ( $\alpha = 0.4, \Delta = 30$ )		<b>SH</b> ( $\alpha = 0.6, \Delta = 20$ )		<b>SH</b> ( $\alpha = 0.8, \Delta = 10$ )	
	<i>Risky</i> <i>Safe</i>		<i>Risky</i> <i>Safe</i>		<i>Risky</i> <i>Safe</i>		<i>Risky</i> <i>Safe</i>
<i>Risky</i>	53, 53   13, 21	<i>Risky</i>	43, 43   13, 25	<i>Risky</i>	36, 36   16, 28	<i>Risky</i>	30, 30   20, 28
<i>Safe</i>	21, 13   21, 21	<i>Safe</i>	25, 13   25, 25	<i>Safe</i>	28, 16   28, 28	<i>Safe</i>	28, 20   28, 28
<b>HD</b> ( $1 - \alpha = 0.8, \Delta = 40$ )		<b>HD</b> ( $1 - \alpha = 0.6, \Delta = 30$ )		<b>HD</b> ( $1 - \alpha = 0.4, \Delta = 20$ )		<b>HD</b> ( $1 - \alpha = 0.2, \Delta = 10$ )	
	<i>Risky</i> <i>Safe</i>		<i>Risky</i> <i>Safe</i>		<i>Risky</i> <i>Safe</i>		<i>Risky</i> <i>Safe</i>
<i>Risky</i>	13, 13   53, 21	<i>Risky</i>	13, 13   43, 25	<i>Risky</i>	16, 16   36, 28	<i>Risky</i>	20, 20   30, 28
<i>Safe</i>	21, 53   21, 21	<i>Safe</i>	25, 43   25, 25	<i>Safe</i>	28, 36   28, 28	<i>Safe</i>	28, 30   28, 28

Predictions in **SH** and **HD** are more subtle. By construction, in all 32 rounds there are two pure-strategy and one mixed-strategy equilibria. Subjects may move from one equilibrium to another, so behavior depends crucially on beliefs about the other player's action and comparative statics should be taken with a grain of salt. However, fixing the belief about the other player's constant, it seems intuitive that *Risky* is more attractive in both **SH** and **HD** as the sure payoff  $S$  becomes closer to  $R_L$ , that is, as  $\alpha$  decreases. Again, the effect of changes in the spread of payoffs  $\Delta$  is more nuanced and depends on the position of  $S$ .

Finally, there are also interesting differences between **SH** and **HD**. **SH** is a coordination game, where risk-taking behavior is a strategic complement. This means that, holding constant the belief about the opponent, a decrease in  $\alpha$  offers the subject more incentives to take risks. Furthermore, the subject realizes that the opponent also has more incentives to take risks, reinforcing the value of playing *Risky*. By contrast, **HD** is an anti-coordination game where risk-taking behavior is a strategic substitute. As  $\alpha$  decreases, the subject has more incentives to choose *Risky* but realizes that the opponent has the same incentives, which decreases the value of risk-taking. Overall, strategic considerations make comparative statics significantly easier to evaluate when incentives of players are aligned (**SH**) than when they are not (**HD**).

### 3. AGGREGATE RESULTS

#### 3.1. Stress

**Figure 2** shows the evolution of cortisol levels throughout the experimental sessions in both treatments. Each dot represents the average level of salivary cortisol samples (ng/mL) taken at baseline, peak, and end of the experiment. We report minutes on the x-axis. Note that the timing of the end sample was different across sessions and we represent the average number of minutes in each treatment. The control and stress groups start with statistically indifferent levels of average cortisol (2.42 vs. 2.75; two-sided Welch  $t$ -test,  $p$ -value = 0.133). The stress group experiences a large and statistically significant increase in average cortisol (2.75 vs. 5.16;  $p$ -value < 0.001). In comparison, the control group experiences a slight and statistically significant decrease in average cortisol (2.42 vs. 2.03;  $p$ -value = 0.022). Higher cortisol levels are also observed in the stress group in the end sample (1.81 vs. 3.14;  $p$ -value < 0.001).

#### 3.2. Allocation between Options

The average proportion of wealth invested in *Safe* is 0.63 in **LO**, 0.53 in **SH** and 0.65 in **HD**. Results between lotteries and games are not directly comparable. By contrast, results between the two games are comparable since the 16 rounds of **SH** involve the same payoff triplets ( $R_L, S, R_H$ ) as the 16 rounds of **HD**. We notice a significantly lower allocation to *Safe* in **SH** than in **HD** (< 0.001).

### 3.3. Testing the Theory

#### 3.3.1. Behavior in Lotteries

Choices in **LO** conformed to the theoretical predictions. Holding  $\Delta$  constant, the proportion allocated to *Safe* increased as  $\alpha$  increased and as  $p$  decreased for all stakes and in both treatments.

**TABLE 4 |** Allocation to *Safe* as a function of  $\alpha$  and  $\Delta$  by game (pooled treatments).

SH	Stakes ( $\Delta$ )					HD	Stakes ( $\Delta$ )				
	40	30	20	10			40	30	20	10	
$\alpha = 0.2$	0.31	0.27	0.19	0.23		$\alpha = 0.2$	0.49	0.49	0.36	0.41	
$\alpha = 0.4$	0.49	0.46	0.41	0.41		$\alpha = 0.4$	0.59	0.58	0.55	0.52	
$\alpha = 0.6$	0.66	0.61	0.62	0.69		$\alpha = 0.6$	0.73	0.72	0.72	0.78	
$\alpha = 0.8$	0.75	0.83	0.80	0.82		$\alpha = 0.8$	0.83	0.83	0.89	0.89	

Overall, subjects were (weakly) risk averse. They invested, on average, 97% of the endowment in *Safe* when the expected value of *Risky* was below the *Safe* option, against 70% when it was equal and 17% when it was above the *Safe* option<sup>5</sup>. Finally, the proportion in *Safe* was significantly lower in the low stakes rounds ( $\Delta \in \{10, 20\}$ ) compared to the high stakes rounds ( $\Delta \in \{30, 40\}$ ) under stress ( $p$ -value = 0.035) but only marginally in the control group ( $p$ -value = 0.051).

#### 3.3.2. Behavior in Games

The proportion of wealth allocated to *Safe* varied with  $\alpha$  as predicted in Section 2.4. In **SH** and keeping beliefs constant, increasing  $\alpha$  makes *Safe* more attractive for a subject and, as the same logic applies for the partner, higher allocation rates in *Safe* are expected. **Table 4** (left) shows that this is exactly how subjects behave for all stake levels. The average fraction allocated to *Safe* was significantly different between all pairs of  $\alpha$  for all  $\Delta$  ( $p$ -values < 0.05). In **HD** and keeping beliefs constant, increasing  $\alpha$  (that is, decreasing  $1 - \alpha$ ) makes again *Safe* more attractive and should push more subjects to invest in *Safe*. However, they should expect their partner to also invest more in *Safe*, which should ultimately reduce the incentives to invest in that option. This implies that the response to an increase in  $\alpha$  in **HD** should be less pronounced than in **SH**. Empirically, **Table 4** (right) shows that increasing  $\alpha$  made subjects invest significantly more in *Safe* for all pairs of  $\alpha$  and all  $\Delta$  ( $p$ -values < 0.05)<sup>6</sup>. Finally, we also computed for each individual the average increase in the fraction allocated to *Safe* between  $\alpha = 0.2$  and  $\alpha = 0.8$  in both **SH** and **HD**. We found a statistically higher increase in **SH** than in **HD** (0.56 vs. 0.43,  $p$ -value < 0.001), suggesting that subjects understood the difference between the strategic complementarity and the strategic substitutability of risk-taking in these two tasks. Last and as noted before, there is no particular reason to observe an aggregate effect of stakes in behavior. Empirically, we found none.

**Result 1.** *On aggregate, subjects behave in accordance with our predictions: the allocation to the safe option is increasing in  $\alpha$  in*

<sup>5</sup>Since virtually no subject exhibited risk-loving preferences, the four **LO** rounds where *Risky* has lower expected value than the *Safe* option contain no extra information. As a robustness check, we conducted the entire analysis of the paper without these four rounds. All the results were statistically identical.

<sup>6</sup>Recall that in **SH**,  $\alpha$  is the probability of playing *Risky* in the mixed strategy equilibrium. In **HD**,  $1 - \alpha$  is the probability of playing *Risky* in the mixed strategy equilibrium.

all three tasks and decreasing in  $p$  in lotteries. Changes in stakes have no systematic effect on behavior.

## 4. STRESS

### 4.1. Stress and Tasks

We noted a slight increase in the average proportion allocated to *Safe* in the stress treatment in all tasks compared to the control treatment (0.64 vs. 0.63 in **LO**, 0.55 vs. 0.52 in **SH**, and 0.65 vs. 0.65 in **HD**). However, the differences were not statistically significant. As presented in **Figure 3**, the cumulative distribution functions of the average amounts allocated to *Safe* were also similar across treatments in all three tasks, with no statistically significant effect according the Kolmogorov-Smirnoff test ( $p$ -value = 0.31 in **LO**,  $p$ -value = 0.31 in **SH**, and  $p$ -value = 0.97 in **HD**). Overall, we found no evidence that stress affected behavior within each task.

The existing literature is ambiguous on this issue. Some studies have found that stress affects behavior in lotteries (Preston et al., 2007; Lighthall et al., 2009; Van Den Bos et al., 2009) whereas others found no effect of stress (von Dawans et al., 2012; Gathmann et al., 2014). Differences in responses to stress may be attributed to differences across studies in risk elicitation methods (BART, IGT, objective lotteries) and experimental procedures (presence/absence of incentives, hypothetical/real choices, different stressors). For instance, it may be that the emotional component contained in the BART experiment (anticipation of the balloon explosion and visual representation of such explosion) is responsible for shifts in behavior. Moreover, in BART and IGT subjects are typically *not* informed of the objective probabilities of the events. This ambiguity component may also trigger different thought processes that are differentially affected by stress (Buckert et al., 2014; Danese et al., 2017).

### 4.2. Stress and Gender

In **Table 5** we present the differences in allocation across gender. In the control condition, females allocate significantly more to *Safe* than males in **LO** and **SH** but not in **HD**. In the stress condition we find no significant gender differences in any task.

Our data contribute to gender research in three ways. First, the fact that women take less risk in **LO** in the control group aligns

with earlier literature (Charness and Gneezy, 2012). Second, finding males in the control group to be more cooperative in **SH** contributes to our understanding of gender differences in coordination games. However, we are hesitant to extrapolate about general inclinations to cooperate since, as suggested by Croson and Gneezy (2009), gender differences seem to be highly sensitive to context. Finally, since the only significant gender differences are found in the control group, we conclude that stress has the capability to diminish differences between genders.

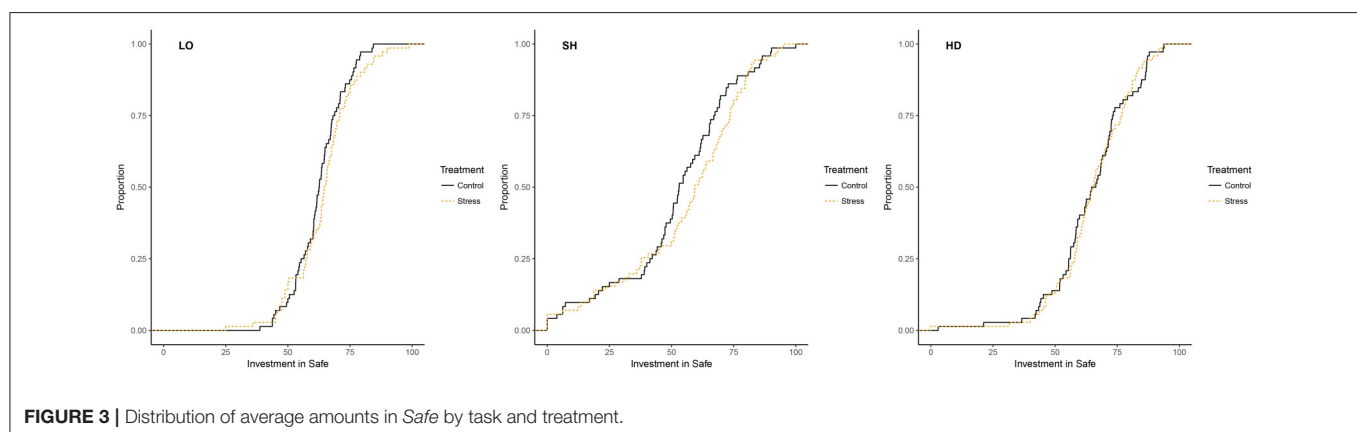
### 4.3. The Effect of Stress on the Relationship between Tasks

Our next question is whether the willingness of individuals to choose *Risky* is correlated across tasks. On the one hand, it seems natural that subjects who are less risk-averse, that is, those who invest more in *Risky* in **LO** (individual lotteries with objective probabilities) are also expected to take more risks in games. On the other hand, this may not be necessarily true since our games have multiple equilibria, so risk-taking in **SH** and **HD** depends crucially on beliefs about the other player's behavior. Furthermore, the two games are fundamentally opposite in the optimal reaction to the other player's choice (coordination vs. anti-coordination). **Table 6** presents the Pearson correlation coefficient ( $\rho$ ) of the proportion allocated to *Safe* by individuals

**TABLE 5 |** Average allocation to *Safe* by gender, treatment and task.

	Control			Stress		
	Female	Male	Difference [ $p$ -value]	Female	Male	Difference [ $p$ -value]
<b>LO</b>	0.65 (0.018)	0.60 (0.014)	0.045	0.66 (0.023)	0.62 (0.020)	0.194
<b>SH</b>	0.59 (0.039)	0.46 (0.037)	0.015	0.60 (0.042)	0.51 (0.040)	0.143
<b>HD</b>	0.66 (0.033)	0.63 (0.022)	0.554	0.68 (0.024)	0.63 (0.027)	0.112

Standard errors in parenthesis.



**FIGURE 3 |** Distribution of average amounts in *Safe* by task and treatment.



**TABLE 6 |** Correlation of individual risk taking behavior across tasks by treatment.

	Control		Stress	
	LO	SH	LO	SH
SH	0.347**	–	0.416***	–
HD	0.147	0.117	0.461***	0.497***

\* $p < 0.05$ ; \*\* $p < 0.01$ ; and \*\*\* $p < 0.001$ .

across tasks, both in the control (left panel) and stress (right panel) conditions.

In the control condition, the amount allocated to *Safe* in **LO** is significantly correlated with the amount allocated to *Safe* in **SH**, suggesting that risk attitude is a reasonably good predictor of behavior in the coordination game. This finding aligns with previous studies showing a correlation between **LO** and **SH** choices (Heinemann et al., 2009; Chierchia and Coricelli, 2015). By contrast, the control condition shows no significant correlation between **LO** and **HD** or between **SH** and **HD**. This may not be surprising given the previous research showing that these tasks activate different areas of the brain (Ekins et al., 2013; Nagel et al., 2014).

By contrast, in the stress condition, the amounts allocated to *Safe* are significantly correlated across all tasks. Correlations are also stronger, suggesting that risk-taking under stress is very similar across tasks, irrespective of the situation. This important result indicates that, even though stress did not have an effect on the overall distribution of risk taking in the population across tasks, it did affect intra-personal decisions. The result was confirmed by a set of robust regressions reported in **Table 7**, which suggests a stronger relationship between the amount allocated to *Safe* in **LO**, **SH** and **HD** under stress than in the control treatment. This effect will be corroborated with the trial-by-trial regression analysis.

We then compared the correlation coefficients across conditions by assessing statistical significance of the Fisher's  $r$  to  $z$  transformations. We found that the correlation between **LO** and **SH** are not significantly different between control and stress conditions. By contrast, correlations between **LO** and **HD** and between **SH** and **HD** respectively are significantly different (with respective  $p$ -values of 0.040 and 0.012). This result further supports the finding that subjects under stress make choices that are more similar across tasks than subjects in the control treatment.

A possible explanation for this result is that subjects under stress (and only those subjects) exhibit contextual blindness, that is, they ignore the context that distinguishes these three tasks. Indeed, **LO** measures an individual's propensity to take risks which has no social context. **SH** captures a tension between risk and cooperation whereas **HD** captures a tension between risk and aggression. The experiment was designed so that these contexts were the only difference between tasks. **Table 6** reveals that the behavior of stressed subject when faced with an objective probability over earnings was strongly and positively correlated with their behavior when faced with a strategic opponent, even if games were opposite in nature. For control subjects there

**TABLE 7 |** Robust regression of the average investment in *Safe* in **SH** and **HD** on the average investment in the safe option in lotteries (*Safe-LO*) by treatment.

	Control		Stress	
	SH	HD	SH	HD
<i>Safe-LO</i>	0.94**	0.52*	1.06***	0.78***
Constant	–4.41	32.86*	–9.33	16.84*
Robust SE	17.7	11.92	16.94	10.13
Adj. $R^2$	0.168	0.126	0.323	0.460

\* $p < 0.05$ ; \*\* $p < 0.01$ ; and \*\*\* $p < 0.001$ .

**TABLE 8 |** OLS of investment in *Safe* in **SH** and **HD** including fixed effects.

	Control		Stress	
	SH	HD	SH	HD
Lottery	0.70* (0.27)	0.23 (0.20)	0.78*** (0.22)	0.55*** (0.13)
Male	–10.07 (5.29)	–1.27 (3.97)	–5.52 (5.46)	–3.64 (3.35)
High Stakes	0.28 (1.68)	0.05 (1.54)	4.87** (1.60)	3.42* (1.46)
$\alpha = 0.8$	55.3*** (2.37)	39.3*** (2.18)	54.9*** (2.27)	44.6*** (2.07)
$\alpha = 0.6$	40.4*** (2.37)	30.6*** (2.18)	38.9*** (2.27)	29.6*** (2.07)
$\alpha = 0.4$	17.5*** (2.37)	13.1*** (2.18)	20.9*** (2.27)	11.3*** (2.07)
Constant	–15.4 (18.0)	30.23* (13.5)	–23.2 (15.1)	8.9 (9.3)
Observations	1,152	1,152	1,136	1,136
FE groups	72	72	71	71
df	9	9	9	9
Log-likelihood	–5,559	–5,447	–5,429	–5,300
BIC	11,182	10,957	10,922	10,663

Standard errors in parenthesis. \* $p < 0.05$ ; \*\* $p < 0.01$ ; and \*\*\* $p < 0.001$ .

was only a relationship between **LO** and **SH**. In other words, control subjects responded more to the differing contexts than stressed subjects. One implication is that the choices of subjects under stress are generally more predictable: knowing the average amount a subject invests into *Safe* in any one task provides significant information about behavior in the other two.

We also ran OLS regressions of the trial-by-trial amounts allocated to *Safe* for each game and in each condition. We used as regressors the individual average amount allocated to *Safe* in **LO** (which captures the risk attitude of each individual), and dummies for stakes (1 = High stakes), for the position of  $S$  relative to  $R_L$  and  $R_H$  ( $\alpha$ ), and for gender (1 = Male). We constructed a fixed effect model by including a dummy variable for each individual. The results are compiled in **Table 8**.

In the *Control* condition, the average allocation to *Safe* in **SH** is predicted by the behavior in **LO**, but the average allocation in

**HD** is not. In the *Stress* condition, the average amounts allocated to *Safe* in both **SH** and **HD** are highly predicted by behavior in **LO**. These regressions further confirm the contextual blindness result<sup>7</sup>. We also notice that gender has no explanatory power and that the allocation to the safe choice is increased for high stakes, but only in the *Stress* condition.

To better assess the significance of the effect of stress in **HD**, we ran a regression of the trial-by-trial amounts allocated to *Safe* in **HD** in both conditions on the same regressors as before as well as the individual difference in cortisol between baseline and peak ( $\Delta\text{Cortisol}$ ) and an interaction term between that measure and the average allocation to *Safe* in **LO**<sup>8</sup>. For comparison, we ran the same regression for **SH** as well. This exercise tests directly whether the coefficients of the average allocation to *Safe* in **LO** in the previous table are significantly different across treatments. The results are reported in the first two columns of **Table 9**. The absence of a significant interaction in the case of **SH** confirms that the amount allocated to *Safe* in **LO** does not predict differentially behavior in **SH** across conditions. By contrast, the interaction term is significant in the case of **HD**, the contribution of the amount allocated to *Safe* in **LO** to behavior in **HD** differs across conditions. We finally ran a full regression over both games using a dummy variable for our games ( $1 = \text{SH}$ ). The results are reported in the last column of **Table 9**. The fact that the three way interaction between the average allocation to *Safe* in **LO**, the treatment and the increase in cortisol is significant indicates that the interaction between *Safe* in **LO** and *Stress* is significantly different across games. The regression also shows a subtle interaction between cortisol increase and games: subjects who exhibit a higher increase in cortisol level tend to increase more their investment to *Safe* in **HD**.

#### 4.4. Cluster Analysis

The fact that stress does not have any visible effect on aggregate behavior (Section 4.1) but reduces gender differences (Section 4.2) and impacts the relationship between tasks (Section 4.3) is puzzling. We therefore decided to study in more detail the behavior of individuals across the three tasks.

We conducted a cluster analysis in each condition to group subjects according to their average allocation to *Safe* in each task. We retained a model-based clustering method to identify the clusters present in our population. A wide array of heuristic clustering methods are commonly used but they typically require the number of clusters and the clustering criterion to be set ex-ante rather than endogenously optimized. Mixture models, on

**TABLE 9** | OLS with interactions of investment in *Safe* in **SH** and **HD** including fixed effects.

	SH	HD	All games
<i>Lottery</i>	0.73*** (0.19)	0.28* (0.12)	0.25 (0.13)
$\Delta\text{Cortisol}$	−0.17 (4.81)	−7.81** (3.22)	−8.33* (3.33)
<i>Game (SH = 1)</i>	–	–	−44.16*** (5.36)
<i>Lottery</i> * $\Delta\text{Cortisol}$	0.01 (0.07)	0.12** (0.05)	0.13** (0.05)
<i>Lottery</i> * <i>Game</i>	–	–	0.51*** (0.08)
$\Delta\text{Cortisol}$ * <i>Game</i>	–	–	8.67*** (2.19)
<i>Lottery</i> * $\Delta\text{Cortisol}$ * <i>Game</i>	–	–	−0.13*** (0.03)
<i>Male</i>	−7.93* (3.85)	−3.64 (2.58)	−5.79* (2.52)
<i>High Stakes</i>	2.56* (1.16)	1.72 (1.06)	2.14* (0.86)
$\alpha = 0.8$	55.11*** (1.64)	41.90*** (1.50)	48.50*** (1.21)
$\alpha = 0.6$	39.66*** (1.64)	30.12*** (1.50)	34.89*** (1.21)
$\alpha = 0.4$	19.19*** (1.64)	12.17*** (1.50)	15.68*** (1.21)
<i>Constant</i>	−19.44 (12.68)	26.94** (9.3)	25.83** (8.73)
Observations	2,288	2,288	4,576
FE groups	143	143	143
df	11	11	15
Log-likelihood	−11,007	−10,764	−22,037
BIC	22,099	21,613	44,200

Standard errors in parenthesis. \* $p < 0.05$ ; \*\* $p < 0.01$ ; and \*\*\* $p < 0.001$ .

the other hand, treat each cluster as a component probability distribution. Thus, the choice between numbers of clusters and models can be made using Bayesian statistical methods (Fraley and Raftery, 2002). We implemented our model-based clustering analysis with the Mclust package in R (Fraley and Raftery, 2006). We considered ten different models with a maximum of nine clusters each, and retained the cluster combination that yielded the minimum Bayesian Information Criterion (BIC). In the *Control* condition, the best model consisted of three clusters (C1, C2, and C3). In the *Stress* condition, four different clusters best summarized behavior (S1, S2, S3, and S4). **Table 10** summarizes the descriptive statistics in each cluster. **Figure 4** provides a visual representation of the clusters across treatments<sup>9</sup>.

In the *Control* condition, the majority of the subjects (C1) exhibited the typical behavior: they invested similar proportions

<sup>7</sup>We also ran the same OLS regressions with the behavior in the other game as an extra regressor. Results and significance were very similar. Furthermore, and confirming the results in **Table 6**, the new variable had a positive and significant coefficient in the *Stress* regressions and a positive but not significant coefficient in the *Control* regressions. Notice that a two-censored non-linear Tobit model would allow for censoring at 0 and 100 but requires analysis at the subject-average level since it cannot account for subject-level fixed effects. The average data was rarely censored at either 0 or 100 which makes such a model inappropriate.

<sup>8</sup>The previous analysis only makes a qualitative comparison of the association between the allocation to *Safe* in **HD** and **SH** and the average allocation to *Safe* in **LO** in the two conditions. A formal analysis of the interactions between money allocation, games and conditions within the same model allows to directly compare the strength of the across conditions and games (see Nieuwenhuis et al., 2011).

<sup>9</sup>To better represent the information, we do not use three-dimensional graphs. Instead, we provide projections of each pair of tasks separately.

in the *Safe* asset in **LO** and **HD** and less in **SH**, suggesting large homogeneity across subjects in this treatment. A few individuals (C2) were an extreme version of this typical play, with overly risky behavior in **SH**. Finally, a minority of all female subjects (C3) allocated significantly more to *Safe* in **LO** and **HD**, but especially in **SH**. This group was responsible for the gender effect detected in **LO** and **SH** in the control condition.

In the *Stress* condition, there were three main clusters (S4 consists of 3 outliers), similar to the clusters obtained in the

control condition. Cluster S1 was the analog of C1, while S2 was similar to C2, except for a safer proportion of choices in **SH**. However, half of the subjects were now grouped in S3, a cluster similar to C3. These subjects allocated a large fraction of their endowment to *Safe* in all tasks. S3 had also the particularity that allocations were extremely similar across tasks (69.1–74.7% with low standard errors). These subjects were responsible for strengthening the relationship between tasks. Moreover, there was no gender supremacy in that cluster, causing the gender effect observed in the control condition to disappear under stress.

Result 2. *Aggregate behavior is similar across treatments whereas individual choices are affected by stress. A significant fraction of participants in the stress condition are subject to contextual blindness, choosing a similar allocation independently of the task.*

**TABLE 10 |** Endogenous clusters in each condition (standard errors in parenthesis).

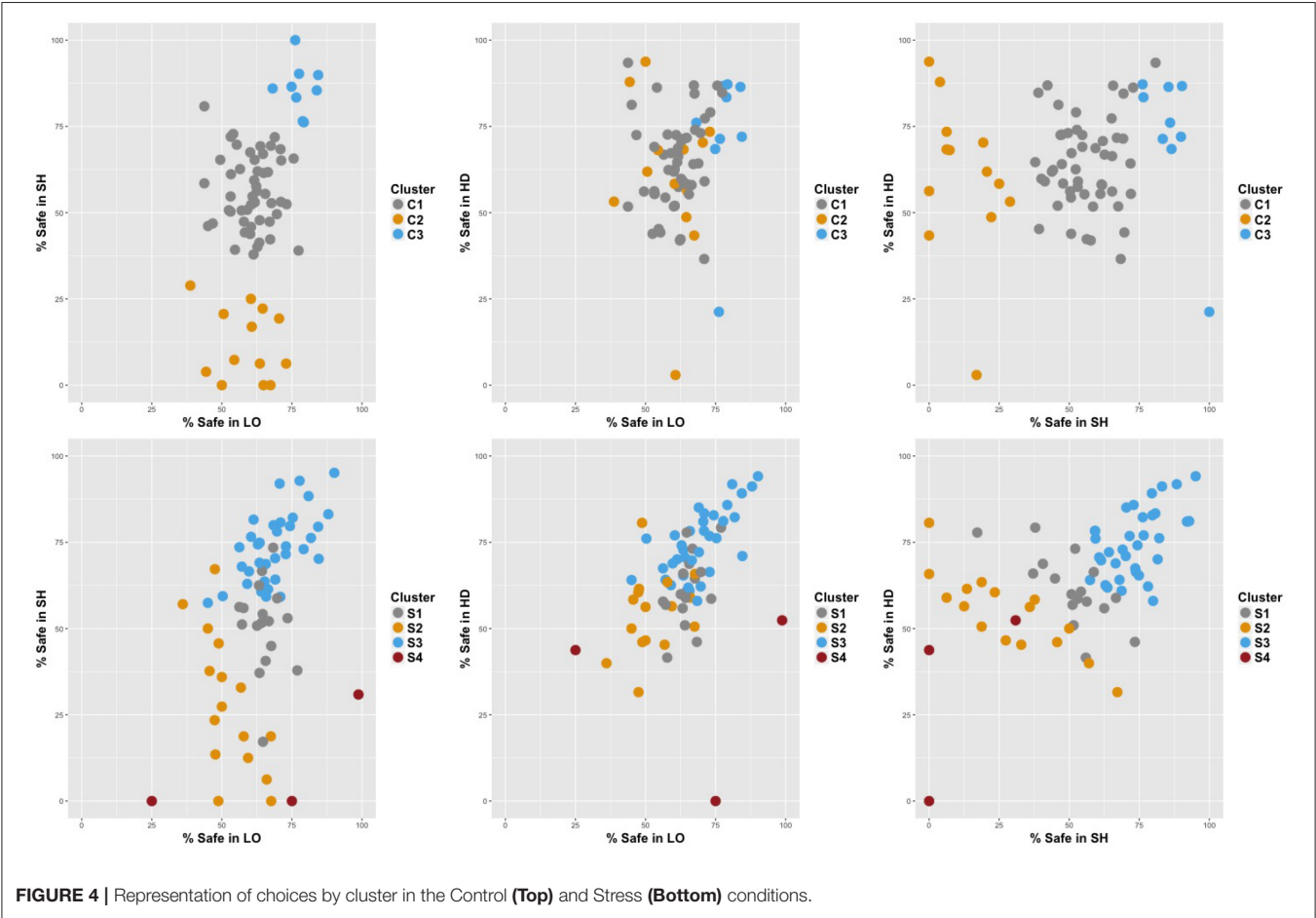
	Control			Stress			
	C1	C2	C3	S1	S2	S3	S4
Male/Female	29/21	9/4	0/9	12/5	10/6	15/20	2/1
% Safe in <b>LO</b>	60.9 (1.11)	58.7 (2.87)	77.7 (1.62)	65.1 (1.30)	52.6 (2.25)	69.1 (1.74)	66.3 (21.7)
% Safe in <b>SH</b>	56.0 (1.49)	12.1 (2.88)	86.0 (2.42)	50.9 (3.13)	27.9 (5.02)	73.4 (1.71)	10.3 (10.3)
% Safe in <b>HD</b>	64.2 (1.87)	60.5 (6.23)	72.5 (6.86)	61.4 (2.46)	54.5 (2.89)	74.7 (1.64)	32.0 (16.2)

5. REACTION TIMES

5.1. Task Difficulty

In Table 11 we report the average reaction time (RT) in seconds separated by task and treatment.

Making choices took more time under stress across all tasks, although the effect was mostly due to **HD**. We also found that RT were longer in **HD** compared to **SH** irrespective of the treatment



**TABLE 11 |** Reaction time by task and treatment.

	LO	SH	HD	All
Control	25.6 (0.71)	24.2 (0.61)	28.2 (0.62)	26.0 (0.37)
Stress	27.3 (0.71)	25.7 (0.65)	31.1 (0.73)	28.0 (0.40)
Difference [p-value]	0.087	0.097	0.002	<0.001

Standard errors in parenthesis.

**TABLE 12 |** Reaction time in lotteries by treatment and expected value of lottery (EV).

	Control	Stress	Difference [p-value]
EV < S	19.8 (1.51)	19.6 (1.59)	0.937
EV = S	28.6 (1.59)	30.3 (1.51)	0.420
EV > S	25.6 (1.51)	28.7 (1.70)	0.173

( $p < 0.001$ ), consistent with the idea that the anti-coordination game is more complex to evaluate than the coordination game.

## 5.2. Attention in Lotteries

As reflected in **Table 12**, risky options with expected value below the safe alternative (EV < S) were quickly discarded. Subjects took significantly more time to choose when the expected value of the risky option was equal (EV = S) or greater (EV > S) than the safe option (*t*-test,  $p$ -value < 0.01 for all paired comparisons in *Control* and *Stress* treatments). For the more complex lotteries (EV > S), subjects took slightly more time under stress, although not significantly so.

## 5.3. Attention in Games

**Table 13** presents the reaction times in **SH** and **HD** as a function of the parameters of the games,  $\alpha$  and  $\Delta$ .

