Psychoneuroendocrinology (2014) 43, 1—10

Available online at www.sciencedirect.com

ScienceDirect

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

Can exposure to prenatal sex hormones (2D:4D) predict cognitive reflection?☆,☆☆

Antoni Bosch-Domènech [a], Pablo Brañas-Garza [b,*], Antonio M. Espín [c]

[a] *Department of Economics and Business, Universitat Pompeu Fabra and Barcelona Graduate School of Economics, 08005 Barcelona, Spain*
[b] *Business School, Middlesex University London, NW4 4BT London, UK*
[c] *GLOBE, Departamento de Teoría e Historia Económica, Universidad de Granada, 18071 Granada, Spain*

Received 10 October 2013; received in revised form 27 January 2014; accepted 28 January 2014

KEYWORDS
Cognitive Reflection Test;
2D:4D;
Prenatal testosterone;
Patience;
Mathematical proficiency;
Sex

Summary The Cognitive Reflection Test (CRT) is a test introduced by Frederick (2005). The task is designed to measure the tendency to override an intuitive response that is incorrect and to engage in further reflection that leads to the correct response. The consistent sex differences in CRT performance may suggest a role for prenatal sex hormones. A now widely studied putative marker for relative prenatal testosterone is the second-to-fourth digit ratio (2D:4D). This paper tests to what extent 2D:4D, as a proxy for the prenatal ratio of testosterone/estrogens, can predict CRT scores in a sample of 623 students. After controlling for sex, we observe that a lower 2D:4D (reflecting a relative higher exposure to testosterone) is significantly associated with a higher number of correct answers. The result holds for both hands' 2D:4Ds. In addition, the effect appears to be stronger for females than for males. We also control for patience and math proficiency, which are significantly related to performance in the CRT. But the effect of 2D:4D on performance in CRT is not reduced with these controls, implying that these variables are not mediating the relationship between digit ratio and CRT.

1. Introduction

The Cognitive Reflection Test (CRT) is a three-item test introduced by Frederick (2005). The task, of an algebraic nature, is designed to measure the tendency to override an intuitive response that is incorrect and to engage in further reflection that leads to the correct response. When answering the test, many people give the first response that comes to mind without thinking further and not realizing that it cannot be the right answer. For instance, the first item from the CRT is: "*A bat and a ball cost $1.10. The bat costs $1.00 more than the ball. How*

☆☆ The authors' names are placed in alphabetic order. All of them contributed equally to the paper.
* Corresponding author at: Business School, Middlesex University London, Hendon, Campus, The Burroughs, London NW4 4BT, UK. Tel.: +44 020 8411 4262.
E-mail address: branasgarza@gmail.com (P. Brañas-Garza).

0306-4530/$ — see front matter
http://dx.doi.org/10.1016/j.psyneuen.2014.01.023

much does the ball cost? _____ Cents." A glib, incorrect, and frequent answer is 10 cents; the correct answer is 5 cents (see the complete test in Appendix). Mathematical ability is no guarantee against making the error. What makes the CRT different from problem-solving or math tests is that the latter tests do not usually trigger a plausible intuitive response that must be overridden.

As Kahneman and Frederick (2002) made clear, the framework of an incorrectly primed initial response that must be overridden fits in nicely with currently popular (in psychology) dual-process frameworks, one emotional/impatient and the second one deliberative/patient (e.g. Bernheim and Rangel, 2004; Fudenberg and Levine, 2006; Alter et al., 2007; Brocas and Carrillo, 2008). The dual process of emotional/deliberative mental systems has received different names: Fast and slow thinking, hot and cold, locomotion and assessment, automatic and controlled thought (see Camerer et al., 2005).

Frederick (2005) observed that with as few as three items his CRT was able to predict performance on measures of temporal discounting, risk preference, and the tendency to choose high-expected-value gambles. Moreover, CTR scores reflect individual differences in cognitive style that predict important daily-life "decisions" such as whether to believe in God/paranormal phenomena (Pennycook et al., 2012; Shenhav et al., 2012) and making utilitarian choices in moral dilemmas (Paxton et al., 2012). A large literature has developed about the relation between CRT and performance, but the data have proved to be inconsistent in some instances (e.g. Cokely and Kelley, 2009; Oechssler et al., 2009; Campitelli and Labollita, 2010; Koehler and James, 2010; Toplak et al., 2011). Yet, the larger number of correct responses to the CRT by males appears to be a robust result (e.g. Frederick, 2005; Oechssler et al., 2009; Brañas-Garza et al., 2012). While many reasons can account for this result, including differences in upbringing and education of males and females, the sex differences in CRT answers may suggest a role for prenatal organizational hormones, particularly testosterone. Traits that may be linked with prenatal exposure to testosterone expression are, among others, spatial/mathematical skills (e.g. Geschwind and Galaburda, 1985; Grimshaw, 1995); performance in computer science (Brosnan et al., 2011); heightened attention to detail, intensified focus, and narrow interests (Baron-Cohen et al., 2005); less emotion recognition, eye contact and social sensitivity, a poorer ability to judge what others are thinking or feeling, lack of empathy (Baron-Cohen et al., 2004).

A now widely studied putative marker for prenatal sex hormones exposure or, more precisely, for the relative exposure to testosterone compared to estrogens while in uterus, is the second-to-fourth digit ratio (2D:4D), such that a lower ratio (i.e. a shorter index finger in comparison with the ring finger) indicates a higher relative exposure to testosterone (e.g. Manning et al., 1998; Zheng and Cohn, 2011; Auger et al., 2013; Manning et al., 2013). Earlier studies that have stood the test of replication have reported that 2D:4D varies by sex and ethnicity but that male 2D:4D tends to be lower than female 2D:4D in all ethnic groups and the effect is strongest in the right hand (Manning, 2002). These differences emerge prenatally and appear to be stable during the developing years (e.g. Manning, 2002; McIntyre et al., 2005; Trivers et al., 2006).

