

Neyse, Levent; Johannesson, Magnus; Dreber, Anna

Article — Published Version

2D:4D does not predict economic preferences: Evidence from a large, representative sample

Journal of Economic Behavior & Organization

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Neyse, Levent; Johannesson, Magnus; Dreber, Anna (2021) : 2D:4D does not predict economic preferences: Evidence from a large, representative sample, Journal of Economic Behavior & Organization, ISSN 0167-2681, Elsevier, Amsterdam, Vol. 185, pp. 390-401,
<https://doi.org/10.1016/j.jebo.2021.02.029>

This Version is available at:

<https://hdl.handle.net/10419/235709>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

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

2D:4D does not predict economic preferences: Evidence from a large, representative sample

Levent Neyse^{a,b,c}, Magnus Johannesson^d, Anna Dreber^{d,e,*}

^a SOEP at DIW, Berlin, Germany

^b WZB, Berlin, Germany

^c IZA, Bonn, Germany

^d Stockholm School of Economics, Department of Economics, Box 6501 Stockholm SE-113 83, Sweden

^e Department of Economics, University of Innsbruck, Innsbruck, Austria

ARTICLE INFO

Article history:

Received 22 December 2020

Revised 22 February 2021

Accepted 28 February 2021

JEL classification:

D03

D87

D9

Keywords:

Economic behavior

Prenatal hormones

Testosterone

Digit ratio

ABSTRACT

The digit ratio (2D:4D) is considered a proxy for testosterone exposure in utero, and there has been a recent surge of studies testing whether 2D:4D is associated with economic preferences. Although the results are not conclusive, previous studies have reported statistically significant correlations between 2D:4D and risk taking, altruism, positive reciprocity, negative reciprocity and trust. However, most previous studies have small sample sizes gathered from university students and there is also no consensus on the type of analysis (e.g., which hand to analyze or subgroup to focus on) or the preference elicitation method. We present results from a pre-registered large sample study testing if 2D:4D is associated with economic preferences. Data were collected in a representative sample of adults in the German Socioeconomic Panel-Innovation Sample (SOEP-IS), in a sample of about 3450 respondents (about 5 times larger than the previously largest study in this field). Using experimentally validated survey questions, we find no statistically significant association between 2D:4D and economic preferences in the largest study to this date on the topic.

© 2021 The Author(s). Published by Elsevier B.V.
This is an open access article under the CC BY license
(<http://creativecommons.org/licenses/by/4.0/>)

1. Introduction

There is substantial variation in economic preferences between individuals. Some of this variation has been linked to cultural or situational factors such as gender norms (e.g. Gneezy et al., 2009) or reference points (e.g. Thaler and Johnson, 1990). Relatively recently, the potential role of hormones has received substantial interest. Hormones have organizational effects on the brain during fetal development (Phoenix et al., 1959; Arnold, 2009; Lombardo et al., 2012), and these effects are often hypothesized to have long lasting effects on preferences and behaviors later in life. In particular testosterone, the sex steroid, has been hypothesized to be associated with a wide range of economic preferences and decisions.

While some studies have explored circulating levels of testosterone or administered testosterone (e.g. Sapienza et al., 2009; Zethraeus et al., 2009; Eisenegger et al., 2011; Bos et al., 2010), with mixed results (see Dreber and Johannesson, 2019 for further discussion), others have focused on prenatal testosterone exposure in utero. Such exposure is hypoth-

* Corresponding author at: Stockholm School of Economics, Department of Economics, Box 6501 SE-113 83 Stockholm, Sweden.
E-mail address: Anna.Dreber@hhs.se (A. Dreber).

esized to impact brain development of the fetus and could thus potentially explain some of the individual heterogeneity in preferences (Baron-Cohen, 2002). Since testosterone exposure in utero cannot be measured directly after birth, researchers have looked for proxies of this exposure. The ratio of the length of the 2nd digit to the length of the 4th digit (2D:4D) on each hand (Manning et al., 1998) has been used as such a proxy.

2D:4D has been linked to a number of traits such as personality, sexual orientation and various cognitive abilities, but the results are often contradictory and there is also mixed evidence on the role of publication bias (see, e.g. Puts et al., 2008; Voracek and Loibl, 2009; Grimbos et al., 2010; Voracek et al., 2011; Hönekopp and Schuster, 2010; Hönekopp and Watson, 2011). Recently, there has been a surge in the number of papers testing for an association between 2D:4D and economic preferences or outcomes (e.g. Coates et al., 2009; Van Honk et al., 2012). Just on risk preferences alone, we count 18 papers. Such relationships suggest that biological factors may partially affect economic activities through decisions. Several studies also investigated direct associations with real economic outcomes. For example, Coates et al. (2009) reported an association between 2D:4D and the profitability of financial traders, which if real could work through to a channel of economic preferences but also through other channels, and Nicolaou et al. (2017) reported an association between 2D:4D and self-employment.

We contribute to this literature by testing the associations between 2D:4D and five economic preference measures in a large, representative sample. To do so, we integrated a 2D:4D and Economic Preferences module in the German Socioeconomic Panel-Innovation Sample (SOEP-IS) (Richter and Schupp, 2015). Our sample contains 3433 adult respondents, which is about 5 times larger than the previously largest study in this field (Brañas-Garza et al., 2018). Our five measures of economic preferences are i) risk taking, ii) altruism, iii) positive reciprocity, iv) negative reciprocity and v) trust. The questions used to measure these economic preferences were adapted from the preference survey module of the Global Preference Survey (Falk et al., 2018) (see the Methods section for further details about the preference measures and Supplementary Information Text for the complete list of items). We chose to focus on these five economic preference measures because of previous attempts to link these measures to 2D:4D, although a large proportion of this literature has focused on risk taking and altruism. All our tests are pre-registered to avoid "researcher degrees of freedom" affecting the results (e.g. from choosing ex-post which hand or subgroup to focus on).

Our hypotheses about the direction of the association between 2D:4D and each economic preferences are based on the previous literature. For risk taking, the hypothesis is a negative relation to 2D:4D, which would indicate a positive association between prenatal testosterone exposure and risk taking. However, even though previous literature has also had this hypothesis, the results in this literature are far from conclusive (see Parslow et al., 2019 for a review). For example, the first paper on this topic reports a negative correlation between risk taking and left hand 2D:4D but finds no statistically significant correlation with the right hand 2D:4D, and the significant correlation can only be found in one sample and not another (Dreber and Hoffman, 2007). There are other examples of statistically significant negative correlations (e.g. Brañas-Garza and Rustichini, 2011; Garbarino et al., 2011; Stenstrom et al., 2011; Sytsma, 2014; Bönnte et al., 2016; Barel, 2019; Brañas-Garza et al., 2018), though many of these only find statistically significant correlations in a subset of the population studied. There are thus also several null results (e.g. Stenstrom et al., 2011; Sytsma, 2014; Bönnte et al., 2016; Barel, 2019; Brañas-Garza et al., 2018; Apicella et al., 2008; Sapienza et al., 2009; Aycinena et al., 2014; Schipper, 2014; Drichoutis and Nayga, 2015; Chicaiza-Becerra and Garcia-Molina, 2017; Lima de Miranda et al., 2018; Alonso et al., 2018; Neyse et al., 2020) and also one finding of a statistically significant positive correlation between 2D:4D and risk taking in a female subsample in one of the papers that found a statistically significant negative correlation in another subsample (Brañas-Garza and Rustichini, 2011). Based on the previous literature we test the hypothesis that there is a statistically significant negative association between 2D:4D and risk taking.

There are also several studies testing the association between 2D:4D and altruism, measured as giving in the dictator game (Brañas-Garza et al., 2013; Galizzi and Nieboer, 2015; Millet and Dewitte, 2010; Buser, 2012; Brañas-Garza et al., 2019). Although these previous results have been mixed, the main hypothesis in the previous literature has been a positive association between 2D:4D and altruism - this is also the hypothesis that we test. In the exploratory analyses we also test for a non-linear relationship between 2D:4D and altruism by also including the squared 2D:4D as done in some of the previous studies (e.g. Brañas-Garza et al., 2013; Galizzi and Nieboer, 2015).