In **SH**, we found that RT were shorter for higher  $\alpha$ : shortest at  $\alpha = 0.8$  and longest at  $\alpha = 0.4$  in both conditions (*t*-tests of difference,  $p < 0.01$  in both conditions). We also found that RT were longer in high stakes than in low stakes rounds (*t*-test of difference,  $p < 0.001$  in *Control* and  $p = 0.012$  in *Stress*). The trend was identical in **HD**, with shortest RT at  $\alpha = 0.8$  and longest at  $\alpha = 0.4$  in the control group and  $\alpha = 0.2$  in the stress group (*t*-tests of difference,  $p < 0.001$  in both conditions). RT were also longest in high stakes trials (*t*-test of difference,  $p < 0.001$  in both groups). It is unclear why  $\alpha$  significantly affects reaction times in the games. In both **SH** and **HD**, increasing  $\alpha$  makes the safe option relatively more valuable. It is plausible that *Safe* becomes easier to evaluate as it becomes more attractive, resulting in a quicker response. As for stakes, we conjecture that subjects find the decision to be more important (hence, more

**TABLE 13 |** Reaction time in games as a function of  $\alpha$  and  $\Delta$ .

	$\alpha$				$\Delta$	
	0.2	0.4	0.6	0.8	High	Low
<b>SH</b>						
Control	24.1 (1.74)	28.2 (1.76)	22.7 (1.53)	21.6 (1.49)	27.7 (1.64)	20.6 (1.21)
Stress	26.6 (1.72)	30.5 (1.93)	23.8 (1.62)	21.6 (1.68)	28.5 (1.64)	22.9 (1.47)
<b>HD</b>						
Control	29.7 (1.54)	33.5 (1.86)	28.4 (1.75)	21.8 (1.31)	31.4 (1.54)	25.0 (1.23)
Stress	35.1 (1.94)	34.8 (1.93)	31.6 (2.24)	24.2 (1.74)	35.0 (1.82)	27.2 (1.36)

**TABLE 14 |** OLS of decision time in **SH** and **HD** including fixed effects.

	SH	HD
Stress	1.64 (2.00)	2.94 (2.00)
Male	-2.24 (2.00)	0.141 (2.00)
High Stakes	6.38*** (0.74)	7.02*** (0.80)
$\alpha = 0.8$	-3.69*** (1.04)	-9.08*** (1.13)
$\alpha = 0.6$	-2.02 (1.05)	-2.49* (1.14)
$\alpha = 0.4$	4.04*** (1.04)	2.18 (1.14)
Constant	22.52*** (1.91)	27.02*** (1.93)
Observations	2,260	2,227
FE groups	143	143
df	9	9
Log-likelihood	-9,810	-9,831
BIC	19,689	19,731

Standard errors in parenthesis. \* $p < 0.05$ ; \*\* $p < 0.01$ ; and \*\*\* $p \leq 0.001$ .

worthy of attention) when, other things being equal, the set of payoffs is more spread out. In any case, the consistency of the reaction time comparative statics across games and conditions is remarkable and deserves further investigation. Finally, in **SH** there was no effect of stress. In **HD**, there was an increase in RT under stress only when  $\alpha = 0.2$  ( $p = 0.030$ ) and when stakes were high ( $p = 0.015$ ), suggesting an interaction between game complexity and difficulty to evaluate alternatives. It is also consistent with studies showing that stress affects working memory and executive decision-making. High levels of cortisol have been associated with more errors in card sorting tasks meant to measure executive functioning (McCormick et al., 2007) as



well as O-span and backwards digit-span tasks meant to measure working memory (Schoofs et al., 2009). While our finding reflects the intuition behind results showing stressed subjects performing worse on more complicated tasks (Schoofs et al., 2009), our contribution shows that more complicated decisions also take longer (in our setting, there are no right or wrong decisions). This finding illustrates an important difference between how stressed subjects reach decisions in strategic games vs. in working memory or executive functioning tasks.

We then conducted a mixed effect OLS regression to better analyze the contribution of each effect to reaction times in both games. For both **SH** and **HD**, we regressed reaction times on a Treatment dummy (1 = Stress), a Gender dummy (1 = Male), a Stakes dummy (1 = High stakes), and dummies identifying the level of  $\alpha$  in each round. The results are reported in **Table 14**. They confirm the effect of high stakes and  $\alpha$  levels reported above. Stress and gender did not have significant effects.

**Result 3.** *Reaction times are higher in the conceptually more difficult game HD, in the more complex rounds of LO, when stakes are high and when the safe option is intrinsically less attractive in SH and HD. Stress (weakly) increases reaction times in those cases.*

## 6. DISCUSSION

In this paper, we examined the effect of stress on decision-making in three tasks: lotteries, Stag Hunt games, and Hawk-Dove games. Previous experiments and neuro-imaging studies suggest that people are responsive to differences in incentives across these tasks, which aligns with our control group. However, a significant portion of subjects under stress do not respond to these different incentives, which we interpret as contextual blindness.

The results contribute to our understanding of the complex relationship between stress and decision-making. In this regard, we found both conflicting and confirming evidence. Unlike some of the recent literature on lottery choice, in our study we did not find that stress had a systematic effect on any of the three tasks. However, our main finding of contextual blindness fits in well with previous work on stress inducing habituation with regard to cognitive inflexibility.

Stress-induced contextual blindness is demonstrated by a predictable pattern where subjects who choose to be relatively risk-seeking in one context also choose to be relatively risk-seeking in other, radically different ones. This predictability can be leveraged in order to reach desirable outcomes in coordination games either through directly modulating stress or by optimizing the pairing of players and games. For example, placing under stress two subjects who are risk-takers

in lotteries may encourage them to be risk-seeking in Stag Hunt, therefore promoting the payoff-dominant equilibrium outcome. Alternatively, in settings where subjects need to be paired together to play coordination games, risk-preference can serve as a guide to create optimal subject-pairings in stressful circumstances. In Stag Hunt situations, optimal pairings would combine subjects with similar risk-seeking behavior in lotteries whereas in Hawk-Dove situations, optimal pairings would combine subjects with opposite risk preferences. Practical applications include team formation in military operations with limited communication.

Finally, it is surprising to observe similar attitudes when facing another individual and a lottery draw. The extent to which contextual blindness contributes to an attributed loss of opponents' agency is unclear. Subjects under stress have been shown to treat other players as less strategic decision-makers (Leder et al., 2013), but this is different from treating them as probabilistic outcomes. Further research may disentangle how stress modulates the level of autonomy attributed to other players. It may be that stress makes humans less likely to incorporate the *intention* of an action, which would have important implications in social contexts.

## AUTHOR CONTRIBUTIONS

IB, JC, and RK contributed to all aspects of this project equally.

## ACKNOWLEDGMENTS

We are grateful to members of the Los Angeles Behavioral Economics Laboratory (LABEL) for their insights and comments in the various phases of the project. This study was carried out in accordance with the recommendations of "University of Southern California, Institutional Review Board" with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the University of Southern California, Institutional Review Board, approval number UP-12-00663. We acknowledge the financial support of the National Science Foundation grant SES-1425062 and the University of Southern California's Provost's Postdoctoral Research Scholar Grant.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnbeh.2017.00236/full#supplementary-material>

## REFERENCES

- Al-Ubaydli, O., Jones, G., and Weel, J. (2013). Patience, cognitive skill, and coordination in the repeated stag hunt. *J. Neurosci. Psychol. Econ.* 6:71. doi: 10.1037/npe0000005
- Buchanan, T. W., and Preston, S. D. (2014). Stress leads to prosocial action in immediate need situations. *Front. Behav. Neurosci.* 8:5. doi: 10.3389/fnbeh.2014.00005
- Buckert, M., Schwieren, C., Kudielka, B. M., and Fiebach, C. J. (2014). Acute stress affects risk taking but not ambiguity aversion. *Front. Neurosci.* 8:82. doi: 10.3389/fnins.2014.00082
- Büyükbayacı, M. (2014). Risk attitudes and the stag-hunt game. *Econ. Lett.* 124, 323–325. doi: 10.1016/j.econlet.2014.06.019
- Charness, G., and Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *J. Econ. Behav. Organ.* 83, 50–58. doi: 10.1016/j.jebo.2011.06.007

- Chierchia, G., and Coricelli, G. (2015). The impact of perceived similarity on tacit coordination: propensity for matching and aversion to decoupling choices. *Front. Behav. Neurosci.* 9:202. doi: 10.3389/fnbeh.2015.00202
- Crosan, R., and Gneezy, U. (2009). Gender differences in preferences. *J. Econ. Literat.* 47, 448–474. doi: 10.1257/jel.47.2.448
- Danese, G., Fernandes, E., Watson, N. V., and Zilioli, S. (2017). Testosterone and cortisol jointly predict the ambiguity premium in an ellberg-urns experiment. *Front. Behav. Neurosci.* 11:68. doi: 10.3389/fnbeh.2017.00068
- Ekins, W. G., Caceda, R., Capra, C. M., and Berns, G. S. (2013). You cannot gamble on others: dissociable systems for strategic uncertainty and risk in the brain. *J. Econ. Behav. Organ.* 94, 222–233. doi: 10.1016/j.jebo.2013.07.006
- Engelmann, J. B., Meyer, F., Ruff, C. C., and Fehr, E. (2017). The neural circuitry of emotion-induced distortions of trust. *bioRxiv* 129130. doi: 10.1101/129130
- Fraley, C., and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *J. Am. Stat. Assoc.* 97, 611–631. doi: 10.1198/016214502760047131
- Fraley, C., and Raftery, A. E. (2006). *Mclust Version 3: An R Package for Normal Mixture Modeling and Model-based Clustering*. Technical report, University of Washington.
- Gathmann, B., Schulte, F. P., Maderwald, S., Pawlikowski, M., Starcke, K., Schäfer, L. C., et al. (2014). Stress and decision making: neural correlates of the interaction between stress, executive functions, and decision making under risk. *Exp. Brain Res.* 232, 957–973. doi: 10.1007/s00221-013-3808-6
- Heinemann, F., Nagel, R., and Ockenfels, P. (2009). Measuring strategic uncertainty in coordination games. *Rev. Econ. Stud.* 76, 181–221. doi: 10.1111/j.1467-937X.2008.00512.x
- Hockey, G. R. J. (1970). Effect of loud noise on attentional selectivity. *Q. J. Exp. Psychol.* 22, 28–36. doi: 10.1080/14640747008401898
- Jessie, D., and Kendall, R. (2015). *Decomposing Models of Bounded Rationality*. Technical report, IMBS technical reports MBS 15-06.
- Kandasamy, N., Hardy, B., Page, L., Schaffner, M., Graggaber, J., Powlson, A. S., et al. (2014). Cortisol shifts financial risk preferences. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3608–3613. doi: 10.1073/pnas.1317908111
- Kassam, K. S., Koslov, K., and Mendes, W. B. (2009). Decisions under distress: stress profiles influence anchoring and adjustment. *Psychol. Sci.* 20, 1394–1399. doi: 10.1111/j.1467-9280.2009.02455.x
- Leder, J., Häusser, J. A., and Mojzisch, A. (2013). Stress and strategic decision-making in the beauty contest game. *Psychoneuroendocrinology* 38, 1503–1511. doi: 10.1016/j.psyneuen.2012.12.016
- Lighthall, N. R., Mather, M., and Gorlick, M. A. (2009). Acute stress increases sex differences in risk seeking in the balloon analogue risk task. *PLoS ONE* 4:e6002. doi: 10.1371/journal.pone.0006002
- Mather, M., and Lighthall, N. R. (2012). Risk and reward are processed differently in decisions made under stress. *Curr. Dir. Psychol. Sci.* 21, 36–41. doi: 10.1177/0963721411429452
- McCormick, C., Lewis, E., Somley, B., and Kahan, T. (2007). Individual differences in cortisol levels and performance on a test of executive function in men and women. *Physiol. Behav.* 91, 87–94. doi: 10.1016/j.physbeh.2007.01.020
- Nagel, R., Brovelli, A., Heinemann, F., and Coricelli, G. (2014). “Neural correlates of risk and strategic uncertainty,” in *Working Paper*.
- Neumann, T., and Vogt, B. (2009). “Do Players’ beliefs or risk attitudes determine the equilibrium selections in 2x2 coordination games?,” in *FEMM Working Paper*.
- Nieuwenhuis, S., Forstmann, B. U., and Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. *Nat. Neurosci.* 14, 1105–1107. doi: 10.1038/nn.2886
- Preston, S., Buchanan, T., Stansfield, R., and Bechara, A. (2007). Effects of anticipatory stress on decision making in a gambling task. *Behav. Neurosci.* 121:257. doi: 10.1037/0735-7044.121.2.257
- Samson, D., Apperly, I., Chiavarino, C., and Humphreys, G. (2004). Left temporoparietal junction is necessary for representing someone else’s belief. *Nat. Neurosci.* 7, 499–500. doi: 10.1038/nn1223
- Schoofs, D., Wolf, O. T., and Smeets, T. (2009). Cold pressor stress impairs performance on working memory tasks requiring executive functions in healthy young men. *Behav. Neurosci.* 123:1066. doi: 10.1037/a0016980
- Schwabe, L., Haddad, L., and Schachinger, H. (2008). Hpa axis activation by a socially evaluated cold-pressor test. *Psychoneuroendocrinology* 33, 890–895. doi: 10.1016/j.psyneuen.2008.03.001
- Schwabe, L., Joëls, M., Roozendaal, B., Wolf, O. T., and Oitzl, M. S. (2012). Stress effects on memory: an update and integration. *Neurosci. Biobehav. Rev.* 36, 1740–1749. doi: 10.1016/j.neubiorev.2011.07.002
- Schwabe, L., and Wolf, O. T. (2009). Stress prompts habit behavior in humans. *J. Neurosci.* 29, 7191–7198. doi: 10.1523/JNEUROSCI.0979-09.2009
- Starcke, K., and Brand, M. (2012). Decision making under stress: a selective review. *Neurosci. Biobehav. Rev.* 36, 1228–1248. doi: 10.1016/j.neubiorev.2012.02.003
- Taylor, S. E., Klein, L. C., Lewis, B. P., Gruenewald, T. L., Gurung, R. A., and Updegraff, J. A. (2000). Biobehavioral responses to stress in females: tend-and-befriend, not fight-or-flight. *Psychol. Rev.* 107:411. doi: 10.1037/0033-295X.107.3.411
- Van Den Bos, R., and Flik, G. (2015). Decision-making under stress: the importance of cortico-limbic circuits. *Front. Behav. Neurosci.* 9:203. doi: 10.3389/fnbeh.2015.00203
- Van Den Bos, R., Harteveld, M., and Stoop, H. (2009). Stress and decision-making in humans: performance is related to cortisol reactivity, albeit differently in men and women. *Psychoneuroendocrinology* 34, 1449–1458. doi: 10.1016/j.psyneuen.2009.04.016
- von Dawans, B., Fischbacher, U., Kirschbaum, C., Fehr, E., and Heinrichs, M. (2012). The social dimension of stress reactivity: acute stress increases prosocial behavior in humans. *Psychol. Sci.* 23, 651–660. doi: 10.1177/0956797611431576
- Yerkes, R. M., and Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *J. Comp. Neurol.* 18, 459–482. doi: 10.1002/cne.920180503
- Yu, R. (2016). Stress potentiates decision biases: a stress induced deliberation-to-intuition (sidi) model. *Neurobiol. Stress* 3, 83–95. doi: 10.1016/j.yynstr.2015.12.006

**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 Brocas, Carrillo and Kendall. 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.



# Digit Ratio (2D:4D) Predicts Self-Reported Measures of General Competitiveness, but Not Behavior in Economic Experiments

Werner Bönke<sup>1,2,3\*</sup>, Vivien D. Procher<sup>1,2,4</sup>, Diemo Urbig<sup>1,2</sup> and Martin Voracek<sup>5</sup>

<sup>1</sup> Jackstädt Center of Entrepreneurship and Innovation Research, University of Wuppertal, Wuppertal, Germany,

<sup>2</sup> Schumpeter School of Business and Economics, University of Wuppertal, Wuppertal, Germany, <sup>3</sup> Institute for Development Studies, School of Public and Environmental Affairs, Indiana University, Bloomington, IN, United States, <sup>4</sup> RWI-Leibniz-Institut für Wirtschaftsforschung, Essen, Germany, <sup>5</sup> Department of Basic Psychological Research and Research Methods, School of Psychology, University of Vienna, Vienna, Austria

## OPEN ACCESS

### Edited by:

Carmen Sandi,  
École Polytechnique Fédérale de  
Lausanne, Switzerland

### Reviewed by:

Leyla Loued-Khenissi,  
École Polytechnique Fédérale de  
Lausanne, Switzerland

Kai Hiraishi,

Keio University, Japan

### \*Correspondence:

Werner Bönke  
boente@wiwi.uni-wuppertal.de

**Received:** 31 August 2017

**Accepted:** 15 November 2017

**Published:** 08 December 2017

### Citation:

Bönke W, Procher VD, Urbig D and Voracek M (2017) Digit Ratio (2D:4D) Predicts Self-Reported Measures of General Competitiveness, but Not Behavior in Economic Experiments. *Front. Behav. Neurosci.* 11:238. doi: 10.3389/fnbeh.2017.00238

The ratio of index finger length to ring finger length (2D:4D) is considered to be a putative biomarker of prenatal androgen exposure (PAE), with previous research suggesting that 2D:4D is associated with human behaviors, especially sex-typical behaviors. This study empirically examines the relationship between 2D:4D and individual competitiveness, a behavioral trait that is found to be sexually dimorphic. We employ two related, but distinct, measures of competitiveness, namely behavioral measures obtained from economic experiments and psychometric self-reported measures. Our analyses are based on two independent data sets obtained from surveys and economic experiments with 461 visitors of a shopping mall (Study I) and 617 university students (Study II). The correlation between behavior in the economic experiment and digit ratios of both hands is not statistically significant in either study. In contrast, we find a negative and statistically significant relationship between psychometric self-reported measures of competitiveness and right hand digit ratios (R2D:4D) in both studies. This relationship is especially strong for younger people. Hence, this study provides some robust empirical evidence for a negative association between R2D:4D and self-reported competitiveness. We discuss potential reasons why digit ratio may relate differently to behaviors in specific economics experiments and to self-reported general competitiveness.

**Keywords:** competitiveness, competition, digit ratio, 2D:4D, prenatal androgen exposure

## INTRODUCTION

Digit ratio (2D:4D), comparing the length of the index finger to the length of the ring finger, is a sexually dimorphic trait with males displaying, on average, a lower digit ratio than females (Manning and Fink, 2008; Hönekopp and Watson, 2010). Since the mid-1990s, digit ratios have attracted research attention because evidence suggests it is related to prenatal androgen exposure (PAE) (Manning et al., 1998; Lutchmaya et al., 2004) and, hence, is often used as a noninvasive retrospective marker for PAE (Ribeiro et al., 2016). Prenatal androgen exposure, with testosterone being the most important androgen, plays an important role in the sexual differentiation of the

mammalian brain, which has an enduring influence on behavior (Lombardo et al., 2012; Auyeung et al., 2013; Hines et al., 2015; Manning et al., 2017). These organizational effects of PAE are critically important for the masculinization and sexually differentiated behaviors across the lifespan (Archer, 2006). Those human behaviors that differ by sex are especially expected to be influenced by PAE (Hines et al., 2015).

Individual competitiveness, describing an individual's general tendency to enter competitive situations (Niederle, 2017), is a behavioral trait that is often viewed as sexually dimorphic. Gender differences in individual competitiveness are gaining increasing attention, with behavioral research indicating that women are less willing than men to enter competitions (Croson and Gneezy, 2009; Niederle, 2017). Endorsing the practical relevance of competitiveness, scholars propose that the heterogeneity in sex-specific individual competitiveness may even play an important role for educational and occupational choices (Bönte and Piegeler, 2013; Buser et al., 2014; Flory et al., 2015; Reuben et al., 2015; Bönte et al., 2017b).

This study investigates the association between individual competitiveness and digit ratio (2D:4D). In doing so, we strictly focus on selection into competitive situations and do not examine individual behavior within competitions. While experimental studies on competitiveness focus on gender differences (Croson and Gneezy, 2009), we study within-sex variation of competitiveness and digit ratios. We hypothesize that individuals—men and women—with lower (more masculine) digit ratios are more likely to enter competitive situations than individuals with higher (more feminine) digit ratios.

Links between 2D:4D and other economic behaviors are empirically examined in several studies, with evidence both for and against such links (Millet, 2011; Voracek, 2011). However, to the best of our knowledge, only one study investigates the relationship between selection into competition and 2D:4D (Apicella et al., 2011). In a sample of 93 men aged 18–23, however, Apicella et al. (2011) fail to find a statistically significant correlation between digit ratios of both hands and a behavioral measure of competitiveness obtained from an economic experiment. However, they do not control for risk preferences, even though risk preferences are argued to affect behavior in such economic experiments (Niederle and Vesterlund, 2007) and are also found to be related to 2D:4D (Bönte et al., 2016; Brañas-Garza et al., 2017). These other results suggest that there should be a relationship between 2D:4D and competitiveness in settings as those studied by Apicella et al., especially due to spurious effects by risk preferences. Hence, further tests of this relationship are warranted.

Our study makes several contributions to the literature: First, while existing studies are usually based on single, and rather small, samples, we make use of two large and independent samples, including men and women of different ages, to increase validity of our findings: a general population sample consisting of 461 visitors to a shopping mall (Study I) and a student sample comprising 617 university students (Study II). Second, we employ behavioral measures of competitiveness derived from an experimental design introduced by Niederle and Vesterlund (2007), along with psychometric self-reported

measures of competitiveness (Bönte et al., 2017a). A similar approach is used by Brañas-Garza et al. (2017) to examine the relationship between experimental and a simple one-dimensional self-reported measures for risk taking and digit ratio (2D:4D). Going beyond Brañas-Garza et al. (2017), however, we follow Bönte et al. (2017a) and, by employing different psychometric measures of competitiveness, thereby account for the potential multidimensionality of individual competitiveness (Smither and Houston, 1992; Newby and Klein, 2014). Third, in our two studies, we measure digit ratios in different ways. In Study I, an electronic caliper is used to measure 2D:4D, whereas Study II employs a self-reported ruler-based measurement of 2D:4D. This allows for checking the robustness of our results with respect to finger-length measurements. Fourth, we go beyond previous studies and account for two other sex-dimorphic traits viewed as important confounds of competitiveness (Niederle and Vesterlund, 2007) and that are found to be correlated with digit ratio: risk taking (Apicella et al., 2015; Brañas-Garza et al., 2017) and confidence (Da Silva et al., 2015; Neyse et al., 2016). Including these two variables in our regression analyses allows us to check for the robustness of our results and to avoid spurious results due to related confounding effects. Fifth, we discuss the influence of age on the relationship between individual competitiveness and digit ratio, arguing and providing empirical evidence that this relationship is stronger for young people.

The rest of the paper is organized as follows. In section Conceptual Background, we present the conceptual background and discuss the potential relationship between individual competitiveness and digit ratio. In sections Method–Study I and Method–Study II, we describe the methodologies employed in Study I and Study II, respectively. In section Results–Studies I and II, we present the results of both studies. We further discuss our findings and conclude in section Discussion and Conclusions.

## CONCEPTUAL BACKGROUND

### Digit Ratio (2D:4D) and Prenatal Androgen Exposure (PAE)

Digit ratio (2D:4D) gained increased interest since Manning et al. (1998) hypothesized that it is related to PAE. Since then, the digit ratio is used in numerous scientific studies as a noninvasive retrospective biological marker for PAE (Ribeiro et al., 2016). More specifically, it is assumed that 2D:4D is negatively correlated with prenatal androgen and positively with prenatal estrogen (Manning et al., 1998, 2017).

The direct link between 2D:4D and prenatal androgen exposure in humans cannot be experimentally demonstrated since ethical constraints ban such experiments. Hence, different attempts are made to provide indirect evidence of the relationship between PAE and 2D:4D. These approaches fall into two groups: correlational studies and experiments with both non-human mammals (Manning et al., 2014) and other vertebrate classes, such as birds (Romano et al., 2005). Correlational studies and quasi-experimental studies are based on three types of evidence (cf., Brañas-Garza et al., 2017): (a) correlation



between digit ratio and sex hormones in amniotic fluid; (b) supposed androgen spillovers in zygotic twins; and (c) digit ratios of individuals with sex hormone related syndromes, like Congenital Adrenal Hyperplasia (CAH), Complete Androgen Insensitivity Syndrome (CAIS), and Klinefelter's Syndrome. The results of these studies provide some evidence for the proposed link between PAE and 2D:4D, but results are often mixed and based on small samples (see Manning et al., 2014; Brañas-Garza et al., 2017 for more detailed surveys). The most compelling evidence may come from experiments with non-human mammals that require, however, buying into the assumption that the effects of PAE on human 2D:4D are similar to those observed in experiments with non-human mammals (Manning et al., 2014). The study by Zheng and Cohn (2011), for instance, provides experimental evidence that the 2D:4D ratio is a lifelong signature of prenatal testosterone exposure. Their study shows that, "sexually dimorphic 2D:4D ratios in mice are similar to those of humans and are controlled by the relative levels of androgen and estrogen signaling in utero" (Zheng and Cohn, 2011, p. 16289). In an experiment with rats, Talarovičová et al. (2009) find that an increase in testosterone during pregnancy reduced 2D:4D in both male and female rats by increasing 4D length (i.e., digit ratio becomes more masculinized). Also experimenting with rats, Auger et al. (2013) exposed male rat fetuses to estrogenic and anti-androgenic disruptors, finding that treated rats had more feminized (higher) digit ratios when compared to a control group. Going beyond mammals, Romano et al. (2005) show that a prenatal testosterone treatment affects digit ratios in birds, too. Overall, these findings support the assumption that varying testosterone levels during embryonic life significantly and causally affects digit ratios.

Below we build on the assumption that 2D:4D, in particular the digit ratio of the *right* hand (R2D:4D), is related to PAE, in order to present potential mechanisms for the link between R2D:4D and individual competitiveness<sup>1</sup>. Although the usefulness of digit ratios as a retrospective marker of PAE is challenged in the more recent literature (Hines et al., 2015; Warrington et al., 2016), this assumption neither restricts nor invalidates our empirical analysis since we only examine whether individual competitiveness is related to digit ratio (2D:4D). The fact that the digit ratio is a sexually dimorphic trait shows that it is determined by sex related biological factors, which can be due to prenatal androgen exposure, but also due to other sex-related biological factors; various candidate genes are discussed, for instance, HOX genes<sup>2</sup>. Thus, in our empirical analysis, we choose to take an "agnostic" perspective by focusing on the relationship between digit ratio and measures of individual competitiveness.

<sup>1</sup>Previous research suggests that, in particular, the digit ratio of the right hand (R2D:4D) is significantly correlated with sex-dependent behavioral traits (Fink et al., 2004; Hampson et al., 2008).

<sup>2</sup>While HOX genes have a fundamental role in embryonic development, with the differentiation of fingers and toes influenced by HOXA and HOXD genes (Manning et al., 2003), in recent genome-wide association studies (GWAS) of 2D:4D no signal emerged that HOX genes would impact 2D:4D (Medland et al., 2010; Lawrance-Owen et al., 2013; Warrington et al., 2016).

## Prenatal Androgens, Brain Development, and Sexually Differentiated Behavior

Embryos are exposed to androgens, estrogens, and other hormones with the resulting balance of sex hormones affecting the nervous system's development. Literature in biology and neuroscience suggests that prenatal androgen exposure has organizing effects on the development of the nervous system and brain in the uterus (Phoenix et al., 1959; Goy and McEwen, 1980; Lombardo et al., 2012; for summaries see Hines, 2010; Auyeung et al., 2013). While the female fetus is exposed to different levels of androgens than the male fetus, there is also considerable variation in prenatal androgen exposure within sexes (Hines, 2010; Auyeung et al., 2013). Previous research suggests that PAE affects behavioral characteristics, such as sexually differentiated childhood behavior in girls and in boys (Auyeung et al., 2009) and some sex-related cognitive, motor, and personality characteristics (Hines, 2010). These organizational effects of PAE on brain development are critically important for the masculinization and sexually differentiated behaviors across the lifespan (Archer, 2006; Hines et al., 2015). Hence, it is expected that, in particular, those behavioral traits showing noticeable gender differences tend to be influenced by PAE and may therefore be correlated with the digit ratio.

## Individual Competitiveness and Digit Ratio

A growing body of literature examines gender differences in individual competitiveness, defined as an individual's general tendency to select into competitive environments (Bönte et al., 2017a)<sup>3</sup>. Reviewing the literature on gender differences in economic experiments, Croson and Gneezy (2009, p. 464) conclude that, "women are more reluctant than men to engage in competitive interactions." A seminal contribution in this field is the experimental study by Niederle and Vesterlund (2007), who introduce a design for measuring individual competitiveness. This experimental design provides a binary behavioral measure of competitiveness, such that participants have to perform a real

<sup>3</sup>The conceptualization of competitiveness as tendency to self-select into competitive environments should be distinguished from three alternative conceptualizations. First, it differs from individuals' responses within a competitive environment (Croson and Gneezy, 2009; Bönte et al., 2017a). For example, willingness to win might trigger individuals to increase their efforts to leverage odds of winning in response to being in a competitive environment, independent of whether or not they seek competitive environments. It also differs from individuals' tendencies to maximize own, relative to others', rewards. While individuals maximizing relative rewards are sometimes considered to be competitive individuals (e.g., van Lange et al., 1997; Fehr and Schmidt, 1999), this defining feature does not relate to the selection into competitive situations, but rather to behavior within competitive environments. Last, we distinguish individual competitiveness as selection into competitive environments from competitiveness as ability to win (physical) competitions or as (physically) best performing (e.g., Manning and Taylor, 2001; Hönekopp et al., 2006). While individuals who believe they will be more likely to win might also be more likely to enter competitions, this would not reflect a unique preference for competition, but only a preference to maximize one's expectancies. Thus it is not only expectations about winning but also individuals' willingness to take risks that might make individuals look as if they favor competitive environments (Gneezy et al., 2003). Consistent with previous research, we distinguish such beliefs and preferences, which may make individuals look like being competitive, from individual competitive preferences.

effort task and have to choose between a non-competitive piece rate payment scheme and a competitive tournament incentive scheme. Niederle and Vesterlund (2007) find that 73% of the male participants in their experiment selecting themselves into a competitive situation compared to no more than 35% of the females. As performance, risk attitudes, and confidence are themselves subject to gender differences and may also affect the observed choice, Niederle and Vesterlund statistically control for these potential confounds in subsequent regression analyses. They stress that the remaining gender difference points to gender differences in the *preference* for competition. This result is confirmed independently in a number of experimental studies that introduced minor modifications to the original design by Niederle and Vesterlund (see Niederle, 2016 for a survey).

In summary, empirical evidence suggests that individual competitiveness is a sexually dimorphic trait and might, therefore, be related to sex-related biological factors. As mentioned above, masculinization of the human brain in utero due to PAE could result in sexually differentiated behaviors later in life. If 2D:4D is a valid retrospective marker of PTE or PAE, then 2D:4D will tend to be negatively related to more masculine behavioral traits, such as the general tendency to enter competitive situations. Hence, we hypothesize that individuals—men and women—with more masculine (i.e., lower) digit ratios are more competitively inclined than individuals with more feminine (i.e., higher) digit ratios.

## Potentially Confounding Factors: Risk Attitudes and Confidence

As mentioned above, competitive preferences revealed in economic experiments may not only reflect competitiveness as a specific behavioral trait but they may also reflect other behavioral traits, such as confidence in one's abilities or risk attitudes (Niederle and Vesterlund, 2007). Empirical evidence suggests that women are more risk averse than men both in laboratory experiments and in investment decisions in the field (Croson and Gneezy, 2009). Men also tend to be more (over)confident than women (Lundeberg et al., 1994). Most experimental studies indicate that controlling for risk attitudes and confidence reduces the gender difference in selection into competition, but does not fully eliminate it (Niederle and Vesterlund, 2011; Niederle, 2016).<sup>4</sup> Moreover, there is some empirical evidence that these two sexually dimorphic confounding variables are correlated with 2D:4D. Several experimental studies investigating the relationship between risk taking and digit ratio provide mixed evidence (Apicella et al., 2015). A more recent study using a large sample ( $n = 704$ ) finds that male and female subjects with lower digit ratios tend to choose riskier lotteries in incentivized experiments, whereas the digit ratio is not associated with self-reported risk attitude (Brañas-Garza et al., 2017). In contrast, Bönte et al. (2016) and Stenstrom et al. (2011) find that digit ratio is negatively associated with self-reported risk attitudes.

<sup>4</sup>The gender difference in tournament entry in stereotypical male tasks persists after controlling for performance, confidence, and risk attitudes (Niederle, 2016). The gender gap tends to be reduced or vanishes if tasks are not male stereotyped and time constraints are removed (Shurchkov, 2012).