The 2D:4D literature is large. While a number of failed replications have been reported, 2D:4D appears to be successfully associated with cognitive abilities (Brañas-Garza and Rustichini, 2011); impulsivity (Hanoch et al., 2012); aggression (Bailey and Hurd, 2005; Coyne et al., 2007; Hampson et al., 2008) and risk-taking (Coates et al., 2009; Sapienza et al., 2009; Brañas-Garza and Rustichini, 2011; Garbarino et al., 2011; Stenstrom et al., 2011), among other effects on personality and cognition.

The purpose of the paper is to test to what extent 2D:4D, as a proxy for prenatal exposure to testosterone, correlates with the CRT results in a non-random sample of 623 students (260 males). Since 2D:4D is lower in males than females and males score higher than females in CRT, our prediction is that 2D:4D and CRT will show a negative correlation. Given that the cognitive mechanisms involved in answering the CRT may share common underlying processes with those engaged in mathematical and time-discounting decisions (Frederick, 2005), we include in the analysis the results of mathematical and time-discounting tests to control for possible confounding factors. Interestingly, our analysis shows that 2D:4D is related to CRT performance beyond patience and math skills. However, as a caution, it should be noted that some papers appear to question the notion that differences in digit ratios solely reflect variation in prenatal androgen exposure (e.g. Berenbaum et al., 2009; Wallen, 2009), while others (Hampson and Sankara, 2013) even question that prenatal androgen exposure is related to the 2D:4D ratio (but see Hönekopp, 2013). If this view prevailed, then the current results would be showing a relationship of cognitive reflection with 2D:4D and not, in a straightforward way, with the relative exposure to prenatal testosterone.

2. Methods

In October 2011, 927 first-year students at the College of Business and Economics of the University of Granada (Spain) were asked to participate in a survey-experiment at the Laboratory of Experimental Economics, EGEO. Participation was voluntary and the number of participants ended up being 659 (71% of the population), distributed in 27 sessions. All subjects gave written informed consent to participate. We excluded from the sample those observations with missing values in any of the variables used in this paper. To ensure ethnic homogeneity, three non-Caucasian subjects were also excluded from the sample. The resulting sample was composed of 623 Caucasian subjects (260 males; age: mean $\pm$ SD = 19.1 $\pm$ 2.3).

During a session, using a computer-based system, participants were asked to complete several questionnaires on their socio-demographic characteristics, were tested for their time-discounting attitudes, and answered a math test with four questions, three of which are straightforward. After responding to the computer-based questionnaires, participants answered the CRT three questions using paper and pencil. No time pressure was imposed in any of the processes. Participants were also asked to play some economic games, not considered in this paper. (For the details of the survey-experiment, with another sample, see Exadaktylos et al., 2013).

To test the participants for their time-discounting attitudes (i.e. their willingness to delay gratification, or "patience"), they were presented with two series of intertemporal decisions involving hypothetical monetary rewards. Previous studies have shown that the distribution of individual choices in time preference tests is not significantly altered by the existence of real (vs. hypothetical) incentives, either within or between subjects (e.g. Johnson and Bickel, 2002; Madden et al., 2004; Lagorio and Madden, 2005; but see Coller and Williams, 1999). Participants faced a total of six decisions in each of the two subtasks. In the first decision of the first subtask, participants had to choose between €5 to be received "today" (sooner option) and €5 to be received "tomorrow" (later option). The remaining five decisions kept the sooner reward constant while increasing the later reward, in this order: €6, €7, €8, €9, €10. The second subtask was identical but now the sooner option was €150 to be received in 1 month time, while the later option went from €150 to €250, in €20 increments, to be received in seven months' time (for similar tasks, see e.g. Coller and Williams, 1999; Harrison et al., 2002; Espín et al., 2012). The total number of "sooner" choices (from 0 to 12) is our measure of impatience. We excluded from the sample the 13 subjects making inconsistent choices in any of the subtasks (i.e. non-monotonic patterns or multiple switching from sooner to later reward).

The questions for the CRT and the math test are presented in Appendix. We describe below the results of these two tests by the number of correct answers to them. The math questions come from "Section K" of *Encuesta de Protección Social* (2009) by the Government of Chile.

After taking the tests, the participants were asked one by one to have their two hands scanned using a high-resolution scanner (Canon Slide 90) and their fingers were measured, in mm, from the middle of the basal crease to the tip of the finger using Photoshop. Computer-assisted measurements of 2D:4D from scanned pictures have been found to be more precise and reliable than measurements using other methods (Allaway et al., 2009; Kemper and Schwerdtfeger, 2009). The 2D:4D of the scanned pictures was measured twice for each hand at an interval of one month by the same experienced measurer (not involved in this paper). These measurements displayed a high repeatability (right hand: intraclass correlation coefficient (ICC) = 0.9566, $P < 0.001$, left hand: ICC = 0.9440, $P < 0.001$) and were averaged to obtain a single value of the 2D:4D ratio for each hand.

2.1. Ethics statement

All participants in the experiments reported in the manuscript were informed about the content of the experiment before they participated and provided written consent. Besides, their anonymity was always preserved (in agreement with the Spanish Law 15/1999 for Personal Data Protection) by assigning them a random numerical code, which would identify them in the system. No association was ever made between their real names and the results. As it is standard in socio-economic experiments, no ethic concerns are involved other than preserving the anonymity of participants. This procedure was checked and approved by the Vice dean of Research of the School of Economics of the University of Granada, the institution hosting the experiment.

3. Results

Descriptive statistics of the 2D:4D measurements, including tests of normality, are presented in Table 1. The results are displayed separately for males and females and for left and right hands. We find no significant departure from normality of the 2D:4D data except in the case of males' right hand, for which the normality test reaches a marginal $P = 0.099$, due to a non-normally skewed distribution ($P = 0.034$).

The digit ratio is significantly higher in the left hand than in the right hand for both men (two-sided t-test: $t_{259} = 3.2708$, $P = 0.001$) and women ($t_{362} = 2.4716$, $P = 0.014$). In line with previous literature (e.g. Phelps, 1952; Williams et al., 2003; Manning et al., 2007), the digit ratio was found to be lower for

Table 1 Descriptive statistics of 2D:4D.