When it comes to 2D:4D, positive reciprocity, negative reciprocity and trust, there are only a handful of earlier studies. Buser (2012), who explored the relation between 2D:4D and altruism, also studies the trust game and finds that low 2D:4D individuals send back statistically significantly less as the second player in the trust game, and thus are less reciprocal.¹ This was the only previous study we were aware of testing for an association between 2D:4D and positive reciprocity before we wrote our pre-analysis plan.² Based on this study we test the hypothesis that there is a statistically significant positive association between 2D:4D and positive reciprocity. Responder behavior in the ultimatum game can be interpreted as a measure of negative reciprocity. Buser (2012) also tests for an association between 2D:4D and responder behavior in the ultimatum game, but finds no statistically significant association. However, another study finds a statistically significant negative correlation between right hand 2D:4D and the minimum acceptance level in the ultimatum game (a higher minimum acceptance

¹ Buser (2012) used a self-reported 2D:4D measure, which was criticized by Brañas-Garza and Kovářík (2013).

² The recent paper by Van Leeuwen et al. (2020), which is more thoroughly discussed in the Discussion section, also studies 2D:4D and among other things the economic behaviors trust, positive reciprocity and negative reciprocity and find no evidence for associations between these variables and 2D:4D.

level suggest stronger negative reciprocity) (Van den Bergh and Dewitte, 2006). These are the only previous studies we are aware of testing for an association between 2D:4D and negative reciprocity. Based on the study with the statistically significant negative correlation (Van den Bergh and Dewitte, 2006) we test the hypothesis that there is a statistically significant negative association between 2D:4D and negative reciprocity. Finally, there is only one previous study testing the relation between 2D:4D and trust. Buser (2012) finds a statistically significant positive association between 2D:4D and trust in the first stage of the trust game. Based on this study we test the hypothesis that there is a statistically significant positive association between 2D:4D and trust.

The aforementioned studies are typically small and with many “researcher degrees of freedom” in the analysis (Simmons et al., 2011; Gelman and Loken, 2013). For example, sometime researchers focus on one hand and not the other, or take the average of the two hands, or look at subsamples and find something statistically significant in one subsample but not the other, and it is not clear ex-ante why there should only be an association in one group and not others. This type of analysis, sometimes referred to as “forking” as in “the garden of forking paths” (Gelman and Loken, 2013), can occur when the researcher has a specific hypothesis but has not specified exactly how the data will be analyzed so they let the data decide how to analyze it. Forking (and more intentionally misleading analyses) can lead to results with p-values that cannot be interpreted as false positive probabilities – statistically significant results can easily be more likely to be false positive than true positive ones (Simmons et al., 2011). The average total sample size for the 2D:4D studies discussed in Parslow et al. (2019) is $N=262$, with a range from 86 to 704 participants. In addition, most of these studies look at smaller subsamples based on gender or ethnicity, thus effective sample sizes tend to be substantially smaller. There is thus a need for a much larger study with a pre-registered analysis plan to rigorously test if 2D:4D is associated with economic preferences.

For our primary hypotheses tests, we focus on the right hand 2D:4D because it has been argued that the right hand 2D:4D is more sexually dimorphic than the left hand (Hönekopp and Watson, 2011). We test if the right hand 2D:4D is associated with the five measures of economic preferences in five regression analyses controlling for gender. We use the left hand 2D:4D and the average of the two hands in our robustness tests. Since previous literature sometimes reports finding correlations in one gender and not the other, we perform a pre-registered exploratory analysis where we interact 2D:4D with gender to test if the associations differ between men and women. As some previous studies also test for a non-linear relationship between 2D:4D and economic preferences, we also carry out a pre-registered exploratory analysis adding 2D:4D squared to the regression analysis.

2. Methods

2.1. Data collection and sample

The study was pre-registered prior to starting the data collection (<https://osf.io/5vpdn/>). The pre-registration included the data, collection, measurement procedures and all statistical tests described below (including the exploratory tests and the robustness tests with the exception of the robustness tests using the corrected sample and the restricted sample, that were not pre-registered). The pre-registration did not include the instructions to the interviewers, but they are now posted on the OSF site. When we make any deviation from the pre-registration this is explicitly mentioned in the text. The data collection of the study was performed between September 2018 and December 2018 in the German Socioeconomic Panel-Innovation Sample (Goebel et al., 2019) and we received the data in May 2019. SOEP is a longitudinal survey study that has been running since 1984 and it has nearly 30,000 participants today. SOEP-IS, on the other hand, was established in 2011 (Richter and Schupp, 2015) and it is open for experimental and survey module suggestions from researchers every year. The survey committee decides on which of those suggested modules to include in the next wave. The selected modules are subsequently integrated to different sub-samples of SOEP-IS. The right to use the collected data belongs to applicant researchers for a year. After the one year embargo period, the data become available for the whole research community. The merged dataset also includes longitudinal socio-economic data, as well as previous years' modules. If a set of individuals participated in Modules A & B in 2014 and C & D in 2016, all their answers to these four modules can thus be merged into a single dataset. According to the 2018 release of SOEP-IS, it has a total number of 5722 participants from 3232 households.

We planned to survey the whole sample of SOEP-IS for the 2D:4D collection module. While the number of participants was predicted around 4500 adults prior to the survey, the actual number was 4,860. As 2D:4D measurement was voluntary, the participants could skip the measurement and continue responding to the survey questions. As a result, our working sample consisted 3482 participants with complete right or left hand measures (3433 individuals with right, 3454 with left, 3405 with both). On top of 2D:4D measurements the participants were also asked to reply to a set of survey questions about their economic and social preferences. The gender distribution of the working sample is balanced (53.8% women) and the mean age is 54.22 (SD=18.39).

2.2. Economic preference measures

The 2D:4D module included a total number of 8 survey questions to elicit risk taking, altruism, positive reciprocity, negative reciprocity and trust (see Supplementary Information for the phrasing of these 8 survey questions). All questions were adapted from the preference survey module of the seminal paper of Falk et al. (2018), who followed both experimental

tests and survey optimization procedures to construct their module (see also Falk et al. (2018) for a detailed description of the preference module construction procedure).

The risk question is the standard risk question that has been included in the past waves of SOEP as well, where participants are asked “How do you rate yourself personally? In general, are you someone who is ready to take risks or do you try to avoid risks (risk-averse)?”. For this question, participants are asked to indicate their willingness to take risks on a Likert scale from 0 to 10 where a higher number indicates more risk taking. The Global Preference Survey involves almost an identical self reported question (“In general how willing are you to take risks?”) and also a staircase procedure with hypothetical lotteries. We chose to use the standard self-reported question which was also experimentally validated in Dohmen et al. (2011).

We measured altruism with two questions and negative reciprocity with three questions (see Supplementary Information for more details). The multiple survey questions were combined to a measure of altruism and negative reciprocity in the same way as done in the Global Preference Survey (Falk et al., 2018). They standardized responses on each survey question to z-scores and estimated weights for aggregating responses into the overall preference measure. We used the same approach and the same weights to aggregate survey questions into preference measures (see Supplementary Information for the weights of Falk et al. (2018) that we used). However, the index resulting from aggregating the z-scores for the individual questions will not in itself be a z-index (as the STD of this index will not be equal to 1). We therefore also standardized the index for altruism and the index for negative reciprocity into z-scores, to ease the interpretation of results. This standardization of the indices was not done in the previous study (Falk et al., 2018) using these two indices and it was not mentioned in the pre-analysis plan (but it does not affect the results per se; but only the units used to report results in). For comparability in results between our five preference measures we standardized responses to z-scores also for risk taking, positive reciprocity, and trust for the regression analyses testing the hypotheses.

To measure positive reciprocity, we integrated the self-assessment question (“When someone does me a favor, I am willing to return it.”) in the global preference module but did not integrate the hypothetical choice item in the data-collection. The respondents were asked whether they agreed with this expression on a 10 point Likert Scale. Finally, trust was also measured by a Likert Scale from 0 to 10, where a higher number is more agreement. The trust item we used was “People are basically honest”. Note that this phrasing of the trust question is slightly different than the one we proposed to include in the data collection and that we included in the pre-analysis plan (“As long as I am not convinced otherwise, I assume that people have only the best intentions”). As the SOEP-IS questionnaire already had the former item included in another module and also three other trust related items, the survey management decided to include the shorter version rather than the version we proposed. As the data for the module using the trust question were collected in only a sub-sample of SOEP-IS, we have a smaller number of observations for trust than planned and a smaller number of observations than for the other preference measures.