The empirical evidence is also mixed for the relation between confidence and 2D:4D. Dalton and Ghosal (2014) find that men with lower digit ratios are less likely to set unrealistically high performance expectations. Da Silva et al. (2015) report that low digit-ratio children (preschoolers) show more overconfidence in fine and gross motor skill tasks. Neyse et al. (2016) find that males with low digit ratios are more overconfident about their performance in a non-incentivized treatment, while males with low digit ratios are less overconfident in an incentivized treatment. In view of this evidence, we cannot fully rule out the possibility that individual competitiveness is not directly related to 2D:4D but only indirectly via its association with confidence and risk attitudes. Thus, in our empirical analysis we will control for confidence and risk attitudes, hypothesizing that 2D:4D is independently related to individual competitiveness.

## Age and Individual Competitiveness

Age might be another factor that affects the relationship between individual competitiveness and 2D:4D. While individual differences and sex differences in 2D:4D already emerge prenatally and digit ratios appear stable over lifetime (Trivers et al., 2006), there are compelling reasons to assume that the association of individuals' general willingness to enter competitive situations and 2D:4D changes across the life span. Individual competitiveness of men and women might be influenced by life experience with respect to education, occupations, and family; in other words, nurture might overwrite nature. Hence, the strength of the association between competitiveness and digit ratio may change because factors other than 2D:4D, like individual experiences, make individuals more or less competitive over the span of life<sup>5</sup>.

Although 2D:4D is stable over lifetime and not associated with adult sex hormone levels (Manning et al., 2004; Hönekopp et al., 2007), hormonal changes across the life span may also influence the relationship between 2D:4D and individual competitiveness. Prenatal testosterone's organizing effects on brain development, in adulthood, moderates the activating effects of current androgen levels (Auyeung et al., 2013; Manning et al., 2014).<sup>6</sup> Hence, it is likely that the strength of the relationship between 2D:4D and competitiveness depends on individuals' current levels of steroid hormones. Specifically, the relationship between 2D:4D and individual competitiveness—moderated by current testosterone—is expected to be stronger when individuals are young, because men's and women's levels of circulating

<sup>5</sup>There is some, but not yet replicated, evidence from a lab-in-the-field experiment conducted by Mayr et al. (2012) that competitiveness of both men and women changes with age and, specifically, displays an inverse U-shaped relationship. Moreover, Mayr et al. (2012) show that age does not notably affect the difference between genders in competitiveness throughout the life span. Using a representative data set of more than 25,000 individuals from 36 countries and a self-reported measure of competitiveness, Bönte (2015) confirms this finding, reporting that gender differences among adult men and women are hardly affected by age. It is also demonstrated that gender differences in competitiveness already exists at a young age in experimental studies focusing on samples consisting of children (Gneezy and Rustichini, 2004; Sutter and Rützler, 2010).

<sup>6</sup>Empirically supporting this view, van Honk et al. (2012) demonstrate that the negative effect of testosterone administration on cognitive empathy in the context of human bargaining behavior is boosted by high levels of PTE.

testosterone gradually decrease with age (Gray et al., 1991; Davison et al., 2005).

To sum up, it is likely that the relationship between 2D:4D and individual competitiveness can be better identified when using samples of young people, because the brain's response to activational steroid hormones decreases with age and because the individual competitiveness of younger people is less likely to be influenced by external factors not related to biology, like experience-based overwriting of individual predispositions (Bönte et al., 2016). Consequently, we hypothesize that individual competitiveness and 2D:4D are more strongly related when using samples of younger people than when using older people.

## Existing Evidence and Own Approach

To the best of our knowledge, the only study examining the relationship between individual competitiveness and digit ratio is Apicella et al. (2011). Based on a sample of 93 men aged 18–23, Apicella et al. (2011) investigate the association between an experimental measure of individuals' preferences to enter competitive situations and four hormonal variables, namely cortisol, circulating testosterone, facial masculinity, and the second-to-fourth digit ratio (2D:4D). Their experimental measure of competitiveness is adapted from Gneezy and Potters (1997): Before conducting a maze solving task, participants are asked to self-select into either a piece rate scheme or a competitive payment scheme (tournament). Apicella et al. (2011) find that the decision to select into a competitive environment is neither significantly correlated with R2D:4D (right hand) nor with L2D:4D (left hand).

Besides the above-mentioned problem that Apicella et al. (2011) do not control for important confound such as risk preferences and confidence, it can also not be ruled out that the relationship between behavioral measures obtained from economic experiments and 2D:4D is influenced by the specific experimental design (context) and, hence, tells us less about an individual's overall competitive disposition. Millet and Dewitte (2009), for instance, demonstrate the relevance of experimental context-specificity for the relationship between economic decision-making and digit ratio. They show that the relationship between 2D:4D and prosocial behavior can turn sign depending on the context, such that the effect might, on average, even disappear.

In order to address the problem that context specificity can alter the relationship between 2D:4D and individual competitiveness, we use two different approaches. First, we use two different real-effort tasks in our two independent studies, respectively. Previous research suggests that different tasks may differently affect the decision to enter competition. For instance, a stronger gender difference in competitiveness is observed if stereotypical male tasks, such as math tasks, are used (Niederle, 2016). Employing different tasks decreases the extent to which our conclusions depend on particularities of a single task. Second, we do not only use behavioral measures of competitiveness, but also self-reported psychometric measures. Following Bönte et al. (2017a), we argue that experimental measures tend to be more context-specific than psychometric scales that are based on general items. The estimated effect of 2D:4D may be stronger

if more general measures that are less influenced by a specific context are used (Bönte et al., 2016).

To increase the validity of our research, we employ two independent samples with a total of 1078 individuals, allowing us to have substantial power in each of these samples and to check whether results hold in both samples. We also statistically control for important confounding variables, that is, risk preferences and confidence.

## METHOD—STUDY I

For Study I, we obtain data from a survey combined with a lab-in-the-field experiment in a shopping mall. Having a general population sample with a large variety in age allows us to investigate the association of 2D:4D and competitiveness conditioned on participants' age.

### Sample and procedures

The survey and lab-in-the-field experiments were conducted in a shopping mall in a large German city for six days in June and October 2014. Visitors were approached and asked whether they would like to participate in a 10–15 min experiment on “decision-making behavior of adults” in return for earnings of at least €5.00. From a total of 488 responses, we exclude 10 due to missing data on finger lengths and 17 due to missing responses to the psychometric measure of competitiveness. In total, 461 responses could be analyzed, including 221 men and 240 women. The average age was 38.26 years (S.D. = 14.37), ranging from 16 to 89 years, with 21 and 58 years marking the tenth and ninetieth percentiles, respectively.

We started with a brief survey on the participant's socioeconomic background, e.g., age and gender, which serve as control variables. Moreover, participants assessed two statements concerning their own competitiveness. Next, mall visitors participated in competition games. To create a low-tech environment, the games were conducted with paper and pencil. Further adapting the experimental environment to the time-constrained shopping mall context, we focused on selection into competition under different treatments but not on effects of competition on performance or behavior within competitive environments (cf. Bönte et al., 2017a). Upon completion and just before paying the earnings from the experiment participants were asked to have measured the lengths of the index fingers (2D) and the ring fingers (4D) of both hands in exchange for another €2.00.

## Measurements

### Behavioral Measure of Competitiveness

All participants performed a task to collect points and chose the way they were paid for participation. We implemented a math task (cf., Niederle and Vesterlund, 2010) and used an implementation inspired by Mayr et al. (2012). For 30 s, participants verify up to 20 simple single-digit equations (e.g., “ $7+2+3-6 = 5$ . Is the result true or false?”). The sets of 20 mathematically equally difficult equations were randomly composed and randomly assigned. One out of two equations was wrong. A correctly verified equation added one point



and an incorrect verification subtracted one point. The task description included examples. Before starting with the actual task, participants chose between a non-competitive payment scheme, i.e., a piece-rate of €0.25 for each point of the overall score, and a competitive payment scheme, i.e., €0.50 for each point if the overall score was better than that of a randomly selected previous anonymous participant, €0 otherwise<sup>7</sup>. The behavioral measure of competitiveness is a dummy variable that is zero for participants choosing the non-competitive piece-rate payment and one for participants choosing the competitive payment scheme.

To reduce problems stemming from participants' potential tendency to be self-congruent with respect to their self-reported competitiveness and their plans for their behavior in the experiment, self-reported competitiveness scales were administered *before* participants knew the content of the experiment. Because the experiment is associated with real payoffs, we believe that behavior in the experiment and, hence, the behavioral measure of competitiveness, is less likely to be affected by earlier self-reported competitiveness than vice versa.

### Psychometric Measure of Competitiveness

To measure individual competitiveness, we use two items to assess perceived enjoyment associated with competitive situations. The first item ("I like situations in which I compete with others") is an adaptation of an item from Helmreich and Spence (1978), which is employed in large international surveys run by the European Union, i.e., the Flash Eurobarometer Entrepreneurship 2009 (Bönte and Piegeler, 2013). Replicating the response mode from the Flash Eurobarometer, participants evaluated this item on a 4-point Likert scale from 1 (strongly disagree) to 4 (strongly agree). A second item ("In career terms, I like situations in which I compete with others") was added to focus more on domains that are of substantial importance to one's professional life. Participants responded on a 7-point scale from 1 (does not apply at all) to 7 (applies strongly). As the scaling of both items varies, we converted the response to the first item to match the range of the second item. The psychometric score for individual competitiveness is the average of these two responses (sample  $\alpha = 0.77$ ).

### Digit Ratio

At the end of the experiment we asked participants, in exchange for additional money (€2), whether they would allow us to measure the lengths of their ring fingers and the index fingers of both hands. We opted for direct measurement and used an electronic caliper to measure finger lengths<sup>8</sup>.

<sup>7</sup>Methodological differences did not affect the behavioral measure of competitiveness: not the experimenter's gender [ $\chi^2_{(1)} = 0.28, p = 0.60$ ], not the day of the experiment [ $\chi^2_{(5)} = 1.71, p = 0.89$ ], not the type of another game they were exposed to [ $\chi^2_{(2)} = 0.92, p = 0.63$ ], and not whether the measurement was taken before or after this other game [ $\chi^2_{(1)} = 0.90, p = 0.34$ ].

<sup>8</sup>The two common methods used in previous research to measure digit ratio 2D:4D are the direct and the indirect approaches. While direct approach measures finger length directly on the finger, the indirect approach is based on indirectly measured fingers from photocopies or scans. Hence, we had to choose between indirect and direct measurement of finger lengths. We opted for a direct measurement of the digit ratio presuming that visitors of a shopping mall are likely to be suspicious of scanning their entire hands.

To distinguish between older and younger participants, we included an indicator that is one if the participant is older than 25 years. This cut-off reflects the 25-percentile (first quartile) of the age distribution. Exploring the effect of 2D:4D for the four age quartiles, we find that there is only a significant effect for the first quartile (see Appendix C).<sup>9</sup> Hence, and to be consistent with age ranges in our Study II, we chose to focus on the first age quartile.

As important additional control variables, we included risk preferences and confidence. To measure *risk preferences* participants responded to the statement, "In general, I am willing to take risks" on a 7-point scale from 1 (does not apply at all) to 7 (does fully apply). The item is validated by Dohmen et al. (2011), who find that the score of this general risk question is the best all-round predictor of actual risk-taking behavior and is demonstrated to be rather robust (Lönnqvist et al., 2015). In order to create a measure for *confidence*, participants were asked to report how many of 10 potential competitors would have less or an equal number of points; if they were correct they earned another 50 cents. Confidence is measured by subtracting this response from 10 and dividing the resulting score by 10, which approximates the perceived winning probability.

## METHOD-STUDY II

For Study II, we targeted students in a classroom with a survey and an embedded experiment. This study focuses on a large sample of young people, the group of people we expect to display the strongest association of 2D:4D and individual competitiveness. Going beyond Study I, and exploiting the classroom context, which allows more comprehensive measures, we included an established psychometric scale for individual competitiveness and explore to what extent different dimensions of competitiveness contribute to a correlation between 2D:4D and competitiveness. The behavioral measure of competitiveness is available only for a subsample of all participants. Furthermore, due to the classroom context and the limited time available, we could not rely on experimenters directly measuring participants' digit lengths. Therefore, we employed a self-reported ruler-based measurement of 2D:4D (Bönte et al., 2016).

### Sample and procedures

In winter-terms 2012/13 and 2014/15, we surveyed first- and second-year undergraduate students who attended economics lectures at a German university. At the beginning of the questionnaire, students were informed that their identities were not recorded to ensure confidentiality and that the data would be used solely for scientific purposes. Participants were not informed about the specific nature of the research. From a total of 886 responses, we exclude 77 with missing data of finger lengths, 33 with missing data for self-reported competitiveness, confidence, age, gender, or risk taking. Further, we excluded 86 observations with implausible or inconsistent measures of finger lengths (see below). As we want to focus on young people, we also excluded

<sup>9</sup>While Appendix C could be interpreted as perhaps indicating an inverse U-shaped moderating effect of age on the link between 2D:4D and self-reported competitiveness, none of further tests of such effects are statistically significant.



72 responses (about 8% of the total sample) from participants older than 25 years. Hence, we employed 618 observations for our analyses. Comparing the restricted (final) and unrestricted sample, we do not find statistically significant differences for our key variables<sup>10</sup>. The majority (82%) of the students were enrolled in business, economics, or related fields such as health economics. The average age was 21.6 years (S.D. = 1.72), ranging from 18 to 25 years, with 20 and 24 years as the tenth and ninetieth percentiles, respectively.

In winter term 2014/2015, we started with a classroom survey, which included questions on self-reported competitiveness, self-efficacy, and risk preferences. There were explicit instructions to wait until all participants had finished this part of the survey. Then participants were provided with a description of an economic experiment. Next all participants chose how they would behave in this experiment. Then participants generated a key that would allow the experimenter to make a random draw of 30 participants who would later participate in the experiment without making public any private information of the participants (like names). Next participants were instructed how to do the measurement of the index, middle, and ring fingers of the right hand and the left hand. After the measurement of the fingers, the participants were asked questions concerning sociodemographic factors, like age and sex. At the end, 30 randomly chosen self-generated keys were listed and these participants performed the experiment and necessary decisions were predetermined based on what they indicated in their survey. The other participants answered questions related to the content of the lecture (economic policy). In winter term 2012/13 the chronology was very similar: first the survey and then the measurement of finger lengths; however, no classroom-experiment was conducted.

## Measurements

### Behavioral Measure of Competitiveness

For a subsample of 150 students (in winter-term 2014/15), we obtained a behavioral measure of individual competitiveness derived from a classroom experiment that was embedded into the survey and related confidence measures. Although conducted in class, participation was voluntary. For the experiment, we adopted a design that is frequently used to measure competitiveness (e.g., Niederle and Vesterlund, 2007; Shurchkov, 2012). Participants had to choose between a noncompetitive compensation scheme ("piece-rate") and a competitive compensation scheme ("tournament") with respect to their performance in a real task. Specifically, participants had to answer 20 trivia questions on various areas of general knowledge within 5 minutes (questions taken from Eberlein et al., 2011). For each question, participants had to choose the one correct answer out of four given options. Before choosing the payment scheme, all participants received 4 example questions, which they were asked to solve (without any incentives) to

familiarize themselves with the task and to gain an impression of the level of difficulty. Students were informed that they could earn up to €20.00 when performing in the task. To save time, however, not all students had to participate in the real task. After the survey, we collected the paperwork with potential participants' decisions and randomly selected 30 of them. The selected students were asked to join the experimenter to perform their task. Questions were presented on a quiz sheet and could be answered in any order. No feedback was provided during the quiz. The payoffs were then paid according to their decisions and the decisions of randomly matched partners. Those participants who previously chose piece-rate, received 50 cents for every correctly answered question in the quiz. The scores of those participants who chose the tournament payment scheme during the survey, were compared to the score of another randomly matched participant<sup>11</sup>. The participant with more correct answers ("the winner") received 100 cents for every correct answer. The other participant received 0 cents. In case of a tie, the winner was determined randomly. The *behavioral measure of competitiveness* is a dummy variable that is zero for participants choosing the non-competitive piece-rate payment and one for participants choosing the competitive tournament payment.

As in Study I and for the same reasons, self-reported competitiveness scales were administered before participants knew the content of the incentivized behavioral measure of competitiveness.

### Aggregate Psychometric Measure of Competitiveness

As the first self-reported measure, we employed an adaptation of the competitiveness subscale of the Work and Family Orientation Scale (WOFO; Helmreich and Spence, 1978). This measure aggregates individuals' enjoyment of interpersonal competition but also individuals' desire to do better than others and their desire to win in interpersonal situations (Houston et al., 2002). To stay within a general context easily applicable to the sample of young students, we replaced the item "I enjoy working in situations involving competition with others" with an item that refers to a general rather than a work-specific context: "I like situations in which I compete with others." The score for this *aggregate measure of competitiveness* is calculated as the average score of responses to the five items of the competitiveness subscale of WOFO ( $\alpha = 0.77$ ).

### Enjoyment of Competition

Empirical studies using larger sets of items confirm that the scale by Helmreich and Spence (1978) does not reflect a unidimensional concept of competitiveness but comprises different dimensions of competitiveness (Houston et al., 2002; Newby and Klein, 2014). To account for the enjoyment one receives from competition, our second measure of competitiveness focuses on the enjoyment of competition.

<sup>10</sup>Behavioral measure of competitiveness (two-sample test of proportions:  $z = 0.38$ ,  $p = 0.70$ ), the two self-reported measures of individual competitiveness (HS:  $t = 0.26$ ,  $p = 0.79$ ; EC:  $t = 0.62$ ,  $p = 0.54$ ), and the right- and left-hand second to fourth digit ratio (right:  $t = 0.29$ ,  $p = 0.77$ ; left:  $t = 1.24$ ,  $p = 0.21$ ).

<sup>11</sup>As the whole study was conducted in class, all participants knew their potential competitors. The matching pool of competitors included only those participants who selected the tournament. Participants were not provided any information regarding the matched competitor.

We included the highest loading item from Newby and Klein's (2014) "general competitiveness" subscale ("*I enjoy competing against others.*") and the highest loading reverse-coded item from Smither and Houston (1992) emotion factor ("*I find competitive situations unpleasant*") (see Appendix B). Participants responded to each item on a 7-point scale from "does not apply at all" (1) to "fully applies" (7). The score for enjoyment of competition is calculated as the average scores of both items ( $\alpha = 0.71$ ).

### Aggregate Competitiveness Not Driven by Enjoyment of Competition

To better differentiate between enjoyment of competition and other dimensions of competitiveness that are captured by the aggregate measure of competitiveness, we employed a residualization technique to partition variation in the aggregate measure into two uncorrelated parts (for a similar approach see Bönte et al., 2017a), where one part is not driven by variation in enjoyment of competition. Residualization is implemented by an ordinary least squares regression where the aggregate score of the HS-Scale is the dependent variable and the aggregate score of the EC-Scale is the only explanatory variable. The measure of "competitiveness not driven by enjoyment of competition" is given by the residual plus the constant (RHS = residualized HS-scale).

### Digit Ratio

We employ a self-reported ruler-based measurement of 2D:4D. On four sheets of the questionnaire, two rulers were displayed which were arranged as a triangle, with the rulers starting with zero at the point where they met (see **Figure 1**). Students marked the length of the ring finger and the length of the middle finger (1st sheet) and then marked the length of the middle finger and length of the index finger (2nd sheet) of the right hand. The same measurement was completed for the left hand (3rd and 4th sheet). Verbal instructions were given on how to do the measurement (e.g., how to position the hand and that the tip of a finger is relevant for measurement, but not the finger nails). We obtained the 2D:4D by dividing the length of the index finger (2D) by the length of the ring finger (4D). Since it is very likely that self-reported measurement of finger length is associated with substantial measurement error, we took measures to detect and drop responses with implausible or unreliable 2D:4D measurements. We extend the measurement approach of Manning and Fink (2008) by exploiting that the middle fingers of both hands are measured twice. We excluded 78 observations where the two measurements of the same middle finger of a hand (once in conjunction with the index and then together with the ring finger) differ by more than 10%, which we interpreted as indicating a substantial lack of reliability for the individually self-measured finger lengths. This is advantageous as the judgment of reliability is based on a finger that does not form the variables of interest. Furthermore, we excluded 8 observations where the 2D:4D did not fall into the usually observed range of 0.8–1.2 (cf., Hönekopp and Watson, 2010; Bönte et al., 2017a; Manning et al., 2017). Visual inspection of the latter observations showed that these outliers tend to

be the result of errors when marking the length of fingers on rulers<sup>12</sup>.

In our regression analyses, we control for gender, risk preference, and confidence. Gender is a dummy indicating female participants, risk taking is measured by participants' agreement (from 1—"does not apply at all" to 7—"applies strongly") with the statement, "In general, I am willing to take risks." Following Bönte et al. (2017a), we measured confidence in four ways: In contrast to Study I, the data of Study II also contain a measure of general confidence (not related to the experiment), measured by participants' agreement (from 1—"does not apply at all" to 7—"applies strongly") with the statement "Generally, when facing difficult tasks, I am certain that I will accomplish them" (see Bönte and Piegeler, 2013, as an adaptation of an item from Chen et al., 2001). Given that in a specific context, participants may employ different heuristics to form beliefs about their own and others' performances when choosing to select into competitions, we include three distinct measures (cf. Bönte et al., 2017a): We asked participants to forecast their own numbers of correctly answered questions (confidence: own performance) and the average score of all other participants (confidence: average performance). Participants also estimated the percentage of other participants who correctly answered more questions than they themselves do; as in our first study, subtracting this number from 100 and dividing the resulting number by 100 provides an approximation of the estimated winning probability (confidence: winning probability).

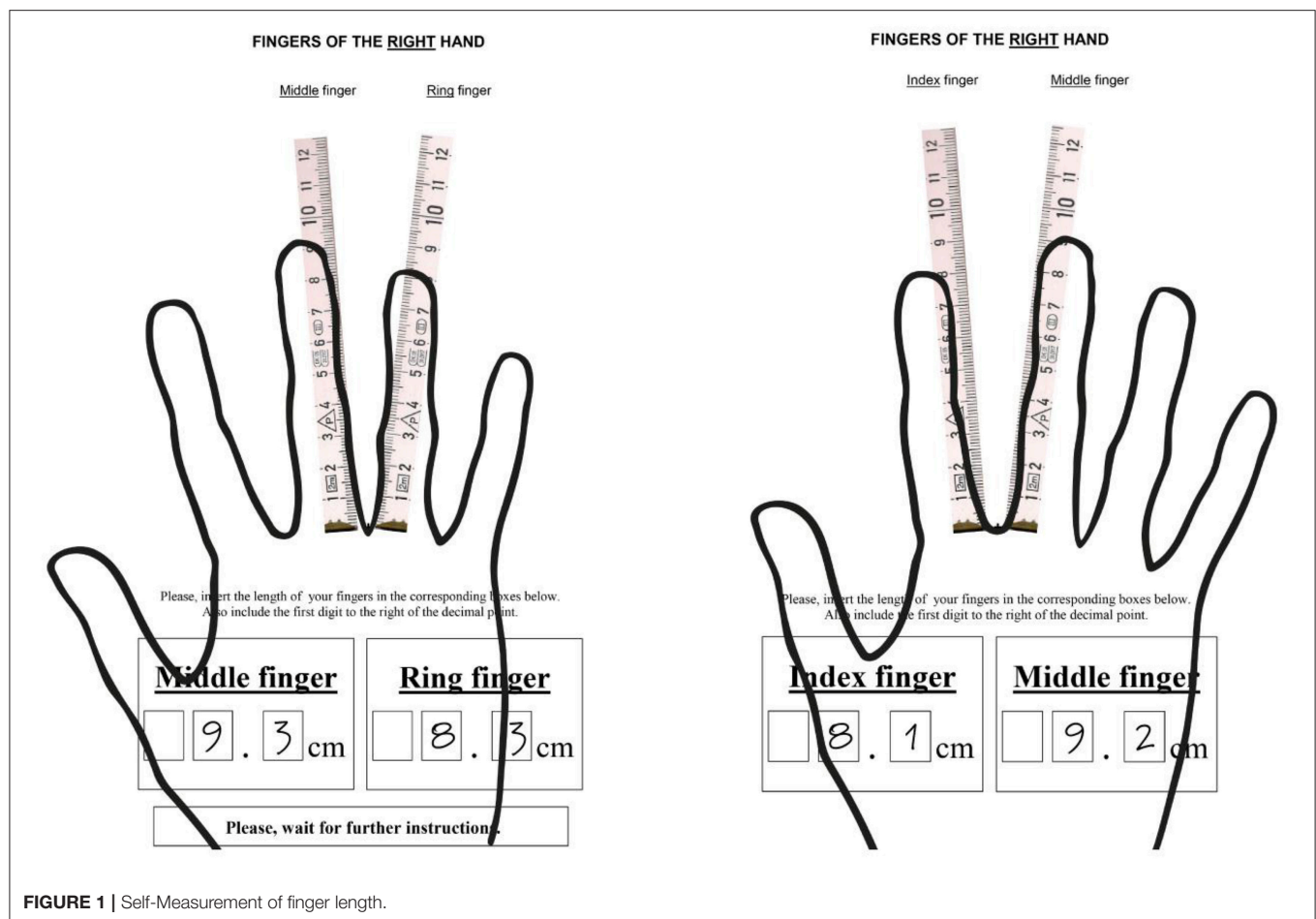
## RESULTS—STUDIES I AND II

### Replication of Stylized Facts Related to Digit Ratio and Individual Competitiveness

We first explore whether we can replicate the finding of previous research indicating that 2D:4D and individual competitiveness are sexually dimorphic. In both studies (see **Tables 1-I, II**), we find that female participants display larger 2D:4D and this effect is stronger for the right than for the left hand (Manning and Fink, 2008; Hönekopp and Watson, 2010). Calculating Cohen's  $d$  for the difference between sexes is larger for the right hand (I:  $d = 0.19$ , II:  $d = 0.42$ ) than for the left hand (I:  $d = 0.15$ , II:  $d = 0.24$ ). While for the general population sample (Study I) the values are lower, the values observed in the student sample (Study II) are not significantly different from values reported by Hönekopp and Watson (2010) for direct measurements of the right hand ( $d = 0.353$ , S.E. = 0.040) and left hand ( $d = 0.284$ , S.E. = 0.044). We further observe in Study I that 2D:4D of the right hand and the left hand do not correlate with age (see **Table 1-I**).

Our experimental and self-reported measures of individual competitiveness also replicate previous findings related to

<sup>12</sup>Note that when considering the descriptive statistics reported in **Tables 1-I, II**, we see that despite the sample means of measures of right-hand and left-hand 2D:4D are of comparable sizes in Studies I and II, the standard errors are substantially larger in Study II, which is based on the self-reported measure of 2D:4D. This observation could indicate that this measure is subject to larger measurement errors.



**FIGURE 1** | Self-Measurement of finger length.

gender differences (e.g., Croson and Gneezy, 2009). Both the behavioral measures and self-reported psychometric measures of competitiveness are negatively correlated with the female dummy variable, suggesting that men, on average, are more competitively inclined than women (see **Tables 1-I, II**).

For the general population sample (Study I) and its self-reported competitiveness, the calculated level of Cohen's  $d$  ( $d = 0.49$ ) is close to the value reported by Bönte (2015, **Table 1**) for a representative sample of German citizens ( $d = 0.41$ ). In both our studies, the behavioral measures and the self-reported measures of competitiveness are significantly correlated, suggesting that both types of measures overlap in measuring an individual's tendency to select into competitive situations (see **Tables 1-I, II**). For Study II, we see that this association is stronger for enjoyment of competition (EC) than for Helmreich and Spence's (1978) aggregate measure of competitiveness (HS) and almost absent for the residualized measure (RHS) not reflecting the variation related to enjoyment of competition. This suggests that selection into competition is not driven by the desire to win or to perform better in competitions. Our following analyses, thus, focus on the narrower measure of enjoyment of competition rather than Helmreich and Spence's multi-faceted measure.

## Correlational Analyses of the Relationships between 2D:4D and Competitiveness

Both correlation tables (**Tables 1-I, II**) show that the association of individual competitiveness with 2D:4D is generally stronger for the right hand than for the left hand. This coincides with previous studies suggesting that the right-hand 2D:4D tends to be more strongly affected by prenatal testosterone than the left-hand ratio (Lutchmaya et al., 2004; Hönekopp and Watson, 2010; Zheng and Cohn, 2011) and that significant correlations between sex-dependent behavioral traits and digit ratio are predominantly found for the right hand (Fink et al., 2004; Hampson et al., 2008).

To explore if—as we expect—the correlations between competitiveness and R2D:4D (right hand) depend on age, we also split the sample of the general population into younger (25 years or less) and older (more than 25 years) participants<sup>13</sup>. The correlation with the behavioral measure is not statistically significant for both age groups ( $\leq 25$ :  $r = -0.014$ ,  $p = 0.883$ ;  $> 25$ :  $r = -0.036$ ,  $p = 0.507$ ). However, we observe that the correlation with the self-reported measure is larger and statistically significant for younger participants, but smaller and

<sup>13</sup>Appendix C reports analyses for further splitting the group of those older than 25 years.

**TABLE 1-I |** Summary statistics and correlations (Study I).

No.	Variable	Mean	S.D.	N	Pearson correlation coefficients						
					1	2	3	4	5	6	7
COMPETITIVENESS MEASURES											
1	Behavioral	0.529	0.500	461	1						
2	Self-reported	4.350	1.534	461	0.234***	(0.77)					
DIGIT RATIOS											
3	R2D:4D (right-hand)	0.991	0.037	461	−0.030	−0.119*	1				
4	L2D:4D (left-hand)	0.989	0.038	461	−0.016	−0.031	0.506***	1			
CONTROL VARIABLES											
6	Female	0.521	0.500	461	−0.157***	−0.238***	0.095*	0.075	1		
7	Age (>25 years)	0.740	0.439	461	−0.044	−0.061	−0.007	0.001	−0.025	1	
8	Risk taking	4.735	1.428	461	0.108*	0.345***	−0.038	−0.045	−0.139***	−0.030	1
9	Confidence: Winning prob. (0-1)	0.522	0.173	461	0.301***	0.109*	0.017	−0.004	−0.285	0.048	0.016

$R(L)2D:4D = 2D:4D$  of right (left) hand. Where available, Cronbach's alpha is reported in parentheses on the diagonal.

To explore if the correlations between competitiveness and 2D:4D depend on age, we also report these correlations conditioned on the age dummy. The correlation with the behavioral measure is statistically insignificant for both age groups ( $\leq 25$ :  $r = -0.014$ ,  $p = 0.883$ ;  $> 25$ :  $r = -0.036$ ,  $p = 0.507$ ). As expected, however, we observe that the correlation with the self-reported measure is large for young and smaller and even statistically not significant for the older participants ( $\leq 25$ :  $r = -0.279$ ,  $p = 0.002$ ;  $> 25$ :  $r = -0.066$ ,  $p = 0.223$ ).

Significance levels: \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$  and \*\*\*\* $p < 0.001$ .

**TABLE 1-II |** Summary statistics and correlations (Study II).

No.	Variable	Mean	S.D.	N	Pearson correlation coefficients										
					1	2	3	4	5	6	7	8	9	10	11
COMPETITIVENESS MEASURES															
1	Behavioral measure (BM)	0.320	0.468	150	1										
2	Self-reported aggregate (HS)	4.586	1.169	618	0.239**	(0.77)									
3	Self-reported enjoyment (EC)	4.453	1.380	618	0.355***	0.585***	(0.71)								
4	Residualized aggregate (RHS)	2.378	0.948	618	0.026	0.811***	0.000	1							
DIGIT RATIOS															
5	R2D:4D (right-hand)	0.994	0.053	618	−0.013	−0.101*	−0.161***	−0.008	1						
6	L2D:4D (left-hand)	0.978	0.056	618	−0.060	−0.115**	−0.107**	−0.065	0.433***	1					
CONTROL VARIABLES															
7	Female	0.560	0.497	618	−0.435***	−0.257***	−0.318***	−0.087*	0.205***	0.116**	1				
8	Risk taking	4.635	1.407	618	0.247**	0.192***	0.267***	0.044	−0.101*	−0.060	−0.151***	1			
9	Conf.: General	4.985	1.277	618	0.192*	0.247***	0.326***	0.070	−0.078+	−0.058	−0.169***	0.260***	1		
10	Conf.: Own perf. (0–20)	10.51	3.521	150	0.361***	0.245**	0.226**	0.124	−0.008	−0.142+	−0.325***	0.253**	0.204*	1	
11	Conf.: Average perf. (0–20)	9.300	2.818	150	−0.017	0.042	0.047	0.015	−0.070	−0.074	−0.234**	−0.006	0.011	−0.350***	1
12	Conf.: Winning prob. (0–1)	0.586	0.186	150	0.222**	0.239**	0.268***	0.087	0.010	−0.109	−0.322***	0.179*	0.203*	0.453***	0.092

$R(L)2D:4D = 2D:4D$  of right (left) hand. Where available, Cronbach's alpha is reported in parentheses on the diagonal.