	Males		Females	
	Left	Right	Left	Right
Mean	0.9651	0.9597	0.9749	0.9717
SD	0.0317	0.0333	0.0316	0.0332
SEM	0.0020	0.0021	0.0017	0.0017
Median	0.9639	0.9585	0.9737	0.9695
Skewness	0.2403	0.321	−0.013	0.180
P-value	0.109	0.034	0.915	0.156
Kurtosis	2.809	3.026	2.932	3.181
P-value	0.617	0.763	0.922	0.394
Normal (Chi^2)	2.84	4.63	0.02	2.75
P-value	0.241	0.099	0.989	0.253

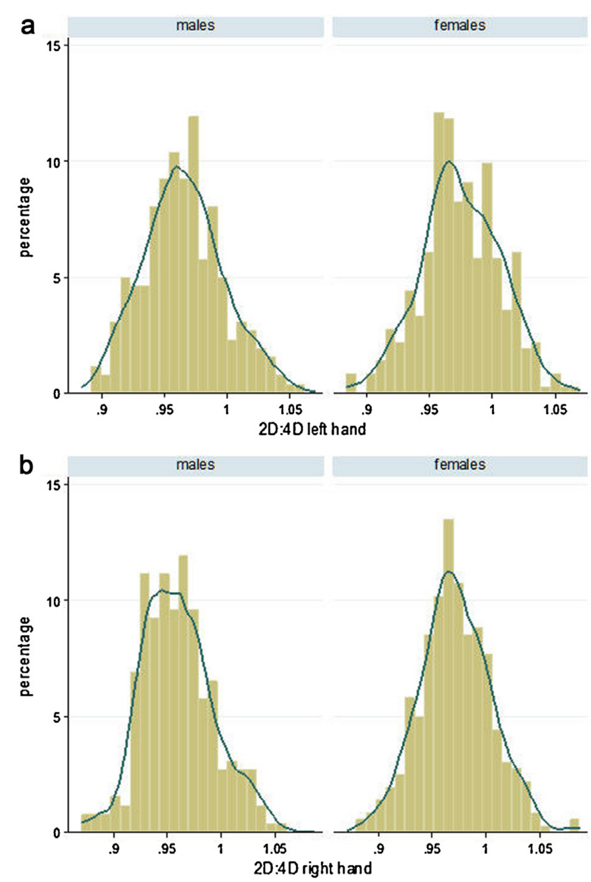

Figure 1 Distribution of 2D:4D: histogram and kernel density. Figure reports the histogram and kernel density estimation of 2D:4D in our sample. The results are displayed separately for males ($n = 260$) and females ($n = 363$) and for the left hand (panel *a*) and right hand (panel *b*). More information is provided in Table 1.

Table 2 CRT: % of correct answers by sex.

	Males (%)	Females (%)	*P*-value
CRT-item 1	35.77	29.20	0.098
CRT-item 2	25.77	10.47	0.000
CRT-item 3	34.23	18.73	0.000
0 correct answers	43.46	61.43	
1 correct answer	28.85	23.97	
2 correct answers	16.15	9.37	
3 correct answers	11.54	5.23	

P-values from two-sided Fisher's exact tests for the difference in proportions.

men than for women (right hand: t_{621} = 4.4661, $P < 0.001$; left hand: t_{621} = 3.8079, $P < 0.001$).

Fig. 1 reports the histogram and kernel density estimation of 2D:4D in our sample. The results are displayed separately for males and females and for the left hand (panel *a*) and right hand (panel *b*).

The results of the CRT appear in Table 2. The upper part of the table reports, for each question, the percentage of males and females who answered it correctly and the significance level of the difference between sexes (two-sided Fisher's exact test). Men were significantly more likely than women to answer correctly each of the three questions (although for question 1 the difference is only marginally significant). The mean (±SEM) number of correct responses in the CRT was 0.958 ± 0.064 for males and 0.584 ± 0.045 for females (Cohen's *d* = 0.3941).

The bottom part of the table reports the distribution of the number of correct answers for males and females: 27.69% of males had two or three correct answers in the CRT, while this percentage shrinks to 14.60% for females, and 11.54% of males and 5.23% of females answered correctly all the three CRT questions. A notable fraction of the subject pool (43.46% of males and 61.43% of females) was unable to solve any of the referred questions.

The relationship between the subjects' performance in the CRT and their 2D:4D is shown in Fig. 2. Smoothed curves were fit using locally weighted regressions (LOWESS smoothing) with a standard, conservative bandwidth of 0.8. For both sexes, we observe a *negative relationship* between the number of correct answers in the CRT and both the left-hand (panel *a*) and the right-hand (panel *b*) 2D:4D. In addition, the effect of 2D:4D on the number of correct answers in the CRT appears to be stronger for females than for males.

Column (1) of Table 3 presents estimates of an ordered probit regression for the effects of 2D:4D and sex on the number of correct answers to the CRT (left panels refer to the left hand and right panels to the right hand). Zero-order correlations between all the variables used are reported (uncorrected for multiple comparisons), separately for males and females, in Table A.1 in Appendix.

A lower 2D:4D is significantly associated with a higher number of correct answers (left hand: *P* = 0.028; right hand: *P* = 0.001), and males had significantly more correct answers than females ($P < 0.001$). Interaction effects are shown in column (2). There is a marginally significant interaction between right-hand 2D:4D and sex (*P* = 0.072), indicating that the negative impact of 2D:4D on CRT is more pronounced for females. Wald tests on the coefficients of that model indicate that the effect is significant for females (Chi^2 = 12.82, $P < 0.001$) but not for males (Chi^2 = 0.77, $P > 0.3$). No significant interaction effect is found for the left-hand 2D:4D ($P > 0.2$), although the sign of the interaction term is the same as for the right hand (i.e. more pronounced effect for females). To put these results into perspective, note that the mean number of correct answers among females in the bottom quartile of 2D:4D is 108% and 75%, respectively for right and left hands, higher than among females in the top quartile (mean ± SEM number of correct answers top vs. bottom, right hand: 0.422 ± 0.084 vs. 0.878 ± 0.112; left hand: 0.444 ± 0.078 vs. 0.778 ± 0.108; *n* = 90 in both groups). For males, these differences are less striking (right hand: 0.892 ± 0.120 vs. 1.015 ± 0.136; left hand: 0.969 ± 0.133 vs. 1.015 ± 0.131; *n* = 65 in both groups).