The distribution for each of the five economic preference measures are presented in the Supplementary Information (see Figures S6-S10).

2.3. 2D:4D measurements

Both left and right hand 2D:4D's of the participants were measured during the household surveys with the help of digital calipers. Our main reasons for choosing a direct measurement method, instead of using scanners or mobile applications, were because of confidentiality of personal information as well as mobility and time efficiency. Scanned images of the hands would contain fingerprints of the participants, which may be considered as delicate personal information. Any discomfort or suspicion of the participants due to using scanned images may also result in dropping out of the study in further waves. Moreover, as the data was collected from a large number of households, digital calipers are easier to carry than flatbed scanners and they are also time and cost effective. While there are several mobile applications being developed for 2D:4D measurements, they are currently available only in demo versions.

263 interviewers were trained for the 2D:4D measurements in August, 2018. To minimize measurement errors, they were given a detailed hand measurement protocol that we prepared. The protocol instructed interviewers about i) Calibrating and using the digital calipers, ii) preparation for measurements (e.g. seating and hand positions, taking off the jewelry, iii) finger measurement, iv) second measurement if the 2D:4D ratio exceeds usual range of 0.8 and 1.1 significantly. This range was determined by the minimum and maximum 2D:4D ratios in previous studies (i.e. Neyse et al., 2020; Neyse et al., 2016; Lima de Miranda et al., 2018). The interviewers were asked to record both first and second measurements if they measured the finger lengths for the second time. The complete hand measurement protocol has been posted as an Appendix to the pre-analysis plan (but it was not posted at the same time as the pre-analysis plan, but after the data collection was completed). This protocol was prepared following a set of tests on different types of digital calipers, measurement methods and multiple measurers. To test the protocol, two research assistants in SOEP measured a number of hands using the calipers and the proposed protocol. These measures obtained by the research assistants were compared with digital measurements carried out by one of the authors of this study, LN. The measures obtained with the final protocol were almost identical to those obtained with digital scanners.

Beside these precautions, we also run a large set of robustness tests. For this purpose, we run the whole analysis with three different outlier criteria. Our main analysis relies on the Full Sample, which is based on all the participants including the outliers. Our “Corrected Sample” uses the second finger measurements done by the interviewer. Note that the

interviewers were instructed to re-measure the finger lengths if they were in doubt of their accuracy. In the “Restricted Sample”, we drop the observations outside the usual range of 0.8 and 1.2. We present our results for corrected and restricted samples in the Appendix. See the Results section for further details and OSF repository for supporting documentation (<https://osf.io/5vpdn/>).

2.4. Statistical analysis

Following the recommendations by Benjamin et al. (2018), we consider results with p -values below 0.005 to be statistically significant and results with p -values below 0.05 as suggestive evidence. We conducted our main analysis with right hand 2D:4D and repeated the same analysis for left hand 2D:4D and the average of the two hands as robustness checks. A large body of the literature focuses on the right hand 2D:4D in their analyses based on the argument that this ratio shows greater sex differences (Hönekopp and Watson, 2011).

The five primary hypotheses were tested in OLS regressions estimated with robust (Huber-White) standard errors. The survey measure of economic preferences was the dependent variable and the right hand 2D:4D and the gender of the participant were the independent variables. The hypothesis tests tested if the coefficient of the right hand 2D:4D was significant in the hypothesized direction. For each hypothesis we included all participants with data on right hand 2D:4D and gender who responded to the survey question used to construct the preference measure tested in the hypothesis. For altruism and negative reciprocity, where the preference measures were based on combining answers to more than one survey question (see Supplementary Information), we included all participants who responded to all the survey questions used to construct the preference measure (thus we did not include potential participants who only answered to one question for these measures).

We included all participants with a recorded right hand 2D:4D measurement and we did not exclude any 2D:4D observations in our main results. However, as explained in more detail in the beginning of the Results section we added two not pre-registered robustness tests to test if our results are sensitive to outliers due to mismeasurement or injured fingers. These robustness tests are referred to as “the corrected sample” and “the restricted sample” and these results are reported in Supplementary Information Tables S2–S7, S11–S16, S20–S25.

2.4.1. Pre-registered exploratory analyses 1: Gender interactions

For the five primary hypotheses we tested if the association between right hand 2D:4D and the survey measures differed between men and women. These tests were done by adding an interaction between gender and the right hand 2D:4D in the five OLS regressions used to test the five primary hypotheses. We tested if this interaction coefficient was statistically significant, but without specifying a hypothesized direction of the interaction coefficient as these analyses were exploratory and would need confirmation in other studies to carry much weight. We also reported the statistical significance of the right hand 2D:4D for men and women separately as part of these analyses, but these tests are only relevant if the interaction coefficient is statistically significant.

2.4.2. Pre-registered exploratory analyses 2: Non-linear relationships

Some studies in the literature have reported non-monotonic findings between 2D:4D and economic preferences, by adding a squared 2D:4D term to a linear regression (see for example Brañas-Garza et al., 2013; Galizzi and Nieboer, 2015). As exploratory analyses we therefore added the squared right hand 2D:4D to the OLS regression equations used for testing the five primary hypotheses, to test for non-linear effects of 2D:4D. In these tests we assessed the significance of the squared right hand 2D:4D, but without specifying a hypothesized direction of the squared term as these analyses are exploratory.

2.4.3. Pre-registered robustness tests: Left hand 2D:4D and average of left and right hand 2D:4D

As a robustness test we also estimated all the analyses above using the left hand 2D:4D and the average of the left hand and the right hand 2D:4D instead of the right hand 2D:4D (in the existing literature the right hand 2D:4D, the left hand 2D:4D and the average of the two have been used). We included all participants with data on left hand 2D:4D in the analyses for left hand 2D:4D in that robustness test (and with data on gender and the preference measure tested in each hypothesis as above) and all participants with data on both left and right hand 2D:4D in that robustness test (and data on gender and the preference measure tested in each hypothesis as above).

3. Results

In this section, we report our results in accordance with our pre-registered pre-analysis plan. In the pre-analysis plan, we did not include any exclusion criteria for excluding outliers (we wrote that all observations would be included). But the instructions to the interviewers for measuring 2D:4D in SOEP included some steps to prevent outliers due to mismeasurement or injured fingers. The interviewers were told that the 2D:4D typically lies between 0.8 and 1.1, and that if the measured ratio significantly exceeded this range they were told to repeat the measurement. For some individuals there therefore exists two recorded measurements of the digit ratio (one first measurement and a second measurement due to a suspected first mismeasurement). As we did not write anything about replacing the first measurement with the second measurement in these cases in the pre-analysis plan, we report results based on the first measurement for all individuals as our main results below. We supplement these results with an additional not pre-registered robustness test where we replace the first

Table 1
Descriptives by Gender (Full Sample).

	All	Men		Women	p-value	e-size
R2D:4D	1.000 (0.055)	0.999 (0.060)	<	1.002 (0.051)	0.151	0.049
L2D:4D	1.001 (0.068)	0.997 (0.062)	<	1.004 (0.073)	0.001	0.112
Mean 2D:4D	1.001 (0.049)	0.998 (0.047)	<	1.003 (0.050)	0.003	0.102
Risk	4.291 (2.320)	4.746 (2.332)	>	3.899 (2.238)	<0.001	0.371
Altruism	0.000 (0.769)	-0.094 (0.784)	<	0.081 (0.746)	<0.001	0.229
Positive Reciprocity	8.948 (1.558)	8.891 (1.550)	<	8.996 (1.564)	0.046	0.068
Negative Reciprocity	0.000 (0.768)	0.104 (0.791)	>	-0.090 (0.737)	<0.001	0.255
Trust	5.555 (1.989)	5.641 (1.987)	>	5.479 (1.989)	0.126	0.081

Note: Standard deviations are in parentheses. Reported tests are based on *t*-tests; effect sizes (*cohen's d*) are reported in e-size column. Reported values of Altruism and Neg. Reciprocity are the indices calculated from *z*-scores of the variables. All values are based on full sample without any exclusion criteria such as hand injuries or outliers.

measurement with the second measurement for all individuals where a second measurement is available in the data (a second measurement of the right/left hand digit ratio was available for 90/149 individuals in the data). We carry out this robustness test for the primary hypotheses, the pre-registered exploratory analyses and the pre-registered robustness tests reported below and these results are reported in Tables S2–S4, S11–S13, and S20–S22. We refer to this robustness test as “the corrected sample”.