Significance levels: + $p < 0.10$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

not even statistically significant for older participants ( $\leq 25$ :  $r = -0.279$ ,  $p = 0.002$ ;  $> 25$ :  $r = -0.066$ ,  $p = 0.223$ ).

## Basic Regression Analyses Controlling for between Sexes Variation

Since individual competitiveness (Croson and Gneezy, 2009) and R2D:4D (Hönekopp and Watson, 2010) are sexually dimorphic, we cannot exclude the possibility that the correlation between them is only driven by the sexual dimorphism of these variables and not by variation within sexes. Therefore, we control for participants' sex in our regressions. For Study I with the general

population sample, we additionally allow the association between 2D:4D and competitiveness to depend on age. Specifically, we include a dummy variable for participants who are older than 25. In both studies, the relationships between 2D:4D with the behavioral measures were analyzed using logistic regression analyses and the relationships with the self-reported measures were analyzed using ordinary least squared regressions analyses (see **Tables 2-I, II**).

In **Tables 2-I, II**, we observe that being female is rather robustly, and independent of the measure of competitiveness, negatively associated with competitiveness. Our regression



**TABLE 2-II |** Basic regression analyses (Study II).

Model	Behavioral measure (logistic regression)		Self-reported measure (EC) (ordinary least squares regression)	
	1	2	3	4
R2D:4D	4.207 (3.749)		−2.597* (1.010)	
L2D:4D		0.195 (3.542)		−1.754+ (0.950)
Female	−2.091*** (0.412)	−1.970*** (0.394)	−0.827*** (0.108)	−0.860*** (0.107)
Constant	−3.796 (3.674)	0.129 (3.443)	7.497*** (0.995)	6.649*** (0.926)
Observations	150	150	618	618
Fit index LL/R <sup>2</sup>	−79.24***	−79.88***	0.111***	0.106***
Fit statistic ( $\chi^2/F$ )	(29.57)	(28.30)	(38.32)	(36.54)

R(L)2D:4D = 2D:4D of right (left) hand. Table reports estimated coefficients and standard errors (in parentheses).

Significance levels: + $p < 0.10$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

analyses consistently demonstrate that the relationships of digit ratios of the right (R2D:4D) and the left (L2D:4D) hand with the behavioral measure of competitiveness are negligibly small and statistically insignificant. However, we consistently observe—across both samples—a negative relationship of the right-hand digit ratio (R2D:4D) with the self-reported measures of competitiveness. For Study I, we observe that this relationship is significantly weaker for the older participants. In fact, calculating the effect for the older participants, we observe that it is statistically not significant (Table 2-I, Model 6:  $-9.862 + 7.831 = -2.031$ , S.E. = 2.198,  $p = 0.356$ ).

## Controlling for Important Confounding Effects

In a next step, we go beyond existing research (Apicella et al., 2011) by taking into account and controlling for risk preferences and confidence. Thereby we can rule out that the omission of these important variables creates spurious correlations between self-reported competitiveness and 2D:4D or suppresses correlations between 2D:4D and the behavioral measure of competitiveness. As explained in section 2, individuals' risk preferences and confidences may influence individuals' decisions to select into competition (Niederle and Vesterlund, 2007), with existing research suggesting that 2D:4D is related to individuals' risk preferences (Apicella et al., 2015; Brañas-Garza et al., 2017) as well as individuals' confidence (Da Silva et al., 2015; Neyse et al., 2016). Therefore, we perform regressions where we also include measures for risk preferences and confidences (see Tables 3-I, II, Models 1, 2, 5, and 6).

We observe rather consistently across the different models that risk taking and confidence affect competitiveness. Risk preferences are positively associated with competitiveness, though it misses statistical significance for the behavioral measure in Study II. With one exception, in each analysis at least one measure of confidence tends to be positively associated with competitiveness. Only in Study I, where we

only have a context-specific measure of confidence available, we do not observe a statistically significant association with self-reported competitiveness (see Table 3-I, Models 5 and 6). This lack of a relationship between specific confidence and general competitiveness may result from violations of the compatibility principle suggesting that predictors and criterion should be specified at the same level of specificity (cf. Ajzen and Fishbein, 2005; Bönte et al., 2017a). Observing that in Study II, the context-specific measures are not, but the general confidence measure is, related to the general self-reported measure of competitiveness supports this reasoning.

Regarding our main explanatory variables, we still do not observe relationships of 2D:4D with the behavioral measure of competitiveness; hence, the confounding effects do not suppress relationships of 2D:4D with behavioral measures of competitiveness. For self-reported measures of competitiveness, we observe that relationships with 2D:4D remain robust for the right hand. For Study II, the relationships with right-hand and left-hand 2D:4D become smaller and, for the left-hand, it does not even reach conventional levels of statistical significance.

As the hand preference displays interactions with effects of 2D:4D (Manning and Peters, 2009), our estimations may be biased, possibly underestimating the effect of 2D:4D. Hence, we complement our analyses with estimations excluding those participants who indicated having a preference for the left-hand (Tables 3-I, II, Models 3, 4, 7, and 8). Our results do not change substantially. While previously not significant effects of 2D:4D on behavioral competitiveness still do not reach any meaningful level of statistical significance, previously significant effects on self-reported competitiveness remain statistically significant.

While the comparison between behavioral and self-reported measures of competitiveness are based on the same sample in Study I, in Study II, the behavioral measure is only available for a subsample of those for whom we have the behavioral measure available. Differences in statistical significance may, hence, result from sample differences. As an additional robustness check, we therefore also estimated the effect on the self-reported measure on the same subsample (see Table 3-II, Model 9 compared with Model 3). We see that the significant results still hold, although on a substantially weaker level; hence, the difference we observe between behavioral and self-reported measures of competitiveness—as in Study I—should not be attributed to sample differences and, particularly, not to the smaller samples size.

As a last more exploratory analysis, we acknowledge that the effects of digit ratios might be gender-specific, such that the relationships differ for men and women. Our estimations testing the gender differences based on an interaction with a gender contrast code, which are reported in Appendix A, however, do not point to gender differences.

## DISCUSSION AND CONCLUSIONS

To investigate the association between individual competitiveness and digit ratio (2D:4D), this study employs two independent samples with a total of 1078 individuals. While Study I is based on a general population sample (461 visitors at a shopping mall), Study II is based on a student sample (618

**TABLE 2-I |** Basic regression analyses (Study I).

Model	Behavioral measure (logistic regression)				Self-reported measure (ordinary least squares regression)			
	1	2	3	4	5	6	7	8
R2D:4D	−0.846 (2.594)	0.417 (5.079)			−4.064* (1.899)	−9.862** (3.686)		
L2D:4D			−0.252 (2.523)	−4.775 (5.268)			−0.521 (1.859)	−0.813 (3.794)
Age (>25 years)		1.483 (5.846)		−6.080 (5.948)		−7.996+ (4.247)		−0.623 (4.303)
R2D:4D × Age		−1.726 (5.895)				7.831+ (4.283)		
L2D:4D × Age				5.915 (6.006)				0.393 (4.347)
Female	−0.629*** (0.190)	−0.637*** (0.191)	−0.633*** (0.190)	−0.644*** (0.191)	−0.702*** (0.139)	−0.702*** (0.139)	−0.728*** (0.140)	−0.733*** (0.140)
Constant	1.286 (2.565)	0.207 (5.032)	0.700 (2.492)	5.350 (5.218)	8.743*** (1.878)	14.664*** (3.652)	5.245** (1.836)	5.710 (3.754)
Observations	461	461	461	461	461	461	461	461
Fit index LL/R <sup>2</sup>	−313.00**	−312.40*	−313.09**	−312.00**	0.066***	0.077***	0.057***	0.061***
Fit statistic ( $\chi^2/F$ )	(11.50)	(12.70)	(11.40)	(13.49)	(16.21)	(9.57)	(13.83)	(7.47)

*R(L)2D:4D = 2D:4D of right (left) hand. Table reports estimated coefficients and standard errors (in parentheses). The effect of R2D:4D on self-reported competitiveness for the older participants (>25 years) is  $-9.862 + 7.831 = -2.031$  with S.E. = 2.198 and  $p = 0.356$ .*

*Significance levels: + $p < 0.10$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .*

students at a university). We use these two independent samples to replicate and validate our findings. Moreover, individual competitiveness is measured in two different ways: by behavioral measures obtained from incentivized behavioral experiments and by self-reported psychometric measures.

The results of both studies suggest that the associations between behavioral measures of competitiveness and digit ratios are not statistically significant. This confirms, using a much larger sample and including men and women, the finding reported by Apicella et al. (2011) for a small sample of 93 young men. Moreover, although we use two different real effort tasks in the incentivized experiments in Study I (math task) and in Study II (quiz task), the results are not affected by these task differences.

In contrast to our results regarding the behavioral measure, we find a negative and statistically significant relationship between psychometric measures and 2D:4D in both studies. Our specific findings suggest that psychometric scales reflecting enjoyment of competition are significantly related to the right-hand digit ratio (R2D:4D). The results remain robust when applying slightly different psychometrics scales reflecting individuals' perceived enjoyment of competition. In Study II, we additionally used a seven-item scale introduced by Helmreich and Spence (1978) that also reflects individuals' desire to perform better than others and their desire to win in interpersonal competitions (Houston et al., 2002). Following Bönte et al. (2017a), we employ a residualization technique to identify the part of the HS-scale that is not driven by variations in enjoyment of competition. Our estimation results show that R2D:4D is not significantly correlated with the residual part that reflects variations in the desire to perform better and to win against others. Hence, our

results imply that the digit ratio is, first and foremost, related to enjoyment of competition, suggesting that individuals with low (more masculine) digit ratios tend to select into competition not primarily for winning a competition but for the sake of competition itself.

Previous research shows that statistically significant associations between sex-dependent behavioral traits and digit ratio are predominantly found for the right hand (Fink et al., 2004; Hampson et al., 2008). Our observation that the left-hand digit ratio is either not or more weakly associated with competitiveness than the right-hand digit ratio confirms this finding. Our theoretical consideration indicate that it is important to additionally control for potentially confounding variables, namely individuals' confidence and risk attitudes (Niederle and Vesterlund, 2007), which tend to be related to both digit ratio (2D:4D) and selection into competition. Our results show that while the estimated effect is robust for the right-hand digit ratio (R2D:4D) in both studies, it is not for the left-hand digit ratio (L2D:4D). More specifically, the estimated coefficient is still statistically significant for R2D:4D even when controlling for individuals' confidence and risk attitudes. In contrast, the estimated coefficient of L2D:4D becomes statistically insignificant in Study II. This result provides further evidence that sex-dependent behaviors, like individual competitiveness, are predominantly associated with the right-hand digit ratio (R2D:4D).

Moreover, our exploratory analyses indicate that the strength of the relationship between digit ratio and individual competitiveness tends to depend on age. Based on a general population sample, we find that the relationship between

**TABLE 3-I |** Regression analyses controlling for important confounding variables (Study I).

Model	Behavioral measure (logistic regression)				Self-reported measure (ordinary least squares regression)			
	1	2	3	4	5	6	7	8
R2D:4D	0.884 (5.294)		2.236 (5.532)		−9.602** (3.489)		−9.027* (3.649)	
L2D:4D		−4.572 (5.537)		−2.087 (5.774)		−1.617 (3.595)		−1.183 (3.771)
Age (> 25 years)	3.007 (6.116)	−5.686 (6.242)	2.772 (6.393)	−3.161 (6.489)	−7.927* (4.021)	−2.172 (4.084)	−7.132+ (4.208)	−0.852 (4.264)
R2D:4D × Age	−3.301 (6.168)		−3.169 (6.443)		7.790+ (4.055)		6.956 (4.241)	
L2D:4D × Age		5.483 (6.306)		2.824 (6.561)		1.989 (4.127)		0.628 (4.312)
Risk taking	0.149* (0.071)	0.153* (0.071)	0.164* (0.074)	0.166* (0.074)	0.339*** (0.046)	0.342*** (0.047)	0.356*** (0.048)	0.357*** (0.049)
Conf.: Wining Prob.	0.376*** (0.066)	0.371*** (0.066)	0.358*** (0.068)	0.354*** (0.068)	0.051 (0.039)	0.048 (0.040)	0.027 (0.041)	0.027 (0.041)
Female	−0.262 (0.207)	−0.272 (0.207)	−0.297 (0.218)	−0.299 (0.217)	−0.518*** (0.138)	−0.553*** (0.139)	−0.595*** (0.145)	−0.629*** (0.146)
Constant	−3.082 (5.287)	2.326 (5.495)	−4.307 (5.535)	−0.014 (5.737)	12.418*** (3.474)	4.516 (3.570)	11.949** (3.638)	4.181 (3.746)
Observations	461	461	418	418	461	461	418	418
Fit index LL/R <sup>2</sup>	−292.08***	−292.00***	−264.81***	−264.84***	0.177***	0.163***	0.196***	0.182***
Fit statistic ( $\chi^2/F$ )	(53.34)	(53.51)	(48.47)	(48.42)	(16.29)	(14.69)	(16.66)	(15.27)

*R(L)2D:4D = 2D:4D of right (left) hand. Table reports estimated coefficients and standard errors (in parentheses). Models 3, 4, 7, and 8 exclude those participants who indicated that their dominant hand is the left hand.*

*Significance levels: + $p < 0.10$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .*

individual competitiveness and the right-hand digit ratio (R2D:4D) is stronger for younger people (age  $\leq 25$ ). This might be explained by the fact that competitive preferences of younger people are less likely to be influenced by external factors not related to digit ratios (e.g., experiences in education, jobs, and family). Moreover, the relationship between individual competitiveness and the digit ratio may be stronger for young people because the average level of circulating testosterone is higher in younger people, males (Gray et al., 1991) and females (Davison et al., 2005) and the strength of this relationship might be positively moderated by the level of circulating testosterone (van Honk et al., 2012). Hence, future research might consider that the effects of digit ratio (2D:4D) on individual competitiveness and other sexually dimorphic behaviors are moderated by both age and, possibly, circulating testosterone.

Our finding that the digit ratio (R2D:4D) is associated with the self-reported psychometric measures of competitiveness but not with the behavioral measures deserves a more detailed discussion. On the one hand, a significant association between R2D:4D and self-reported enjoyment of competition might be spurious due to confounding effects related to self-reported measures. While we already go beyond previous studies by controlling for risk taking and confidence as the most important confounding variables, there might be other more subtle confounding effects. If participants, despite anonymization, want to display specific

characteristics, then the significant association might indicate that individuals with low R2D:4D want to display enjoyment with competition. While this could theoretically be the case, controlling for risk taking and confidence and not identifying a related effect for the HS-scale, which includes an individual's declared wish to perform better than others and their willingness to win, any potentially confounding effect must be rather specific to self-reported enjoyment of competition.

On the other hand, and as a more substantive explanation for the asymmetric effect, one could argue that in economic experiments, participants have to make decisions in very specific experimental settings and empirical evidence suggests that, for instance, variation in the type of real effort tasks influences an individual's decision to select into competition (Niederle, 2016). Moreover, the results reported by Millet and Dewitte (2009) show that context in experiments can affect the relationship between behavior in experiments and the digit ratio. Although employing two different real effort tasks, performing a classroom and a lab-in-the-field experiment, and make use of a student and a general population sample, the finding of both an insignificant relation between the digit ratio and behavioral measures as well as a significant relationship between the digit ratio and self-reported measures of competitiveness is robust with respect to different contexts and samples.

Our finding that the digit ratio is significantly correlated with the self-reported measures of competitiveness but not with the

**TABLE 3-II |** Regression analyses controlling for important confounding variables (Study II).

Model	Behavioral measure (logistic regression)				Self-reported measure (EC) (ordinary least squares regression)				
	1	2	3	4	5	6	7	8	9
R2D:4D	3.122 (3.845)		2.625 (4.165)		−1.998* (0.958)		−1.994* (0.978)		−3.580+ (2.069)
L2D:4D		1.296 (3.928)		5.347 (4.428)		−1.352 (0.898)		−1.356 (0.937)	
Risk taking	0.156 (0.176)	0.155 (0.174)	0.222 (0.189)	0.229 (0.189)	0.158*** (0.037)	0.161*** (0.037)	0.172*** (0.038)	0.175*** (0.038)	0.088 (0.089)
Conf: General	0.130 (0.178)	0.142 (0.176)	0.169 (0.189)	0.165 (0.187)	0.256*** (0.041)	0.257*** (0.041)	0.261*** (0.041)	0.263*** (0.041)	0.173+ (0.094)
Conf: Own perf.	0.180* (0.080)	0.182* (0.080)	0.165+ (0.086)	0.177* (0.087)					0.005 (0.043)
Conf: Average perf.	−0.026 (0.088)	−0.023 (0.088)	−0.089 (0.094)	−0.087 (0.094)					−0.015 (0.048)
Conf: Winning prob.	−0.310 (1.359)	−0.192 (1.342)	−1.006 (1.543)	−0.909 (1.510)					0.927 (0.756)
Female	−1.701*** (0.493)	−1.612*** (0.474)	−1.924*** (0.548)	−1.911*** (0.533)	−0.662*** (0.104)	−0.686*** (0.103)	−0.658*** (0.106)	−0.686*** (0.105)	−0.659* (0.281)
Constant	−5.924 (4.011)	−4.314 (4.332)	−4.666 (4.260)	−7.506 (4.771)	4.800*** (0.991)	4.132*** (0.920)	4.721*** (1.012)	4.056*** (0.955)	6.701** (2.152)
Observations	150	150	131	131	618	618	581	581	131
Fit index LL/R <sup>2</sup>	−73.27***	−73.55***	−64.42***	63.88***	0.208***	0.206***	0.213***	0.210***	0.193***
Fit statistic ( $\chi^2/F$ )	(41.52)	(40.97)	(33.98)	(35.05)	(40.32)	(39.66)	(39.03)	(38.38)	(4.20)

R(L)2D:4D = 2D:4D of right (left) hand. Table reports estimated coefficients and standard errors (in parentheses). Models 3, 4, 7, and 8 exclude those participants who indicated that their dominant hand is the left hand. Model 9 additionally excludes participants for whom the behavioral measure of competitiveness is not available.

Significance levels: + $p < 0.10$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

behavioral measures does not imply, however, that self-reported measures are, *per se*, more strongly correlated with the digit ratio. Rather, our results, especially Study II, show that it is important to understand the factors driving the correlations between different measures of competitiveness and the digit ratio. Study II shows that those elements of competitiveness that are not related to enjoyment of competition, e.g., the desire to perform better and to win against others, are neither significantly correlated with the behavioral measure nor with the digit ratio. Consequently, these facets of competitiveness do not seem to explain the observed patterns of correlation between different measures of competitiveness and the digit ratio. Hence, psychometric scales that do not focus on enjoyment of competition may lead to different conclusions regarding the relationship between competitiveness and digit ratios.

Follow-up studies could more comprehensively examine the different facets of competitiveness by employing behavioral measures and psychometric measures of competitiveness reflecting more facets of competitiveness. Since our findings suggest that the digit ratio is related to enjoyment of competition, we would expect that significant correlations between digit ratio and behavioral measures might be found if the latter is obtained from experimental designs that provide more opportunities for enjoyment of competition. Moreover, future research could examine the potential role of moderators for selection into competition. Moderating variables may also explain

seemingly conflicting findings related to the relationship between hormones and behavior. Existing studies suggest, for instance, that interactions between hormones and contextual cues affect individuals' decisions to cooperate (e.g., Sanchez-Pages and Turiegano, 2010; Millet, 2011; Declerck et al., 2014). However, the decision to cooperate in environments characterized by elements of competition is better classified as behavior *within* competition rather than individuals' tendencies to *select into* competitive environments (Bönte et al., 2017a). Future research related to contextual cues might also more thoroughly build on demonstrated differences induced by specific cultural environments (e.g., Gneezy et al., 2009; Cárdenas et al., 2012).

Examining different behavioral and experimental measures might also be a fruitful approach for empirical studies investigating relationships between the digit ratio and other sex-dependent behaviors. For example, Brañas-Garza et al. (2017) report that their experimental measure of risk taking is significantly correlated with the digit ratios of both hands, whereas the correlation between their self-reported (single item) measure of risk taking and the digit ratio is statistically insignificant. As outlined above, the results reported by Brañas-Garza et al. also do not imply that experimental measures of risk taking are, *per se*, more strongly correlated with digit ratio than self-reported measures. Their single-item measure might be confounded by facets of risk taking that are, generally or in their specific context, not related to the digit ratio. In sum, and as



already demonstrated by Bönte et al. (2017a), combining various experimental measures with different self-reported measures of competitiveness allows for a better understanding of the facets of competitiveness that are reflected by behavioral and psychometric measures and our study suggests that this approach is also useful for investigating the relation between the digit ratio and sex-dependent behaviors, like individual competitiveness.

It is a limitation of our study that we do not fully understand the causal links between digit ratio and individual competitiveness. While we discuss a potential link through prenatal testosterone exposure as well as indirect links via risk taking and confidence, there might be other sexually dimorphic behavioral traits that could be related to selection into competition or behavior in competition and that are also correlated with the digit ratio; candidates could be aggressiveness and sensation-seeking (Hampson et al., 2008). The potential causal link between competitiveness and digit ratio that we present is based on the assumption that 2D:4D is a proxy for PAE, which influences individual competitiveness through its effect on the masculinization of the brain. While the validity of 2D:4D as marker for PAE is supported by a number of studies (e.g., Manning et al., 1998; Manning, 2002; Lutchmaya et al., 2004; McIntyre et al., 2006; Hönekopp and Watson, 2010), the usefulness of 2D:4D as a proxy for PAE is also challenged in the literature. It is argued that the link between finger ratios and PAE appears too weak or absent (Hines et al., 2015; Warrington et al., 2016) and 2D:4D might be affected by other factors than PAE (cf. Medland et al., 2010; Dressler and Voracek, 2011). In any case, our results indicate that individual competitiveness is related to a sexually dimorphic biological trait, namely 2D:4D.

Another relevant limitation of our study is the measurement error that is introduced by our measurements of 2D:4D. In previous studies, numerous methods are used to measure 2D:4D and the ongoing debate about the reliability of different approaches has not yet reached consensus (e.g., Allaway et al., 2009; Ribeiro et al., 2016). We use two different measurement approaches. In Study I, the finger lengths were measured with an electronic caliper and a self-reported ruler-based measurement of 2D:4D was used in Study II. In particular, the reliability of self-measured finger lengths is an issue (Hönekopp and Watson, 2010). To address this problem, we eliminate unreliable observations by extending the measurement method of Manning and Fink (2008). Specifically, middle finger length is measured twice for each hand (once in conjunction with the index finger, then again with the ring finger), which allows us to exclude observations where the two measurements for the middle finger strongly differ. While this approach helps to increase the reliability, we still find that the standard error of the digit ratio (R2D:4D) in Study II (0.053) is somewhat higher than in Study I (0.037), while the mean value is very similar in Study I (0.991),

and Study II (0.994). These potential measurement errors in our two measures tend to result in a downward (attenuation) bias of estimated effect sizes. Consequently, the estimated effect sizes of R2D:4D in both studies, and particularly in Study II, may only represent the lower bound of the true effect size.

To conclude, our study provides empirical evidence for a negative association between right-hand digit ratio (R2D:4D) and individual competitiveness, while identifying age as an important moderator. We hope that our work stimulates future research that further elaborates on the role that biological factors play for selection into competition, thereby searching for causal explanations that may guide and improve empirical research in this field.

## ETHICS STATEMENT

We followed standard rules for Germany, in general, and for our university, in particular, which do not require participants' signatures, but allow an informed consent implied by behavior. That is, directly at the beginning of the survey or experiment, participants were informed in writing about the content of the survey or experiment as well as about how the data would be used, i.e., for demonstration in the following teaching within this course (for parts of the survey) and for scientific research (all data). They were explicitly informed that participation was anonymous and fully voluntary; that is they could decide to not participate and also to stop participation at any point of the survey or experiment. For Study II, which was composed of a survey and a separated experiment for a subsample, the information was repeated before they decided to participate in the experiment that accompanied the survey.

## AUTHOR CONTRIBUTIONS

WB, VP, and DU: Contributed substantially to the conception and design of the work, the analysis, and interpretation of data for the work; drafted the work and revisited it critically for important intellectual content; and approved the version to be published and agrees to be accountable for all aspects of the work. MV: Contributed substantially to the interpretation of data for the work; revisited the work critically for important intellectual content; and approves the version to be published and agrees to be accountable for all aspects of the work.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnbeh.2017.00238/full#supplementary-material>

## REFERENCES

Ajzen, I., and Fishbein, M. (2005). The influence of attitudes on behavior. *Handb. Attit.* 173:31.

Allaway, H. C., Blooski, T. G., Pierson, R. A., and Lujan, M. E. (2009). Digit ratios (2D: 4D) determined by computer-assisted analysis are more reliable than those using physical measurements, photocopies, and printed scans. *Am. J. Hum. Biol.* 21, 365–370. doi: 10.1002/ajhb.20892

- Apicella, C. L., Carré, J. M., and Dreber, A. (2015). Testosterone and economic risk taking. *Adapt. Hum. Behav. Physiol.* 4, 358–385. doi: 10.1007/s40750-014-0020-2
- Apicella, C. L., Dreber, A., Gray, P. B., Hoffman, M., Little, A. C., and Benjamin, B. C. (2011). Androgens and competitiveness in men. *J. Neurosci. Psychol. Econ.* 4, 54–62. doi: 10.1037/a0021979
- Archer, J. (2006). Testosterone and human aggression: an evaluation of the challenge hypothesis. *Neurosci. Biobehav. Rev.* 30, 319–345. doi: 10.1016/j.neubiorev.2004.12.007
- Auger, J., Le Denmat, D., Berges, R., Doridot, L., Salmon, B., Canivenc-Lavier, M. C., et al. (2013). Environmental levels of oestrogenic and antiandrogenic compounds feminize digit ratios in male rats and their unexposed male progeny. *Proc. R. Soc. Lond. B Biol. Sci.* 280:20131532. doi: 10.1098/rspb.2013.1532
- Auyeung, B., Baron-Cohen, S., Ashwin, E., Knickmeyer, R., Taylor, K., Hackett, G., et al. (2009). Fetal testosterone predicts sexually differentiated childhood behavior in girls and in boys. *Psychol. Sci.* 20, 144–148. doi: 10.1111/j.1467-9280.2009.02279.x
- Auyeung, B., Lombardo, M. V., and Baron-Cohen, S. (2013). Prenatal and postnatal hormone effects on the human brain and cognition. *Eur. J. Physiol.* 465, 557–571. doi: 10.1007/s00424-013-1268-2
- Bönte, W. (2015). Gender Differences in competitive preferences: new cross-country empirical evidence. *Appl. Econ. Lett.* 22, 71–75. doi: 10.1080/13504851.2014.927560
- Bönte, W., and Piegeler, M. (2013). Gender gap in latent and nascent entrepreneurship: driven by competitiveness. *Small Busin. Econ.* 41, 961–987. doi: 10.1007/s11187-012-9459-3
- Bönte, W., Lombardo, S., and Urbig, D. (2017a). Economics meets psychology: experimental and self-reported measures of individual competitiveness. *Person. Individ. Differ.* 116, 179–185. doi: 10.1016/j.paid.2017.04.036
- Bönte, W., Procher, V., and Urbig, D. (2016). Biology and selection into entrepreneurship: the relevance of prenatal testosterone exposure. *Entrepreneurship Theory Pract.* 50, 1121–1148. doi: 10.1111/etap.12165
- Bönte, W., Procher, V., and Urbig, D. (2017b). Gender differences in selection into self-competition. *Appl. Econ. Lett.* doi: 10.1080/13504851.2017.1343441. [Epub ahead of print].
- Brañas-Garza, P., Galizzi, M., and Nieboer, J. (2017). Experimental and self-reported measures of risk taking and digit ratio (2D:4D): evidence from a large, systematic study. *Int. Econ. Rev.* Available on line at: <http://eprints.lse.ac.uk/80747/>
- Buser, T., Niederle, M., and Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *Q. J. Econ.* 129, 1409–1447. doi: 10.1093/qje/qju009
- Cárdenas, J. C., Dreber, A., Von Essen, E., and Ranehill, E. (2012). Gender differences in competitiveness and risk taking: comparing children in Colombia and Sweden. *J. Econ. Behav. Organ.* 83, 11–23. doi: 10.1016/j.jebo.2011.06.008
- Chen, G., Gully, S. M., and Eden, D. (2001). Validation of a new general self-efficacy scale. *Organ. Res. Methods* 4, 62–83. doi: 10.1177/109442810141004
- Croson, R., and Gneezy, U. (2009). Gender differences in preferences. *J. Econ. Lit.* 47, 448–474. doi: 10.1257/jel.47.2.448
- Da Silva, S., Moreira, B., and Da Costa, N. Jr. (2015). Handedness and digit ratio predict overconfidence in cognitive and motor skill tasks in a sample of preschoolers. *Economics Bull.* 35, 1087–1097.
- Dalton, P. S., and Ghosal, S. (2014). *Self-Confidence, Overconfidence and Prenatal Testosterone Exposure: Evidence from the Lab*. Tilburg: Center Discussion Paper Series No. 2014-014.
- Davison, S. L., Bell, R., Donath, S., Montalto, J. G., and Davis, S. R. (2005). Androgen levels in adult females: changes with age, menopause, and oophorectomy. *J. Clin. Endocrinol. Metab.* 90, 3847–3853. doi: 10.1210/jc.2005-0212
- Declerck, C. H., Boone, C., and Kiyonari, T. (2014). The effect of oxytocin on cooperation in a prisoner's dilemma depends on the social context and a person's social value orientation. *Soc. Cogn. Affect. Neurosci.* 9, 802–809. doi: 10.1093/scan/nst040
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, S., and Wagner, G. G. (2011). Individual risk attitudes: measurement, determinants, and behavioral consequences. *J. Eur. Econ. Assoc.* 9, 522–550. doi: 10.1111/j.1542-4774.2011.01015.x
- Dressler, S. G., and Voracek, M. (2011). No association between two candidate markers of prenatal sex hormones: digit ratios (2D:4D and Other) and finger-ridge counts. *Dev. Psychobiol.* 53, 69–78. doi: 10.1002/dev.20488
- Eberlein, M., Ludwig, S., and Nafziger, J. (2011). The effects of feedback on self-assessment. *Bull. Econ. Res.* 63, 177–199. doi: 10.1111/j.1467-8586.2009.00345.x
- Fehr, E., and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Q. J. Econ.* 114, 817–868. doi: 10.1162/003355399556151
- Fink, B., Manning, J. T., and Neave, N. (2004). Second to fourth digit ratio and the “Big Five” personality factors. *Pers. Individ. Dif.* 37, 495–503. doi: 10.1016/j.paid.2003.09.018
- Flory, J. A., Leibbrandt, A., and List, J. A. (2015). Do competitive workplaces deter female workers? A large-scale natural field experiment on job entry decisions. *Rev. Econ. Stud.* 82, 122–155. doi: 10.1093/restud/rdu030
- Gneezy, U., and Potters, J. (1997). An experiment on risk taking and evaluation periods. *Q. J. Econ.* 112, 631–645. doi: 10.1162/003355397555217
- Gneezy, U., and Rustichini, A. (2004). Gender and competition at a young age. *Am. Econ. Rev.* 94, 377–381. doi: 10.1257/0002828041301821
- Gneezy, U., Leonard, K. L., and List, J. A. (2009). Gender differences in competition: evidence from a matrilineal and a patriarchal society. *Econometrica* 77, 1637–1664. doi: 10.3982/ECTA6690
- Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in competitive environments: gender differences. *Q. J. Econ.* 118, 1049–1074. doi: 10.1162/00335530360698496
- Goy, R. W., and McEwen, B. S. (1980). *Sexual Differentiation of the Brain*. Cambridge, MA: MIT Press.
- Gray, A., Berlin, J. A., McKinlay, J. B., and Longcope, C. (1991). An examination of research design effects on the association of testosterone and male aging: results of a meta-analysis. *J. Clin. Epidemiol.* 44, 671–684. doi: 10.1016/0895-4356(91)90028-8
- Hampson, E., Ellis, C. L., and Tenk, C. M. (2008). On the relation between 2D:4D and sex-dimorphic personality traits. *Arch. Sex. Behav.* 37, 133–144. doi: 10.1007/s10508-007-9263-3
- Helmreich, R. L., and Spence, J. T. (1978). Work and family orientation questionnaire: an objective instrument to assess components of achievement motivation and attitudes toward family and career. *JSAS Catalog Select. Documents Psychol.* 8:1677.
- Hines, M. (2010). Sex-related variation in human behavior and the brain. *Trends Cogn. Sci.* 14, 448–456. doi: 10.1016/j.tics.2010.07.005
- Hines, M., Constantinescu, M., and Spencer, D. (2015). Early androgen exposure and human gender development. *Biol. Sex Differ.* 6:3. doi: 10.1186/s13293-015-0022-1
- 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., Manning, J. T., and Müller, C. (2006). Digit ratio (2D:4D) and physical fitness in males and females: evidence for effects of prenatal androgens on sexually selected traits. *Horm. Behav.* 49, 545–549. doi: 10.1016/j.yhbeh.2005.11.006
- 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
- Houston, J., Harris, P., McIntire, S., and Francis, D. (2002). Revising the competitiveness index using factor analysis. *Psychol. Rep.* 90, 31–34. doi: 10.2466/pr0.2002.90.1.31
- Lawrance-Owen, A. J., Bargary, G., Bosten, J. M., Goodbourn, P. T., Hogg, R. E., and Mollon, J. D. (2013). Genetic association suggests that SMOC1 mediates between prenatal sex hormones and digit ratio. *Hum. Genet.* 132, 415–421. doi: 10.1007/s00439-012-1259-y
- Lombardo, M. V., Ashwin, E., Auyeung, B., Chakrabarti, B., Taylor, K., Hackett, G., et al. (2012). Fetal testosterone influences sexually dimorphic gray matter in the human brain. *J. Neurosci.* 32, 674–680. doi: 10.1523/JNEUROSCI.4389-11.2012
- Lönnqvist, J. E., Verkasalo, M., Walkowitz, G., and Wichardt, P. C. (2015). Measuring individual risk attitudes in the lab: task or ask? An empirical comparison. *J. Eco. Behav. Organ.* 119, 254–266. doi: 10.1016/j.jebo.2015.08.003
- Lundeberg, M. A., Fox, P. W., and Punčohar, J. (1994). Highly confident but wrong: gender differences and similarities in confidence judgments. *J. Educ. Psychol.* 86, 114–121. doi: 10.1037/0022-0663.86.1.114