As mentioned, the negative impact of 2D:4D on CRT is more pronounced for females than for males. Frederick (2005) observes that CRT scores are more highly correlated

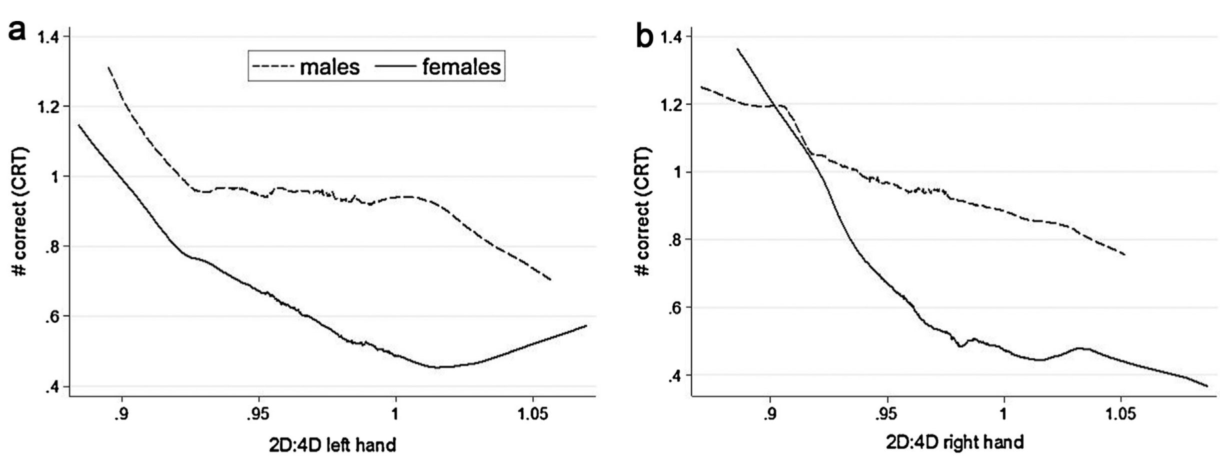

Figure 2 LOWESS smoothing: cognitive reflection as a function of 2D:4D. Figure shows cognitive reflection as a function of 2D:4D. Smoothed curves were fit using locally weighted regressions (LOWESS smoothing) with a standard, conservative bandwidth of 0.8. For both sexes, we observe a *negative relationship* between the number of correct answers in the CRT and both the left-hand (panel *a*) and the right-hand (panel *b*) 2D:4D. In addition, the effect of 2D:4D on the number of correct answers in the CRT appears to be stronger for females than for males.

Table 3 The impact of 2D:4D on CRT.

	(a) Left hand				(b) Right hand			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
2D:4D	-3.225^{**}	-1.550	-3.829^{***}	-1.869	-4.572^{***}	-1.827	-4.977^{***}	-2.431
	(1.465)	(2.174)	(1.483)	(2.177)	(1.410)	(2.076)	(1.424)	(2.082)
Female	-0.424^{***}	2.543	-0.336^{***}	3.186	-0.407^{***}	4.499^{*}	-0.321^{***}	4.276
	(0.094)	(2.850)	(0.096)	(2.870)	(0.094)	(2.733)	(0.096)	(2.751)
2D:4D $\times$ female		-3.062		-3.635		-5.090^{*}		-4.771^{*}
		(2.940)		(2.961)		(2.834)		(2.854)
Math			0.265^{***}	0.268^{***}			0.265^{***}	0.265^{***}
			(0.061)	(0.061)			(0.061)	(0.061)
Impatience			-0.041^{**}	-0.041^{**}			-0.041^{**}	-0.039^{**}
			(0.019)	(0.019)			(0.019)	(0.019)
Log likelihood	-695.863	-695.321	-683.461	-682.707	-692.993	-691.377	-680.641	-679.241
Chi^2	28.57^{***}	29.65^{***}	53.37^{***}	54.88^{***}	34.31^{***}	37.54^{***}	59.01^{***}	61.81^{***}
Pseudo R^2	0.0201	0.0209	0.0376	0.0386	0.0242	0.0264	0.0415	0.0435
N	623	623	623	623	623	623	623	623

Note: Ordered probit estimates. Columns on the left refer to left hand (a) while columns on the right focus on the right hand (b). In all regressions, the dependent variable is the CRT score (four categories, from 0 to 3 correct answers). In column (1), the explanatory variables are 2D:4D and sex, while their interaction is added in column (2). Columns (3) and (4) repeat the same regressions, respectively, controlling for math ability and impatience. Standard errors in brackets.

* Significance at the 0.1 level.

** Significance at the 0.05 level.

*** Significance at the 0.01 level.

with time preferences for women than for men. This may suggest that some of the effect of 2D:4D on the CRT is due to time preference or impatience. After all, according to a dual-process approach, answering correctly the CRT appears to require that the deliberative/patient mind overrules the intuitive/impatient response. Similarly one could posit that some of the effect of 2D:4D on the CRT may signal mathematical ability, since the CRT questions, although simple, have an algebraic content. To disentangle whether the effect of 2D:4D on CRT is in fact capturing the impact of *mathematical ability* or a *degree of impatience*, we extend our analysis to account for these two factors.

We now estimate the effects of 2D:4D and sex, as before, but controlling for the effect of math proficiency, as measured by the number of correct answers to the mathematical test, and for the effect of impatience, as measured by the number of impatient answers in the time preference task. The results appear in columns (3) and (4) of Table 3 (for both the left and the right hands).