If the interviewee had a missing or severely injured second digit (2D) or fourth digit (4D), the interviewers were instructed not to measure the injured hand and continue with the other hand. But from comments added by interviewers in the data set, it is evident that in some cases they measured the injured hand and then added a comment about an injured finger. In our main results below, we include all the measured digit ratios even if there was a comment about an injured finger in the data set (as that is most consistent with our pre-analysis plan). But we have added an additional not-pre-registered robustness test where we have excluded digit ratios outside the range of 0.8–1.2 (this range corresponds to about +/- four STDs away from the mean in our data), as it is very unusual to observe values outside this range in previous studies (in the data set we observed 14/30 digit ratios outside this range for the right/left hand digit ratio). This will imply that digit ratios that are outliers due to injured fingers will be excluded from the analysis. We chose this approach instead of trying to identify every recorded measurement in the data set due to injured fingers as it is not always obvious from the interviewers' comments if fingers are severely injured or not. We carry out this robustness test for the primary hypotheses, the pre-registered exploratory analyses and the pre-registered robustness tests reported below and these results are reported in Tables S5–S7, S14–S16, and S23–S25. We refer to this robustness test as “the restricted sample”. For our restricted sample mean 2D:4D analyses we generated the dependent variable, m2D:4D, by calculating the average of restricted right and restricted left hand 2D:4Ds.

When we below report our results for the pre-registered tests, we comment in the text if the findings differ importantly in the not pre-registered tests in “the corrected sample” or “the restricted sample” (and we refer to the results below as “the full sample”). Following the recommendations by a recent paper (Benjamin et al., 2018), we consider results with *p*-values below 0.005 to be statistically significant and results with *p*-values below 0.05 as suggestive evidence; and all statistical tests are based on two-sided tests (these *p*-value cut-offs were mentioned already in the pre-registration).

3.1. Descriptives

Descriptives of the dependent and independent variables are shown in Table 1 (and in Table S1 for the corrected sample and the restricted sample), with results reported separately for men and women (the results for the preference variables are based on the $n=3,482$ individuals with at least one of the right or left hand 2D:4D data in the full sample). We also report the *p*-value of the gender difference of each variable based on an independent samples *t*-test, but these tests were not part of our pre-registered hypothesis tests. As conjectured in the 2D:4D literature, 2D:4D tends to be lower for men than for women. This difference is statistically significant for the left hand digit ratio and the mean of the left hand and right hand digit ratio, but not for the right hand digit ratio in the full sample. Note that the Full Sample includes all the participants including the outliers and the ones with hand injuries. In the corrected sample and the restricted sample there is suggestive evidence ($p<0.05$) for a lower digit ratio for men also for the right hand digit ratio.

Table 2
Regression analysis: primary hypotheses.

	(1) Risk	(2) Altruism	(3) Positive Recip.	(4) Negative Recip.	(5) Trust
R2D:4D	0.040 (0.326)	0.251 (0.282)	0.063 (0.294)	0.037 (0.307)	-0.234 (0.512)
female	-0.361** (0.034)	0.222** (0.035)	0.066 (0.034)	-0.247** (0.034)	-0.085 (0.054)
constant	0.157 (0.326)	-0.372 (0.283)	-0.099 (0.295)	0.095 (0.309)	0.273 (0.512)
N	3431	3350	3422	3378	1391
adj. R ²	0.032	0.012	0.001	0.015	0.001

OLS regressions; robust standard errors in parentheses.
* $p < 0.05$, ** $p < 0.005$.

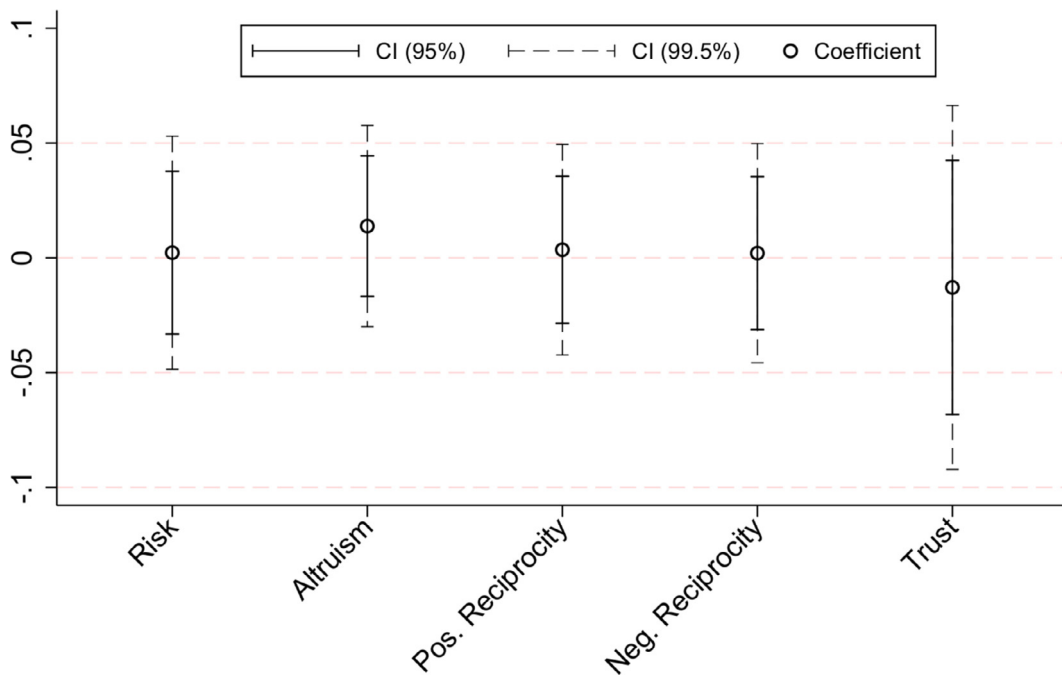


Fig. 1. Standardized Regression Coefficients of Primary Hypotheses.

In line with previous literature on gender differences in economic preferences (Croson and Gneezy, 2009), men are on average more risk taking than women and less altruistic. Using the same measures as we do, Falk et al. investigated the economic and social preferences in 76 countries and with almost 80,000 participants. They found systematic gender differences in positive reciprocity, negative reciprocity and trust, with women on average being more likely to positively reciprocate, less likely to negatively reciprocate and more trusting (Falk et al., 2018). We confirm that finding for negative reciprocity ($p < 0.001$) and positive reciprocity ($p < 0.05$) whereas trust do not vary statistically significantly between men and women in our data.

3.2. Primary hypotheses

The primary hypothesis tests, robustness tests, and exploratory analyses reported below were pre-registered prior to starting the data collection. In Table 2 we present the results for our primary hypotheses in regressions with the economic preference measure as a function of 2D:4D and gender. Note that in the regressions we have standardized all the five preference measures to z-scores so that one unit in the dependent variable represents a standard deviation of the preference measure. We cannot reject the null hypothesis of no association between 2D:4D and economic preferences for any of the five preference measures; and the sign of the coefficient is in the opposite direction of the hypothesis for three of the five hypotheses. The female coefficients confirm the gender differences in preferences observed in the descriptives data.

Table 3
Regression Analysis: Exploratory Analysis 1.