- 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. Develop.* 77, 23–28. doi: 10.1016/j.earlhumdev.2003.12.002
- Manning, J. T. (2002). *Digit Ratio: A Pointer to Fertility, Behavior, and Health*. New Brunswick, NJ: Rutgers University Press.
- Manning, J. T., Callow, M., and Bundred, P. E. (2003). Finger and toe ratios in humans and mice: implications for the aetiology of diseases influenced by *HOX* genes. *Med. Hypotheses* 60, 340–343. doi: 10.1016/S0306-9877(02)00400-0
- Manning, J. T., and Fink, B. (2008). Digit ratio (2D:4D), dominance, reproductive success, asymmetry, and sociosexuality in the BBC internet study. *Am. J. Hum. Biol.* 20, 451–461. doi: 10.1002/ajhb.20767
- Manning, J. T., and Peters, M. (2009). Digit ratio (2D:4D) and hand preference for writing in the BBC Internet Study. *Laterality* 14, 528–540. doi: 10.1080/13576500802637872
- Manning, J. T., and Taylor, R. P. (2001). Second to fourth digit ratio and male ability in sport: implications for sexual selection in humans. *Evol. Hum. Behav.* 22, 61–69. doi: 10.1016/S1090-5138(00)00063-5
- Manning, J. T., Kilduff, L., Cook, C., Crewther, B., and Fink, B. (2014). Digit ratio (2D:4D): a biomarker for prenatal sex steroids and adult sex steroids in challenge situations. *Front. Endocrinol.* 5:9. doi: 10.3389/fendo.2014.00009
- 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
- Manning, J. T., Trivers, R., and Fink, B. (2017). Is digit ratio (2D: 4D) related to masculinity and femininity? Evidence from the BBC internet study. *Evol. Psychol. Sci.* 3, 316–324. 1–9. doi: 10.1007/s40806-017-0098-4
- Manning, J. T., Wood, S., Vang, E., Walton, J., Bundred, P. E., van Heyningen, C., et al. (2004). Second to fourth digit ratio (2D:4D) and testosterone in men. *Asian J. Androl.* 6, 211–215.
- Mayr, U., Wozniak, D., Davidson, C., Kuhns, D., and Harbaugh, W. T. (2012). Competitiveness across the life span: the feisty fifties. *Psychol. Aging*, 27, 278–285. doi: 10.1037/a0025655
- McIntyre, M. H., Cohn, B. A., and Ellison, P. T. (2006). Sex dimorphism in the digital formulae of children. *Am. J. Phys. Anthropol.* 129, 143–150. doi: 10.1002/ajpa.20240
- Medland, S. E., Zayats, T., Glaser, B., Nyholt, D. R., Gordon, S. D., Wright, M. J., et al. (2010). A variant in LIN28B is associated with 2D:4D finger-length ratio, a putative retrospective biomarker of prenatal testosterone exposure. *Am. J. Hum. Genet.* 86, 519–525. doi: 10.1016/j.ajhg.2010.02.017
- Millet, K. (2011). An interactionist perspective on the relation between 2D:4D and behavior: an overview of (moderated) relationships between 2D:4D and economic decision making. *Person. Ind. Differ.* 51, 397–401. doi: 10.1016/j.paid.2010.04.005
- Millet, K., and Dewitte, S. (2009). The presence of aggression cues inverts the relation between digit ratio (2D: 4D) and prosocial behaviour in a dictator game. *Br. J. Psychol.* 100, 151–162. doi: 10.1348/000712608X324359
- Newby, J. L., and Klein, G. K. (2014). Competitiveness reconceptualized: psychometric development of the competitiveness orientation measure as a unified measure of trait competitiveness. *Psychol. Rec.* 64, 879–895. doi: 10.1007/s40732-014-0083-2
- Neyse, L., Boworth, S., Ring, P., and Schmidt, U. (2016). Overconfidence, incentives and digit ratio. *Sci. Rep.* 6:23294. doi: 10.1038/srep23294
- Niederle, M. (2016). “Gender,” in *Handbook in Experimental Economics, 2nd Edn.*, eds J. Kagel and A. E. Roth (Princeton: Princeton University Press), 481–553.
- Niederle, M. (2017). A gender agenda: a progress report on competitiveness. *Am. Economic Rev.* 107, 115–119. doi: 10.1257/aer.p20171066
- Niederle, M., and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Q. J. Economics* 122, 1067–1101. doi: 10.1162/qjec.122.3.1067
- Niederle, M., and Vesterlund, L. (2010). Explaining the gender gap in math test scores: the role of competition. *J. Econ. Perspect.* 24, 129–144. doi: 10.1257/jep.24.2.129
- Niederle, M., and Vesterlund, L. (2011). Gender and competition. *Annu. Rev. Econom.* 3, 601–630. doi: 10.1146/annurev-economics-111809-125122
- Phoenix, C. H., Goy, R. W., Gerall, A. A., and Young, W. C. (1959). Organizing action of prenatally administered testosterone propionate on the tissues mediating mating behavior in the female guinea pig. *Endocrinology* 65, 369–382. doi: 10.1210/endo-65-3-369
- Reuben, E., Sapienza, P., and Zingales, L. (2015). *Taste for Competition and the Gender Gap Among Young Business Professionals*. Cambridge, MA: NBER Working Paper No. 21695.
- Ribeiro, E., Neave, N., Morais, R. N., and Manning, J. T. (2016). Direct versus indirect measurement of digit ratio (2D:4D): a critical review of the literature and new data. *Evol. Psychol.* 14, 1–8. doi: 10.1177/1474704916632536
- Romano, M., Rubolini, D., Martinelli, R., Bonisoli Alquati, A., and Sainom, N. (2005). Experimental manipulation of yolk testosterone affects digit length ratios in the ring-necked pheasant (*Phasianus colchicus*). *Horm. Behav.* 48, 342–246. doi: 10.1016/j.yhbeh.2005.03.007
- 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
- Shurchkov, O. (2012). Under pressure: gender differences in output quality and quantity under competition and time constraints. *J. Eur. Econ. Assoc.* 10, 1189–1213. doi: 10.1111/j.1542-4774.2012.01084.x
- Smith, R. D., and Houston, J. M. (1992). The nature of competitiveness: the development and validation of the competitiveness index. *Educ. Psychol. Meas.* 52, 407–418. doi: 10.1177/0013164492052002016
- Stenstrom, E., Saad, G., Nepomuceno, M. V., and Mendenhall, Z. (2011). Testosterone and domain-specific risk: digit ratios (2D:4D and rel2) as predictors of recreational, financial and social risk-taking behaviors. *Pers. Individ. Dif.* 51, 412–416. doi: 10.1016/j.paid.2010.07.003
- Sutter, M., and Rützel, D. (2010). *Gender Differences in Competition Emerge Early in Life*. IZA Discussion Paper No. 5015. Available online at: <https://ssrn.com/abstract=1631480>
- Talarovičová, A., Kršková, L., and Blažeková, J. (2009). Testosterone enhancement during pregnancy influences the 2D: 4D ratio and open field motor activity of rat siblings in adulthood. *Horm. Behav.* 55, 235–239. doi: 10.1016/j.yhbeh.2008.10.010
- Trivers, R., Manning, J., and Jacobson, A. (2006). A longitudinal study of digit ratio (2D:4D) and other finger ratios in Jamaican children. *Horm. Behav.* 49, 150–156. doi: 10.1016/j.yhbeh.2005.05.023
- 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
- van Lange, P. A., De Bruin, E., Otten, W., and Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: theory and preliminary evidence. *J. Pers. Soc. Psychol.* 73, 733–746. doi: 10.1037/0022-3514.73.4.733
- Voracek, M. (2011). Special issue preamble: digit ratio (2D:4D) and individual differences research. *Person. Ind. Differ.* 51, 367–370. doi: 10.1016/j.paid.2011.04.018
- Warrington, N., Hemani, G., Hysi, P., Mangino, M., McMahon, G., Hickey, M., et al. (2016). “Genome-wide association study of 6,939 individuals identifies five novel loci and suggests that prenatal exposure to testosterone is not a major determinant of individual differences in 2D: 4D finger ratio,” in *Behavior Genetics. 46th Annual Meeting of the Behavior-Genetics-Association* (Brisbane, QLD).
- Zheng, Z., and Cohn, M. J. (2011). Development basis of sexually dimorphic digit ratios. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16289–16294. doi: 10.1073/pnas.1108312108

**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.

The reviewer LLK and handling Editor declared their shared affiliation.

Copyright © 2017 Bönte, Procher, Urbig and Voracek. 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.



# Differential Effects of Oxytocin on Visual Perspective Taking for Men and Women

Tong Yue<sup>1,2</sup>, Yuhang Jiang<sup>3</sup>, Caizhen Yue<sup>4</sup> and Xiting Huang<sup>1\*</sup>

<sup>1</sup> Faculty of Psychology, Southwest University, Chongqing, China, <sup>2</sup> Post-doctoral Station of Mathematics, Southwest University, Chongqing, China, <sup>3</sup> School of Humanities, Shandong Management University, Jinan, China, <sup>4</sup> Department of Education, Chongqing University of Arts and Sciences, Chongqing, China

Although oxytocin (OXT) has been shown to lead to reduced self-orientation, no study to date has directly and effectively weakened the egocentric tendencies in perspective taking tasks for both men and women. In this double-blind, placebo-controlled, mixed design study we investigated the effects of OXT on men and women in visual perspective taking tasks. The results showed that OXT shortened the differences in response time between men and women in all experimental conditions. In addition, after OXT administration, the difference in reaction time between judging from one's own perspective and judging from others' perspectives decreased in female participants; however, this effect was not present in males. This may indicate that under OXT treatment, women have a higher tendency to overcome interference from their position and mindset when judging others' perspectives. However, OXT did not affect participants' accuracy, which is possibility because the used task was not suited to detect performance improvements caused by OXT. In summary, the above results may indicate that OXT could increase perspective-taking abilities through reducing self-bias and increasing the perception of others; furthermore, this trend mainly affected women rather than men.

**Keywords:** oxytocin, theory of mind, perspective taking, egocentric biases, sex differences

## INTRODUCTION

Perspective taking is the psychological process of contemplating and inferring other perspectives (Galinsky et al., 2005). The essential characteristic of the process is to set aside one's own perspective in order to see through the others' eyes, to imagine what others might think or feel, or to achieve what is sometimes colloquially referred to as "putting oneself in another's shoes." However, previous research has shown that resisting interference from one's own perspective is not easily achieved. For example, children under the age of four cannot distinguish their own mental state from others; in the false-belief task, they often respond according to their own mental state (Moore et al., 1995; Wellman et al., 2001). Even when adults reason about others' beliefs or thinking, egocentric biases are common (Keysar et al., 2003; Royzman et al., 2003; Bernstein et al., 2004; Apperly et al., 2009), particularly when under cognitive load (Epley et al., 2004). Many researchers believe that self-centeredness is a default choice when inferring other people's mental states (Decety and Sommerville, 2003), although this bias could provide a reasonable starting point and reference for understanding others' mental states (Epley, 2008). However, self-centeredness sometimes renders

## OPEN ACCESS

### Edited by:

Levent Neyse,  
Institut für Weltwirtschaft, Germany

### Reviewed by:

Gökhan Aydoğan,  
Arizona State University, United States  
Katie Daughters,  
Cardiff University, United Kingdom

### \*Correspondence:

Xiting Huang  
xthuang@swu.edu.cn

**Received:** 05 September 2017

**Accepted:** 02 November 2017

**Published:** 15 November 2017

### Citation:

Yue T, Jiang Y, Yue C and Huang X  
(2017) Differential Effects of Oxytocin  
on Visual Perspective Taking for Men  
and Women.  
*Front. Behav. Neurosci.* 11:228.  
doi: 10.3389/fnbeh.2017.00228



people unable to distinguish between themselves and others effectively, which results in difficulties in communication and interaction (Keysar et al., 2000) and thus often requires correction or constraint when attempting to adopt someone else's perspective.

The effects of oxytocin (OXT) have become a major focus of research in modern biological psychology (Heinrichs and Domes, 2008; Heinrichs et al., 2009; Meyer-Lindenberg et al., 2011; Kumsta and Heinrichs, 2013). While there is ongoing debate concerning the precise nature and mechanisms of the effects of OXT in humans, it is generally considered that it may primarily operate as an enhancement of the salience of social stimuli and affiliative behaviors (Bartz et al., 2011; Shamay-Tsoory and Abu-Akel, 2016). In recent years, interest has also been increasing in its potential role in influencing the ability of perspective taking. Intranasal dosing of OXT, which is believed to cross the blood-brain barrier and achieve access to the CNS (Neumann et al., 2013; Striepens et al., 2013), has been found to increase the ability of perspective taking. From the perspective of strategic judgments, Domes et al. (2007) reported that OXT could improve the ability to infer the mental state of others from the eye region. Aydogan et al. (2017) further reported that participants who received OXT were significantly better at predicting the actions of others, which indicated that OXT could enhance perspective taking in strategic interactions. Moreover, Shamay-Tsoory et al. (2013) reported that intranasal OXT led to a remarkable increase in empathy for the pain of even adversary out-group members, demonstrating its important role for promoting perspective taking in emotional judgments.

In fact, the positive effect of OXT on perspective taking may be related to its potential role in influencing aspects of self-processing and in particular distinctions between self and other. Colonnello et al. (2013) reported that OXT reduced the threshold to distinguish between one's own face and an unfamiliar face in a morphing paradigm, indicating its role in sharpening the self-other perceptual boundary. A study about the empathy for pain reported that OXT only increased the empathy for pain ratings toward others when participants had been instructed to adopt the perspective of another, but not when they adopted a self-perspective (Abu-Akel et al., 2015). Further study also reported that OXT could reduce the sense of agency in anxiously attached individuals, indicating OXT's effects of reduced self-orientation (Bartz et al., 2015). In addition, OXT's established role in promoting affiliative behavior and social bonds (Bartz et al., 2011; Striepens et al., 2011; Bethlehem et al., 2013) seems consistent with results reporting a decrease in self-interest and an increase in interest in others. Considering these characteristics of OXT raises the question of whether it can effectively weaken egocentric tendencies in perspective taking. We investigated this question in our study, which, to our knowledge, has not yet been explored.

Our discussion of the problem is aided by the basic visual perspective taking process, as a typical paradigm through which to explore the cognitive process of distinction between the self and others. The experimental paradigm we used in this study originates from a study by Samson et al. (2010), where participants were presented with a picture of a room in which

either one or two walls displayed red discs. A human avatar stands facing one of the walls on which red discs are displayed. During the consistent perspective condition, both the participant and the avatar could see the same number of discs. However, in the inconsistent perspective condition, the participant and the avatar saw a different number of discs (some of the discs were not visible to the avatar). Participants were then asked to identify whether they were able to see the same number of discs as the avatar. Perhaps, due to interference by the ego in perspective taking, many studies found that, in the inconsistent conditions, judging from the perspective of the avatar resulted in slower response times and more errors compared to when participants judged from their own perspective, which is a typical example of egocentric bias (Samson et al., 2010; Wang et al., 2015). On this basis, we speculate that, if OXT decreased self-centeredness, it may also affect subjects in the visual perspective taking task; i.e., by reducing egocentric bias.

In addition, differences in sex are an important factor when examining the effects of OXT on human social cognition. Many previous research results, including social judgment (Hoge et al., 2014; Gao et al., 2016), social approach/avoidance (Theodoridou et al., 2013; Preckel et al., 2014), social cooperation/competition (Fischer-Shofty et al., 2013; Scheele et al., 2014), and the ability to maintain social relations (Yao et al., 2014) have found inconsistent and even opposing results in the effects of OXT in different genders. Currently however, there is no definite explanation for why such gender differences exist and it is difficult to predict under what type of social situation the effects of OXT will cause such a difference. Following this, in our exploration of the effects of OXT on the visual perspective taking task, we wondered whether this study could also provide insight into differences between males and females. To address this question, both men and women were recruited and the results were compared and analyzed to explore the role of OXT in the information processing system of self and others and to examine potential differences in results across male and female participants.

## METHODS

### Participants and Treatment

Subjects from the Southwest University and the Chongqing University of Arts and Sciences, in China, were recruited through local advertisements. Each subject was provided with a written informed consent form prior to study enrollment. Eighty-five students (39 males and 46 females, with a mean age of 21.2 years; S.D. = 1.76) participated in the study. None of the subjects were taking any form of medication or reported having had neurological problems or psychiatric illnesses prior to the start of the study. None of the female subjects were menstruating, which is important because the menstrual cycle may influence the effectiveness of OXT administration (Bakermans-Kranenburg and van IJzendoorn, 2013) and no women were pregnant or using oral contraceptives. Before the formal experiment, we asked the participants to maintain their regular sleep pattern and abstain from caffeine, alcohol, and smoking for at least 12 h prior to the experiment. After a detailed explanation of

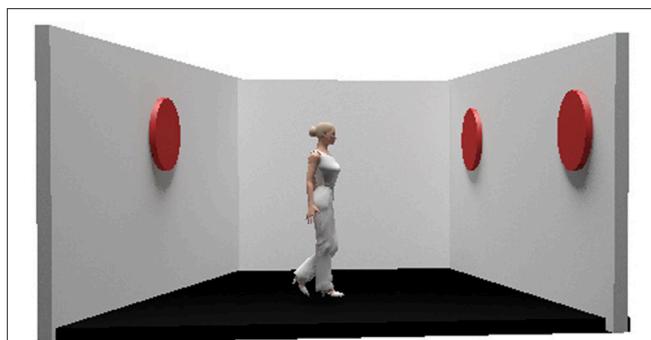
the study protocol, all subjects were asked to sign a written informed consent form. The study was approved by the Ethics Committee of the Southwest University and the Chongqing University of Arts and Sciences, and all involved procedures were in accordance with the sixth revision of the Declaration of Helsinki.

The study used a double-blind, placebo-controlled, mixed design. In the experiment, all subjects first received a single intranasal dose of 24 IU OXT (Syntocinon Spray, Sichuan Meike Pharmacy Co. Ltd, China; three puffs of 4 IU per nostril with 30 s between each puff) or PLC (with the identical type of bottle from the same pharmaceutical company, containing all of the same ingredients as the OXT nasal spray except the neuropeptide, i.e., sodium chloride and glycerin; also, three puffs were administered per nostril). In line with a previous study (Striepens et al., 2011), the formal experiments started 45 min after OXT or PLC treatments. This time lapse was used because, within that time limit, the peptide will increase its concentrations within the cerebrospinal fluid. Of the total number of participants, 21 female and 15 male subjects were treated with OXT. The remaining 49 subjects received PLC treatment. During post-experiment interviews, subjects could not identify (with any better degree of accuracy than by chance) whether they had received the OXT or PLC treatment.

## Experimental Design

The participants were then presented with a picture showing a lateral view into a room with the left, back, and right walls visible and with red discs displayed on one or two of the walls. Female and male avatars were created with the 3D cartoon software Poser 6 (e frontier, Scotts Valley, California, USA), and were positioned in the center of the room, facing either the left or the right wall. On either side of the room or on opposite sides of the wall, either 0, 1, 2, or 3 red discs were displayed randomly (**Figure 1**). During the experiment, female subjects were presented with female model avatars and male subjects were presented with male model avatars. In 50% of the experimental sequences, the number of red discs seen by the avatar and the subjects was identical (consistent condition). In the remaining 50% of the experimental sequences, the avatar was positioned in such a way that he or she could not see some of the discs that were visible to the participants (inconsistent condition). In both conditions, the position of the discs changed while the position of the avatar remained constant.

The experiment was controlled by the E-Prime program. At the beginning of the experiment, subjects were familiarized with the process of the task and how to respond to the cues. Each stimulus sequence consisted of four stages (see **Figure 2**). The stages were as follows: First, a fixation cross appeared for a duration of 750 ms. Second, after an interval of 500 ms, Chinese characters “you” or “him/she?” (male/female avatar) appeared for 750 ms to prompt participants whether to adopt their own perspective (self-condition) or that of the avatar (other condition). After another 500 ms interval, a number ranging between 0 and 3 was displayed, lasting for 750 ms, which specified the number of discs the subject was required to judge. Finally, the image of the room appeared until participants reacted with a the



**FIGURE 1** | Example of the visual stimulus used for the experiment.

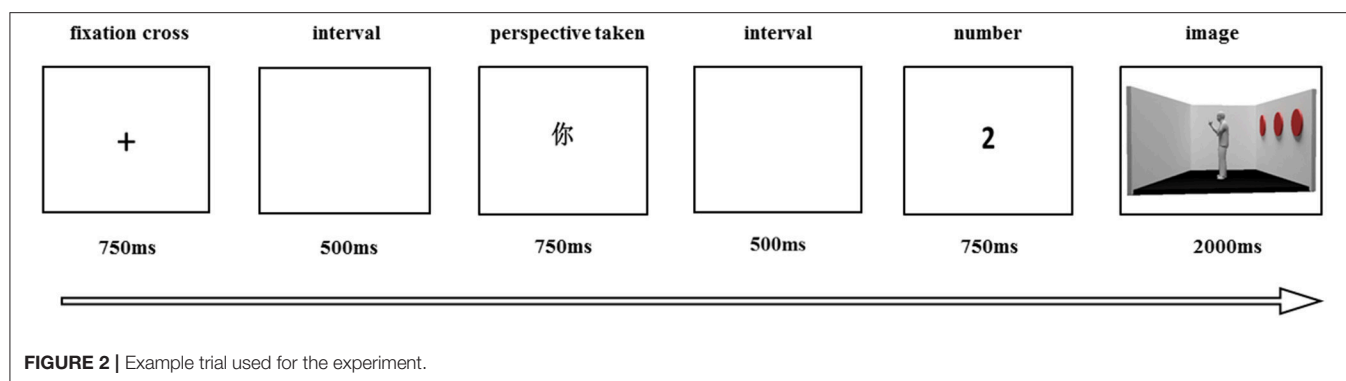
“yes” (matched) or “no” (mismatched) on the keyboard from the given perspective, then went to the next sequence. If the subjects still remained unresponsive at 2000 ms, the next trial appeared automatically.

The experiment included a total of 208 trials, 104 of which required “yes” responses and the 104 remaining stimuli required “no” responses. In the trials of 104 “yes” responses, 48 stimulus trials required the subjects to verify their own perspective (including 24 consistent trials and 24 inconsistent trials) and 48 trials the avatar’s perspective (with 24 consistent perspective trials and 24 inconsistent perspective trials). There was an equal number of mismatching (“no”) responses. The experiment also added 16 filler trials where no discs were displayed on the wall. Therefore, the answer “0” was sometimes the correct response. These filler trials included an equal number of self and other trials, consistent and inconsistent trials and “yes” and “no” trials. The experiment was divided into four blocks, each with 52 test trials (48 test trials and four filler tests). Prior to the formal experiment, 26 practice trials were presented. Within each block, the sequence of tests was pseudo-random, then fixed across participants so that there were no more than three consecutive trials of the same type and self and other trials were equally preceded by the same perspective (no shift of perspective) and by a different perspective (shift of perspective). The order of presentation of the blocks was counterbalanced across participants.

Our results were analyzed via SPSS 16.0. Some principles in the analyses of variance have been reported below. When the sphericity hypothesis was violated, we used Greenhouse-Geisser corrections. When follow-up tests were required, Bonferroni corrections were applied.

## RESULTS

We performed a  $2 \times 2$  repeated measure analysis of variance (ANOVA) with treatment type (OXT vs. PLC), gender (male vs. female) as between-subjects factors and the type of perspective taken (self vs. other) and the reaction condition (consistent vs. inconsistent) as within subject variables. Response time and accuracy were used as dependent variables (**Table 1**).



**TABLE 1 |** The mean and standard deviation of the reaction time and accuracy between male and female subjects at different perspectives and different reaction conditions in both OXT and PLC groups.

		RT (ms)				ACC (%)			
		OXT		PLC		OXT		PLC	
		Male	Female	Male	Female	Male	Female	Male	Female
Self	Inconsistent	737.72 (34.24)	751.54 (28.94)	663.09 (27.07)	758.97 (26.52)	93.72 (1.53)	94.88 (1.30)	92.50 (1.14)	94.46 (1.18)
	Consistent	722.79 (33.02)	736.96 (27.90)	640.36 (26.10)	723.74 (25.57)	94.74 (1.04)	95.46 (0.88)	94.95 (0.82)	96.92 (0.81)
Other	Inconsistent	771.58 (33.06)	767.29 (27.94)	699.76 (26.13)	808.66 (25.61)	89.40 (1.59)	90.48 (1.35)	90.12 (1.30)	92.00 (1.23)
	Consistent	721.29 (32.74)	731.01 (26.67)	657.26 (25.89)	736.40 (25.36)	94.41 (1.05)	96.57 (0.88)	95.51 (0.83)	95.92 (0.81)

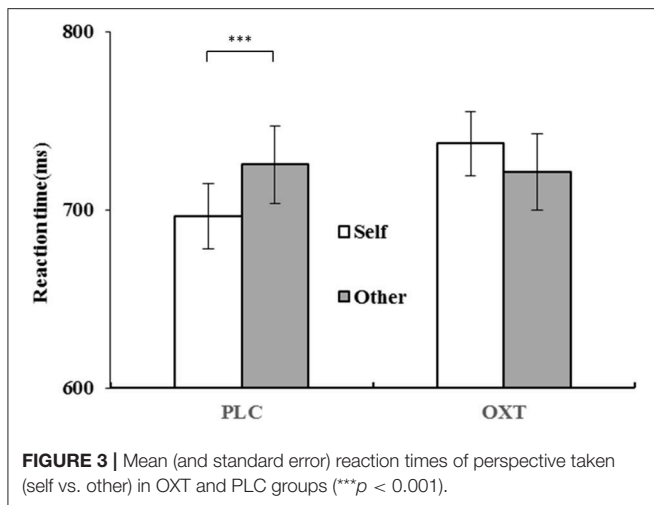
## Reaction Time Analysis

The ANOVA analysis revealed a significant primary effect of reaction condition [ $F_{(1, 81)} = 87.68, p < 0.001, \eta^2_p = 0.52$ ] with RTs being overall slower in the inconsistent condition ( $M = 744.83$  ms) when compared to the consistent condition ( $M = 702.73$  ms). The main effect of perspective taking was also significant [ $F_{(1, 81)} = 21.92, p < 0.001, \eta^2_p = 0.21$ ]; participants were significantly quicker when judging from their own perspective ( $M = 717.90$  ms) than the avatar's perspective ( $M = 736.66$  ms). There was a significant reaction condition  $\times$  perspective interaction effect [ $F_{(1, 81)} = 40.04, p < 0.001, \eta^2_p = 0.33$ ]; a simple effect test showed that under the condition of inconsistency, the judgment of the self-perspective was significantly faster than the judgment of other people's perspective; however, no effect was noted under the consistent condition. We also found the interaction effect between treatment and perspective,  $F_{(1, 81)} = 23.63, p < 0.001, \eta^2_p = 0.23$ , with participants responding faster in the self-perspective judgments ( $M = 696.54$  ms) than in other-perspective judgments ( $M = 725.52$  ms) in the PLC group; however, no effect was found in the OXT group (Figure 3).

We were most interested to investigate whether OXT could affect the egocentric biases in the visual perspective taking in response to sex differences; thus, the simple effect test of the perspective taken with the other three factors fixed (treatment, gender, and reaction condition) were performed. The analysis results showed that: (1) For female subjects in the PLC group, participants responded faster in the self-perspective than in the other-perspective under the condition of inconsistency [ $F_{(1, 81)}$

$= 26.20, p < 0.001, \eta^2_p = 0.24$ ]; however, they showed no effect in the consistent condition between both perspectives; in the OXT group, no effect was found in the two conditions between the two perspectives. (2) For male subjects, there was a significant perspective effect on both the OXT group [ $F_{(1, 81)} = 7.30, p < 0.01, \eta^2_p = 0.08$ ] and the PLC group [ $F_{(1, 81)} = 13.70, p < 0.001, \eta^2_p = 0.15$ ] under the condition of inconsistency, with participants being quicker at judging their own perspective than that of the avatar. No effect was found under the consistent condition in both groups between the perspectives (Figure 4).

To examine the effects of OXT on the response time of four experimental conditions, we analyzed the data for male and female participants separately. However, there were no differences between the drug effects on all four conditions for both males and females. I.e., the effects of OXT on the response time of visual perspective taking have not reached the significance level. Then, we analyzed the sex differences in the four experimental conditions for males and females. The results also showed that there were significant gender effects on all four conditions within the subject experimental condition in the PLC group (consistent-self [ $F_{(1, 81)} = 5.21, p < 0.05, \eta^2_p = 0.06$ ], consistent-other [ $F_{(1, 81)} = 4.77, p < 0.05, \eta^2_p = 0.06$ ], inconsistent-self [ $F_{(1, 81)} = 6.40, p < 0.05, \eta^2_p = 0.07$ ], inconsistent-other [ $F_{(1, 81)} = 8.86, p < 0.01, \eta^2_p = 0.10$ ], respectively), with males responding faster than females. However, no gender differences were found in all conditions in the OXT group. Based on these results, we concluded that OXT did not lead to a statistically



**FIGURE 3 |** Mean (and standard error) reaction times of perspective taken (self vs. other) in OXT and PLC groups (\*\* $p < 0.001$ ).

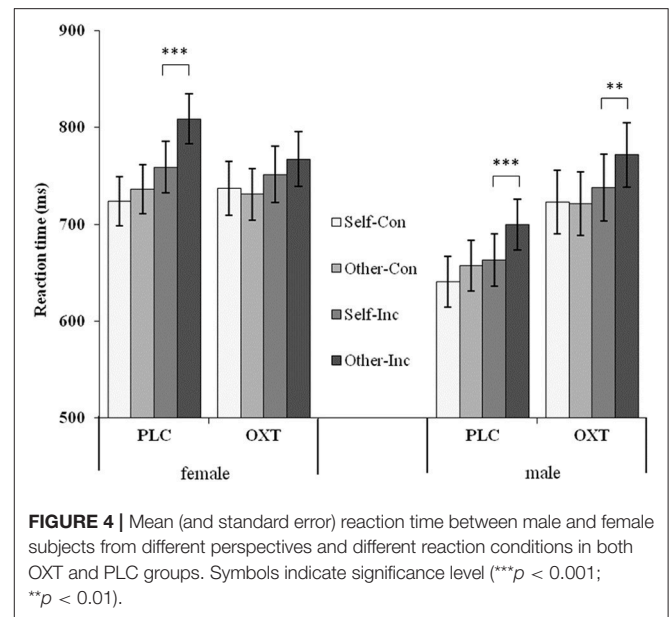
different reaction time of male participants compared to female participants.