As in Frederick (2005), we find that impatience is negatively and significantly related to performance in the CRT ($Ps < 0.05$). As expected, mathematical ability is a positive and strong determinant of CRT scores ($Ps < 0.01$). Yet, there is an interesting insight obtained from these regressions: The effect of 2D:4D on CRT *is not reduced* (it even increases slightly; right hand: $P < 0.001$, left hand: $P = 0.010$; column (3)) when controlling for the performance in the math and impatience tests. This implies that these variables are not mediating the relationship between 2D:4D and CRT. It appears, therefore, that the effect of 2D:4D captures a component of the determinants of the subjects' performance in the CRT that is *different* from the effect of sex, performance in a simple mathematical test, and impatience. Notice here that it could be argued for instance that being more reflective, as measured by the CRT, leads to less impatient behavior in the time preferences task, rather than the opposite causal way. To alleviate this concern, we performed partial correlations between CRT scores and each of the explanatory variables, while keeping the other variables constant: the significance levels remain nearly identical to those reported in Table 3 (available upon request from the authors). And, clearly, the causality of the main relationship (that is, prenatal hormone exposure impacts on CRT scores) cannot be reversed.

4. Discussion

The results presented above indicate that prenatal hormone exposure, expressed in its putative marker 2D:4D, has a significant and positive effect on how females and, to a more ambiguous degree, males answer the CRT. Moreover, such effect is not mediated by impatience and math proficiency. In plain words, we observe an association between 2D:4D and CRT scores, which suggests a relation between relative higher levels of prenatal testosterone and attention, concentration, diligence or whatever traits that, beyond competence in algebra and impatience, facilitate overriding the intuitive but incorrect responses to the test. In this regard, the attention to detail observed in autism (in which 2D:4D is particularly low; Manning et al., 2001) has been related to low 2D:4D in typically developing samples, sometimes in a sex-dependent manner (Baron-Cohen et al., 2005). Further research should try to test whether other factors, like enhanced persistence in an effort, or increased ability not to be distracted by irrelevant information, or higher "need for achievement" (Millet, 2009), may mediate the effect of prenatal sex hormones on CRT.

Based on an observed negative correlation between financial traders' 2D:4Ds and their long-term success in a high-frequency market, Coates et al. (2009) suggested that prenatal androgen exposure increases risk-preferences and promotes more rapid visuomotor scanning and physical reflexes. Considering our results, it can be suggested that long-term success under the high-volatility conditions of the financial markets might also require a high level of reflective cognition in order to rapidly process new information *in an analytical manner*, therefore overriding automatic/intuitive maladaptive responses. Interestingly, low 2D:4D has been associated with increased risk-taking in a number of studies (see e.g. Brañas-Garza and Rustichini, 2011; Garbarino et al., 2011). If one considers risk-taking as an impulsive/maladaptive behavior, those findings might seem to contradict ours. However, the Coates et al.'s result provides a nice example of risk-taking representing a long-term profitable behavior, far from impulsive. The studies referred above show that low-2D:4D individuals are less prone to avoid risks in situations where the *optimal* strategy is, precisely, taking more risk: In other words, risks are taken in situations where the expected value of the high-risk option exceeds that of the low-risk one (see Frederick, 2005 for a discussion on how this may relate to cognitive reflection).

In our large sample of first-year college students some do think through the intuitive answer while others do not. 2D:4D can help to predict who will and who will not, especially among women. Our results show that women with a lower prenatal testosterone/estrogens ratio do poorly compared with women with a higher relative prenatal exposure to testosterone. A differential impact of 2D:4D between sexes has often been reported in the literature: on visual-spatial abilities (Poulin et al., 2004; Bull and Benson, 2006); on musical abilities (Sluming and Manning, 2000); on numerical ability/literacy (Brookes et al., 2007; Brosnan, 2008); on sensation seeking (Austin et al., 2002; but see Voracek et al., 2010).

Since male fetuses have higher testosterone/estrogens ratios, the lower size effect of 2D:4D for males compared to females could perhaps be an indication of the existence of ceiling effects or non-linearities on the influence exerted by prenatal androgen exposure (see e.g. Fink et al., 2006; Hampson et al., 2008; Valla and Ceci, 2011). Or that males' and females' prenatal brain organization processes are affected differently by the same prenatal hormones (Valla and Ceci, 2011). A number of papers observe this differential effect (e.g. Finegan et al., 1992; Romano et al., 2006; Valla et al., 2010), and sex-dependent effects are indeed gaining traction in the literature on neural organization (see e.g. Kempel et al., 2005; Lenroot and Giedd, 2010; Elton et al., 2013).

It appears, then, that early androgen surges exert an organizational influence on brain development, indicating that prenatal testosterone in humans may act as a programming mechanism that influences behavior later in life (see e.g. Lombardo et al., 2012). Admittedly, trying to pin down differences in the CRT answers to one single factor, prenatal testosterone/estrogens ratio, is simplistic and might eventually lead to conflicting, erratic or inconclusive results (indeed, from the pseudo-R^2 values reported in Table 3, it can be observed that much of the variation remains unexplained in our regressions). While 2D:4D is a fixed and predetermined variable, other processes influencing behavior may have occurred or may even be occurring while subjects take the test. Coates (2012) conjectures a "preparation for the test effect" and a "winner effect" (that in our test may result from the satisfaction of answering correctly the first question in the CRT) resulting in a variation in circulating hormones that may distort the predictive power of the 2D:4D biometric measurements.

Finally, it is important to note that in our sample 2D:4D does not correlate significantly with the number of correct answers in the math test ($Ps > 0.2$; see Table A.1), except in the case of females' left hand ($P = 0.034$). That the latter relationship is positive may explain why the negative impact of 2D:4D on the CRT score is even stronger when controlling in the regressions of Table 3 for the number of correct answers in the math test. It could be argued that the different procedure used (the math test was embedded in a long questionnaire while the CRT was presented as a separate task), or the simplicity of the math test may have influenced the results. Indeed, it has been hypothesized that higher prenatal exposure to testosterone might predict a higher "need for achievement" (Millet, 2009), which could be more prominent in more self-motivating, complicated or salient tasks.

All in all, the robust effect of both hands' 2D:4D ratios on subjects' answers to the CRT, which is not mediated by their answers to the impatience or basic math tests, should encourage further controlled experiments to pin down why individuals exposed to a larger than average relative amount of testosterone in utero offer better, more reasoned, solutions in the CRT twenty years after the fact.

Role of the funding source

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of interest statement

None declared.

Acknowledgements

This paper has been benefited by comments and suggestions of Mark Coulson, Tom Dickins, Roberto Hernán, Praveen Kujal, Kobe Millet and Martin Voracek. The authors thank Levent Neyse for his help in data collection. The first author acknowledges financial aid from Spanish Ministerio de Educación y Ciencia (ECO2011-2529) and from the Spanish Ministerio de Economía through the Severo Ochoa Program for Centers of Excellence in R&D (SEV-2011-0075). The second and the third authors acknowledge financial support from Ministerio Educación y Ciencia (ECO2010-17049), Fundación Ramón Areces R + D 2011 and Junta de Andalucia-Excelencia (P07.SEJ.02547).

Appendix

The questions in the tests were asked in Spanish. We provide the Frederick's (2005) original CRT questions and an English translation of the math test.

CRT questions

Spanish:

1. *Un bate y una pelota cuestan 1,10 euros en total. El bate cuesta 1 euro más que la pelota, ¿cuántos céntimos cuesta la pelota?*
2. *Se necesitan 5 máquinas durante 5 minutos para hacer 5 objetos, ¿cuántos minutos tardarían 100 máquinas en hacer 100 objetos?*
3. *En un lago hay un conjunto de nenúfares. Cada día, el conjunto se duplica. Si se tardan 48 días en que el conjunto de nenúfares cubra el lago entero, ¿cuántos días tarda el conjunto de nenúfares en cubrir la mitad del lago?*

English (Frederick, 2005):

1. *A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? _____ cents*
2. *If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes*
3. *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? _____ days*

Math questions

Spanish:

1. *Si la probabilidad de contraer una enfermedad es de un 10 por ciento, ¿cuántas personas de 1.000 contraerían la enfermedad?*
2. *Si 5 personas tienen el número premiado de la lotería y el premio a repartir es de dos millones de euros, ¿cuánto recibiría cada una?*
3. *Supongamos que tienes 100€ en una cuenta de ahorro, y la tasa de interés que ganas por estos ahorros es de 2% por año. Si mantienes el dinero por 5 años en la cuenta, ¿cuánto tendrá al término de estos 5 años?:*
   a. *Más de 102€*
   b. *Exactamente 102€*
   c. *Menos de 102€*
   d. *NS/NR*
4. *Digamos que tienes 100€ ahorrados en una cuenta de ahorro. La cuenta acumula un 10% de interés por año. ¿Cuánto tendrás en la cuenta al cabo de dos años?*

English:

1. *If the probability of being infected by an illness is 10%, how many persons of a group of 1000 would be infected by that kind of illness?*
2. *If there are 5 persons that own the winning lottery ticket and the prize to be shared is two million euros, how much money would each person receive?*
3. *Suppose that you have 100€ in a savings account and the rate of interest that you earn from the savings is 2% per year. If you keep the money in the account for 5 years, how much money would you have at the end of these 5 years?:*
   a. *More than 102€*
   b. *102€ exactly*
   c. *Less than 102€*
   d. *S/he cannot/do not want to answer*
4. *Suppose that you have 100€ in a savings account. The account accumulates a 10% rate of interest per year. How much money would you have in your account after two years?*

Table A.1 Pairwise correlations between variables (by sex).

	CRT	CRT-1	CRT-2	CRT-3	2D:4D right	2D:4D left	Impatience
Males							
CRT-item 1	0.7101***						
CRT-item 2	0.7346***	0.2575***					
CRT-item 3	0.7712***	0.2903***	0.4090***				
2D:4D right	−0.0630	0.0215	−0.0700	−0.0936			
2D:4D left	−0.0502	−0.0003	−0.0533	−0.0593	0.6580***		
Impatience	−0.0201	0.0101	−0.0178	−0.0374	−0.0161	0.0249	
Math	0.1258**	0.0405	0.0665	0.1702***	0.0793	0.0530	−0.0743
Females							
CRT-item 1	0.7802***						
CRT-item 2	0.6759***	0.2752***					
CRT-item 3	0.7716***	0.3438***	0.3893***				
2D:4D right	−0.1834***	−0.1789***	−0.1602***	−0.0713			
2D:4D left	−0.1322**	−0.0825	−0.1683***	−0.0641	0.7088***		
Impatience	−0.1630***	−0.1547***	−0.1035**	−0.0990*	0.0768	0.0253	
Math	0.1772***	0.2190***	0.1179**	0.0441	0.0431	0.1114**	−0.0283

Note: Pearson correlations.
* Significance at the 0.1 level.
** Significance at the 0.05 level.
*** Significance at the 0.01 level.

References

Allaway, H.C., Bloski, T.G., Pierson, R.A., 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.

Alter, A.L., Oppenheimer, D.M., Epley, N., Eyre, R.N., 2007. Overcoming intuition: metacognitive difficulty activates analytic reasoning. J. Exp. Psychol. Gen. 136, 569—576.

Auger, J., Le Denmat, D., Berges, R., Doridot, L., Salmon, B., Canivenc-Lavier, M.C., Eustache, F., 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. 280, 20131532.

Austin, E.J., Manning, J.T., McInroy, K., Mathews, E., 2002. A preliminary investigation of the associations between personality, cognitive ability and digit ratio. Pers. Indiv. Differ. 33, 1115—1124.

Bailey, A.A., Hurd, P.L., 2005. Finger length ratio (2D:4D) correlates with physical aggression in men but not in women. Biol. Psychol. 68, 215—222.

Baron-Cohen, S., Knickmeyer, R., Belmonte, M.K., 2005. Sex differences in the brain: implications for explaining autism. Science 310, 819—823.

Baron-Cohen, S., Lutchmaya, S., Knickmeyer, R., 2004. Prenatal Testosterone in Mind: Studies of Amniotic Fluid. MIT Press/Bradford Books, Cambridge, MA.

Berenbaum, S.A., Bryk, K.K., Nowak, N., Quigley, C.A., Moffat, S., 2009. Fingers as a marker of prenatal androgen exposure. Endocrinology 150, 5119—5124.

Bernheim, B.D., Rangel, A., 2004. Addiction and cue-triggered decision processes. Am. Econ. Rev. 94, 1558—1590.

Brañas-Garza, P., Rustichini, A., 2011. Organizing effects of testosterone and economic behavior: not just risk taking. PLoS ONE 6, e29842.

Brañas-Garza, P., García-Muñoz, T., Hernán-González, R., 2012. Cognitive effort in the Beauty contest game. J. Econ. Behav. Organ. 83, 254—260.

Brocas, I., Carrillo, J.D., 2008. The brain as a hierarchical organization. Am. Econ. Rev. 98, 1312—1346.

Brookes, H., Neave, N., Hamilton, C., Fink, B., 2007. Digit ratio (2D:4D) and lateralization for basic numerical quantification. J. Ind. Diff. 28, 55—63.

Brosnan, M.J., 2008. Digit ratio as an indicator of numeracy relative to literacy in 7-year-old British school children. Br. J. Psychol. 99, 75—85.

Brosnan, M., Gallop, V., Iftikhar, N., Keogh, E., 2011. Digit ratio (2D:4D), academic performance in computer science and computer-related anxiety. Pers. Indiv. Differ. 51, 371—375.

Bull, R., Benson, P.J., 2006. Digit ratio (2D:4D) and the spatial representation of magnitude. Horm. Behav. 50, 194—199.
Camerer, C., Loewenstein, G., Prelec, D., 2005. Neuroeconomics: how neuroscience can inform economics. J. Econ. Lit. 43, 9—64.
Campitelli, G., Labollita, M., 2010. Correlations of cognitive reflection with judgments and choices. Judgm. Decis. Mak. 5, 182—191.
Coates, J., 2012. The Hour Between Dog and Wolf. Risk Taking, Gut Feelings, and the Biology of Boom and Bust. Penguin Press USA, New York.
Coates, J.M., Gurnell, M., Rustichini, A., 2009. Second- and fourth digit ratio predicts success among higher frequency financial traders. Proc. Natl. Acad. Sci. U.S.A. 106, 623—628.
Cokely, E.T., Kelley, C.M., 2009. Cognitive abilities and superior decision making under risk: a protocol analysis and process model evaluation. Judgm. Decis. Mak. 4, 20—33.
Coller, M., Williams, M.B., 1999. Eliciting individual discount rates. Exp. Econ. 2, 107—127.
Coyne, S.M., Manning, J.T., Ringer, L., Bailey, L., 2007. Directional asymmetry in digit ratio (2D:4D) predict indirect aggression in women. Pers. Indiv. Differ. 43, 865—872.
Elton, A., Tripathi, S.P., Mletzko, T., Young, J., Cisler, J.M., James, G.A., Kilts, C.D., 2013. Childhood maltreatment is associated with a sex-dependent functional reorganization of a brain inhibitory control network. Hum. Brain Mapp., http://dx.doi.org/10.1002/hbm.22280.
Espín, A.M., Brañas-Garza, P., Herrmann, B., Gamella, J.F., 2012. Patient and impatient punishers of free-riders. Proc. R. Soc. Lond. B. Biol. 279, 4923—4928.
Exadaktylos, F., Espín, A.M., Brañas-Garza, P., 2013. Experimental subjects are not different. Sci. Rep. 3, 1213, http://dx.doi.org/10.1038/srep01213.
Finegan, J., Niccols, G.A., Sitarenios, G., 1992. Relations between prenatal testosterone levels and cognitive abilities at 4 years. Dev. Psychol. 28, 1075—1089.
Fink, B., Brookes, H., Neave, N., Manning, J.T., Geary, D.C., 2006. Second to fourth digit ratio and numerical competence in children. Brain Cogn. 61, 211—218.
Frederick, S., 2005. Cognitive reflection and decision making. J. Econ. Perspect. 19, 25—42.
Fudenberg, D., Levine, D.K., 2006. A dual-self model of impulse control. Am. Econ. Rev. 96, 1449—1476.
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—26.
Geschwind, N., Galaburda, A.M., 1985. Cerebral lateralization—biological mechanisms, associations and pathology: a hypothesis and a program for research. Arch. Neurol. 42, 428—459.
Grimshaw, G.M., 1995. Relations between prenatal testosterone and cerebral lateralization in children. Neuropsychology 9, 74—75.
Hanoch, Y., Gummerum, M., Rolison, J., 2012. Second-to-fourth digit ratio and Impulsivity: a comparison between offenders and non-offenders. PLoS ONE 7 (10) e47140.
Hampson, E., Ellis, C.L., Tenk, C.M., 2008. On the relation between 2D:4D and sex-dimorphic personality traits. Arch. Sex. Behav. 37, 133—144.
Hampson, E., Sankara, J.S., 2013. Re-examining the Manning hypothesis: androgen receptor polymorphism and the 2D:4D digit ratio. Evol. Hum. Behav. 33, 557—561.
Harrison, G.W., Lau, M.I., Williams, M.B., 2002. Estimating individual discount rates in Denmark: a field experiment. Am. Econ. Rev. 92, 1606—1617.
Hönekopp, J., 2013. No evidence that 2D:4D is related to the number of CAG repeats in the androgen receptor gene. Front. Endocrinol. 4, 185.
Johnson, M.W., Bickel, W.K., 2002. Within-subject comparison of real and hypothetical money rewards in delay discounting. J. Exp. Anal. Behav. 77, 129—146.
Kahneman, D., Frederick, S., 2002. Representativeness revisited: attribute substitution in intuitive judgment. In: Gilovich, T., Griffin, D., Kahneman, D. (Eds.), Heuristics and Biases: The Psychology of Intuitive Judgment. Cambridge University Press, New York, pp. 49—81.
Kempel, P., Gohlke, B., Klempau, J., Zinsberger, P., Reuter, M., Hennig, J., 2005. Second-to-fourth digit length, testosterone and spatial ability. Intelligence 33, 215—230.
Kemper, C.J., Schwerdtfeger, A., 2009. Comparing indirect methods of digit ratio (2D:4D) measurement. Am. J. Hum. Biol. 21, 188—191.
Koehler, D.J., James, G., 2010. Probability matching and strategy availability. Mem. Cogn 38, 667—676.
Lagorio, C.H., Madden, G.J., 2005. Delay discounting of real and hypothetical rewards III: steady-state assessments, forced-choice trials, and all real rewards. Behav. Process. 69, 173—187.
Lenroot, R.K., Giedd, J.N., 2010. Sex differences in the adolescent brain. Brain Cogn. 72, 46—55.
Lombardo, M.V., Ashwin, E., Auyeung, B., Chakrabarti, B., Lai, M.C., Taylor, K., Hackett, G., Bullmore, E.T., Baron-Cohen, S., 2012. Fetal programming effects of testosterone on the reward system and behavioral approach tendencies in humans. Biol. Psychiat. 72, 839—847.
Madden, G.J., Raiff, B.R., Lagorio, C.H., Begotka, A.M., Mueller, A.M., Hehli, D.J., Wegener, A.A., 2004. Delay discounting of potentially real and hypothetical rewards II: between- and within-subject comparisons. Exp. Clin. Psychopharm. 12, 251—261.
Manning, J.T., 2002. Digit Ratio: A Pointer to Fertility, Behavior, and Health. Rutgers University Press, New Jersey.
Manning, J.T., Baron-Cohen, S., Wheelwright, S., Sanders, G., 2001. The 2nd to 4th digit ratio and autism. Dev. Med. Child Neurol. 43, 160—164.
Manning, J.T., Churchill, A.J.G., Peters, M., 2007. The effects of sex, ethnicity, and sexual orientation on self-measured digit ratio (2D:4D). Arch. Sex. Behav. 36, 223—233.
Manning, J.T., Kilduff, L.P., Trivers, R., 2013. Digit ratio (2D:4D) in Klinefelter's syndrome. Andrology 1, 94—99.
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 levels of testosterone, LH and oestrogen. Hum. Reprod. 13, 3000—3004.
McIntyre, M.H., Ellison, P.T., Lieberman, D.E., Demerath, E., Towne, B., 2005. The development of sex differences in digital formula from infancy in the Fels Longitudinal Study. Proc. R. Soc. Lond. B. Biol. 272, 1473—1479.
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.
Oechssler, J., Roider, A., Schmitz, P.W., 2009. Cognitive abilities and behavioral biases. J. Econ. Behav. Organ. 72, 147—152.
Paxton, J.M., Ungar, L., Greene, J.D., 2012. Reflection and reasoning in moral judgment. Cogn. Sci. 36, 163—177.
Pennycook, G., Cheyne, J.A., Seli, P., Koehler, D.J., Fugelsang, J.A., 2012. Analytic cognitive style predicts religious and paranormal belief. Cognition 123, 335—346.
Phelps, V.R., 1952. Relative index finger length as a sex-influenced trait in man. Am. J. Hum. Genet. 4, 72—89.
Poulin, M., O'Connell, R.L., Freeman, L.M., 2004. Picture recall skills correlate with 2D:4D ratio in women but not men. Evol. Hum. Behav. 25, 174—181.
Romano, M., Leoni, B., Saino, N., 2006. Examination marks of male university students positively correlate with finger length ratios (2D/4D). Biol. Psychol. 7, 175—182.
Sapienza, P., Zingales, L., 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.

Shenhav, A., Rand, D.G., Greene, J.D., 2012. Divine intuition: cognitive style influences belief in God. J. Exp. Psychol. Gen. 141, 423—428.

Sluming, V.A., Manning, J.T., 2000. Second to fourth digit ratio in elite musicians: evidence for musical ability as an honest signal of male fitness. Evol. Hum. Behav. 21, 1—9.

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. Pers. Indiv. Differ. 51, 412—416.

Toplak, M.E., West, R.F., Stanovich, K.E., 2011. The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. Mem. Cogn. 39, 1275—1289.

Trivers, R., Manning, J., Jacobson, A., 2006. A longitudinal study of digit ratio (2D:4D) and other finger ratios in Jamaican children. Horm. Behav. 49 (2) 150—156.

Valla, J.M., Ganzel, B.L., Yoder, K.J., Chen, G.M., Lyman, L.T., Sidari, A.P., 2010. More than maths and mindreading: sex differences in empathising/systemising covariance. Autism Res. 3, 174—184.

Valla, J., Ceci, S.J., 2011. Can sex differences in science be tied to the long reach of prenatal hormones?. Brain organization theory, digit ratio (2D/4D), and sex differences in preferences and cognition. Perspect. Psychol. Sci. 6, 134—136.

Voracek, M., Tran, U.S., Dressler, S.G., 2010. Digit ratio (2D:4D) and sensation seeking: new data and meta-analysis. Pers. Indiv. Differ. 48, 72—77.

Wallen, K., 2009. Does finger fat produce sex differences in second to fourth digit ratios? Endocrinology 150, 4819—4822.

Williams, J.H.G., Greenhalgh, K.D., Manning, J.T., 2003. Second to fourth finger ratio and developmental psychopathology in pre-school children. Early Hum. Dev. 72, 57—65.

Zheng, Z., Cohn, M.J., 2011. Developmental basis of sexually dimorphic digit ratios. Proc. Natl. Acad. Sci. U.S.A. 108, 16289—16294.