	(1) Risk	(2) Altruism	(3) Positive Recip.	(4) Negative Recip.	(5) Trust
R2D:4D	−0.551 (0.525)	0.128 (0.358)	−0.079 (0.410)	−0.540 (0.366)	0.622 (0.717)
female	−1.638* (0.700)	−0.043 (0.573)	−0.240 (0.600)	−1.489* (0.583)	1.661 (1.011)
R2D:4D X female	1.277 (0.699)	0.265 (0.572)	0.306 (0.598)	1.241* (0.581)	−1.749 (1.011)
constant	0.748 (0.524)	−0.250 (0.359)	0.043 (0.410)	0.671 (0.370)	−0.579 (0.716)
R2D:4D (F)	0.726 (0.463)	0.394 (0.446)	0.228 (0.436)	0.702 (0.451)	−1.128 (0.712)
N	3431	3350	3422	3378	1391
adj. R ²	0.033	0.012	0.000	0.016	0.002

OLS regressions; robust standard errors in parentheses; * $p < 0.05$, ** $p < 0.005$. R2D:4D (F) presents the 2D:4D effects for women, which were calculated by summing the coefficients of R2D:4D and R2D:4D X female variables.

In Fig. 1 we graphically illustrate our results for the five primary hypotheses, by showing 95% and 99.5% confidence intervals of the estimated regression coefficients for the association between 2D:4D and economic preferences. To ease the interpretation of effects we report these results as standardized regression coefficients so that the results show how many standard deviation units economic preferences change for a one standard deviation change in 2D:4D. For risk taking, altruism, positive reciprocity and negative reciprocity the upper and lower limit of the 99.5% confidence interval is within about a ± 0.05 standard deviation change in economic preferences for a one standard deviation change in 2D:4D. The confidence intervals for trust are somewhat wider due to the lower sample size for trust (the trust question was only included in a sub-sample of SOEP-IS). These results thus show precisely estimated null results, suggesting that 2D:4D is not associated with important effects on economic preferences. These results are reinforced by our high power to find significant associations. Based on the observed standard errors we have 80% power to detect a standardized regression coefficient at the 0.005 (0.05) p -value level of 0.066 (0.051) for risk taking, 0.057 (0.044) for altruism, 0.060 (0.046) for positive reciprocity, 0.062 (0.048) for negative reciprocity, and 0.103 (0.079) for trust.

3.3. Exploratory analyses 1: Gender interactions

We pre-registered two exploratory analyses. The first of these tests for a gender difference in the association between 2D:4D and economic preferences by adding an interaction between gender and 2D:4D in the regressions. In testing for the gender interaction coefficient we also report if 2D:4D is statistically significantly associated with economic preferences among men or women. The results, including the interaction effect tests, are reported in Table 3. The gender interaction is not statistically significant in any of the five regressions ($p > 0.05$). For negative reciprocity there is suggestive evidence ($p = 0.033$) of a positive gender interaction with a negative point estimate of the association between 2D:4D and positive reciprocity for men and a positive point estimate for women. However, these point estimates do not provide statistically significant or suggestive evidence for an association between 2D:4D and positive reciprocity among either men or women. The suggestive evidence for the gender interaction is thus likely to be due to randomness. This is reinforced by that there is no statistically significant or suggestive evidence of this gender interaction in the corrected sample or the restricted sample. In the restricted sample there is instead suggestive evidence of a positive gender interaction for risk taking with a negative point estimate of the association between 2D:4D and risk taking for men and a positive point estimate for women. However, also in this case the point estimate is not statistically significant for either men or women (and there is no suggestive evidence of an association either) (see Table S6).

3.4. Exploratory analyses 2: non-linear relationships

Our second exploratory analysis tests for a non-linear relationship between 2D:4D and economic preferences, by adding 2D:4D squared to the regression equation. Such non-linear relationships have been tested in several previous studies (e.g. Brañas-Garza et al., 2013; Galizzi and Nieboer, 2015; Parslow et al., 2019; Lima de Miranda et al., 2018). These results are reported in Table 4. The squared 2D:4D coefficient is not statistically significant in the five regression equations, with the exception of negative reciprocity. For negative reciprocity the squared 2D:4D term is statistically significant with a negative coefficient (and the 2D:4D coefficient has a positive coefficient). This result is also observed in the corrected sample, but not in the restricted sample suggesting that this result is due to outliers due to injured fingers (see Tables S4 and S7).

Table 4
Regression analysis: exploratory analysis 2.

	(1) Risk	(2) Altruism	(3) Positive Recip.	(4) Negative Recip.	(5) Trust
R2D:4D	−0.768 (1.233)	0.316 (0.982)	−1.023 (0.951)	2.293* (0.993)	1.467 (5.709)
female	−0.361** (0.034)	0.222** (0.035)	0.067 (0.034)	−0.249** (0.034)	−0.085 (0.054)
R2D:4D ²	0.370 (0.500)	−0.030 (0.346)	0.498 (0.340)	−1.032** (0.346)	−0.832 (2.756)
constant	0.594 (0.755)	−0.408 (0.644)	0.488 (0.624)	−1.125 (0.654)	−0.593 (2.965)
N	3431	3350	3422	3378	1391
adj. R ²	0.032	0.012	0.000	0.015	−0.000

OLS regressions; robust standard errors in parentheses.

* $p < 0.05$, ** $p < 0.005$.

In checking the data, there is one extreme right hand digit ratio of 2.175 in the full sample and in the corrected sample, but not in the restricted sample. According to the comments by the interviewers this is due to an injury of the 4D finger. If this outlier is removed the squared 2D:4D term is no longer statistically significant in the full sample (and there is no suggestive evidence either). We thus find no robust evidence of a non-linear relationship between 2D:4D and economic preferences. Without this extreme outlier the suggestive evidence for the 2D:4D gender interaction and negative reciprocity in Table 3 above also disappears.

3.5. Robustness tests: Left hand 2D:4D and average of right and left hand 2D:4D

As pre-registered robustness tests we estimated the primary hypotheses and exploratory analyses also for the left hand 2D:4D and the average of the left hand and right hand 2D:4D. For the full sample there is no statistically significant or suggestive evidence of an association between 2D:4D and any of the economic preferences in any of the robustness tests for the primary hypotheses or the exploratory analyses (see Tables S8–S10 and S17–S19).

We also carried out the analyses using the left hand 2D:4D and the average of the left hand and right and 2D:4D using the corrected sample and the restricted sample (see Tables S11–S16 and S20–S25). As noted above these additional robustness tests were not pre-registered. In these analyses there is no statistically significant or suggestive evidence in the robustness tests of the primary hypotheses or the exploratory tests of a gender interaction. But in three cases the squared 2D:4D coefficient is statistically significant, in the robustness tests of a non-linear relationship. For negative reciprocity the squared 2D:4D term for the average 2D:4D of both hands is statistically significant in the corrected sample with a negative coefficient (and the 2D:4D coefficient has a positive coefficient) (see Table S22). As above for right hand 2D:4D this result seem to be due to outliers as the squared term is not statistically significant in the restricted sample (see Table S25), and as above the result is not robust to removing the one extreme outlier due to an injured finger noted above (there is not even suggestive evidence for the squared term without this observation).

The other two cases with a statistically significant squared term is for risk taking in the restricted sample; here the squared term is statistically significant in both the regression for left hand 2D:4D and the regression for the average of the left hand and right hand 2D:4D (see Tables S16 and S25). In both these regressions the squared term has a negative coefficient (and the 2D:4D has a positive coefficient). This implies that risk taking initially increases with 2D:4D until the function peaks, and then it decreases. For left hand 2D:4D the function peaks at 0.97 and for the average of the right hand and the left hand 2D:4D the function peaks at 0.98. The hypothesized sign of 2D:4D and risk taking is a negative sign, which implies that for a sizeable part of the 2D:4D distribution the direction of the association is in the opposite direction of this hypothesis in these models with the squared term (the lower part of the 2D:4D distribution until the function peaks). The results for these two regressions with the squared term are thus not fully consistent with the hypothesized negative association, and we think they are likely to be chance findings. These analyses are also part of the exploratory analyses, and would need confirmation in future confirmatory studies to provide substantive evidence of a non-linear association between 2D:4D and risk taking.

4. Discussion

Our null results suggest that variation in 2D:4D does not explain any important variation in economic preferences between individuals - at least as measured by the measures included here. The point estimates for the five measures of economic preferences are close to zero and are precisely estimated with 99.5% confidence intervals within about a +/- 0.05 standard deviation change in economic preferences for a one standard deviation change in 2D:4D (with the exception of trust, which has somewhat wider confidence intervals due to the smaller sample size).

Why do we find null results while several smaller studies have found indications of 2D:4D correlating with economic preferences? This could be due to the “usual suspects” increasing the risk of false positive results; small sample sizes, publication bias and “researcher degrees of freedom” such as “forking”. For example, in most previous papers there is room to try whether there is a statistically significant association between the studied behavior for the right hand, or for the left hand, or for the average of the two hands, or for men only, or for women only, or for the pooled sample of men and women, etc. Even in studies where researchers are testing specific hypotheses, there are typically some degrees of freedom in how these hypotheses are tested, and it is easy to convince oneself that the statistically significant result is the one that was always intended as the focus. (We do not think that most researchers are “p-hacking” in the sense that they are actively trying to find only statistically significant results somewhere and only communicate these to the world, but that “forking” is very natural and hard to avoid without a pre-analysis plan.) Our sample size was about 5 times larger than the previously largest study in this field, and the pre-registration of our hypothesis tests leaves little room for “researcher degrees of freedom” affecting our results.

Another interpretation of our findings is that testosterone exposure in utero is unimportant for economic preferences. Recent evidence by [Van Leeuwen et al. \(2020\)](#) suggest that this may be the case – using direct measures of testosterone and estrogen from umbilical cord blood at birth they find no evidence for an association of these hormones and economic preferences (risk preferences, competitiveness, time preferences and social preferences) in a sample of 217 adults. In this study the authors also find no evidence for associations between 2D:4D and economic preferences, including risk preferences, positive reciprocity, negative reciprocity and trust, in a sample of 597 adults. Another possibility is that 2D:4D might not be a valid proxy for testosterone exposure in utero. The evidence of a link between 2D:4D and testosterone exposure, while coming from several strands of literature, is mainly indirect and the results are often inconsistent. First, it has been argued that 2D:4D is lower in men than in women. While men on average have a lower 2D:4D also in our study, the difference is small, and this result is not always found (for example [Apicella et al. \(2015\)](#) find no gender difference in a large sample of the Hadza population). Second, the most direct evidence comes from a study on the testosterone-to-estradiol ratio in amniotic fluid in a sample of 29 children ([Lutchmaya et al., 2004](#)). They report a statistically significant negative correlation between this ratio and right hand 2D:4D, even after controlling for gender, while the correlation for the left hand is reported as insignificant. Similarly, another study finds a weak negative correlation between the testosterone levels in the mother's plasma and 2D:4D of 106 newborns for both hands and both sexes ([Ventura et al., 2013](#)). Yet, in a considerably larger study (341 children) the result is not statistically significant ([Hollier et al., 2015](#)). Third, based on the theory of sex hormone transfer in utero where the idea is that female fetuses receive transfers of testosterone from their male co-twin ([Miller, 1994](#)), some studies compare the 2D:4D of same sex and opposite sex twins. While one study finds that women with a male co-twin have lower left hand 2D:4D than women with a female co-twin, there is no statistically significant result for the right hand ([van Anders et al., 2006](#)). Another study similarly reports a significant result for mean 2D:4D ([Voracek and Dressler, 2007](#)), whereas studies with larger sample sizes report no significant differences ([Hiraishi et al., 2012](#); [Medland et al., 2008](#)).

There are also studies that have attempted to establish a link between 2D:4D and congenital adrenal hyperplasia (CAH, a disease that results in an excess production of testosterone) as well as CAG repeat polymorphism (which affects the transcriptional activity of the androgen receptor). For CAH the results consist of a mix of positive and null ones for CAH (see, e.g. [Brown et al., 2002](#); [Okten et al., 2002](#); [Buck et al., 2003](#)) whereas for CAG the meta-analytic results are null results (see e.g. [Voracek et al., 2019](#)). In a study with a sample size of 51 men with Klinefelter's Syndrome, in which testosterone deficiency is one of the main features, these individuals were observed to have higher 2D:4D than the control group ([Manning et al., 2013](#)).

Finally, it should also be noted that we use self-reported measures in our study, and that many of the other experiments on 2D:4D and economic preferences have been on incentivized measures. While the hypothetical (self-reported) measures used here were validated ([Falk et al., 2018](#); [Dohmen et al., 2011](#)) and they predict incentivized decisions successfully, it remains to be shown whether monetary incentives have an interaction role between 2D:4D and economic preferences. The evidence from previous studies is not clear on this matter. On the one hand, [Neyse et al. \(2016\)](#) showed that without incentives, 2D:4D and overconfidence were statistically significantly negatively correlated but the relationship disappeared when monetary incentives were introduced ($N=281$). In terms of risk attitudes on the other hand, the potential role of incentives is more puzzling. [Brañas-Garza et al. \(2018\)](#) found that 2D:4D and risk preferences were statistically significantly negatively correlated in an incentivized task but no statistically significant relationship was found with the self-reported question ($N=704$). Yet, in another study that tested the relationship between risk taking and 2D:4D ([Brañas-Garza and Rustichini, 2011](#)) an unincentivized risk elicitation task yielded a significant negative relationship ($N=188$).

In sum, while there is some evidence in support of a link between 2D:4D and testosterone exposure in utero, the results do not provide strong evidence and it should be a research priority to establish if such a link actually exists based on a large sample pre-registered study. We would also encourage pre-registered large sample replication studies on the association between 2D:4D and other traits previously linked to 2D:4D such as personality, sexual orientation and spatial ability.

5. Open practices statement

The pre-analysis plan and hypotheses of this study were formally pre-registered in Open Science Framework prior to data collection (<https://osf.io/5vdpdn/>). The analysis code and the finger length measurement protocol can be found in the

same link. The dataset from SOEP-IS is open to whole research community. Please contact LN (lneyse@diw.de) for further information.

Declaration of Competing Interest

None.

Acknowledgments

We thank Johannes Koenig, David Richter and Michael Kraemer for their support during the data collection and analysis. We thank Martin Voracek for his comments on the manuscript. For financial support we can thank the Swedish Foundation for Humanities and Social Sciences (NHS14-1719:1), the Jan Wallander and Tom Hedelius Foundation, the Knut and Alice Wallenberg Foundation and the Marianne and Marcus Wallenberg Foundation (A.D. is a Wallenberg Scholar) and the Austrian Science Fund (FWF, SFB F63). L.N. acknowledges the financial support of German Research Foundation (DFG) through project CRC TRR 190 Rationality and Competition and Leibniz Association through project SAW-10868. L.N. is a member of Berlin Centre for Consumer Policies (BCCP).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.jebo.2021.02.029](https://doi.org/10.1016/j.jebo.2021.02.029).

References

- Alonso, J., Di Paolo, R., Ponti, G., Sartarelli, M., 2018. Facts and misconceptions about 2d:4d, social and risk preferences. *Front. Behav. Neurosci.* 12. doi:[10.3389/fnbeh.2018.00022](https://doi.org/10.3389/fnbeh.2018.00022).
- van Anders, S.M., Vernon, P.A., Wilbur, C.J., 2006. Finger-length ratios show evidence of prenatal hormone-transfer between opposite-sex twins. *Horm. Behav.* 49 (3), 315–319. doi:[10.1016/j.yhbeh.2005.08.003](https://doi.org/10.1016/j.yhbeh.2005.08.003).
- Apicella, C.L., Dreber, A., Campbell, B., Gray, P.B., Hoffman, M., Little, A.C., 2008. Testosterone and financial risk preferences. *Evolut. Hum. Behav.* 29 (6), 384–390. doi:[10.1016/j.evolhumbehav.2008.07.001](https://doi.org/10.1016/j.evolhumbehav.2008.07.001).
- Apicella, C.L., Tobolsky, V.A., Marlowe, F.W., Miller, K.W., 2015. Hadza hunter-gatherer men do not have more masculine digit ratios (2d:4d). *Am. J. Phys. Anthropol.* 159 (2), 223–232. doi:[10.1002/ajpa.22864](https://doi.org/10.1002/ajpa.22864).
- Arnold, A.P., 2009. The organizational–activational hypothesis as the foundation for a unified theory of sexual differentiation of all mammalian tissues. *Horm. Behav.* 55 (5), 570–578. doi:[10.1016/j.yhbeh.2009.03.011](https://doi.org/10.1016/j.yhbeh.2009.03.011).
- Aycinena, D., Baltaduonis, R., Rentschler, L., 2014. Risk preferences and prenatal exposure to sex hormones for ladinos. *PLoS One* 9 (8), e103332. doi:[10.1371/journal.pone.0103332](https://doi.org/10.1371/journal.pone.0103332).
- Barel, E., 2019. 2D: 4d, optimism, and risk taking. *Curr. Psychol.* 38 (1), 204–212. doi:[10.1007/s12144-017-9598-4](https://doi.org/10.1007/s12144-017-9598-4).
- Baron-Cohen, S., 2002. The extreme male brain theory of autism. *Trends Cogn. Sci. (Regul. Ed.)* 6 (6), 248–254. doi:[10.1016/S1364-6613\(02\)01904-6](https://doi.org/10.1016/S1364-6613(02)01904-6).
- Benjamin, D.J., Berger, J.O., Johannesson, M., Nosek, B.A., Wagenmakers, E.-J., Berk, R., Bollen, K.A., Brembs, B., Brown, L., Camerer, C., et al., 2018. Redefine statistical significance. *Nat. Hum. Behav.* 2 (1), 6. doi:[10.1038/s41562-017-0189-z](https://doi.org/10.1038/s41562-017-0189-z).
- Van den Bergh, B., Dewitte, S., 2006. Digit ratio (2d:4d) moderates the impact of sexual cues on men's decisions in ultimatum games. *Proc. R. Soc. B* 273 (1597), 2091–2095. doi:[10.1098/rspb.2006.3550](https://doi.org/10.1098/rspb.2006.3550).
- Bönte, W., Procher, V.D., Urbig, D., 2016. Biology and selection into entrepreneurship—the relevance of prenatal testosterone exposure. *Entrep. Theory Pract.* 40 (5), 1121–1148. doi:[10.1111/etap.12165](https://doi.org/10.1111/etap.12165).
- Bos, P.A., Terburg, D., Van Honk, J., 2010. Testosterone decreases trust in socially naive humans. *Proc. Natl. Acad. Sci.* 107 (22), 9991–9995. doi:[10.1073/pnas.0911700107](https://doi.org/10.1073/pnas.0911700107).
- Brañas-Garza, P., Espín, A.M., García-Muñoz, T., Kovářík, J., 2019. Digit ratio (2d: 4d) and prosocial behaviour in economic games: no direct correlation with generosity, bargaining or trust-related behaviours. *Biol. Lett.* 15 (8), 20190185. doi:[10.1098/rsbl.2019.0185](https://doi.org/10.1098/rsbl.2019.0185).
- Brañas-Garza, P., Galizzi, M.M., 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. (Philadelphia)* 59 (3), 1131–1157. doi:[10.1111/iere.12299](https://doi.org/10.1111/iere.12299).
- Brañas-Garza, P., Kovářík, J., 2013. Digit ratios and social preferences: a comment on Buser (2012).
- Brañas-Garza, P., Kovářík, J., Neyse, L., 2013. Second-to-fourth digit ratio has a non-monotonic impact on altruism. *PLoS One* 8 (4), e60419. doi:[10.1371/journal.pone.0060419](https://doi.org/10.1371/journal.pone.0060419).
- Brañas-Garza, P., Rustichini, A., 2011. Organizing effects of testosterone and economic behavior: not just risk taking. *PLoS One* 6 (12), e29842. doi:[10.1371/journal.pone.0029842](https://doi.org/10.1371/journal.pone.0029842).
- Brown, W.M., Hines, M., Fane, B.A., Breedlove, S.M., 2002. Masculinized finger length patterns in human males and females with congenital adrenal hyperplasia. *Horm. Behav.* 42 (4), 380–386. doi:[10.1006/hbeh.2002.1830](https://doi.org/10.1006/hbeh.2002.1830).
- Buck, J.J., Williams, R.M., Hughes, I.A., Acerini, C.L., 2003. In-utero androgen exposure and 2nd to 4th digit length ratio-comparisons between healthy controls and females with classical congenital adrenal hyperplasia. *Hum. Reprod.* 18 (5), 976–979.
- Buser, T., 2012. Digit ratios, the menstrual cycle and social preferences. *Games Econ. Behav.* 76 (2), 457–470. doi:[10.1016/j.geb.2012.07.006](https://doi.org/10.1016/j.geb.2012.07.006).
- Chicaiza-Becerra, L.A., García-Molina, M., 2017. Prenatal testosterone predicts financial risk taking: evidence from latin america. *Pers. Individ. Differ.* 116, 32–37. doi:[10.1016/j.paid.2017.04.021](https://doi.org/10.1016/j.paid.2017.04.021).
- Coates, J.M., Gurnell, M., Rustichini, A., 2009. Second-to-fourth digit ratio predicts success among high-frequency financial traders. *Proc. Natl. Acad. Sci.* 106 (2), 623–628. doi:[10.1073/pnas.0810907106](https://doi.org/10.1073/pnas.0810907106).
- Crosen, R., Gneezy, U., 2009. Gender differences in preferences. *J. Econ. Lit.* 47 (2), 448–474. doi:[10.1257/jel.47.2.448](https://doi.org/10.1257/jel.47.2.448).
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G., 2011. Individual risk attitudes: measurement, determinants, and behavioral consequences. *J. Eur. Econ. Assoc.* 9 (3), 522–550. doi:[10.1111/j.1542-4774.2011.01015.x](https://doi.org/10.1111/j.1542-4774.2011.01015.x).
- Dreber, A., Hoffman, M., 2007. Portfolio selection in utero.
- Dreber, A., Johannesson, M., 2019. Sex Hormones and Economic Decision Making in the Lab: a Review of the Causal Evidence. *Routledge International Handbook of Social Neuroendocrinology*.
- Drichoutis, A.C., Nayga, R.M., 2015. Do risk and time preferences have biological roots? *South Econ. J.* 82 (1), 235–256. doi:[10.4284/0038-4038-2013.246](https://doi.org/10.4284/0038-4038-2013.246).
- Eisenegger, C., Haushofer, J., Fehr, E., 2011. The role of testosterone in social interaction. *Trends Cogn. Sci. (Regul. Ed.)* 15 (6), 263–271. doi:[10.1016/j.tics.2011.04.008](https://doi.org/10.1016/j.tics.2011.04.008).

- Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., Sunde, U., 2018. Global evidence on economic preferences. *Q. J. Econ.* 133 (4), 1645–1692. doi:10.1093/qje/qjy013.
- Galizzi, M.M., Nieboer, J., 2015. Digit ratio (2d:4d) and altruism: evidence from a large, multi-ethnic sample. *Front. Behav. Neurosci.* 9. doi:10.3389/fnbeh.2015.00041.
- Garbarino, E., Slonim, R., Sydnor, J., 2011. Digit ratios (2d:4d) as predictors of risky decision making for both sexes. *J. Risk Uncertain.* 42 (1), 1–26. doi:10.1007/s11166-010-9109-6.
- Gelman, A., Loken, E., 2013. The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There Is No “Fishing Expedition” or “p-hacking” and the Research Hypothesis Was Posited Ahead of Time.
- Gneezy, U., Leonard, K.L., List, J.A., 2009. Gender differences in competition: evidence from a matrilineal and a patriarchal society. *Econometrica* 77 (5), 1637–1664. doi:10.3982/ECTA6690.
- Goebel, J., Grabka, M.M., Liebig, S., Kroh Martin, D., Schröder, C., Schupp, J., 2019. The German socio-economic panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik* 239 (2), 345–360. doi:10.1515/jbnst-2018-0022.
- Grimbos, T., Dawood, K., Burriss, R.P., Zucker, K.J., Puts, D.A., 2010. Sexual orientation and the second to fourth finger length ratio: a meta-analysis in men and women. *Behav. Neurosci.* 124 (2), 278. doi:10.1037/a0018764.
- Hiraishi, K., Sasaki, S., Shikishima, C., 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 (3), 711–724. doi:10.1007/s10508-011-9889-z.
- Hollier, L.P., Keelan, J.A., Jannadass, E.S., Maybery, M.T., Hickey, M., Whitehouse, A.J., 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 (2), 111–117. doi:10.1016/j.earlhumdev.2014.12.011.
- Hönekopp, J., Schuster, M., 2010. A meta-analysis on 2d: 4d and athletic prowess: substantial relationships but neither hand out-predicts the other. *Pers. Individ. Differ.* 48 (1), 4–10. doi:10.1016/j.paid.2009.08.009.
- Hönekopp, J., Watson, S., 2011. Meta-analysis of the relationship between digit-ratio 2d: 4d and aggression. *Pers. Individ. Differ.* 51 (4), 381–386. doi:10.1016/j.paid.2010.05.003.
- Lombardo, M.V., Ashwin, E., Auyeung, B., Chakrabarti, B., Taylor, K., Hackett, G., Bullmore, E.T., Baron-Cohen, S., 2012. Fetal testosterone influences sexually dimorphic gray matter in the human brain. *J. Neurosci.* 32 (2), 674–680. doi:10.1523/JNEUROSCI.4389-11.2012.
- Lutchmaya, S., Baron-Cohen, S., Raggatt, P., Knickmeyer, R., Manning, J., 2004. 2Nd to 4th digit ratios, fetal testosterone and estradiol. *Early Hum. Dev.* 77 (1), 23–28. doi:10.1016/j.earlhumdev.2003.12.002.
- Manning, J.T., Kilduff, L.P., Trivers, R., 2013. Digit ratio (2d:4d) in Klinefelter’s syndrome. *Andrology* 1 (1), 94–99. doi:10.1111/j.2047-2927.2012.00013.x.
- Manning, J.T., Scutt, D., Wilson, J., 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 (11), 3000–3004. doi:10.1093/humrep/13.11.3000.
- Medland, S.E., Loehlin, J.C., Martin, N.G., 2008. No effects of prenatal hormone transfer on digit ratio in a large sample of same-and opposite-sex dizygotic twins. *Pers. Individ. Differ.* 44 (5), 1225–1234. doi:10.1016/j.paid.2007.11.017.
- Miller, E.M., 1994. Prenatal sex hormone transfer: a reason to study opposite-sex twins. *Pers. Individ. Differ.* 17 (4), 511–529. doi:10.1016/0191-8869(94)90088-4.
- Millet, K., Dewitte, S., 2010. The presence of aggression cues inverts the relation between digit ratio (2d:4d) and prosocial behaviour in a dictator game. *Br. J. Psychol.* 100 (1), 151–162. doi:10.1348/000712608X324359.
- Lima de Miranda, K., Neyse, L., 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.
- Neyse, L., Bosworth, S., Ring, P., Schmidt, U., 2016. Overconfidence, incentives and digit ratio. *Sci. Rep.* 6, 23294.
- Neyse, L., Vieider, F.M., Ring, P., Probst, C., Kaernbach, C., van Eimeren, T., Schmidt, U., 2020. Risk attitudes and digit ratio (2d: 4d): evidence from prospect theory. *J. Risk Uncertain.* 1–23. doi:10.1007/s11166-020-09321-w.
- Nicolaou, N., Patel, P.C., Wolfe, M.T., 2017. Testosterone and tendency to engage in self-employment. *Manag. Sci.* 64 (4), 1825–1841. doi:10.1287/mnsc.2016.2664.
- Okten, A., Kalyoncu, M., Yariş, N., 2002. The ratio of second- and fourth-digit lengths and congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Early Hum. Dev.* 70 (1–2), 47–54. doi:10.1016/S0378-3782(02)00073-7.
- Parslow, E., Ranehill, E., Zethraeus, N., Blomberg, L., von Schoultz, B., Hirschberg, A.L., Johannesson, M., Dreber, A., 2019. The digit ratio (2d:4d) and economic preferences: no robust associations in a sample of 330 women. *J. Econ. Sci. Assoc.* doi:10.1007/s40881-019-00076-y.
- Phoenix, C.H., Goy, R.W., Gerall, A.A., 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.
- Puts, D.A., McDaniel, M.A., Jordan, C.L., Breedlove, S.M., 2008. Spatial ability and prenatal androgens: meta-analyses of congenital adrenal hyperplasia and digit ratio (2d: 4d) studies. *Arch. Sex Behav.* 37 (1), 100. doi:10.1007/s10508-007-9271-3.
- Richter, D., Schupp, J., 2015. The soep innovation sample (SOEP) IS. *Schmollers Jahrbuch: Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts- und Sozialwissenschaften* 135 (3), 389–400. doi:10.3790/schm.135.3.389.
- Sapienza, P., Zingales, L., Maestriperi, D., 2009. Gender differences in financial risk aversion and career choices are affected by testosterone. *Proc. Natl. Acad. Sci.* 106 (36), 15268–15273. doi:10.1073/pnas.0907352106.
- Schipper, B., 2014. *Sex Hormones and Choice under Risk*. Technical Report. University of California, Davis, Department of Economics.
- Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* 22 (11), 1359–1366. doi:10.1177/0956797611417632.
- Stenstrom, E., Saad, G., Nepomuceno, M.V., Mendenhall, Z., 2011. Testosterone and domain-specific risk: digit ratios (2d:4d and rel2) as predictors of recreational, financial, and social risk-taking behaviors. *Digit Ratio (2D:4D) Individ. Differ. Res.* 51 (4), 412–416. doi:10.1016/j.paid.2010.07.003.
- Sytsma, T., 2014. Handling Risk: Testosterone and Risk Preference, Evidence from Dhaka, Bangladesh.
- Thaler, R.H., Johnson, E.J., 1990. Gambling with the house money and trying to break even: the effects of prior outcomes on risky choice. *Manag. Sci.* 36 (6), 643–660. doi:10.1287/mnsc.36.6.643.
- Van Honk, J., Montoya, E.R., Bos, P.A., Van Vugt, M., Terburg, D., 2012. New evidence on testosterone and cooperation. *Nature* 485 (7399), E4. doi:10.1038/nature11136.
- Van Leeuwen, B., Smeets, P., Bovet, J., Nave, G., Stieglitz, J., Whitehouse, A., 2020. Do sex hormones at birth predict later-life economic preferences? evidence from a pregnancy birth cohort study. *Proc. R. Soc. B* 287 (1941), 20201756.
- Ventura, T., Gomes, M., Pita, A., Neto, M., Taylor, A., 2013. Digit ratio (2d:4d) in newborns: influences of prenatal testosterone and maternal environment. *Early Hum. Dev.* 89 (2), 107–112. doi:10.1016/j.earlhumdev.2012.08.009.
- Voracek, M., Dressler, S.G., 2007. Digit ratio (2d:4d) in twins: heritability estimates and evidence for a masculinized trait expression in women from opposite-sex pairs. *Psychol. Rep.* 100 (1), 115–126. doi:10.2466/pr0.100.1.115-126.
- Voracek, M., Kossmeier, M., Tran, U.S., 2019. Which data to meta-analyze, and how? *Z. Psychol.* 227, 64–82. doi:10.1027/2151-2604/a000357.
- Voracek, M., Loibl, L.M., 2009. Scientometric analysis and bibliography of digit ratio (2d:4d) research, 1998–2008. *Psychol. Rep.* 104 (3), 922–956. doi:10.2466/PRO.104.3.922-956.
- Voracek, M., Pietschnig, J., Nader, I.W., Stieger, S., 2011. Digit ratio (2d: 4d) and sex-role orientation: further evidence and meta-analysis. *Pers. Individ. Differ.* 51 (4), 417–422. doi:10.1016/j.paid.2010.06.009.
- Zethraeus, N., Kocoska-Maras, L., Ellingsen, T., Von Schoultz, B., Hirschberg, A.L., Johannesson, M., 2009. A randomized trial of the effect of estrogen and testosterone on economic behavior. *Proc. Natl. Acad. Sci.* 106 (16), 6535–6538. doi:10.1073/pnas.0812757106.