## Accuracy Analysis

The analysis revealed a significant main effect on reaction condition [ $F_{(1, 81)} = 57.14$ ,  $p < 0.001$ ,  $\eta^2_p = 0.41$ ] with less accuracy when inconsistent ( $M = 92.17\%$ ) than when consistent ( $M = 95.56\%$ ). The main effect of perspective was also significant [ $F_{(1, 81)} = 12.44$ ,  $p = 0.001$ ,  $\eta^2_p = 0.13$ ]; participants were significantly more accurate when judging from their own perspective ( $M = 94.60\%$ ) than when judging from the avatar's perspective ( $M = 93.13\%$ ). There was a significant treatment  $\times$  reaction condition  $\times$  perspective interaction effect [ $F_{(1, 81)} = 5.26$ ,  $p < 0.05$ ,  $\eta^2_p = 0.06$ ]. A further simple effect test showed that, under the condition of inconsistency on both OXT and PLC groups, the accuracy was higher in the self-perspective judgments ( $M_{PLC} = 93.06\%$ ,  $M_{OXT} = 94.30\%$ ) than other-perspective judgments ( $M_{PLC} = 91.39\%$ ,  $M_{OXT} = 89.94\%$ ). However, there was no effect under the consistent condition between the two perspectives on both OXT and PLC groups.

Further ANOVAs were performed for males and females separately to elucidate the effects of OXT on visual perspective taking. The results showed that, under the condition of consistency, there was no significant difference between the two perspectives in both females and males. However, judging from their own perspective always yielded an ACC advantage when compared to judging from the avatar's perspective under the inconsistent condition in both females and males in the two groups [ $F_{\text{female-PLC}(1, 81)} = 4.44$ ,  $p < 0.05$ ,  $\eta^2_p = 0.05$ ;  $F_{\text{female-OXT}(1, 81)} = 11.94$ ,  $p = 0.001$ ,  $\eta^2_p = 0.13$ ;  $F_{\text{male-PLC}(1, 81)} = 4.28$ ,  $p < 0.05$ ,  $\eta^2_p = 0.05$ ;  $F_{\text{male-OXT}(1, 81)} = 8.79$ ,  $p < 0.01$ ,  $\eta^2_p = 0.10$ ] (see Figure 5). Therefore, the egocentric biases performed in the consistency condition were not affected by OXT.

Corresponding to the analysis of reaction time, we analyzed the effects of OXT and sex for four experimental conditions separately. With regard to the effect of OXT, there were no significant differences in all four subject experimental conditions



**FIGURE 4 |** Mean (and standard error) reaction time between male and female subjects from different perspectives and different reaction conditions in both OXT and PLC groups. Symbols indicate significance level (\*\* $p < 0.001$ ; \*\*\* $p < 0.01$ ).

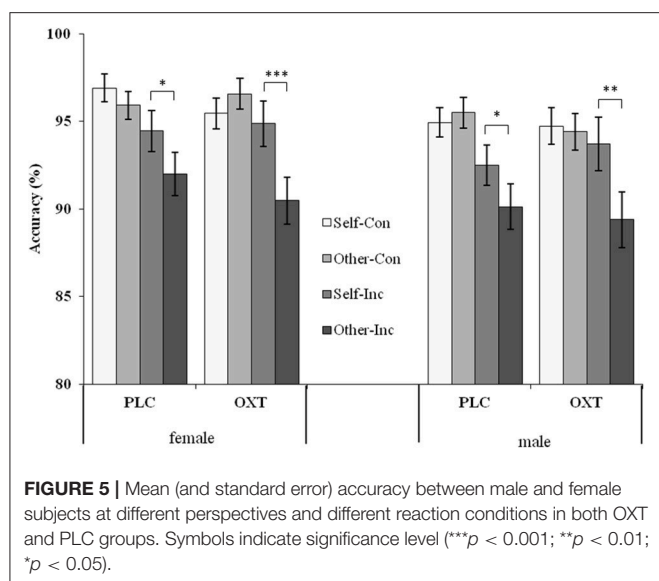
neither in males nor in females. Furthermore, no sex differences existed in all the conditions regardless of group (OXT group or PLC group). It seemed that OXT had no effect on the accuracy of participants.

## DISCUSSION

In this study, we investigated the different effects of OXT on visual perspective taking for men and women. The results showed that OXT reduced differences in reaction time between judging from one's own perspective and judging others' perspectives in female participants under the inconsistent condition. In contrast to females, male participants who received intranasal OXT still showed a significantly slower reaction time when taking on the perspective of another compared to their own perspective, independent of whether the reaction condition was consistent or inconsistent; however, OXT yielded similar reaction times in males than in females across all four experimental conditions. With regard to accuracy, male and female participants were significantly less accurate when taking on the perspective of another compared to their own perspective in the inconsistent condition, which was true for both drug conditions.

The results of the PLC group validated the normal state performance in the visual perspective taking task of previous studies. Firstly, egocentric biases were also present in our PLC group. More precisely, under the consistent condition, there were no significant differences in response time and accuracy between the participants' own perspective and the avatar's perspective; however, under the condition of inconsistency, self-perspective judgments had a significant advantage for both response time and accuracy when compared to other-perspective judgments. These results indicate that in the reasoning process of others' mental states, the information from the perspective of the self plays an important role, which is consistent with a considerable body of previous research (Keysar et al., 2003; Bernstein et al., 2004; Birch





and Bloom, 2007; Apperly et al., 2009). This might be because people tend to anchor from their own point of view and then only adjust from self-perspective to other-perspective, overcoming self-centeredness to arrive at a final judgment (Epley et al., 2004). Thus, under the inconsistent condition, the participants were thrown off by varying information in the self-perspective when judging from the avatar's perspective than when compared with the consistent condition, in which they required more time to adjust and correct egocentric bias. Secondly, we also observed the sex differences in performance in the PLC group. Although the differences between males and females were absent for participants' accuracy, the responses of male subjects were significantly faster than those of female subjects throughout all experimental conditions. Our results were in line with a previous relevant study by Mohr et al. (2010), which showed that women experienced increased reaction times compared to men when performing an avatar perspective task. Mohr et al. (2010) suggested that these results may be caused by a difference in processing strategies between men and women: the object-based spatial strategies may be more prevalent in men, which renders them good at spatial/mental rotation and enables them to spend less time on such tasks; however, women may be inclined to adopt a social perspective taking strategy, which is supported by the link between high empathy and faster reaction times; therefore, this strategy is comparatively more time-consuming when they finish the task.

Partially consistent with our hypothesis, we found no significant differences between the two perspectives in response time after OXT administration in female participants, regardless whether the condition was consistency or inconsistency. Compared to the results of the PLC group, it seems that OXT shortened the reaction time difference between self- vs other-perspective taking in the inconsistent condition. According to the previous discussion, the results may indicate that OXT has the potential to allow female participants to effectively avoid the interference of their own perspective when judging from the

avatar's perspective, which leads to a more rapid reaction. This is in line with OXT's function reported in previous studies; i.e., a shift in focus from self to others (Abu-Akel et al., 2015; Bartz et al., 2015). In general, the results of this study support our hypothesis: OXT administration may decrease self-centeredness and increase focus in others, which renders female participants quicker in inferring the mental state of others. However, the results did not show that OXT has an effect on accuracy between both perspectives. Specifically, the participants' information from their self-perspectives still influenced the process of adopting the views of the avatars, lending more accuracy when judging from one's own perspective than from others' perspectives. The results agree with similar results and explanations provided by other studies; i.e., Hubble et al. (2017) and Di Simplicio et al. (2009), who also reported a lack of effects of OXT on the accuracy in their tasks. This may indicate that the weakening effect of OXT on the egocentric bias in perspective taking may also have limitations, which indicates that self-centeredness is, to a great degree, still a default choice when accurately inferring other people's mental states. However, the behavioral tasks may be less sensitive to OXT triggered changes or just too low in difficulty. Indeed, in our results show that the accuracy was very close to 100%, even if men would start thinking harder in this task, there is almost no room left to improve. Thus, although we found that OXT could enhance individual's attention to others' perspectives and accelerate the process of suppression and correction of egocentric tendencies, future studies are required to further explore the effects of OXT in the process of perspective taking.

Another important foundation of this research was the differential effect of OXT on visual perspective taking between men and women. It seemed that OXT had no effect on male egocentric tendencies in the visual perspective taking task, because they were still advantageous with significantly quicker reaction time and higher accuracy when taking on their own perspective compared to that of the avatar in the inconsistent condition. However, after inhaling OXT, the differences in the response time between sexes on all experimental conditions disappeared; however, it still existed for the accuracy index. The following two reasons can explain the reducing effect of OXT on the response time between males and females: Firstly, we noticed that the response time of OXT group had an accelerated trend in females compared to the PLC group; however, this was not significant (756.94 ms vs. 746.70 ms). The effects of OXT on promoting the empathy ability in women has been verified in the study of Mohr et al. (2010), who reported that women with higher empathy scores responded faster in the perspective taking task. Secondly, OXT led to a trend of reducing the response time in men compared to the PLC group (665.12 ms vs. 738.35 ms; also not significant). This result is consistent with the results of Theodoridou et al. (2013), which showed that male participants in the OXT group responded as slowly as females. According to the explanation by Theodoridou et al. (2013), this may indicate that OXT promotes the attempts of male subjects to adopt similar perspective taking processing strategies as female subjects, and as such, there is a slowing trend in response time.

The question remains why differential effects of OXT on visual perspective taking for men and women still exist. Unfortunately, no definite explanation exists up to now. As far as our research is concerned, this difference may be caused by a variety of factors. Firstly, females are better at taking on the views of others compared to males. While OXT could increase both male and female perspective taking, women may be more affected than men, and thus easier to remain in their initial behaviors. In addition, previous studies found that steroid hormones, such as estradiol and progesterone, can modulate the OXT receptor (Gimpl and Fahrenholz, 2002; Choleris et al., 2008). Essentially, women differ from men with regard to gonadal steroid hormones (Hawkins and Matzuk, 2008). Therefore, the modulation of OXT by gonadal steroids, which affects the differences in the sensitivity to the OXT system, might be an explanation for the inconsistent findings (between men and women) in our tasks.

## LIMITATIONS

This study has several limitations, which can be addressed in future studies. First, although the results showed that women are likely to be more sensitive to OXT than men, the results were concluded based on a relatively small sample size. Comparing male and female performance in larger samples is thus necessary to draw definite conclusions about differences between sexes. Next, the task in our study may be too simple for participants and the ceiling effects appeared to make our results very complicated; i.e., the used task was not suited to detect performance improvements due to OXT. Thus, future studies should overcome this insufficiency using a better experiment task. Then, we used a between-subjects design for drug administration, while individual differences, such as psychological factors, may also moderate the effects of

OXT (Daughters et al., 2015). Controlling for these variables may advance our understanding of the OXT's effect in visual perspective taking in future studies. Ultimately, the behavioral indicators we used in this study (response time and accuracy), may be insensitive to the changes induced by OXT, thus future researcher should select more sensitive indicators, such as event-related potentials studies, to further explore this topic.

## CONCLUSION

In summary, this study investigated the effects of OXT on men and women in a visual perspective taking task. The results showed that OXT shortened the differences of response time between men and women regardless of whether they were taking the perspective of self or others. In addition, after OXT administration, the difference in reaction time between judging from one's own perspective and judging from others' perspectives also decreased in female participants, but this effect was not present in male participants. The above results may indicate that OXT could increase perspective taking abilities through reducing self-bias and increasing the perception of others and this trend is mainly reflected in women rather than in men.

## AUTHOR CONTRIBUTIONS

TY and XH designed experiments; TY, YJ, and CY carried out experiments. TY analyzed sequencing data and wrote the manuscript.

## FUNDING

This study is supported by the General Financial Grant from the China Postdoctoral Science Foundation (2016M602619).

## REFERENCES

- Abu-Akel, A., Palgi, S., Klein, E., Decety, J., and Shamay-Tsoory, S. (2015). Oxytocin increases empathy to pain when adopting the other- but not the self-perspective. *Soc. Neurosci.* 10, 7–15. doi: 10.1080/17470919.2014.948637
- Apperly, I. A., Samson, D., and Humphreys, G. W. (2009). Studies of adults can inform accounts of theory of mind development. *Dev. Psychol.* 45, 190–201. doi: 10.1037/a0014098
- Aydogan, G., Jobst, A., D'Ardenne, K., Müller, N., and Kocher, M. G. (2017). The detrimental effects of oxytocin-induced conformity on dishonesty in competition. *Psychol. Sci.* 28, 751–759. doi: 10.1177/0956797617695100
- Bakermans-Kranenburg, M. J., and van IJzendoorn, M. H. (2013). Sniffing around oxytocin: review and meta-analyses of trials in healthy and clinical groups with implications for pharmacotherapy. *Transl. Psychiatry* 3, e258. doi: 10.1038/tp.2013.34
- Bartz, J. A., Lydon, J. E., Kolevzon, A., Zaki, J., Hollander, E., Ludwig, N., et al. (2015). Differential effects of oxytocin on agency and communion for anxiously and avoidantly attached individuals. *Psychol. Sci.* 26, 1177–1186. doi: 10.1177/0956797615580279
- Bartz, J. A., Zaki, J., Bolger, N., and Ochsner, K. N. (2011). Social effects of oxytocin in humans: context and person matter. *Trends. Cogn. Sci.* 15, 301–309. doi: 10.1016/j.tics.2011.05.002
- Bernstein, D. M., Atance, C., Loftus, G. R., and Meltzoff, A. (2004). We saw it all along: visual hindsight bias in children and adults. *Psychol. Sci.* 15, 264–267. doi: 10.1111/j.0963-7214.2004.00663.x
- Bethlehem, R. A., van Honk, J., Auyeung, B., and Baron-Cohen, S. (2013). Oxytocin, brain physiology, and functional connectivity: a review of intranasal oxytocin fMRI studies. *Psychoneuroendocrinology* 38, 962–974. doi: 10.1016/j.psyneuen.2012.10.011
- Birch, S. A., and Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychol. Sci.* 18, 382–386. doi: 10.1111/j.1467-9280.2007.01909.x
- Choleris, E., Devidze, N., Kavaliers, M., and Pfaff, D. W. (2008). Steroidal/neuropeptide interactions in hypothalamus and amygdala related to social anxiety. *Prog. Brain. Res.* 170, 291–303. doi: 10.1016/S0079-6123(08)00424-X
- Colonnello, V., Chen, F. S., Panksepp, J., and Heinrichs, M. (2013). Oxytocin sharpens self-other perceptual boundary. *Psychoneuroendocrinology* 38, 2996–3002. doi: 10.1016/j.psyneuen.2013.08.010
- Daughters, K., Manstead, A. S., Hubble, K., Rees, A., Thapar, A., and van Goozen, S. H. (2015). Salivary oxytocin concentrations in males following intranasal administration of oxytocin: a double-blind, cross-over study. *PLoS ONE* 10:e0145104. doi: 10.1371/journal.pone.0145104
- Decety, J., and Sommerville, J. A. (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends Cogn. Sci.* 7, 527–533. doi: 10.1016/j.tics.2003.10.004
- Di Simplicio, M., Massey-Chase, R., Cowen, P. J., and Harmer, C. J. (2009). Oxytocin enhances processing of positive versus negative emotional information in healthy male volunteers. *J. Psychopharmacol.* 23, 241–248. doi: 10.1177/0269881108095705

- Domes, G., Heinrichs, M., Michel, A., Berger, C., and Herpertz, S. C. (2007). Oxytocin improves “mind-reading” in humans. *Biol. Psychiatry* 61, 731–733. doi: 10.1016/j.biopsych.2006.07.015
- Epley, N. (2008). Solving the (real) other minds problem. *Social and Pers. Psychol. Compass* 2, 1455–1474. doi: 10.1111/j.1751-9004.2008.00115.x
- Epley, N., Keysar, B., Van Boven, L., and Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *J. Pers. Soc. Psychol.* 87, 327–339. doi: 10.1037/0022-3514.87.3.327
- Fischer-Shofty, M., Brüne, M., Ebert, A., Shefet, D., Levkovitz, Y., and Shamay-Tsoory, S. (2013). Improving social perception in schizophrenia: the role of oxytocin. *Schizophr. Res.* 146, 357–362. doi: 10.1016/j.schres.2013.01.006
- Galinsky, A. D., Ku, G., and Wang, C. S. (2005). Perspective-taking and self-other overlap: fostering social bonds and facilitating social coordination. *Group Process. Intergr. Relat.* 8, 109–124. doi: 10.1177/1368430205051060
- Gao, S., Becker, B., Luo, L. Z., Geng, Y. Y., Zhao, W. H., Yin, Y., et al. (2016). Oxytocin, the peptide that bonds the sexes also divides them. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7650–7654. doi: 10.1073/pnas.1602620113
- Gimpl, G., and Fahrenholz, F. (2002). Cholesterol as stabilizer of the oxytocin receptor. *BBA* 1564, 384–392. doi: 10.1016/S0005-2736(02)00475-3
- Hawkins, S. M., and Matzuk, M. M. (2008). The menstrual cycle. *Ann. N. Y. Acad. Sci.* 1135, 10–18. doi: 10.1196/annals.1429.018
- Heinrichs, M., and Domes, G. (2008). Neuropeptides and social behaviour: effects of oxytocin and vasopressin in humans. *Prog. Brain. Res.* 170, 337–350. doi: 10.1016/S0079-6123(08)00428-7
- Heinrichs, M., von Dawans, B., and Domes, G. (2009). Oxytocin, vasopressin, and human social behavior. *Front. Neuroendocrinol.* 30, 548–557. doi: 10.1016/j.yfrne.2009.05.005
- Hoge, E. A., Anderson, E., Lawson, E. A., Bui, E., Fischer, L. E., Khadge, S. D., et al. (2014). Gender moderates the effect of oxytocin on social judgments. *Hum. Psychopharmacol.* 29, 299–304. doi: 10.1002/hup.2402
- Hubble, K., Daughters, K., Manstead, A. S., Rees, A., Thapar, A., and van Goozen, S. H. (2017). Oxytocin reduces face processing time but leaves recognition accuracy and eye-gaze unaffected. *J. Int. Neuropsychol. Soc.* 23, 23–33. doi: 10.1017/S1355617716000886
- Keysar, B., Barr, D. J., Balin, J. A., and Brauner, J. S. (2000). Taking perspective in conversation: the role of mutual knowledge in comprehension. *Psychol. Sci.* 11, 32–38. doi: 10.1111/1467-9280.00211
- Keysar, B., Lin, S., and Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition* 89, 25–41. doi: 10.1016/S0010-0277(03)00064-7
- Kumsta, R., and Heinrichs, M. (2013). Oxytocin, stress and social behavior: neurogenetics of the human oxytocin system. *Curr. Opin. Neurobiol.* 23, 11–16. doi: 10.1016/j.conb.2012.09.004
- Meyer-Lindenberg, A., Domes, G., Kirsch, P., and Heinrichs, M. (2011). Oxytocin and vasopressin in the human brain: social neuropeptides for translational medicine. *Nat. Rev. Neurosci.* 12, 524–538. doi: 10.1038/nrn3044
- Mohr, C., Rowe, A. C., and Blanke, O. (2010). The influence of sex and empathy on putting oneself in the shoes of others. *Br. J. Psychol.* 101, 277–291. doi: 10.1348/000712609X457450
- Moore, C., Jarrold, C., Russell, J., Lumb, A., Sapp, F., and MacCallum, F. (1995). Conflicting desire and the child's theory of mind. *Cognitive Dev.* 10, 467–482. doi: 10.1016/0885-2014(95)90023-3
- Neumann, I. D., Maloumy, R., Beiderbeck, D. I., Lukas, M., and Landgraf, R. (2013). Increased brain and plasma oxytocin after nasal and peripheral administration in rats and mice. *Psychoneuroendocrinology* 38, 1985–1993. doi: 10.1016/j.psyneuen.2013.03.003
- Preckel, K., Scheele, D., Kendrick, K. M., Maier, W., and Hurlmann, R. (2014). Oxytocin facilitates social approach behavior in women. *Front. Behav. Neurosci.* 8:191. doi: 10.3389/fnbeh.2014.00191
- Royzman, E. B., Cassidy, K. W., and Baron, J. (2003). “I know, you know”: epistemic egocentrism in children and adults. *Rev. Gen. Psychol.* 7, 38–65. doi: 10.1037/1089-2680.7.1.38
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., and Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *J. Exp. Psychol. Hum. Percept. Perform.* 36:1255. doi: 10.1037/a0018729
- Scheele, D., Kendrick, K. M., Khouri, C., Kretzer, E., Schläpfer, T. E., Stoffel-Wagner, B., et al. (2014). An oxytocin-induced facilitation of neural and emotional responses to social touch correlates inversely with autism traits. *Neuropsychopharmacology* 39, 2078–2085. doi: 10.1038/npp.2014.78
- Shamay-Tsoory, S. G., Abu-Akel, A., Palgi, S., Sulieman, R., Fischer-Shofty, M., Levkovitz, Y., et al. (2013). Giving peace a chance: oxytocin increases empathy to pain in the context of the Israeli–Palestinian conflict. *Psychoneuroendocrinology* 38, 3139–3144. doi: 10.1016/j.psyneuen.2013.09.015
- Shamay-Tsoory, S. G., and Abu-Akel, A. (2016). The social salience hypothesis of oxytocin. *Biol. Psychiatry* 79, 194–202. doi: 10.1016/j.biopsych.2015.07.020
- Striepen, N., Kendrick, K. M., Hankin, V., Landgraf, R., Wüllner, U., Maier, W., et al. (2013). Elevated cerebrospinal fluid and blood concentrations of oxytocin following its intranasal administration in humans. *Sci. Rep.* 3:3440. doi: 10.1038/srep03440
- Striepen, N., Kendrick, K. M., Maier, W., and Hurlmann, R. (2011). Prosocial effects of oxytocin and clinical evidence for its therapeutic potential. *Front. Neuroendocrinol.* 32, 426–450. doi: 10.1016/j.yfrne.2011.07.001
- Theodoridou, A., Rowe, A. C., and Mohr, C. (2013). Men perform comparably to women in a perspective taking task after administration of intranasal oxytocin but not after placebo. *Front. Hum. Neurosci.* 7:197. doi: 10.3389/fnhum.2013.00197
- Wang, Y., You, X., Jiao, J., and Chen, P. (2015). Perspective taking: making inferences based on oneself and related individual differences. *Acta Psychol. Sin.* 47, 1039–1049. doi: 10.3724/SP.J.1041.2015.01039
- Wellman, H. M., Cross, D., and Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child. Dev.* 72, 655–684. doi: 10.1111/1467-8624.00304
- Yao, S., Zhao, W., Cheng, R., Geng, Y., Luo, L., and Kendrick, K. M. (2014). Oxytocin makes females, but not males, less forgiving following betrayal of trust. *Int. J. Neuropsychopharmacol.* 17, 1785–1792. doi: 10.1017/S146114571400090X

**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 Yue, Jiang, Yue and Huang. 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.



# The Effect of Testosterone Administration and Digit Ratio (2D:4D) on Implicit Preference for Status Goods in Healthy Males

Yin Wu<sup>1,2,3\*</sup>, Samuele Zilioli<sup>4,5</sup>, Christoph Eisenegger<sup>6</sup>, Luke Clark<sup>7</sup> and Hong Li<sup>1,8\*</sup>

<sup>1</sup>Research Center for Brain Function and Psychological Science, Shenzhen University, Shenzhen, China, <sup>2</sup>Shenzhen Key Laboratory of Affective and Social Cognitive Science, Shenzhen University, Shenzhen, China, <sup>3</sup>Behavioural and Clinical Neuroscience Institute, Department of Psychology, University of Cambridge, Cambridge, United Kingdom, <sup>4</sup>Department of Psychology, Wayne State University, Detroit, MI, United States, <sup>5</sup>Department of Family Medicine and Public Health Sciences, Wayne State University, Detroit, MI, United States, <sup>6</sup>Neuropsychopharmacology and Biopsychology Unit, Department of Basic Psychological Research and Research Methods, Faculty of Psychology, University of Vienna, Vienna, Austria, <sup>7</sup>Centre for Gambling Research at UBC, Department of Psychology, University of British Columbia, Vancouver, BC, Canada, <sup>8</sup>Center for Language and Brain, Shenzhen Institute of Neuroscience, Shenzhen, China

## OPEN ACCESS

### Edited by:

Pablo Brañas-Garza,  
Middlesex University,  
United Kingdom

### Reviewed by:

Antonio M. Espín,  
Middlesex University,  
United Kingdom  
Jeroen Nieboer,  
London School of Economics and  
Political Science, United Kingdom

### \*Correspondence:

Yin Wu  
yinwu0407@gmail.com  
Hong Li  
lihongszu@szu.edu.cn

**Received:** 31 May 2017

**Accepted:** 02 October 2017

**Published:** 16 October 2017

### Citation:

Wu Y, Zilioli S, Eisenegger C, Clark L and Li H (2017) The Effect of Testosterone Administration and Digit Ratio (2D:4D) on Implicit Preference for Status Goods in Healthy Males. *Front. Behav. Neurosci.* 11:193. doi: 10.3389/fnbeh.2017.00193

Testosterone has been linked to social status seeking in humans. The present study investigated the effects of testosterone administration on implicit and explicit preferences for status goods in healthy male participants ( $n = 64$ ), using a double-blind, placebo-controlled, between-subjects design. We also investigated the interactive effect between second-to-fourth digit ratio (2D:4D; i.e., a proximal index of prenatal testosterone) and testosterone treatment on status preferences. Results showed that testosterone administration has no discernable influence on self-reported willingness-to-pay (i.e., the explicit measure) or implicit attitudes towards status goods. Individuals with lower 2D:4D (i.e., more masculine) had more positive attitudes for high-status goods on an Implicit Association Task, and this association was abolished with testosterone administration. These data suggest interactive effects of acute testosterone administration and prenatal testosterone exposure on human social status seeking, and highlight the utility of implicit methods for measuring status-related behavior.

**Keywords:** steroid hormones, social status, conspicuous consumption, implicit association test, prenatal priming

## INTRODUCTION

Testosterone, a steroid hormone produced primarily by the gonads, is implicated in dominant behaviors and decision-making process. For instance, lower second-to-fourth digit ratio (2D:4D; a proximal index of high exposure to prenatal testosterone in the womb) is associated with a higher number of correct answers in the Cognitive Reflection Test (CRT; Bosch-Domenech et al., 2014; but see Nave et al., 2017), a task measuring the tendency to override an intuitive response that is incorrect. Recent research suggests that the role of testosterone in human social interaction is best understood in terms of the search for, and maintenance of, social status (Eisenegger et al., 2011). In the Ultimatum Game (UG), the proposer faces the threat of rejection if he or she makes an unfair offer. By making a fair offer, the proposer can prevent being turned down, and the rejection rate is usually high for unfair offers (Güth et al., 1982). Testosterone increases the concern for



status in the UG such that the proposers perceive a rejection of their offers as more aversive, leading them to make fairer offers (Eisenegger et al., 2010).

Possessions and goods contribute to defining the self and become an extension of one's identity. Individuals can acquire and signal their status within social hierarchies by purchasing and displaying luxury goods, a phenomenon termed "conspicuous consumption" (Veblen, 1899; Sivanathan and Pettit, 2010). Previous research has demonstrated a link between testosterone and consumer behavior. For example, individuals with lower 2D:4D (i.e., more masculinized) were more responsive to the status-related consumption experience such that they were more interested in luxury goods after being primed by mate attraction goals or status display goals (Cornelissen and Palacios-Fenech, 2016). Lower 2D:4D (i.e., more masculinized) was also associated with greater desire to offer erotic gifts to a romantic partner among men with high mating confidence (Nepomuceno et al., 2016a). In one study, salivary testosterone levels increased after driving an expensive sport car (compared to an old station wagon), and this effect was stronger if the experiment took place in a busy downtown area (compared to a semi-deserted highway). Furthermore, the effect of car-induced testosterone increase was enhanced when men's social status was threatened by the wealth displays of a male confederate in the face of a female moderator (Saad and Vongas, 2009). Taken together, these data suggest a link between testosterone levels and displays of high status. However, whether and how testosterone causally influences attitudes and consumption of status-related goods has not been empirically tested.

The aim of the present study was to investigate the effects of a single dose of testosterone on preference for status goods, in a double-blind, placebo-controlled, between-subjects design. Preference for goods can be measured by the Implicit Association Test (i.e., IAT), which has been employed in recent psychopharmacological studies (De Dreu et al., 2011; Terbeck et al., 2012). The IAT is a reliable technique to assess implicit social evaluation, and has been used extensively in the study of attitudes (e.g., racial bias and stereotype; Greenwald et al., 1998). The IAT has also been utilized in the consumer research such that IAT-measured attitudes could predict brand preference, usage and recognition (Maison et al., 2004). In the current version of the IAT, participants categorized positive words and high-status goods with one key, and negative words and low-status goods with another key. In a different task block, the pairings were reversed such that positive words and low-status goods were categorized together. Participants who hold more positive attitudes towards high-status goods (and/or negative attitudes towards low-status goods) should respond faster in the first block compared to the second block. We hypothesized that this difference would be enhanced following testosterone administration. We further tested whether these effects of testosterone were moderated by second-digit-to-fourth digit ratio (2D:4D; van Honk et al., 2011; Carré et al., 2015), a putative indicator of prenatal testosterone exposure obtained by scanning participants' right hands, which plays a large role in brain organization and gendered behavior.

Lastly, we measured participants' explicit evaluations of the status goods by obtaining willingness-to-pay ratings in a standard consumer psychology procedure (Rucker and Galinsky, 2008).

## MATERIALS AND METHODS

### Participants

Sixty-four healthy males (mean age = 22.6 years,  $SD = 1.7$ ; age range = 20–27) were recruited through university advertisements. All participants were screened during a telephone interview to exclude individuals taking psychotropic medications, or having any psychiatric or neurological disorders. We only recruited males, as the dosing and pharmacokinetics associated with single dose Androgel administration are only established for men (Eisenegger et al., 2013). Participants were instructed to abstain from alcohol, caffeine intake and smoking for 24 h before the testing session. Each participant received a single dose of Androgel or placebo gel in a double blind, placebo-controlled, between-subjects design. This study was carried out in accordance with Declaration of Helsinki and was approved by Shenzhen University Medical Research Ethics Committee. Written informed consent was obtained from all participants. Participants were paid 200 Chinese Yuan (~\$30) as their reimbursement.

### Testosterone Administration

All sessions started at 13:00 and lasted approximately 4 h. Participants in the testosterone group received a single dose of testosterone gel, containing 150 mg testosterone [Androgel®]. Participants in the placebo group received colorless hydroalcoholic gel. The gels were applied on the shoulders and upper arms by a male research assistant who was blind to the purpose of the study. Given the 3 h time lag for effects with testosterone gel administration in healthy males (we have corroborated that salivary testosterone levels peaked 3 h after gel administration in an independent sample, not reported here), we began our experimental tasks 3 h post-dosing (Eisenegger et al., 2013). Cognitive testing also involved two further decision-making tasks, not reported here. During the waiting period, participants rested in the laboratory.

### Validation of the Stimulus Set

We validated the experimental stimuli in an independent male sample ( $N = 27$ ). These participants rated the prestige associated with a series of cars (1 = lowest, 9 = highest). As predicted, our high-status cars ( $M = 7.47$ ,  $SD = 1.27$ ; i.e., Porsche, BMW, Ferrari, Maserati, Mercedes-Benz) were rated as more prestigious than low-status cars ( $M = 2.65$ ,  $SD = 1.04$ ; i.e., BYD, Cherry, Dongfeng, Geely, Great Wall) on average,  $t_{(26)} = 17.12$ ,  $p < 0.001$ .

### Implicit Association Test

The IAT (Greenwald et al., 1998) involved two target categories (high-status vs. low-status car stimuli) and two attribute categories (positive vs. negative). The order of congruent and incongruent blocks was randomly assigned. The IAT data were

analyzed using the algorithm from Greenwald et al. (1998). The first two trials of each block were excluded due to typically long response latencies. Next, we excluded latencies below 300 ms and above 3000 ms as outliers due to anticipation or inattention. The average error rate was 3.73% ( $SD = 2.91\%$ ), ranging between 0% and 12.50%. Response latencies were log-transformed for analysis. The IAT effect was calculated as the difference between response latencies for incongruent blocks (high-status stimuli + negative words, low-status stimuli + positive words) compared to congruent blocks (high-status stimuli + positive words, low-status stimuli + negative words; Greenwald et al., 1998), such that higher scores indicate more positive attitudes for high-status goods and/or more negative attitudes for low-status goods.

## Explicit Valuation Measure

For the explicit measure, we presented participants the same car stimuli, and asked them “How much would you be willing to pay for the product featured?”, with 1 = 10% of the retail price of the item, 2 = 20% of the retail price of the item, and increasing intervals of 10% up to 12 = 120% of the retail price. We calculated a difference score between willingness to pay for high-status vs. low-status goods as the dependent variable, with more positive values representing greater explicit preferences for high status goods.

## Digit Ratio Measurement

Digit ratio was measured from an image scan of the right hand, measuring the length of the index (2D) and ring (4D) fingers from the ventral proximal crease to the tip of the finger using Adobe Photoshop. The scan was performed at the start of each testing session, and each participant provided consent for his fingers to be scanned. Two research assistants, who were blind to the purpose of the experiment, measured the 2D:4D ratios on three occasions, and the mean value was used for analysis. Inter-rater reliability was high,  $r = 0.94$ ,  $p < 0.001$ .

## Mood Measurement

We used the Positive Affect and Negative Affect Scale (PANAS; Watson et al., 1988) to measure state mood before and after testosterone administration.

## Statistical Analysis

We first compared the IAT and WTP scores between the testosterone and placebo conditions using independent-samples  $t$  tests. We then looked at the interactive effect between testosterone treatment and 2D:4D ratio on these two dependent variables by using linear regression model.

## RESULTS

Participants did not differ from chance in guessing whether they had received testosterone or placebo in the experiment,  $\chi^2 = 0.016$ ,  $df = 1$ ,  $p > 0.1$ . On the PANAS mood ratings, testosterone had no effect on positive affect (testosterone group,  $M = -0.23$ ,  $SD = 0.58$ ; placebo group,  $M = -0.17$ ,  $SD = 0.42$ ),  $t_{(62)} = 0.49$ ,  $p = 0.62$ , or negative affect (testosterone group,

$M = 0.02$ ,  $SD = 0.28$ ; placebo group,  $M = -0.11$ ,  $SD = 0.35$ ),  $t_{(62)} = -1.62$ ,  $p = 0.11$ .

We first investigated whether the testosterone treatment influenced preferences for status goods. Independent samples  $t$ -test revealed no significant difference in either IAT scores,  $t_{(62)} = -0.63$ ,  $p = 0.53$ , or self-reported willingness-to-pay,  $t_{(62)} = -0.47$ ,  $p = 0.64$ .

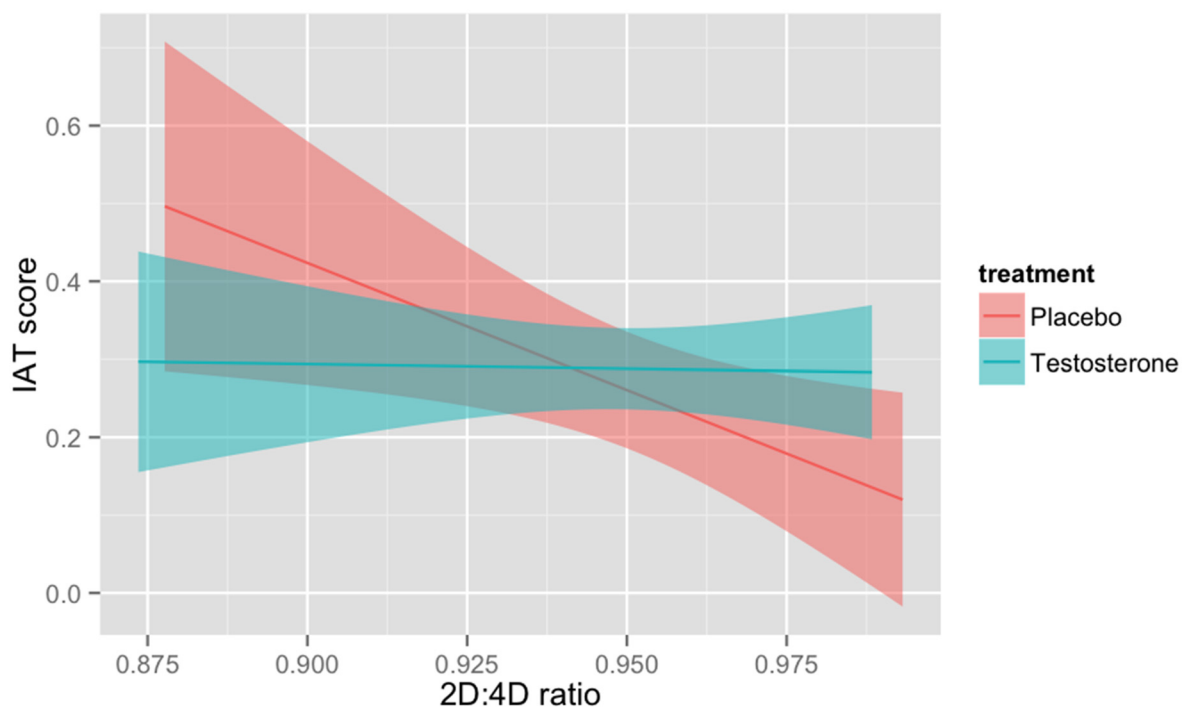
Next, in order to investigate the interaction between testosterone administration and 2D:4D ratio, we first regressed IAT scores against testosterone treatment and 2D:4D using linear regression model as Model 1. There was a significant main effect of 2D:4D,  $b = -1.56$ ,  $SE = 0.80$ ,  $t = -1.96$ ,  $p = 0.05$ . The main effect of testosterone treatment was not significant,  $b = 0.03$ ,  $SE = 0.05$ ,  $t = 0.59$ ,  $p = 0.56$ . In Model 2, the interactive term between treatment and 2D:4D was entered. The overall linear regression model was significant ( $R^2 = 0.12$ , adjusted  $R^2 = 0.08$ ,  $F_{(3,60)} = 2.85$ ,  $p = 0.04$ ). Adding 2D:4D into the model significantly increased the amount of variance explained,  $\Delta F_{(1,60)} = 3.91$ ,  $p = 0.05$ ,  $\Delta R^2 = 0.12$ . We decomposed the interaction by looking at the relationship between 2D:4D and IAT score in the testosterone and placebo groups separately (see **Figure 1**). In the placebo group, the association between 2D:4D and IAT score was significant,  $b = -3.26$ ,  $SE = 1.33$ ,  $t = -2.45$ ,  $p = 0.02$ , suggesting individuals with lower 2D:4D had stronger preference for status-goods. Importantly, this relationship was absent in testosterone administration group,  $b = -0.12$ ,  $SE = 0.85$ ,  $t = -0.14$ ,  $p = 0.89$ .

To further interpret the significant interaction, we also conducted a simple slope analyses for digit ratio 1  $SD$  below the mean and 1  $SD$  above the mean (Aiken and West, 1991; Cohen et al., 2013). Testosterone marginally increased IAT scores among individuals scoring relatively high (1  $SD$  above the mean) on 2D:4D,  $b = 0.12$ ,  $SE = 0.06$ ,  $t = 1.86$ ,  $p = 0.06$ , and testosterone had no reliable effect among individuals low (1  $SD$  below the mean) on 2D:4D,  $b = -0.06$ ,  $SE = 0.06$ ,  $t = -1.01$ ,  $p = 0.32$ . As an additional approach to understand the treatment by digit ratio interaction, we created low and high 2D:4D groups by conducting median splits (median split is a valid robustness check). For individuals in the high 2D:4D group, there was a significant main effect of treatment,  $b = 0.12$ ,  $SE = 0.06$ ,  $t = 1.99$ ,  $p = 0.05$ . For the low 2D:4D group, the main effect of treatment was not significant,  $b = -0.05$ ,  $SE = 0.07$ ,  $t = -0.75$ ,  $p = 0.46$ . Thus the median split analyses showed the same pattern as simple slope analyses.

For the explicit measurement, there were no significant main effects of testosterone treatment,  $b = 2.75$ ,  $SE = 14.44$ ,  $t = 0.19$ ,  $p = 0.85$ , or 2D:4D,  $b = -1.13$ ,  $SE = 11.17$ ,  $t = -0.10$ ,  $p = 0.92$ , and the interaction term was also not significant,  $b = -2.69$ ,  $SE = 15.19$ ,  $t = -0.18$ ,  $p = 0.86$ . There was no significant correlation between the explicit measurement and the IAT score,  $t_{(62)} = 1.05$ ,  $p = 0.30$ .

## DISCUSSION

The present study investigated the effect of testosterone on implicit and explicit preferences for status goods in healthy males. Exogenous testosterone increased IAT scores for status



**FIGURE 1 |** Interactive effect of second-to-fourth digit ratio (2D:4D) and testosterone administration on implicit association test (IAT) score (log-transformed).

goods among individuals with higher 2D:4D ratios (i.e., less masculine), consistent with past work showing the interaction between testosterone administration and prenatal testosterone exposure in human social interaction (van Honk et al., 2012). The status theory of testosterone predicts that, while in social contexts where status is threatened by perceived provocation (e.g., unfair offers in the UG), this motivation may lead to increased aggression (rejection behavior); in the other case, non-aggressive behavior such as generosity, will be more appropriate for increasing social status (Eisenegger et al., 2011). Using the UG, previous research has found that participants treated with testosterone were more likely to punish the proposer who made unfair offers, and more likely to reward the proposer who made fair offers, consistent with a causal role of testosterone in status-enhancing behaviors dependent on the social context (Dreher et al., 2016). Notably, in the current study, the effect of acute testosterone was driven by individuals with higher 2D:4D ratio (lower prenatal testosterone exposure), consistent with the proposal that the effects of testosterone on social behavior are largely due to metabolism to estradiol, and individuals who are prenatally more primed by estradiol (higher 2D:4D) could metabolize more testosterone into estradiol (van Honk et al., 2012).

The main effect of lower 2D:4D on status preferences on the IAT also corroborates previous research showing that high prenatal testosterone in men predicts courtship-related consumption (i.e., display resources and status as to impress women; Nepomuceno et al., 2016b). Lower 2D:4D ratio is associated with more risky choice and more masculine traits

such as aggression, dominance and better performance in sports competition (Coates et al., 2009; Sapienza et al., 2009; Apicella et al., 2015). For instance, 2D:4D ratio is significantly associated with risk preferences over lotteries with real monetary incentives (Brañas-Garza et al., in press). Recent research also showed that 2D:4D ratio correlates with social network centrality (Kovářík et al., 2017). In the current study, this association was abolished by testosterone treatment, possibly due to the enhancing effect of testosterone for status-goods among individuals with higher 2D:4D ratio (less masculine).

In the present study, testosterone has no observable effect on self-reported willingness-to-pay. It has been suggested that human social-status seeking often takes various implicit forms rather than being overly explicit, i.e., physical aggression (Eisenegger et al., 2011). The present study used the IAT to measure attitudes for status goods, a technique that is less susceptible to social desirability biases and demand characteristics. This extends recent testosterone research that employs implicit measures such as implicit power motivation, an indirect measure of individual differences in dominance disposition (Stanton and Schultheiss, 2009), in investigating the relationship with dominance behavior. The present data highlight the utility of using IAT as an indirect measure of human status preference.

Some limitations of the study should be noted. First, our experiment tested exclusively male participants since the pharmacokinetic data on testosterone gel is clear in healthy young males (Eisenegger et al., 2013) and social status seeking is more prevalent among males (Eisenegger et al., 2011). Future

work would benefit from including both genders in the same design to enable direct comparisons to be tested. Second, we selected the car stimuli only based on the “status” dimension, future work should more precisely control the possible confound of quality (i.e., speed of the cars) or familiarity of the stimuli upon implicit and explicit of evaluation. Third, WTP is a kind of self-report in hypothetical scenario, thus it has no consequence for the decisions the participants made. We encourage future research to utilize incentive-compatible paradigms in measuring preference, e.g., Becker-DeGroot-Marschak auction (Becker et al., 1964).

## AUTHOR CONTRIBUTIONS

YW, SZ, CE, LC and HL developed the concepts for the study. YW collected the data. YW analyzed the data. All authors contributed to the manuscript and approved the final version of the manuscript for submission.

## REFERENCES

- Aiken, L. S., and West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. London: Sage.
- Apicella, C. L., Carré, J. M., and Dreber, A. (2015). Testosterone and economic risk taking: a review. *Adapt. Hum. Behav. Physiol.* 1, 358–385. doi: 10.1007/s40750-014-0020-2
- Becker, G. M., DeGroot, M. H., and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behav. Sci.* 9, 226–232. doi: 10.1002/bs.3830090304
- Bosch-Domènech, A., Brañas-Garza, P., and Espín, A. M. (2014). Can exposure to prenatal sex hormones (2D:4D) predict cognitive reflection? *Psychoneuroendocrinology* 43, 1–10. doi: 10.1016/j.psyneuen.2014.01.023
- Brañas-Garza, P., Galizzi, 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.*
- Carré, J. M., Ortiz, T. L., Labine, B., Moreau, B. J. P., Viding, E., Neumann, C. S., et al. (2015). Digit ratio (2D:4D) and psychopathic traits moderate the effect of exogenous testosterone on socio-cognitive processes in men. *Psychoneuroendocrinology* 62, 319–326. doi: 10.1016/j.psyneuen.2015.08.023
- Coates, J. M., Gurnell, M., and Rustichini, A. (2009). Second-to-fourth digit ratio predicts success among high-frequency financial traders. *Proc. Natl. Acad. Sci. U S A* 106, 623–628. doi: 10.1073/pnas.0810907106
- Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Abingdon: Routledge.
- Cornelissen, G., and Palacios-Fenech, J. (2016). Status-relevant experiences and conspicuous consumption—the moderating role of prenatal androgen exposure. *Span. J. Psychol.* 19:E54. doi: 10.1017/sjp.2016.52
- De Dreu, C. K. W., Greer, L. L., Van Kleef, G. A., Shalvi, S., and Handgraaf, M. J. J. (2011). Oxytocin promotes human ethnocentrism. *Proc. Natl. Acad. Sci. U S A* 108, 1262–1266. doi: 10.1073/pnas.1015316108
- Dreher, J., Dunne, S., Pazderska, A., Frodl, T., Nolan, J. J., and O'Doherty, J. P. (2016). Testosterone causes both prosocial and antisocial status-enhancing behaviors in human males. *Proc. Natl. Acad. Sci. U S A* 113, 11633–11638. doi: 10.1073/pnas.1608085113
- Eisenegger, C., Haushofer, J., and Fehr, E. (2011). The role of testosterone in social interaction. *Trends Cogn. Sci.* 15, 263–271. doi: 10.1016/j.tics.2011.04.008
- 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
- Eisenegger, C., von Eckardstein, A., Fehr, E., and von Eckardstein, S. (2013). Pharmacokinetics of testosterone and estradiol gel preparations in healthy

## FUNDING

This work was supported by National Natural Science Foundation of China (31600923), Shenzhen University Natural Science Research Fund (2016073), Shenzhen University Social and Humanity Science Research Fund (17QNFC44) and Treherne Studentship in Biological Sciences (Downing College, Cambridge). CE was supported by the Vienna Science and Technology Fund (WWTF VRG13-007). HL was supported by Shenzhen Peacock Plan (KQTD2015033016104926). The funding sources had no further role in the study design, data collection, analysis, interpretation, or decision to submit this manuscript for publication.

## ACKNOWLEDGMENTS

We are grateful to Dr. Jinting Liu for her help with data collection.

- young men. *Psychoneuroendocrinology* 38, 171–178. doi: 10.1016/j.psyneuen.2012.05.018
- Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74, 1464–1480. doi: 10.1037/0022-3514.74.6.1464
- 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
- Kovářík, J., Branas-Garza, P., Davidson, M. W., Haim, D. A., Carcelli, S., and Fowler, J. H. (2017). “Digit ratio (2D:4D) and social integration: an effect of prenatal sex hormones,” in *Network Science*, ed. A. McCranie (Cambridge, MA: Cambridge University Press), 1–14.
- Maison, D., Greenwald, A. G., and Bruin, R. H. (2004). Predictive validity of the implicit association test in studies of brands, consumer attitudes, and behavior. *J. Consum. Psychol.* 14, 405–415. doi: 10.1207/s15327663jcp1404\_9
- Nave, G., Nadler, A., Zava, D., and Camerer, C. (2017). Single dose testosterone impairs cognitive reflection in men. *Psychol. Sci.* doi: 10.1177/0956797617709592 [Epub ahead of print].
- Nepomuceno, M. V., Saad, G., Stenstrom, E., Mendenhall, Z., and Iglesias, F. (2016a). Testosterone at your fingertips: digit ratios (2D:4D and rel2) as predictors of courtship-related consumption intended to acquire and retain mates. *J. Consum. Psychol.* 26, 231–244. doi: 10.1016/j.jcps.2015.05.007
- Nepomuceno, M. V., Saad, G., Stenstrom, E., Mendenhall, Z., and Iglesias, F. (2016b). Testosterone and gift-giving: mating confidence moderates the association between digit ratios (2D:4D and rel2) and erotic gift-giving. *Pers. Individ. Dif.* 91, 27–30. doi: 10.1016/j.paid.2015.11.017
- Rucker, D. D., and Galinsky, A. D. (2008). Desire to acquire: powerlessness and compensatory consumption. *J. Consum. Res.* 35, 257–267. doi: 10.1086/588569
- Saad, G., and Vongas, J. G. (2009). The effect of conspicuous consumption on men's testosterone levels. *Organ. Behav. Hum. Decis. Process.* 110, 80–92. doi: 10.1016/j.obhdp.2009.06.001
- Sapienza, P., Zingales, L., and Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proc. Natl. Acad. Sci. U S A* 106, 15268–15273. doi: 10.1073/pnas.0907352106
- Sivanathan, N., and Pettit, N. C. (2010). Protecting the self through consumption: status goods as affirmational commodities. *J. Exp. Soc. Psychol.* 46, 564–570. doi: 10.1016/j.jesp.2010.01.006
- Stanton, S. J., and Schultheiss, O. C. (2009). The hormonal correlates of implicit power motivation. *J. Res. Pers.* 43, 942–949. doi: 10.1016/j.jrp.2009.04.001



- Terbeck, S., Kahane, G., McTavish, S., Savulescu, J., Cowen, P. J., and Hewstone, M. (2012). Propranolol reduces implicit negative racial bias. *Psychopharmacology* 222, 419–424. doi: 10.1007/s00213-012-2657-5
- 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
- van Honk, J., Schutter, D. J., Bos, P. A., Kruijt, A.-W., Lentjes, E. G., and Baron-Cohen, S. (2011). Testosterone administration impairs cognitive empathy in women depending on second-to-fourth digit ratio. *Proc. Natl. Acad. Sci. U S A* 108, 3448–3452. doi: 10.1073/pnas.1011891108
- Veblen, T. (1899). *The Theory of the Leisure Class*. New York, NY: Macmillan.
- Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070. doi: 10.1037//0022-3514.54.6.1063

**Conflict of Interest Statement:** LC: The Centre for Gambling Research at UBC is supported by funding from the British Columbia Lottery Corporation and the Province of British Columbia.

The other 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.

The reviewer AME and handling Editor declared their shared affiliation.

Copyright © 2017 Wu, Zilioli, Eisenegger, Clark and Li. 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.



# Testosterone and Cortisol Jointly Predict the Ambiguity Premium in an Ellsberg-Urns Experiment

Giuseppe Danese<sup>1\*</sup>, Eugénia Fernandes<sup>2</sup>, Neil V. Watson<sup>3</sup> and Samuele Zilioli<sup>4</sup>

<sup>1</sup> Católica Porto Business School and CEGE, Universidade Católica Portuguesa, Porto, Portugal, <sup>2</sup> Neuropsychophysiology Lab, Centro de Investigação em Psicologia, Escola de Psicologia, Universidade do Minho, Braga, Portugal, <sup>3</sup> Behavioral Endocrinology Laboratory, Department of Psychology, Simon Fraser University, Burnaby, BC, Canada, <sup>4</sup> Department of Family Medicine and Public Health Sciences, Wayne State University, Detroit, MI, USA

Previous literature has tried to establish whether and how steroid hormones are related to economic risk-taking. In this study, we investigate the relationship between testosterone (T) and cortisol (C) on one side and attitudes toward risk and ambiguity on the other. We asked 78 male undergraduate students to complete several tasks and provide two saliva samples. In the task “Reveal the Bag,” participants expressed their beliefs on an ambiguous situation in an incentivized framework. In the task “Ellsberg Bags,” we elicited from the participants through an incentive-compatible mechanism the reservation prices for a risky bet and an ambiguous bet. We used the difference between the two prices to calculate each participant’s ambiguity premium. We found that participants’ salivary T and C levels jointly predicted the ambiguity premium. Participants featuring comparatively lower levels of T and C showed the highest levels of ambiguity aversion. The beliefs expressed by a subset of participants in the “Reveal the Bag” task rationalize (in a revealed preference sense) their choices in the “Ellsberg Bags” task.

**Keywords:** testosterone, cortisol, ambiguity, Ellsberg paradox, dual hormone hypothesis

## OPEN ACCESS

### Edited by:

Levent Neyse,  
Institut für Weltwirtschaft, Germany

### Reviewed by:

Patrick Ring,  
Institut für Weltwirtschaft, Germany  
Matteo M. Galizzi,  
London School of Economics and  
Political Science (LSE), UK

### \*Correspondence:

Giuseppe Danese  
gdanese@porto.ucp.pt

**Received:** 21 December 2016

**Accepted:** 03 April 2017

**Published:** 21 April 2017

### Citation:

Danese G, Fernandes E, Watson NV  
and Zilioli S (2017) Testosterone and  
Cortisol Jointly Predict the Ambiguity  
Premium in an Ellsberg-Urns  
Experiment.  
Front. Behav. Neurosci. 11:68.  
doi: 10.3389/fnbeh.2017.00068

## INTRODUCTION

Many papers study the relationship between hormones and economic risk-taking. Comparatively, fewer papers in behavioral endocrinology consider the fact that humans face different types of risk, and that these different types of risk might have different endocrine correlates. In economics and the management sciences, however, the distinction between risk proper and uncertainty (or ambiguity) has been customary ever since Knight (1921) first discussed the difference, followed decades later by Ellsberg (1961).

In one of Ellsberg’s famous thought experiments, the decision maker can place bets on a black marble being drawn either from a bag with a known proportion of black and white marbles (the “risky” bag), or from a bag with unknown proportions (the “ambiguous” or “uncertain” bag). Once the participant chooses the bag, one marble is drawn. The color of the marble extracted determines whether the payoff is positive (if a black marble is drawn) or zero. Ellsberg speculated that decision makers would prefer to bet on the risky bag. He also speculated that this preference would likely hold regardless of the winning color (black or white), a finding confirmed in human and even in primate studies (cf. e.g., Hayden et al., 2010). These choices are inconsistent with the rational model of decision under uncertainty (Savage, 1972) and have given rise to many behavioral models that try to explain the preference for known-odds gambles (e.g., Gilboa and Schmeidler, 1989; Klibanoff et al., 2005; Seo, 2009).

The participants' ambiguity premium—the difference between the price the participants set to sell the bet on the risky bag minus the price they set for the bet on the ambiguous bag—provides a discrete measure of the strength of the participants' preference for known odds. A positive ambiguity premium is consistent with Ellsberg's insight that many participants might prefer known odds. A zero premium is consistent with a decision maker who does not differentiate between an equiprobable win or loss and complete lack of information about the chances of winning or losing. A negative premium implies a preference for ambiguous decisional situations. The most intuitive way to understand what kind of information the ambiguity premium conveys is the participants' willingness to pay to go from multiple possible scenarios about the content of the bag to one possible scenario only (equal probability of winning or obtaining zero).

Similarly to what has happened for other decisions that are inconsistent with economic theory, there has been an increasing effort to identify both the neural and hormonal correlates of anomalous behavior in risky and ambiguous situations. Functional magnetic resonance imaging studies have found that the representation of the subjective value of the risky and the ambiguous options seem to take place in the same area of the brain (the striatum and the medial prefrontal cortex; cf. Hsu et al., 2005; Levy et al., 2010). As for the role of hormones, the “dual-hormone hypothesis” (DHH) proposed that several types of human behaviors are explained by an interaction between Testosterone (T) and Cortisol (C). Mehta and Josephs (2010) suggest that status-seeking behaviors are to be expected among individuals with simultaneously high T and low C. The DHH seems to account for a growing number of results from studies on human aggressiveness, empathy, risk-seeking, status-seeking behavior, and overbidding in auctions (cf. Mehta and Prasad, 2015; Pfattheicher, 2017).

The theoretical forerunners of the DHH are the earlier studies that found that T correlated with aggressive behavior only in low-C offenders (Dabbs et al., 1991; Popma et al., 2007; Tackett et al., 2014). Many more papers studied the behavioral correlates of C only (levels or changes), of T only, or of both T and C, without controlling for the presence of interaction effects of C and T. Concerning C, low levels of this hormone were associated with fearlessness and reduced sensitivity to punishment and threats (Van Honk et al., 2003). On the other hand, high levels of C seemed to predict higher anxiety (Brown et al., 1996). T was found to be positively associated with dominance in social hierarchies, status-seeking behavior and success in competition both in animals and humans (the “challenge hypothesis,” cf. Mazur and Booth, 1998; Oliveira and Oliveira, 2014; Casto and Edwards, 2016; Wingfield, 2016).

The rationale for studying the endocrine correlates of economic risk is that risk-taking might have evolved as a way to increase status (Daly and Wilson, 1997; Ellis et al., 2012). Several studies have found a positive relation between risk-taking and T (cf., e.g., Apicella et al., 2008; Sapienza et al., 2009; Zilioli and Watson, 2012; but cf. the null results in Schipper, 2014; Cueva et al., 2015). Coates and Herbert (2008) found that traders in the City of London have significantly higher T levels on days

when they made more than their 1-month daily average. The authors also found a strong positive correlation between the traders' daily C levels and the volatility of their net earnings on the day of the study. Van Honk et al. (2003) found that basal C negatively correlates with risky choices. Kandasamy et al. (2014) found that chronic (i.e., cumulative over several days) C exposure increased risk aversion. Mehta et al. (2015) found that basal testosterone is associated with higher financial risk-taking behaviors, but only for low C subjects, as predicted by the DHH.

Regarding the endocrine correlates of different types of risk, to the best of our knowledge T and C have not been addressed together in the same study. Stanton et al. (2011) found that neither the risk premium nor the ambiguity premium had a significant linear relationship with T (the predictor), and there were instead significant non-linearities in the relationship. Specifically, individuals that were risk and ambiguity averse were the ones who presented intermediate levels of T and individuals neutral to risk and ambiguity were at the two extremes of the distribution of T. Interestingly, their ambiguity task measured the participants' preferences between a situation of radical uncertainty vs. a situation of complete certainty. Their measure of the ambiguity premium is therefore not consistent with Ellsberg's thought experiment, i.e., a situation of known odds vs. unknown odds. Buckert et al. (2014) studied the relation between C, stress and decisions under risk and ambiguity. They found that after undergoing a stress induction protocol, the cortisol response did not affect the percentage of choices of the ambiguous option.

Any conclusion about the sign of the relationship between T and C and decisions under risk and ambiguity is complicated by the variety of tasks used in the literature (Schonberg et al., 2011); the differences between measuring circulating hormones from saliva, allocating participants to receive T (cf. e.g., Zethraeus et al., 2009, failing to find any effect of administering T on a variety of economic tasks) or proxying prenatal T by the 2D:4D finger ratio (Brañas-Garza and Rustichini, 2011); the sample used (the role of gender in particular, cf. e.g., Borghans et al., 2009). We re-examine the relation between T and C and ambiguity attitudes in Ellsberg's original framework. We use an incentive compatible elicitation mechanism to obtain a numeric measure of the participants' ambiguity premium. We design a novel task to elicit the beliefs of the players about an ambiguous situation, as these beliefs are not directly observable in Ellsberg-type experiments.

In line with previous behavioral economics research that links beliefs and choices (cf. e.g., Gilboa and Schmeidler, 1989), we expect to find a strong relationship between the beliefs of the players and their choices in the Ellsberg experiment. In addition, our design allows us to explore if the result in the literature concerning the positive association between T and risk-seeking behavior holds when different types of risks are involved. No previous study allows us to predict whether someone characterized by higher T would prefer knowing the odds and place a higher reservation price for the risky bet or prefer the ambiguous bet. It is also not clear *ex-ante* whether an ambiguity averse individual should exhibit endocrine correlates

of higher stress, as we would expect based on some previous studies linking high C to pronounced risk aversion (Kandasamy et al., 2014; but cf. also Buckert et al., 2014). In our case, the outcome of both bets is unpredictable, and we lack the risk-free (degenerate) lottery that is often used to ascertain whether an individual is risk averse, risk seeking, or risk neutral. On the issue of the relationship between C and T, and their interaction, and the ambiguity premium, our analysis by necessity will be exploratory.

## MATERIALS AND METHODS

### Participants

Seventy-eight students participated in our experiments after they responded to a public announcement. The study was reviewed by the Office of Research Ethics of Simon Fraser University, and all participants provided written consent before the start of the experimental procedures. Exclusion criteria for the participation included (i) eating, drinking liquids other than water, smoking or brushing teeth in the hour prior to the session; (ii) consuming alcohol or drugs in the previous 12 h; (iii) intense physical activity on the day of the experiment; (iv) having a recent history of smoking more than 5 cigarettes a day, or of taking a medication that affects hormonal levels; (v) having bleeding gums and an oral infection. All participants were male undergraduate and graduate students of Simon Fraser University (mean age = 22.60,  $SD = 4.44$ , range = 18–42 years). The participants earned on average \$19 Canadian during the experiment, the sum of their earnings in all the tasks they performed. Earnings were paid in cash at the end of the experiment. The entire experiment lasted on average 1 h 30 min.

### Procedure

The experiment took place in the afternoon (mean time: 2:30 PM,  $SD = 1$  h 31 min). At the beginning of the session, participants completed a survey about their socio-demographic features as well as their recent health state. While completing the questionnaires, participants provided a salivary sample (see below). Afterward, they were tested in the three economic tasks explained in details in the next section (“Reveal the Bag” task or RB, “Ellsberg Bags” task or EB, “Monty Hall” task or MH). Tasks RB and EB were offered in random order, with 35% of the participants taking the RB task first. The MH task was always offered last. Instructions for all tasks are provided in the Supplementary Material (SM) to the article available online. Tasks RB and EB were implemented without the help of computers, using bags filled with real marbles. The bags used were always randomly extracted from a shelf protected by a curtain visible to the participants, before they made their choices. This procedure was dictated by the desire to limit “malicious experimenter effects,” whereby the participant might believe that the experimenter (or the machine) filled the bags after having learned of the bets of the participants (cf. e.g., Kadane, 1992; Kühberger and Perner, 2003; Pulford, 2009). We provided the instructions of the tasks one at a time, and therefore subjects could not formulate at the beginning of the experimental session a strategy for each of the three tasks. To control for order effects of the tasks, we included a dummy for the order in which the

RB task was offered as a robustness check (see below). Given the modest amounts of money at stake, we believe it is unlikely, but cannot ultimately exclude, that the amounts won in a previous task created an “endowment effect” (Kahneman et al., 1991) which affected the ensuing choices.

Interspersed with the three economic tasks, participants completed the BIS/BAS (Carver and White, 1994), Levenson’s LSRP (Levenson et al., 1995) and Rotter’s Internal-External Locus of Control Scale (Rotter, 1966) questionnaires. After all tasks and questionnaires were completed, participants provided a face picture and a scan of both their hands (not used for the analysis in this paper). Afterwards, a second salivary sample was collected. Subjects were then paid their earnings in each of the three tasks. The exchange rate was communicated in the instructions of each task (1 experimental point was always worth \$0.20 Canadian). Subjects at this point left the experimental room.

### “Reveal the Bag” Task (RB)

We devised this novel task to elicit the beliefs of the players about an uncertain situation. The experimenter presented to each participant a bag. Participants were informed that the bag contained 10 marbles, and each marble could be either white or black. Participants were asked to guess the bag’s content. The experimenter randomly picked one bag from the shelf described above. Bags were replaced behind the curtain at the end of every participant’s RB task. The extraction of the bag from behind the curtain was done in front of the participant.

Participants had 10 opportunities (or trials) to guess the composition of the bag and at the end of the 10th trial, they received a monetary reward according to the accuracy of their guesses. Each participant expressed his guesses on a white sheet that featured a printed grid of 11 columns and 10 rows (a sample sheet is reproduced in the SM). The rows represented the trials and the 11 columns represented all the possible scenarios for the composition of the bag, from all black to all white marbles. On each trial, the participant could bet on the bag composition by distributing 11 folder markers (small round stickers) along one row. When asked for the first time to guess the bag composition (first trial), the participant had no information regarding the black/white marbles ratio. After the participant had placed the 11 stickers in the first row, the experimenter started the second trial by extracting, without replacing, one marble from the bag. In the second row of the grid, the participant would write the color that had just been extracted and would express his new guess, by placing again 11 stickers on the sheet. On each trial, the experimenter revealed the color of one new marble. The task ended when the experimenter showed the color of the 10th marble, which completely revealed the composition of the bag. The monetary payoff was then calculated by summing the markers that the participant placed in the column that corresponded to the actual composition of the bag. In this way, we tried to ensure that the participants would use the available information and give some thought to the new information the experimenter provided about the content of the bag.

This task allows us to elicit, in an incentive-compatible framework, the beliefs of the subjects regarding an ambiguous situation that gradually becomes less ambiguous with the



revelation of the color of the marbles. This task allows us to test whether the subjects' choices between the risky bag and the ambiguous bag in task EB, described next, are influenced by their beliefs about the contents of the ambiguous bag, as commonly assumed in behavioral models of ambiguity aversion (cf. e.g., Gilboa and Schmeidler, 1989; Klibanoff et al., 2005).

### “Ellsberg Bags” Task (EB)

The participants were presented with two bags. The first bag contained 10 marbles, either white or black, in equal proportions. The content of this bag was shown to the participants. The second bag contained 10 marbles, either white or black, but in unknown proportions. The second bag was randomly chosen from the same shelf described in the RB task. The second bag was replaced on the shelf at the end of every participant's EB task.

Each participant was asked to choose a winning color for the two bags (white or black). The participant in this game had a right to extract a marble from each of the bags. If the marble extracted was of the same color that the participant has chosen, the participant won 15 experimental points. If the guess was incorrect, the participant won nothing. In an attempt to elicit the certainty equivalent for each lottery, the participants were asked to write two minimum selling prices for their two bets: *PA*, the price for the bet on the ambiguous bag, and *PR*, the price for the bet on the risky bag. The buyer of each bet was the experimenter, whose buying price for each bet was determined through a random physical mechanism (a number between 0 and 15 was drawn from yet another bag, with replacement). If the buying price for a bet was higher than, or equal to, the selling price stated by the participant, the participant pocketed the buying price, and no extraction took place from the corresponding bag. If the buying price was lower than the selling price the participant chose, the extraction of the marble took place. To ensure that the extraction of the random buying number for the first bet (the risky one) did not influence choices in the sale of the second bet, the experimenter's buying values were given only after the participant had stated both his prices. Instructions carefully explained that it was best for subjects to state the true value of the bets and that the price of the bets should reflect the desirability of the bets. If the subject thought the bet was very valuable, meaning that he thought the marble extracted would be very likely of the same color he chose, he should have stated a high selling price (close to 15). Conversely, a subject who believed that the bet was too close to call should have chosen a low price (close to zero), maximizing the chances that the bet would be bought by the experimenter at the random buying price. This elicitation procedure is known in experimental economics as the BDM method (Becker et al., 1964), and it was first adapted to the Ellsberg bags, to the best of our knowledge, by Halevy (2007). Further details on the pros and cons of the BDM method applied to lotteries can be found in Halevy's paper.

This task allows us not only to know if the participant is prey to the “Ellsberg paradox,” stating a higher price for the bet on the risky bag than for the bet on the ambiguous bag, but also to quantify each subject's aversion to ambiguity (or preference for ambiguity, if the price for the risky bag is lower than the price for the ambiguous bag).

### “Monty Hall” Task (MH)

In the MH task participants were presented with three flipped cups and the experimenter stated that under one of the cups there was a black marble which could be exchanged for 15 experimental points. Participants were asked to indicate the cup they wanted to top flip. Next, the experimenter flipped one of the other cups, always one without any marble under it. The participants were then offered the possibility to stand by their initial choice of the cup to flip, or switch. This task is part of a project on the endocrine correlates of Bayesian updating, a topic we might study elsewhere, and with no hypothesized implication for the subjects' ambiguity attitudes studied here.

### Hormonal Assays

Saliva samples were collected using Salimetrics Oral swabs (SOS; Salimetrics LLC, State College PA) placed under the tongue, according to vendor usage instructions for T determinations. According to the vendor, the SOS device consists of “an inert food-grade polymer” individually validated for use in specific assays that include salivary T and C determinations. Participants were instructed to place the oral swab beneath their tongue for at least 4 min. Samples were chilled immediately following collection, and then frozen within one h and held at  $-20^{\circ}\text{C}$  until assay. Samples were assayed at the SFU Neuroendocrinology laboratory using competitive enzyme immunoassays for T and C (Salimetrics kits). For both steroids, the average intra- and inter-assay coefficients of variation were lower than 10%. The two samples provided by two participants were misplaced, and three participants were excluded due to them reporting in the demographics questionnaire that they were using medications (antibiotics hydrocortisone and medication for acne), leaving a final sample size of 73 participants. In all the statistical analyses used in this paper, we average the two measurements of T and C, to have a better proxy for the level of circulating hormones around the time of the experiment. Statistical tests presented in the SM show that differences in each participant's two measurements are not statistically significant.

## RESULTS

### “Reveal the Bag (RB)” Task

In this task, participants expressed their second-order beliefs regarding the contents of an ambiguous bag (10 marbles, either white or black). A second-order belief assigns a probability to a certain scenario for the ambiguous bag (e.g., six black marbles, four white marbles). We collected information about the second-order beliefs of the participants as we gradually revealed to them, marble after marble, the content of the bag.

In an attempt to quantify the dispersion of beliefs about the content of the bag we computed an 11-bin histogram of all response possibilities. Afterward, we estimated the normalized entropy of the individual histograms, according to Shannon's formula (Shannon, 1948; Shannon and Weaver, 1949; cf. also Bennett et al., 2015):

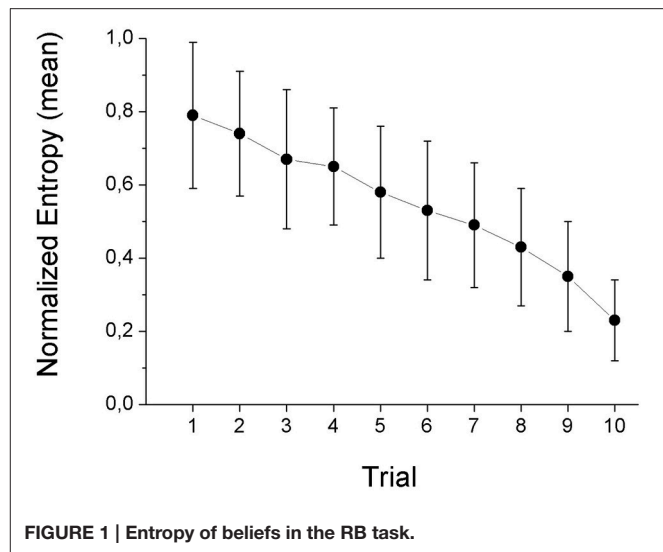
$$H = - \frac{\sum_{i=1}^{11} (p_i \times \log_2 p_i)}{\log_2 11}$$

where  $p_i$  is the relative frequency at bin  $i$ . The normalized entropy reflects the degree of belief uncertainty regarding the bag composition. The H scores range from 0—when the second-order probability mass lies all on one scenario (e.g., six white and four black), to 1—when all scenarios (from all-black to all-white) are believed to be equally likely. The mean entropy (over participants) from trial 1 (the color of zero marbles has been revealed, complete ambiguity) to trial 10 (only the color of one marble remains to be revealed) is shown in **Figure 1**. The downward trend in the dispersion of beliefs was clear. Entropy followed a constant rate of decay. To confirm this finding, we regressed average entropy in each trial on the trial number, finding a significant negative relation ( $p < 0.001$ ). About one-quarter of the participants in the first trial had an entropy of 1, i.e., they thought all the scenarios were equally likely (the so-called Laplace Principle of Insufficient Reason, cf. Gilboa, 2009, p. 14). Only 3% of the participants thought there was only one possible scenario for the bag (entropy of zero). The remaining cases fell in between (cf. **Figure 1** and the histograms in the SM). In about 8% of the total number of cases the participants committed a mistake, by placing positive probability mass on scenarios that were ruled out by the available information (e.g., attributing positive probability to the “10 black, 0 white marbles” after one white marble had already been revealed to them). This provides evidence that subjects understood the task and considered the information that was presented to them in order to make their choices.

### “Ellsberg Bags (EB)” Task

**Table 1** shows the descriptive statistics for PR, PA, and the ambiguity premium (PR-PA).

The reservation prices of the two lotteries were close, which resulted in an ambiguity premium of small positive magnitude, consistent with a modest degree of ambiguity aversion in the sample. The modal choice of premium is zero (31.5% of the subjects). As in Stahl (2014), we find that ambiguity preferences are heterogeneous and that a high degree of aversion to ambiguity might not be the most common finding, as instead the earlier literature supposed (cf. e.g., Halevy, 2007). Both a paired  $t$ -test and a non-parametric Wilcoxon signed-rank test rejected the null that the average of PR is equal to the average of PA ( $p$ -value is in both cases  $< 0.05$ ). The two prices were very close to the expected value of the bet on the risky bag, i.e., 7.5 points (regardless of the color chosen). The fact that PR was on average above the expected value of the lottery implies that participants were on average modestly risk-loving, as in Halevy (2007). When we calculated the average premium as a percentage  $\frac{(PR-PA)}{PA} \times 100$ , the result, 7.5%, was well-below the figure reported in Halevy (2007), i.e., 20%. Borghans et al. (2009), who also used the BDM design, reported a percentage figure of around 15%. The proportion of participants who were ambiguity neutral is comparable to Halevy's finding (current study: 30%; Halevy: 22%). PA and PR were positively correlated ( $r = 0.61$ ): this implies that typically participants displayed either a general distaste for seeing the realization of their random bets (when they chose low prices for



the two lotteries) or a general taste for seeing the realization of their bets.

### Regression and Interaction Analysis

We used regression analysis to determine whether centered average T ( $\bar{t}_i - \bar{t}$ ), centered average C ( $\bar{c}_i - \bar{c}$ ) and the interaction term between the two average centered hormonal measurements predicted our dependent variable  $y$  (the ambiguity premium). Each  $\bar{t}_i$  ( $\bar{c}_i$ ) is the average of the participant's two T (C) measurements (cf. also the SM for robustness checks using only the first measurement). The regression model is shown in Equation 1 ( $i$  is the identifier of the participant).

$$y_i = \alpha + \gamma([\bar{c}_i - \bar{c}]) + \delta([\bar{t}_i - \bar{t}]) + \theta([\bar{c}_i - \bar{c}] * ([\bar{t}_i - \bar{t}])) + \varepsilon_i \quad (1)$$

The reason for subtracting the mean across participants of the average hormonal measurements ( $\bar{t}$  and  $\bar{c}$ ) from each individual's average measurement, a procedure known as “centering,” was that, when using uncentered variables, average C and T were highly correlated with the T\*C interaction term, creating a multicollinearity problem. Aiken and West (1991) suggested centering as a solution to this issue, and the variance inflation factor for the interaction term went from 48 in the uncentered model to 1 in the centered model. **Table 2** shows the regression output of regression model (1), estimated through Ordinary Least Squares, with robust standard errors (R-squared = 0.063, model is significant at 5%).

The interaction term between T and C had a significant, positive relation with the ambiguity premium. Several robustness checks presented in the SM confirm this finding. We also show in the SM that the significance of the interaction term in **Table 2** can be probably attributed to the strong relation between the two hormones and PA.

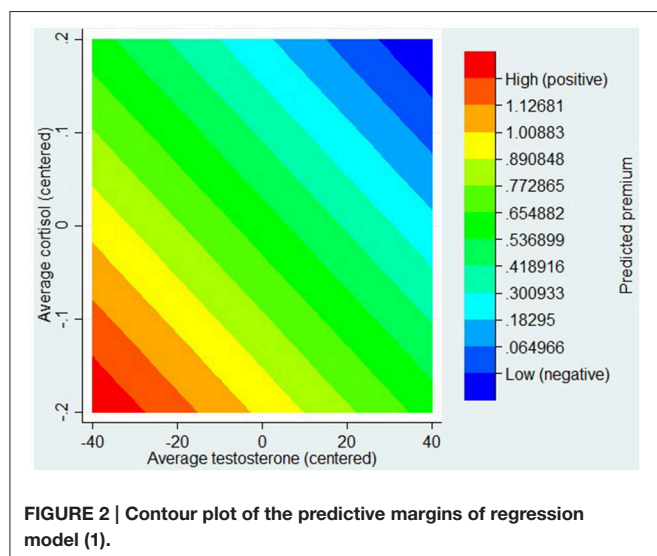
The positive sign of the interaction term, together with the negative signs of T and C, implied an overall negative relationship between the two hormone levels and the ambiguity premium. In **Figure 2**, we show a contour plot of the predictive margins of

**TABLE 1 | Descriptive statistics of the EB task.**

Variable	Obs	Mean	Std. Dev.	Min.	Max.
PR	73	7.87	2.87	0	15
PA	73	7.27	2.81	0	13
Premium (PR-PA)	73	0.59	2.52	-6.5	8

**TABLE 2 | Linear regression predicting the ambiguity premium based on centered hormones.**

Dependent variable: premium	Coef.	Robust Std. Err.
AvgC_centered	-1.939	2.972
AvgT_centered	-0.009	0.009
CrossCT_centered	0.179***	0.067
Constant	0.413	0.309

\*\*\* $p \leq 0.01$ .

the regression model (1). On the axes, we plot T and C about two standard deviations below and above the (zero) mean. The color bands show different levels of (predicted) premium. The Figure shows that a group of participants, specifically those with comparatively lower levels of T and C, exhibited comparatively higher aversion to ambiguity. The aversion to ambiguity declined as T increased, both for the low C and the high C group. The significance of the interaction term in **Table 2** ensures that this pattern is statistically significant.

## Beliefs and Ambiguity Attitudes

We used the responses of the participants in the first trial of the RB task to analyze the choice of reservation prices for the lotteries in the EB task. The participants were not given any signal that the ambiguous bag used in the EB task was the same as the ambiguous bag in the RB task. Given that, however, we were in the realm of complete uncertainty, it seems likely that the participants might have held the same beliefs regarding the ambiguous bag they faced in the first trial of the RB task and the ambiguous bag in the EB task. Using a

revealed preference approach, PR would be greater than PA if the Expected Utility of the lottery defined over the risky bag ( $EU(L_R)$ ) was greater than the (Subjective) Expected Utility of the lottery defined over the ambiguous bag ( $SEU(L_A)$ ). Details of the expected utility calculations for the two lotteries are given in the SM.

The average across participants of the difference between the expected utilities of the two lotteries, which we call  $\pi$  ( $\pi = [EU(L_R) - SEU(L_A)]$ ) is positive (the estimate is  $\tilde{\pi} = 0.094$ ). Together with the finding that the ambiguity premium in the EB task is on average positive, this finding shows that participants found on average bets on a risky bag more attractive than bets on an ambiguous bag. For the participants with entropy equal to 1 in the first trial of the RB task the expected utility of the two EB lotteries was the same, and  $\tilde{\pi} = 0$  was the modal estimate. Together with the finding that the modal value for premium is zero, these two results show that neutrality was the most common attitude to ambiguity in our experiment. We then carried a direct comparison between each participant's  $\pi_i$  and his ambiguity premium. These are two different ways to express the desirability of the bet on the risky bag vs. the bet on the ambiguous bag. Thirty-two percent of the participants passed this test of coherency, i.e., the sign of the variable premium is the same as the sign of  $\pi$ , or they are both zero. Of particular interest is a group of participants, 16% of the total, who featured a  $\tilde{\pi} = 0$  and also a premium equal to zero. These participants expressed their neutrality to ambiguity in a remarkably consistent way. We found no evidence that hormones played a role in determining the responsiveness of the choices of the risky lottery to  $\pi$  in a softmax model like the one used by Frydman et al. (2011). Finally, we do not find any role of T and C in explaining the degree of entropy of the participants' choices in trial 1 of the RB task.

## DISCUSSION

We established that there were instances in which the beliefs of the players translated into choices of one bet vs. the other. Moreover, we found cases in which subjects had an ambiguity premium equal to zero and derived, in our armchair calculations that involved some parameter and functional form choices, the same expected utility from the two bets (risky and ambiguous). This congruency is to be expected if the beliefs of the players about the bags are related to their ambiguity attitudes, as we hypothesized. Yet this congruency is not assured for most participants, contrary to our expectations. Possible reasons are that we used responses from two different tasks in our expected utility computations, assuming that the beliefs in round 1 of the RB task were the same as the beliefs about the ambiguous bag in the EB task, an assumption that might not be valid for all participants. Another possible explanation is that some parameter and functional form choices had to be made *ex-ante* and we did not build around the expected utility estimates an interval that allows for perceptual mistakes about the lotteries and the bags.

We found a significant interaction effect between T and C and the ambiguity premium in an Ellsberg experiment. The participants displaying the highest premium were those with lower C and lower T. These participants showed a preference



for known odds of winning compared to ignorance about the odds. This preference attenuated as cortisol and testosterone jointly increased. This finding supports *some* aspects of the DHH. This hypothesis has two parts: one is methodological, in the sense that it recommends that regression models using C and T should also control for the interaction effects of the two hormones, a suggestion we use and which yields some insights into the endocrine correlates of risk and ambiguity. The second part of the DHH is substantive, and it posits that T should positively correlate with status-seeking behavior only in low C individuals. No consensus exists on the substantive claim (cf. e.g., Welker et al., 2014, finding that testosterone is positively related to aggression only for high C individuals), and we do not find evidence in its favor. The substantive claim of Mehta and Josephs (2010) is, however, not easily applicable to our design, given the presence in our study of both risk and ambiguity. The part of behavioral endocrinology concerned with economic risk-taking is most likely not impermeable to “garden of forking paths” issues (Gelman and Loken, 2014), a problem that might be due to the low number of observations in some studies, affecting the power of the statistical testing procedures. We have tried to ease this problem writing pre-analysis plans (for the beliefs part of the analysis) while acknowledging where our analysis becomes exploratory due to the novelty of the study.

Unlike in Stanton et al. (2011), we did not find any evidence of a non-linear relation between T and the ambiguity premium (cf. also the robustness checks in the SM and earlier work by Schipper, 2014). Comparisons between the results of Stanton et al. (2011) and ours are, however, complicated by the differences in the design.

T has been associated to outperforming in competitions and to status-seeking behavior (Zilioli and Watson, 2014). It could play a role in ambiguous decisions involving monetary gains because most competition situations are ambiguous, in the sense that beliefs about the skills and threat posed by the opponent might be difficult to formulate (cf. Oliveira and Oliveira, 2014, on cognitive appraisal of competitive situations). We would expect therefore participants with higher levels of T to prefer situations that are more ambiguous and potentially more rewarding, displaying a lower premium, as shown in our study. C is the end product of the hypothalamic–pituitary–adrenal stress axis (Dedovic et al., 2009). Higher C might be related to higher sensitivity to stressors in the decision context, and therefore it seems sensible that individuals with high trait T (who preferred ambiguous situations) might also be characterized by higher levels of C. In the current study ambiguity only surrounded the probability of winning (rather than e.g., also the probability of losing, or smaller vs. greater gains), and a high T individual might have preferred the ambiguous bag out of confidence that the ambiguous bag offered higher-probability gains than the risky one. It is left for future research to establish if T and C are positively correlated with a preference for ambiguity when ambiguity entails potentially bigger gains compared to the risky situation. A question we have not addressed is gender-effects in ambiguity and risk attitudes (cf. e.g., Borghans et al., 2009; Lighthall et al., 2009; Boksem et al., 2013; Kandasamy et al., 2014; Schipper, 2014). Future studies

might ask whether our results from a male population extend also to females.

We contribute new evidence to the behavioral endocrinology literature, in particular the branch that focuses on choices over lotteries and their link to T and C. This field has to this date not converged to a consensus about the significance and direction of either T or C, or both, for risk-taking behavior, a situation that invites new studies and replication of existing ones. We hope future research will also bear in mind Ellsberg (1961)’s remark that “not all risks are the same” when discussing risk-taking and its hormonal correlates.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Ethical Conduct for Research Involving Humans guidelines (TCPS-2), Office of Research Ethics of Simon Fraser University, with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Office of Research Ethics of Simon Fraser University.

## AUTHOR CONTRIBUTIONS

GD conceived of the study, drafted the manuscript, and carried out the statistical analyses. EF helped draft the manuscript, participated in the statistical analysis. SZ participated in the design of the study, coordinated the study, and supervised the hormonal analyses. NW provided advice and facilities for the hormonal analyses. All authors gave final approval for publication.

## FUNDING

We wish to acknowledge a Teaching and Learning Development Grant from Simon Fraser University (number L-G0029) to GD and a Discovery Grant 0194522 from the Natural Sciences and Engineering Research Council of Canada (NSERC) to NW.

## ACKNOWLEDGMENTS

We would like to thank Erik Drysdale, Andreas Hovland, Justin Jagore, and Lindsay Cooper for excellent research assistance. We thank Jasmina Arifovic, Erik Kimbrough, and audiences at the University of Trento, Simon Fraser University, the 2014 meeting of the Society for the Advancement of Behavioral Economics (Lake Tahoe, NV), the 2016 Social and Biological Roots of Economics Workshop (Kiel, Germany) for helpful comments. The usual disclaimer applies.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnbeh.2017.00068/full#supplementary-material>



## REFERENCES

- Aiken, L. S., and West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA: Sage Publications.
- 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
- Becker, G. M., DeGroot, M. H., and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behav. Sci.* 9, 226–232. doi: 10.1002/bs.3830090304
- Bennett, D., Murawski, C., and Bode, S. (2015). Single-trial event-related potential correlates of belief updating. *eNeuro* 2:ENEURO.0076-15.2015. doi: 10.1523/ENEURO.0076-15.2015
- Boksem, M. A. S., Mehta, P. H., Van den Bergh, B., van Son, V., Trautmann, S. T., Roelofs, K., et al. (2013). Testosterone inhibits trust but promotes reciprocity. *Psychol. Sci.* 24, 2306–2314. doi: 10.1177/0956797613495063
- Borghans, L., Golsteyn, B. H. H., Heckman, J. J., and Meijers, H. (2009). Gender differences in risk aversion and ambiguity aversion. *J. Eur. Econ. Assoc.* 7, 649–658. doi: 10.3386/w14713
- 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
- Brown, L. L., Tomarken, A. J., Orth, D. N., Loosen, P. T., Kalin, N. H., and Davidson, R. J. (1996). Individual differences in repressive-defensiveness predict basal salivary cortisol levels. *J. Pers. Soc. Psychol.* 70, 362–371. doi: 10.1037/0022-3514.70.2.362
- Buckert, M., Schwieren, C., Kudiella, B., and Fiebach, C. J. (2014). Acute stress affects risk-taking but not ambiguity aversion. *Front. Neurosci.* 8:82. doi: 10.3389/fnins.2014.00082
- Carver, C. S., and White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *J. Pers. Soc. Psychol.* 67, 319–333. doi: 10.1037/0022-3514.67.2.319
- Casto, K. V., and Edwards, D. A. (2016). Testosterone, cortisol, and human competition. *Horm. Behav.* 82, 21–37. doi: 10.1016/j.yhbeh.2016.04.004
- Coates, J. M., and Herbert, J. (2008). Endogenous steroids and financial risk-taking on a London trading floor. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6167–6172. doi: 10.1073/pnas.0704025105
- Cueva, C., Roberts, R. E., Spencer, T., Rani, N., Tempest, M., Tobler, P. N., et al. (2015). Cortisol and testosterone increase financial risk-taking and may destabilize markets. *Sci. Rep.* 5:11206. doi: 10.1038/srep11206
- Dabbs, J. M. Jr., Jurkovic, G. J., and Frady, R. L. (1991). Salivary testosterone and cortisol among late adolescent offenders. *J. Abnorm. Child Psychol.* 19, 469–478. doi: 10.1007/BF00919089
- Daly, M., and Wilson, M. (1997). Crime and conflict: homicide in evolutionary psychological perspective. *Crime Justice* 22, 51–100. doi: 10.1086/449260
- Dedovic, K., Duchesne, A., Andrews, J., Engert, V., and Pruessner, J. C. (2009). The brain and the stress axis: the neural correlates of cortisol regulation in response to stress. *Neuroimage* 47, 864–871. doi: 10.1016/j.neuroimage.2009.05.074
- Ellis, B. J., Del Giudice, M., Dishion, T. J., Figueredo, A. J., Gray, P., Griskevicius, V., et al. (2012). The evolutionary basis of risky adolescent behavior: implications for science, policy, and practice. *Dev. Psychol.* 48, 598. doi: 10.1037/a0026220
- Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Q. J. Econ.* 75, 643–669. doi: 10.2307/1884324
- Frydman, C., Camerer, C., Bossaerts, P., and Rangel, A. (2011). MAOA-L carriers are better at making optimal financial decisions under risk. *Proc. R. Soc. Lond. B Biol. Sci.* 278, 2053–2059. doi: 10.1098/rspb.2010.2304
- Gelman, A., and Loken, E. (2014). The Statistical Crisis in Science Data-dependent analysis—a “garden of forking paths”—explains why many statistically significant comparisons don’t hold up. *Am. Sci.* 102:460. doi: 10.1511/2014.111.460
- Gilboa, I. (2009). *Theory of Decision under Uncertainty*. Cambridge: Cambridge University Press.
- Gilboa, I., and Schmeidler, D. (1989). Maxmin expected utility with a non-unique prior. *J. Math. Econ.* 18, 141–153. doi: 10.1016/0304-4068(89)90018-9
- Halevy, Y. (2007). Ellsberg revisited: an experimental study. *Econometrica* 75, 503–536. doi: 10.1111/j.1468-0262.2006.00755.x
- Hayden, B. Y., Heilbronner, S. R., and Platt, M. L. (2010). Ambiguity aversion in rhesus macaques. *Front. Neurosci.* 4:166. doi: 10.3389/fnins.2010.00166
- Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., and Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310, 1680–1683. doi: 10.1126/science.1115327
- Kadane, J. B. (1992). Healthy scepticism as an expected-utility explanation of the phenomena of Allais and Ellsberg. *Theor. Decis.* 32, 57–64. doi: 10.1007/BF00133627
- Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1991). Anomalies: the endowment effect, loss aversion, and status quo bias. *J. Econ. Perspect.* 5, 193–206. doi: 10.1257/jep.5.1.193
- Kandasamy, N., Hardy, B., Page, L., Schaffner, M., Graggaber, J., Powlson, A. S., et al. (2014). Cortisol shifts financial risk preferences. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3608–3613. doi: 10.1073/pnas.1317908111
- Klibanoff, P., Marinacci, M., and Mukerji, S. (2005). A smooth model of decision making under ambiguity. *Econometrica* 73, 1849–1892. doi: 10.1111/j.1468-0262.2005.00640.x
- Knight, F. H. (1921). *Risk, Uncertainty and Profit*. Boston, MA: Houghton Mifflin Company.
- Kühberger, A., and Perner, J. (2003). The role of competition and knowledge in the Ellsberg task. *J. Behav. Decis. Mak.* 16, 181–191. doi: 10.1002/bdm.441
- Levenson, M., Kiehl, K., and Fitzpatrick, C. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *J. Pers. Soc. Psychol.* 68, 151–158. doi: 10.1037/0022-3514.68.1.151
- Levy, I., Snell, J., Nelson, A. J., Rustichini, A., and Glimcher, P. W. (2010). Neural representation of subjective value under risk and ambiguity. *J. Neurophysiol.* 103, 1036–1047. doi: 10.1152/jn.00853.2009
- Lighthall, N. R., Mather, M., and Gorlick, M. A. (2009). Acute stress increases sex differences in risk seeking in the balloon analogue risk task. *PLoS ONE* 4:e6002. doi: 10.1371/journal.pone.0006002
- Mazur, A., and Booth, A. (1998). Testosterone and dominance in men. *Behav. Brain Sci.* 21, 353–363. doi: 10.1017/S0140525X98001228
- Mehta, P. H., and Josephs, R. A. (2010). Testosterone and cortisol jointly regulate dominance: evidence for a dual-hormone hypothesis. *Horm. Behav.* 58, 898–906. doi: 10.1016/j.yhbeh.2010.08.020
- Mehta, P. H., and Prasad, S. (2015). The dual-hormone hypothesis: a brief review and future research agenda. *Curr. Opin. Behav. Sci.* 3, 163–168. doi: 10.1016/j.cobeha.2015.04.008
- Mehta, P. H., Welker, K. M., Zilioli, S., and Carré, J. M. (2015). Testosterone and cortisol jointly modulate risk-taking. *Psychoneuroendocrinology* 56, 88–99. doi: 10.1016/j.psyneuen.2015.02.023
- Oliveira, G. A., and Oliveira, R. F. (2014). Androgen responsiveness to competition in humans: the role of cognitive variables. *Neurosci. Neuroecon.* 3, 19–32. doi: 10.2147/NAN.S55721
- Pfafftheicher, S. (2017). Illuminating the dual-hormone hypothesis: about chronic dominance and the interaction of cortisol and testosterone. *Aggress. Behav.* 43, 85–92. doi: 10.1002/ab.21665
- Popma, A., Vermeiren, R., Geluk, C. A., Rinne, T., van den Brink, W., Knol, D. L., et al. (2007). Cortisol moderates the relationship between testosterone and aggression in delinquent male adolescents. *Biol. Psychiatry* 61, 405–411. doi: 10.1016/j.biopsych.2006.06.006
- Pulford, B. D. (2009). Is luck on my side? Optimism, pessimism, and ambiguity aversion. *Q. J. Exp. Psychol.* 62, 1079–1087. doi: 10.1080/17470210802592113
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychol. Monogr.* 80, 1–28. doi: 10.1037/h0092976
- Sapienza, P., Zingales, L., and Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15268–15273. doi: 10.1073/pnas.0907352106
- Savage, L. (1972). *The Foundations of Statistics*. New York, NY: Dover Publications.
- Schipper, B. C. (2014). Sex hormones and competitive bidding. *Manage. Sci.* 61, 249–266. doi: 10.1287/mnsc.2014.1959
- Schönberg, T., Fox, C. R., and Poldrack, R. A. (2011). Mind the gap: bridging economic and naturalistic risk-taking with cognitive neuroscience. *Trends Cogn. Sci.* 15, 11–19. doi: 10.1016/j.tics.2010.10.002
- Seo, K. (2009). Ambiguity and second-order belief. *Econometrica* 77, 1575–1605. doi: 10.3982/ECTA6727
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

- Shannon, C. E., and Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.
- Stahl, D. O. (2014). Heterogeneity of ambiguity preferences. *Rev. Econ. Stat.* 96, 609–617. doi: 10.1162/REST\_a\_00392
- Stanton, S. J., O'Dhanial, M.-G., McLaurin, R. E., Kuhn, C. M., LaBar, K. S., Platt, M. L., et al. (2011). Low- and high-testosterone individuals exhibit decreased aversion to economic risk. *Psychol. Sci.* 22, 447–453. doi: 10.1177/0956797611401752
- Tackett, J. L., Herzhoff, K., Paige, K., Page-Gould, E., and Josephs, R. A. (2014). Personality  $\times$  hormone interactions in adolescent externalizing psychopathology. *Pers. Disord.* 5, 235–246. doi: 10.1037/per0000075
- Van Honk, J., Schutter, D. J., Hermans, E. J., and Putman, P. (2003). Low cortisol levels and the balance between punishment sensitivity and reward dependency. *Neuroreport* 14, 1993–1996. doi: 10.1097/00001756-200310270-00023
- Welker, K. M., Lozoya, E., Campbell, J. A., Neumann, C. S., and Carré, J. M. (2014). Testosterone, cortisol, and psychopathic traits in men and women. *Physiol. Behav.* 129, 230–236. doi: 10.1016/j.physbeh.2014.02.057
- Wingfield, J. C. (2016). The challenge hypothesis: where it began and relevance to humans. *Horm. Behav.* doi: 10.1016/j.yhbeh.2016.11.008. [Epub ahead of print].
- Zethraeus, N., Kocoska-Maras, L., Ellingsen, T., Von Schoultz, B. O., Hirschberg, A. L., and Johannesson, M. (2009). A randomized trial of the effect of estrogen and testosterone on economic behavior. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6535–6538. doi: 10.1073/pnas.0812757106
- Zilioli, S., and Watson, N. V. (2012). The hidden dimensions of the competition effect: basal cortisol and basal testosterone jointly predict changes in salivary testosterone after social victory in men. *Psychoneuroendocrinology* 37, 1855–1865. doi: 10.1016/j.psyneuen.2012.03.022
- Zilioli, S., and Watson, N. V. (2014). Testosterone across successive competitions: evidence for a 'winner effect' in humans? *Psychoneuroendocrinology* 47, 1–9. doi: 10.1016/j.psyneuen.2014.05.001

**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.

The reviewer PR and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2017 Danese, Fernandes, Watson and Zilioli. 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.

# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

**Visit us:** [www.frontiersin.org](http://www.frontiersin.org)

**Contact us:** [info@frontiersin.org](mailto:info@frontiersin.org) | +41 21 510 17 00



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

@frontiersin



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership