serie ad

WP-AD 2015-02

Cognitive (Ir)reflection: New Experimental Evidence

Carlos Cueva, Iñigo Iturbe-Ormaetxe, Esther Mata-Pérez, Giovanni Ponti, Marcello Sartarelli, Haihan Yu and Vita Zhukova

Ivie

Working papers

Working papers

Working papers

Versión: febrero 2015 / Version: February 2015

Edita / Published by:
Instituto Valenciano de Investigaciones Económicas, S.A.
C/ Guardia Civil, 22 esc. 2 1º - 46020 Valencia (Spain)

WP-AD 2015-02

Cognitive (Ir)reflection: New Experimental Evidence*

Carlos Cueva, Iñigo Iturbe-Ormaetxe, Esther Mata-Pérez, Giovanni Ponti, Marcello Sartarelli, Haihan Yu and Vita Zhukova**

Abstract

We study whether cognitive ability explains choices in a wide variety of behavioral tasks, including risk and social preferences, by collecting evidence from almost 1,200 subjects across eight experimental projects. Since Frederick (2005)'s Cognitive Reflection Test (CRT) has been administered to all subjects, our dataset is one of the largest in the literature. We divide the subjects pool into three groups depending on their CRT performance. Reflective subjects are those answering at least two of the three CRT questions correctly. Impulsive ones are those who are unable to suppress the instinctive impulse to follow the intuitive although incorrect answer in at least two 2 questions, and the remaining subjects form a residual group. We find that females score significantly worse than males in the CRT, and in their wrong answers impulsive ones are observed more frequently. The 2D-4D ratio, which is higher for females, is correlated negatively with subject's CRT score. In addition, we find that differences between CRT groups in risk aversion depend on the elicitation method used. Finally, impulsive subjects have higher social preferences, while reflective subjects are more likely to satisfy basic consistency conditions in lottery choices.

Keywords: behavioral economics, cognitive reflection, gender, laboratory experiment, personality.

JEL classification numbers: C91, D81, J16.

---

* We thank Enrica Carbone, Xavier Del Pozo, Daniela Di Cagno, Arianna Galliera, Glenn Harrison, Raffaele Miniaci, Ismael Rodriguez-Lara, Iryna Sikora and Josefa Tomás for letting us use data from projects carried out with their direct involvement at the Laboratory of Theoretical and Experimental Economics (LaTEx) of the Universidad de Alicante and the Centro di Economia Sperimentale A Roma Est (CESARE) of LUISS Guido Carli Roma. The usual disclaimers apply. Financial support from the Spanish Ministries of Education and Science and Economics and Competitiveness (SEJ 2007-62656, ECO2011-29230 and ECO2012-34928), MIUR (PRIN 20103S5RN3002), Generalitat Valenciana (Research Projects Gruposo3/086 and PROMETEO/2013/037) and Instituto Valenciano de Investigaciones Económicas (Ivie) is gratefully acknowledged.

** University of Alicante. Corresponding author: G. Ponti (gponti@luiss.it).

1 Introduction

There is a growing literature that studies the link between various aspects of socio-economic behavior, such as risk, time, or social preferences, and proxies of cognitive ability under various formats. These measures vary from school and college students' performance, such as the Grade Point Average (GPA, Kirby *et al.*, 2005), or college entry standardized school test scores, such as GRE or SAT (Dohmen *et al.*, 2010; Chen *et al.*, 2013), up to more customized protocols, from the classical "IQ" test (Borghans *et al.*, 2008b) to the Wonderlic test (Ben-Ner *et al.*, 2004). All these contributions stress the importance of individual heterogeneity with specific reference to cognitive abilities, as a fundamental factor to understand and predict individual and social behavior.

Cognitive ability is also a fundamental component of all theories that advocate a dual and parallel cognitive deliberation process (Evans, 1984; Kahneman, 2011): one ("System 1", or intuitive, heuristic...) fast, automatic, associated with a low cognitive load, the other ("System 2", or controlled, analytic...) more cognitively demanding. The Cognitive Reflection Test (CRT hereafter, (Frederick, 2005)) illustrates the interaction between these two cognitive domains. It is a simple test of a quantitative nature especially designed to elicit the "predominant cognitive system at work", either 1 or 2, in respondents' reasoning:

CRT1. A bat and a ball cost 1.10 dollars. The bat costs 1.00 dollars more than the ball. How much does the ball cost? ___ cents.

CRT2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ___ minutes.

CRT3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ___ days.

The correct answers are in a footnote below.[1] The beauty of the test is that, to each question, is associated an immediate impulsive answer (10, 100 and 24, respectively) that, although incorrect, may be selected by those subjects who do

[1] The correct answers are 5, 5, 47 to questions CRT1 to CRT3, respectively.

not think carefully enough. As Frederick (2005) puts it, *"...the three items on the CRT are easy in the sense that their solution is easily understood when explained, yet reaching the correct answer often requires the suppression of an erroneous answer that springs "impulsively" to mind."* (p. 27). In addition, individuals' scores in the CRT test (number of correct answers to the test) are positively related with numerical literacy, mathematical skills, and to psychological dimensions related with impulsiveness (Morsanyi *et al.*, 2014; Toplak *et al.*, 2011).

Frederick (2005) shows that subjects' performance in the CRT significantly correlates with subjects' risk and time preferences, thus suggesting that cognitive ability matters to explain heterogeneity in behavioral tasks. Recent studies also document that the CRT is associated with subjects' gender-specific exposure to testosterone (Bosch-Domènech *et al.*, 2014). In addition, it helps to explain behavioral biases, such as the base rate fallacy (Bergman *et al.*, 2010; Hoppe and Kusterer, 2010; Oechssler *et al.*, 2009). The CRT has also gained attention for the fact that, contrary to most alternative proxies of cognitive abilities, such as the SAT or the Wonderlic Test, females score significantly worse than males. This stylized fact has been established in a wide variety of studies (Frederick, 2005) and has also been confirmed by the evidence reported in this paper.

In the last five years the CRT has been administered to the participants in a number of experimental sessions, both at LaTEx and CESARE, the experimental labs of the Universidad de Alicante and LUISS Guido Carli in Roma, respectively. As a result of this we have been able to assemble a large dataset, with nearly 1,200 observations. To get directly into the discussion around which this paper is built, let us anticipate a first piece of evidence from our dataset.

Figure 1 reports the distribution of CRT answers, collected from 8 different studies carried out at LaTEx and CESARE (see Section 2 for a detailed description). As Figure 1 shows, in none of the cases the modal response is the correct answer. Instead, the mode (10, 100 and 24, respectively) is associated with the "intuitive answer", the"erroneous answer that springs "impulsively" to mind ...". In this respect, our evidence is perfectly in line with what is reported in the literature: for all three questions, the "intuitive" (System-1) response is much more frequent than the "reflective" (System-2) one (Gill and Prowse, 2014).

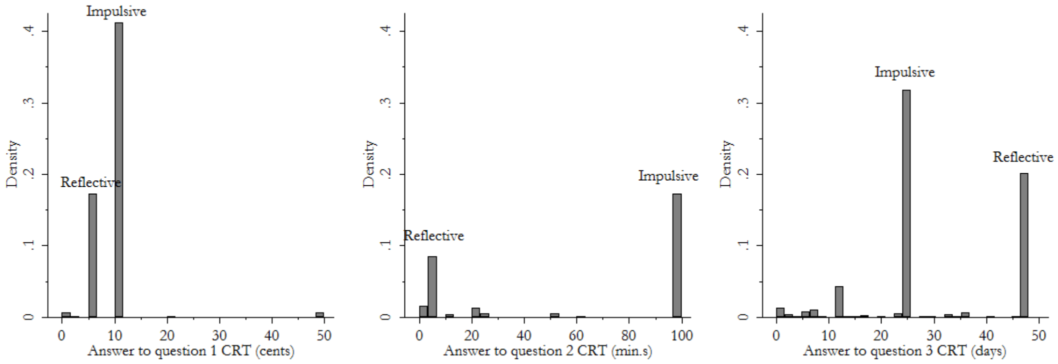


Figure 1: CRT answer distributions

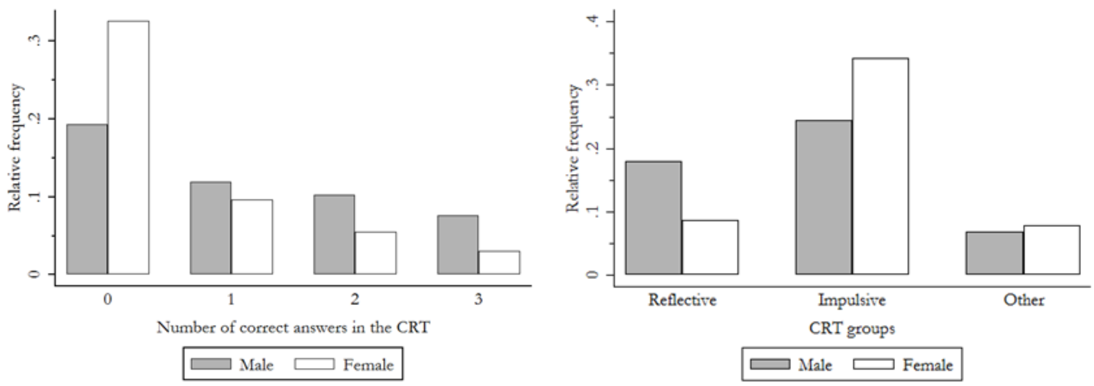
Figure 1 also shows that the response distribution is not completely polarized between these two references: there are also alternatives neither reflective, nor impulsive, that are selected by a non-negligible fraction of individuals. These subjects' answers fall short with respect to the dichotomy "reflective-impulsive" along which the discussion on CRT performance has often focused upon (see, e.g., (Frederick, 2005; Brañas-Garza *et al.*, 2012; Grimm and Mengel, 2012)).

In order to further investigate this issue, this paper puts forward an additional index, labelled as iCRT which is meant to measure cognitive "irreflection" (or "impulsiveness") by means of the three CRT questions:

$$iCRT = 1(CRT1 = 10) + 1(CRT2 = 100) + 1(CRT3 = 24),$$

where 1(.)=1 if condition (.) is satisfied, and 0 otherwise. By analogy with the standard CRT, iCRT is meant to measure the inability to *suppress the erroneous intuitive answer*, which in our view provides as important information as the CRT in characterizing our subject pool. As our previous discussion suggests, we expect women to have, on average, higher iCRT scores, but additional behavioral dimensions need to be explored.

Panel A in Figure 2 reports the distribution of CRT scores disaggregated by gender. The mode is zero for both genders, but the fraction of females who fail the three questions is much higher than the corresponding fraction of males. They have a higher average score in CRT than females (1.12 *vs.* .58, $p < 0.001$), while the opposite holds for iCRT (1.46 *vs.* 1.93, $p < 0.001$). However, there is a also significant fraction of subjects (19% of our subject pool) who scored "low" (i.e., not more than 1 correct answer) in both tests, thus suggesting that cognitive (ir-)reflection does not seem to fully explain their cognitive processing. Figure 2, Panel B, shows the relative frequency of the reflective, impulsive and residual group by gender. We observe substantive gender differences in the reflective and impulsive groups, which are instead hidden if cognitive ability is measured using the CRT score, as Panel A shows.



Panel A Panel B

Figure 2: Panel A: CRT score frequencies by gender. Panel B: CRT groups by gender

The remainder of this paper follows the basic layout of Frederick (2005), in that we provide additional evidence on risk aversion, gender differences, or the relation between CRT and alternative proxies of cognitive ability, around which the original debate on cognitive reflection has been developing over the last 10 years. In addition, we enrich the discussion along less explored dimensions, such as social preferences or a detailed analysis on the determinants of CRT scores and groups.

More specifically, Section 2 provides a brief description of the structure of our dataset, and the associated experimental projects. Section 3 treats the CRT score as the dependent variable and correlates it with individuals' observable characteristics, grouped in Section 3 in three broad categories: *physiological, psychological and socio-demographic*. We find that the large gender difference in CRT performance is significant even after including a number of individual controls.

Sections 4 and 5 make use of our behavioral evidence to look into the link between cognitive reflection, risk and social preferences, respectively. Our evidence shows that the negative correlation between CRT performance and risk aversion crucially depends on the elicitation protocol used, thus confirming the evidence in Andersson *et al.* (2013). We also find that using the CRT performance to split subjects into reflective, impulsive and residual uncovers novel evidence in the literature: impulsive subjects have greater distributional concerns than the other two groups: they are more envious than reflective and more guilty than either of the remaining groups.

In Section 6 we relate CRT performance to alternative measures of rationality and cognitive ability. Here we find that reflective subjects are more likely to satisfy basic conditions for rationality in their lottery choices and have on average higher grades at college (confirming similar findings in (Frederick, 2005)). Finally, Section 7 concludes discussing our main findings, followed by an Appendix containing supplementary statistical evidence.

2 Data and methods

In this section we shall briefly describe the main features and structure of our dataset. We collect data from eight experimental studies carried out at the Laboratory of Theoretical and Experimental Economics (LaTEx) of the Universidad de Alicante and the Center for Experimental Studies At Roma Est (CESARE) of LUISS Guido Carli in Rome, from 2009 to present. The objects of interest in these studies vary broadly and include risk, social preferences, mechanism design and behavioral finance. All experimental protocols also include an extensive debriefing questionnaire regarding socio-demographics, ability and cognitive indicators, psychological and physiological indicators.

2.1 Individual characteristics

Table 1 summarizes the structure of our dataset. The behavioral content of the 8 projects is divided into two broad categories: (IND)ividual and (STR)ategic, depending on the nature of the experimental environment. As we shall report in Sections 4 and 5, this paper is mainly devoted to establish a link between cognitive reflection and individual (as opposed to strategic) behavior, the latter being studied elsewhere, or still in progress (see Section 7 for a "sneak preview" of our preliminary findings).

As we mentioned earlier, we treat observable heterogeneity by grouping individual characteristics into three broad categories: *physiological, psychological* and *socio-demographic*. Our physiological measures are gender and second-to-fourth digit ratio (2D:4D). It has been shown that 2D:4D correlates negatively with prenatal exposure to testosterone (Manning *et al.*, 1998). Due to the simplicity and non-invasiveness of this physiological measure, 2D:4D has been extensively used as a predictor of various individual characteristics, such as risk aversion, competitiveness, prosocial preferences, cognitive ability or career choices (Apicella *et al.*, 2008; Coates *et al.*, 2009; Sapienza *et al.*, 2009; Pearson and Schipper, 2012; Bosch-Domènech *et al.*, 2014).[2]

We use a reduced version of the Big Five personality inventory (Benet-Martinez and John, 1998; John and Srivastava, 1999) as a measure of individual psychological traits. In its various forms, the Big Five questionnaire is among the most relied-upon measures of personality in psychology (see e.g. (Digman, 1990; John *et al.*, 2008)). It measures personality according to five broad dimensions or "traits": *Openness, Conscientiousness, Extraversion, Agreeableness* and *Neuroticism* .[3] The Big Five test has received increasing attention by economists as a useful tool in explaining heterogeneity in individual preferences (Borghans *et al.*, 2009; Daly *et al.*, 2009) and as a predictor of labor market performance and academic achievement (Barrick and Mount, 1991; Judge *et al.*, 1999; Heckman and Rubinstein, 2001; Zhao and Seibert, 2006; Heckman *et al.*, 2006; Borghans *et al.*, 2008a; Heckman and LaFontaine, 2010).

Proj.	Reference	Obs.	IND/ STR	Topic	Quest	2D:4D	BIG5	Risk	Soc. pref.s	Fin. lit.
1	Ponti and Carbone (2009)	48	IND	Herding	Yes	No	Yes	MPL	No	No
2	Di Cagno *et al.* (2014)	192	IND	Risk/soc. preferences	Yes	No	No	N/A	N/A	No
3	Del Pozo *et al.* (2013)	192	IND	Risk/soc. preferences	Yes	No	No	RLP	Yes	No
4	Ponti *et al.* (2014b)	336	STR	Entrepreneurship	Yes	Yes	Yes	MPL	Yes	No
5	Ponti *et al.* (2014a)	192	IND	Risk/Time preferences	Yes	No	No	N/A	N/A	No
6	Ferrara *et al.* (2014)	32	STR	Public good/sleep depr.	Yes	No	No	RLP	Yes	No
7	Albano *et al.* (2014)	92	STR	Procurement auctions	Yes	No	No	No	No	No
8	Cueva *et al.* (2014)	96	STR	Behavioral finance	Yes	Yes	Yes	MPL	No	Yes
Obs.		1,180			1,180	432	480	704	560	96

Table 1: Structure of the meta-dataset

Two socio-demographic measures which turn out to be relevant in our analysis

[2]Figure A1 in the Appendix shows the distribution of 2D:4D in our sample. We also collected information on our subjects' height and weight, and the associated Body Mass Index (BMI). As it turns out, BMI has never been found a significant factor in all the statistical exercises contained in this paper, and therefore, excluded by the set of regressors.

[3]See Table A1 in the Appendix for details.

relate to family education and the number of languages in which the individual is fluent. In particular, *family education* is a dummy variable that takes value 1 if either parent holds a university degree. The dummy variable *languages* takes value 1 if the subject is fluent in more than two languages.[4]

2.2 Behavioral evidence

As we have anticipated, this paper focuses especially on *risk* and *social preferences*, which are two behavioral outcomes common to 5 and 3 studies in our dataset, respectively.

Risk preferences. Subjects' risk attitudes have been elicited either by means of a Multiple Price List (MPL, (Holt and Laury, 2002), Projects 1, 4 and 8) or a Random Lottery Pair (RLP, (Hey and Orme, 1994), Projects 3 and 6) protocol. However, in Sections 4 and 6 we shall not include the evidence from Projects 1 and 6 because the former used hypothetical payoffs and the latter had insufficient observations.

Panel A   Panel B

Figure 3: Panel A: user interface of the MPL (Projects 4 and 8); Panel B: user interface of the RLP (Project 3)

Figure 3 reports the user interface used for MPL (Panel A) and RLP (Panel B), respectively. In Panel A (MPL) there is a sequence of 21 lotteries. The option on the left ("Option A") corresponds to a sure payment whose value increases along the sequence from 0 to 1000 pesetas. The option on the right ("Option B") is constant across the sequence and corresponds to a 50/50 chance to win 1000 pesetas. In Panel B (RLP) the protocol consists of a sequence of 24 binary choices between lotteries

[4]Our study was conducted in a bilingual region. Thus, we wanted to measure whether a subject was fluent in any other language in addition to Spanish and Catalan.

involving four fixed monetary prizes (0, 5, 10 and 15 euros). Lotteries are selected from Hey and Orme (1994)'s original design. For both MPL and RLP, one of the binary choices is selected randomly for payment at the end of the experiment.

Social preferences. As for social preferences, the data analyzed in this paper are taken from Project 4, which replicates Stage 1 in Cabrales *et al.* (2010), a sequence of 24 distributional decisions whose user interface is reported in Figure 4. During 24 rounds individuals face an option set as the one represented in Figure 4. Each individual has to choose one among four options. Then, one of the two individuals is chosen randomly as "dictator" and his/her preferred choice is implemented. This is the "Random Dictator" protocol (Harrison and McDaniel, 2008).

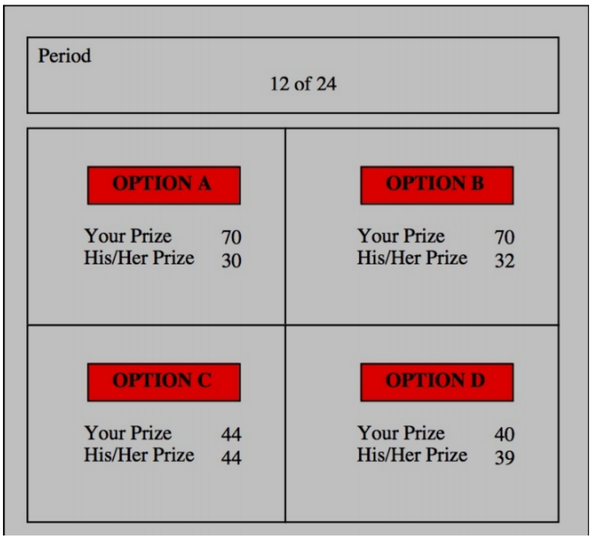


Figure 4: Distributional task, user interface

3 CRT and individual characteristics

In this section we explore the relationship between CRT scores and a set of individual characteristics obtained from the questionnaires administered at the end of each experiment. CRT scores and CRT groups are treated as *dependent variables*, and we study their main predictors among the individual characteristics measured in our subject pool. As our previous discussion suggests, gender effects will be looked at with special care. Table 2 presents mean values of individual characteristics for each CRT group. It also provides p-values from Kruskal-Wallis tests that each individual characteristic across the three CRT groups comes from the same distribution.[5]

As Table 2 shows, subjects belonging to different CRT groups vary significantly with respect to gender, 2D:4D, Neuroticism, Openness and Agreeableness.[6]

	Mean			Kruskal-Wallis	N. obs.
	Reflective	Impulsive	Other	P-value	
Female	0.324	0.583	0.538	<0.001	1178
Left hand 2D:4D	0.970	0.981	0.987	0.015	431
Right hand 2D:4D	0.967	0.978	0.981	0.064	432
Neuroticism	0.435	0.507	0.483	0.009	479
Extraversion	0.582	0.608	0.565	0.175	479
Openness	0.725	0.697	0.655	0.009	479
Agreeableness	0.694	0.685	0.639	0.022	479
Conscientiousness	0.689	0.688	0.671	0.485	479
N. languages > 2	0.440	0.368	0.387	0.462	432
Family educ.	0.311	0.296	0.387	0.377	432

Table 2: Mean values of individuals' characteristics by CRT groups and p-value of the Kruskal-Wallis test

In what follows, we shall investigate the relationship between CRT and individual characteristics in more detail.

3.1 Physiological

We begin by looking at two physiological measures, gender and 2D:4D. As we know from Figure 2, both CRT scores and groups have a strong gender component, with the exception of the residual CRT group. The distributions of both CRT scores and groups are significantly different across gender (Mann-Whitney U test $p < 0.001$ and Chi-square test $p < 0.001$, respectively).

Figure 5 plots mean 2D:4D for each CRT score and group. 2D:4D is lowest for men and women with maximum CRT scores and consequently for those subjects belonging to the reflective group. This relationship seems stronger for males: Kruskal-Wallis tests reject the null hypothesis of no difference in left hand 2D:4D across CRT scores and groups for males ($p = 0.034$ and $p = 0.050$, respectively), but not for females ($p = 0.217$ and $p = 0.668$, respectively). With respect to right hand 2D:4D, Kruskal-Wallis tests cannot reject the null hypothesis of no difference

[5]The Kruskal-Wallis test is a multiple-sample generalization of the Mann-Whitney U-test. Tables A2 and A3 in the Appendix present further mean values and tests disaggregated by gender.

[6]We also consider grades and financial literacy later in the paper (see Section 6).

across CRT scores and groups for both males (p = *0.096* and p =*0.365*, respectively) and females (p = *0.297* and p = *0.494*, respectively).

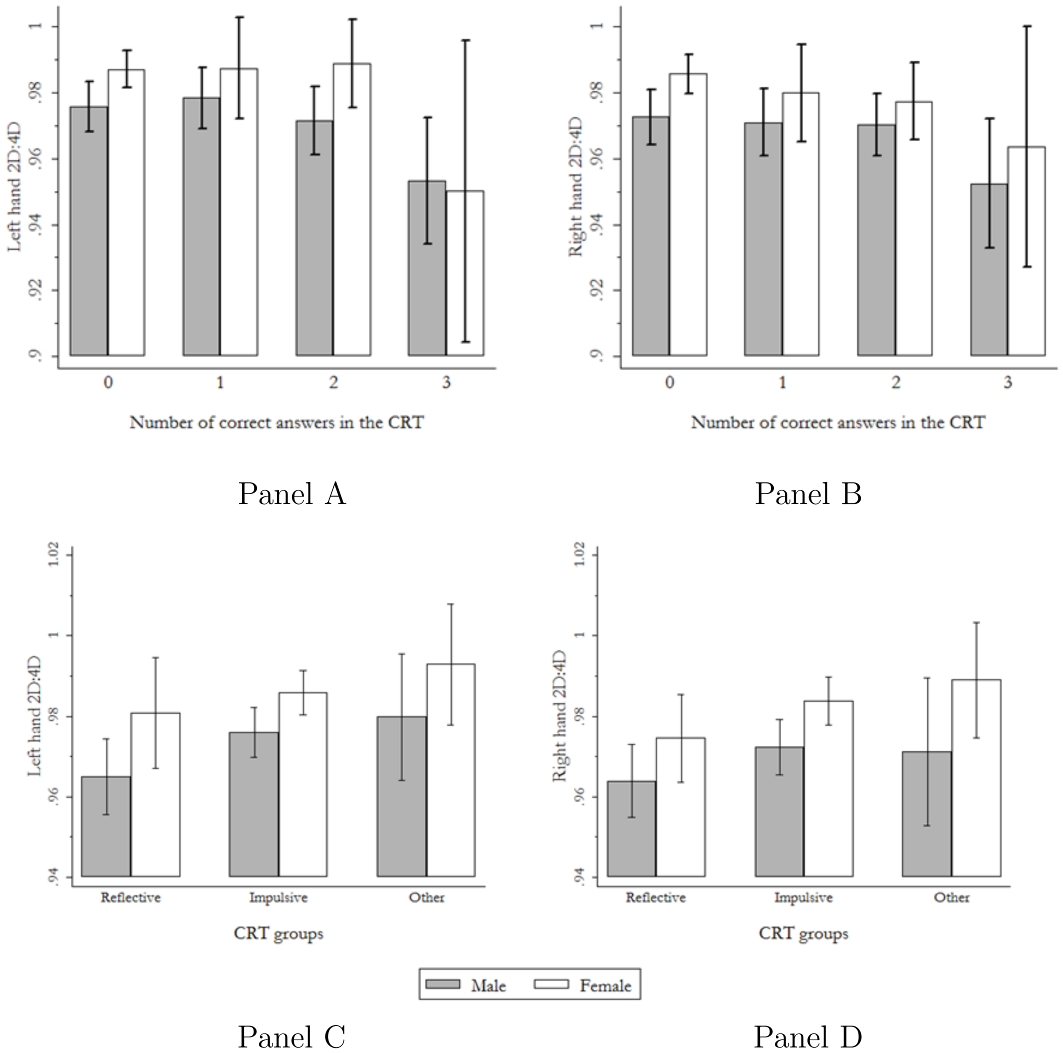


Figure 5: CRT and 2D:4D with 95% confidence intervals. Panel A (B): LH (RH) 2D:4D and CRT. Panel C (D): LH (RH) 2D:4D and CRT groups.

Our finding that males score significantly higher than females in CRT adds further support to growing evidence in the related literature (Frederick, 2005; Oechssler *et al.*, 2009; Brañas-Garza *et al.*, 2012; Bosch-Domènech *et al.*, 2014). Fewer studies have explored the relationship between 2D:4D and cognitive ability. Brañas-Garza and Rustichini (2011) measured performance in the Raven Progressive Matrices task a test of abstract reasoning ability and found results similar to ours: a negative and significant correlation between 2D:4D and Raven test scores for males and no significant correlation for females. More recently, Bosch-Domènech *et al.* (2014) studied the correlation between 2D:4D and CRT scores and found a negative and significant correlation, particularly with the right hand 2D:4D. However, in contrast with our findings, their correlation was stronger for females.

3.2 Psychological

Table 3 presents ordered logit regression results in which Big Five scores are included as predictors of CRT scores. In every regression, Neuroticism and Extraversion are significant (negative) predictors of CRT.[7] As Table 3 shows, our results do not suggest any significant interactions between gender and personality traits.

	(1)	(2)	(3)	(4)	(5)	(6)
		Left Hand			Right Hand	
2D:4D	-0.181	-0.152	-0.191	-0.220**	-0.190*	-0.148
	(0.111)	(0.113)	(0.148)	(0.104)	(0.105)	(0.135)
Female	-1.117***	-1.028***	-0.973***	-1.111***	-1.020***	-0.939***
	(0.205)	(0.209)	(0.312)	(0.206)	(0.210)	(0.315)
Family education	0.0690	0.0397	-0.0568	0.0652	0.0357	-0.0553
	(0.202)	(0.205)	(0.272)	(0.204)	(0.206)	(0.273)
Languages	0.441**	0.439**	0.606**	0.437**	0.434**	0.613**
	(0.201)	(0.204)	(0.271)	(0.201)	(0.205)	(0.272)
Project 8	-0.228	-0.247	-0.275	-0.253	-0.267	-0.296
	(0.220)	(0.230)	(0.242)	(0.223)	(0.232)	(0.244)
Neuroticism		-0.235**	-0.257*		-0.237**	-0.268**
		(0.100)	(0.131)		(0.0998)	(0.131)
Extraversion		-0.198**	-0.262*		-0.198**	-0.261*
		(0.101)	(0.139)		(0.100)	(0.140)
Openness		0.175	0.110		0.172	0.109
		(0.114)	(0.162)		(0.114)	(0.164)
Agreeableness		-0.0287	-0.0443		-0.0340	-0.0593
		(0.114)	(0.127)		(0.114)	(0.128)
Conscientiousness		-0.0682	-0.108		-0.0636	-0.0966
		(0.106)	(0.151)		(0.106)	(0.150)
Female*2D:4D			0.122			-0.101
			(0.234)			(0.227)
Female*Family education			0.206			0.200
			(0.420)			(0.424)
Female*Languages			-0.382			-0.421
			(0.414)			(0.417)
Female*Neuroticism			0.0502			0.0599
			(0.209)			(0.210)
Female* Extraversion			0.183			0.165
			(0.207)			(0.206)
Female*Openness			0.189			0.163
			(0.242)			(0.242)
Female*Agreeableness			0.0125			0.0529
			(0.249)			(0.255)
Female*Conscientiousness			0.135			0.123
			(0.216)			(0.217)
Observations	431	431	431	432	432	432

Table 3: Ordered Logit Model, dependent variable: number of correct answers to CRT. Robust standard errors in parentheses. All explanatory variables except female, languages, family education and project standardized to zero mean and unit standard deviation. *** p <0.01, ** p<0.05, * p<0.1.

As for the related literature, Borghans *et al.* (2008a) examine the impact of per-

[7] Here we use only projects 4 and 8, since these are the only ones in which we have data on the Big Five.

sonality traits on scores in various cognitive tests, including CRT, in a sample of 128 students. Consistently with our evidence, they found that *Extraversion* correlated negatively with the probability of correctly answering the test. In their data, *Openness* correlated positively with CRT, whereas in our regressions, the coefficient on *Openness* is also positive, but not significant. Similarly, *Neuroticism* correlated negatively but not significantly in their data, whereas in our regressions, this negative correlation was significant.

3.3 Socio-demographic

We use two socio-economic indicators in our analysis: whether the subject speaks more than two languages and whether at least one parent holds a university degree. Controlling for other variables, speaking more than two languages turns out to be a highly significant predictor of the number of correct answers to CRT, whereas *family education* is not (see Table 3). Fluency in more than two languages very likely indicates a relatively high socio-economic status in Spain, where the average student is unlikely to be fluent in more than two languages without additional family investment in private education.

3.4 CRT: nature or nurture?

We have used biological, psychological and socio-economic measures to predict performance of our subject pool in CRT. The significant effect of *languages* in all regressions lends support to the idea that nurture matters for performance in the CRT. However, our findings that 2D:4D correlates significantly with CRT, together with those reported in Brañas-Garza and Rustichini (2011) and Bosch-Domènech *et al.* (2014), also lend support to the idea that biological factors affect performance in specific cognitive tests.

Finally, we found certain psychological measures to be correlated with CRT. Even though the relative importance of biological and social determinants of personality is less clear, evidence suggests substantial heritability in *Big Five* scores. For instance, twin studies have estimated that genetic influence can account for around 50% of the variance in *Neuroticism* or *Extraversion* (Loehlin, 1992; Jang *et al.*, 1996; Loehlin *et al.*, 1998).

To give an idea of the relative size of the different determinants of cognitive ability, we compute predicted probabilities of having zero correct answers to CRT according to the estimates in column (5) of Table 3.[8] The probability that males answer zero questions correctly is 0.47, controlling for all other covariates, whereas females have a probability of 0.70. Subjects with right hand 2D:4D one standard deviation below average have a probability of 0.56 of having zero correct answers, whereas those with 2D:4D one standard deviation above average have a probability of 0.60. A score one standard deviation above rather than below average in Neuroticism leads to a 9% difference (0.61 and 0.56, respectively). Similarly, a score one standard deviation above rather than below average in Extraversion leads to a 7% difference (0.60 and 0.56, respectively). Finally, subjects speaking more than two languages are 13% less likely to have zero correct answers to CRT than those who don't (0.53 vs 0.62).

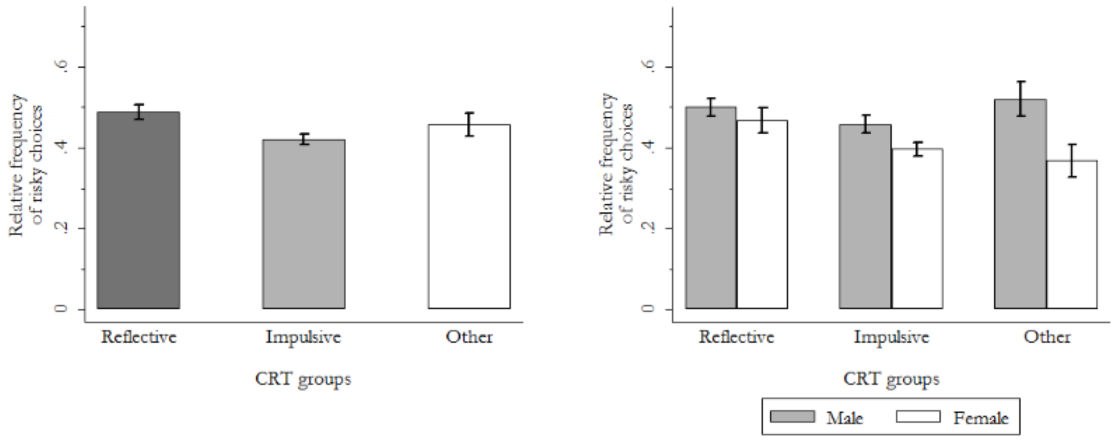
In sum, our results highlight the large gender difference in performance in CRT that remains after controlling for other individual variables: females were almost 50% more likely than males to answer all CRT questions wrong. Variations in personality scores or in the digit ratio of two standard deviations led to much more moderate changes in the predicted probability of giving zero correct answers in CRT (7-9%). Finally, our evidence suggests that educational investment (as proxied by the number of languages spoken) could play a more important role than the psychological and physiological characteristics considered here.

4 CRT and risk preferences

We shall now turn our attention to the behavioral evidence, starting with the analysis on how cognitive reflection relates with risk attitudes. As we already discussed in Section 2, we rely on two different data formats: RLP (Project 3) and MPL (Projects 1, 4 and 8). We look at the RLP data first, which correspond to a sequence of 24 binary choices between lotteries built as probability distributions over the constant prize set. Contrary to MPL, in RLP lotteries are neither ordered with respect to their associated *profitability* (proxied by the expected return), nor with

[8]Note that the modal number of correct answers to CRT is zero for both males and females (see Figure 2).

respect to their associated *risk* (proxied by the variance). Instead, the presentation of each lottery pair is artificially manipulated precisely to control for possible order effects.



Panel A Panel B

Figure 6: CRT and relative frequency of risky choices in RLP (Project 3), with 95% confidence intervals. Panel A (B): full sample (disaggregated by gender).

Figure 6 displays the relative frequency of risky choices in RLP (Panel A), disaggregated by gender (Panel B), where risky choices are identified by the elicited preference for the higher-variance lottery within the pair. It shows that our evidence, by and large, confirms the evidence in the literature that reflective subjects are, on average, less risk averse than the other groups (Donkers *et al.*, 2001; Frederick, 2005; Benjamin and Shapiro, 2005). In addition, Panel B in Figure 6 shows that once we split our subject pool by gender, females tend to be more risk averse than males in the same CRT group and the the average frequency of risk averse males varies negligibly across groups, while, for females, it is higher for reflective group than for the residual one. Our evidence suggests that both cognitive ability and gender play an important role in explaining risk aversion, and omitting one of them may lead to biased estimates of risk aversion.

However, the description of Figure 6 neglects relevant features of the underlying economic decision at stake. When selecting a lottery, subjects most likely compare the profitability of each decision, not simply its associated risk. Put differently, the relative frequency of risky choices does not characterize precisely the economic trade-off underlying the RLP protocol. For this reason, we test the robustness of the preliminary evidence in Figure 6 by estimating subjects' degree of (Constant Relative) Risk Aversion, ρ. Subjects' choices are assumed to maximize the expected value of the utility function $u(x)$ over monetary prizes, x,

$$u(x) = \frac{x^{1-\rho}}{1-\rho}, \rho \neq 1, \tag{1}$$

which we estimate structurally by using standard maximum likelihood methods (Andersen *et al.*, 2008).

The estimates in Table 4 are highly significant and greater than zero, which shows that risk aversion is the representative preference for all CRT groups. When we test the difference in risk aversion at the aggregate level, in column (1), the p-values at the bottom of the table show that it is only significant between reflective and impulsive. When we test differences by gender, we find that the overall difference between reflective and impulsive is mainly driven by females, and we also find a significant difference between reflective females and others, which is hidden in the aggregate estimates.[9]

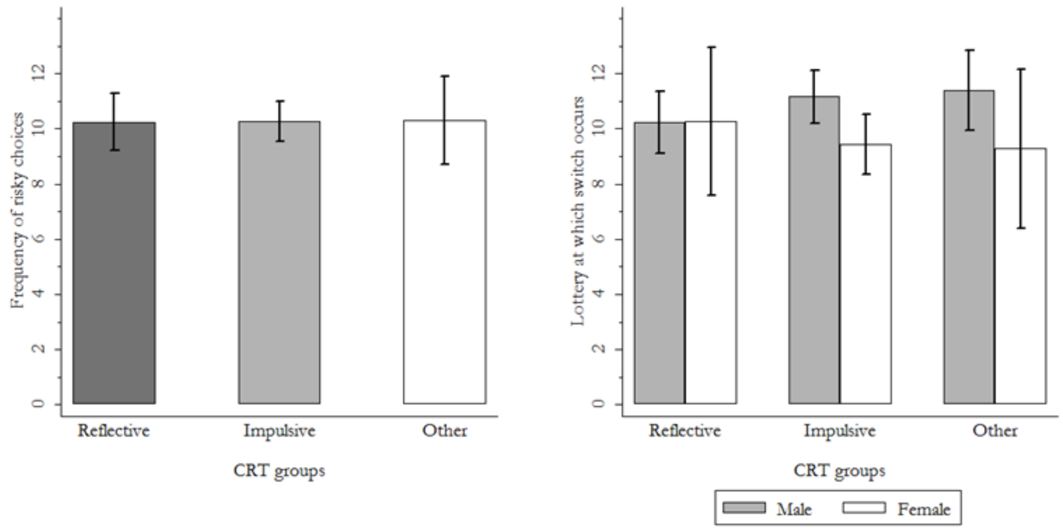
	(1) All	(2) Male	(3) Female
Reflective (R)	0.508***	0.481***	0.545***
	(0.023)	(0.029)	(0.035)
Impulsive (I)	0.571***	0.506***	0.609***
	(0.015)	(0.031)	(0.016)
Other (O)	0.502***	0.394***	0.627***
	(0.047)	(0.080)	(0.031)
Obs.	4,608	2,184	2,424
P-val. diff. R - I	0.012**	0.512	0.081*
P-val. diff. R - O	0.914	0.297	0.065*
P-val. diff. I - O	0.154	0.179	0.592

Table 4: Structural estimation of risk aversion (ρ) using RLP data. Maximum likelihood estimates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The total number of observations is the product between the number of subjects and the number of lottery choices per subject

With this evidence in mind, we now move to analyzing MPL data (projects 4 and 8). Figure 7 reports the frequency of risky choices, disaggregated by CRT group and gender, which we measure as the "average switching choice" from the risky to

[9]The lowest estimate of risk aversion is not associated with the reflective group, but the residual one, which also displays generally a higher standard error, thus indicating a "noisier" behavior compared with the other two groups.

the safe lottery in the sequence. The rationale for Figure 7 is the following: the closer to the bottom subjects switch from the risky Option B, whose profitability is higher at the beginning of the sequence, the lower their revealed risk aversion is. In Figure 7 we only consider subjects whose behavior satisfies minimal "consistency conditions", which will be explained and discussed in Section 6.2.



Panel A Panel B

Figure 7: CRT and frequency of risky choices in MPL (Projects 4 and 8), with 95% confidence intervals. Panel A (B): full sample (disaggregated by gender).

Panel A in Figure 7 shows that at the aggregate level all CRT groups show almost identical behavior. The only detectable difference is, once again, higher variability for the residual group. When we disaggregate by gender, we see that the relation between CRT and risk attitudes has a strong gender component: while risk aversion slightly decreases moving from the reflective to the residual group for males, this pattern is exactly reversed for females. We also observe that the frequency of risky choices for reflective subjects is the same for males and females on average, although females' choices have higher variability.

By analogy with Table 4, Table 5 shows estimates of risk aversion obtained by using MPL data. The p-values at the bottom of the table show that at the aggregate level, in column (1), differences in risk aversion across CRT groups are not significant, thus confirming the preliminary evidence in Figure 7. The same result also holds when we disaggregate by gender, suggesting that the trends we observe in Figure 7 are not statistically significant. Estimates in Table 4 and 5 are obtained from protocols with a rather different structure (MPL vs RLP). However, they both apply the "contextual utility approach" (Wilcox, 2011) and therefore should be scale and task-neutral.

	(1) All	(2) Male	(3) Female
Reflective (R)	0.217***	0.198***	0.223**
	(0.0540)	(0.064)	(0.113)
Impulsive (I)	0.188***	0.0683	0.296***
	(0.045)	(0.064)	(0.058)
Other (O)	0.179**	0.103	0.264**
	(0.078)	(0.081)	(0.128)
Obs.	3,969	2,184	1,785
P-val. diff. R - I	0.643	0.117	0.538
P-val. diff. R - O	0.667	0.322	0.801
P-val. diff. I - O	0.914	0.709	0.806

Table 5: Structural estimation of risk aversion (ρ) using MPL data. Maximum likelihood estimates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The total number of observations is the product between the number of subjects and the number of lottery choices per subject, after excluding subjects making inconsistent lottery choices (see Section 6.2)

5 CRT and Social Preferences

The relation between cognitive ability and social preferences is, to some extent, yet to be explored. Chen *et al.* (2013) find that subjects who perform better in the Math portion of the SAT are more generous in both the Dictator game and in a series of small-stakes "dictatorial" (i.e., unilateral) decisions, known as *Social Value Orientation* (SVO). In contrast, Ben-Ner *et al.* (2004) find that the performance in the *Wonderlic test* has a weak negative effect on giving, which is mainly driven by females. In partial contrast with these results, Benjamin *et al.* (2013) find that school test scores do not affect giving.[10] Somewhat related, Hauge *et al.* (2009) study the relation between attitudes to give in different pro-social tasks (e.g., charitable giving, Dictator Games, etc. . .) and "cognitive load", which they measure by asking subjects to memorize numbers of 7 digits, some of which are easy (hard) to remember, e.g., 1111111 or 1234567 (9325867 or 7591802). They find that the effect

[10]The Wonderlic test is based on problem-solving ability and consists of 50 questions. The score in the test is highly consistent with various measures of intelligence (Hawkins *et al.*, 1990).

of cognitive load on giving is small.

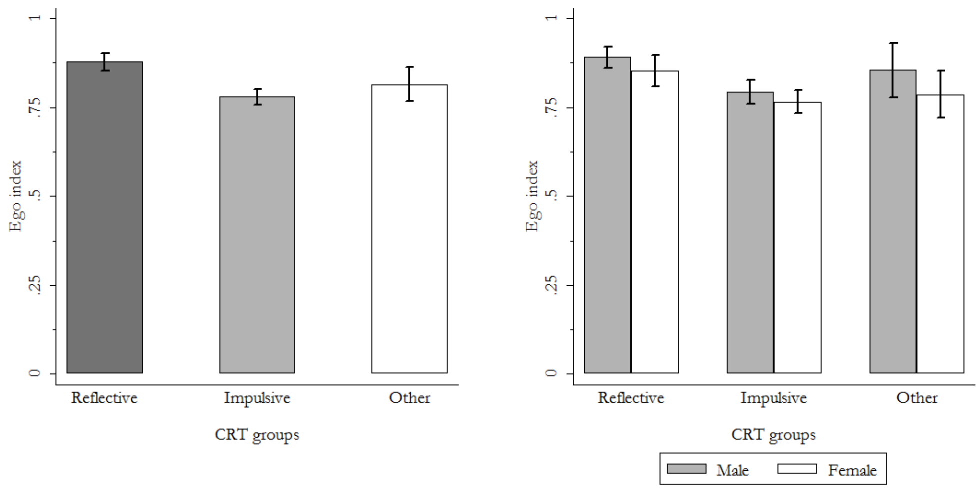
Prompted by the limited (and far from univocal) evidence in this field, Ponti and Rodriguez-Lara (2014) use data from Project 2 on a Linear Dictator Game of 98 subjects and condition the estimates of Fehr and Schmidt (1999) classic structural model of social preferences, which identifies *envy* (i.e., the aversion to inequality experienced from an disadvantaged position) and *guilt* (i.e., the aversion to inequality experienced from an advantaged position), to the same CRT groups used in this paper. They find that *inequality aversion* (often considered as the "standard" distributional attitude) is *typical of impulsive subjects* (especially, in standard Dictator Games). In contrast, reflective subjects are associated with negligible social concerns, with the exception of an *unconditional altruistic attitude*, i.e., negative envy and positive guilt, in situations where the Dictator's payoff is held constant.

Our distributional data are from Project 4 and consist of a sequence of distributional decisions over four monetary payoff pairs in which the identity of the best-paid player is constant across choices (see Section 2.2). Since choice are not naturally ordered, we provide descriptive evidence of this experimental environment by introducing an ad hoc index, borrowed from Project 6, which measures the share of the Dictator's available pie s/he allocates to him/herself (conditional on the specific round choice set):

$$EgoIndex(k) = \frac{x_D(k) - min(x_D(h))}{max_h(x_D(h)) - min_h(x_D(h))}, \quad (2)$$

where $x_D(k)$ denotes the monetary payoff the Dictator allocates to herself when selecting option k. In other words, if the Dictator gives him/herself the maximum (minimum) prize available (regardless of what the Recipient obtains), the value of the $EgoIndex(.)$ is 1 (0), respectively.

Figure 8 reports descriptive statistics of the distribution of *EgoIndex*, disaggregated by CRT group and gender. It shows that impulsive (especially males) subjects have higher distributional concerns, with no noticeable difference between reflective subjects and the remainder.



Panel A Panel B

Figure 8: *EgoIndex* and CRT

Before assessing the empirical content of this preliminary evidence, let us remark that, similarly to what we have just discussed for risky choices in Section 5, Figure 8 captures the economic trade-off underlying Dictators' decisions only partially, as it is calculated looking at the Dictator's payoffs only, and not at the Recipient's. This contrasts with the common view which models social preferences by measuring *relative comparisons* between the Dictator's and the Recipient's payoffs. For this reason, we shall frame Dictators' behavior within the realm of the classic Fehr and Schmidt (1999) model of social preferences, according to which the Dictator's utility $u(.)$ not only depends on her own monetary payoff, but also on that of the Recipient, $x_R(\gamma)$, as follows:

$$u(x_D(\gamma), x_R(\gamma)) = x_D(\gamma) - \alpha max\left\{x_R(\gamma) - x_D(\gamma), 0\right\} - \beta max\left\{x_D(\gamma) - x_R(\gamma), 0\right\}, \quad (3)$$

where the values of α and β determine the Dictator's *envy* (i.e. aversion to inequality when receiving less than the Recipient) and *guilt* (i.e., aversion to inequality when receiving more than the Recipient), respectively.[11] Fehr and Schmidt (1999) follow Loewenstein *et al.* (1989) by assuming $0 \leq \beta < 1$ and $\alpha \geq \beta$. This implies that Dictators are inequality averse ($\alpha \geq \beta \geq 0$), and more so when they get less than Recipients. We estimate α and β without imposing these restrictions and by using a

[11] Our data format seems ideal to identify envy and guilt, in that the identity of the best (worst) paid agent is constant across options.

(static) multinomial logit model in which the utility associated with the Dictator's choice of allocation, γ, follows equation (3). We obtain the estimates by maximum likelihood and by clustering standard errors at the subject level.

	All		Male		Female	
	α	β	α	β	α	β
Reflective (R)	0.116**	0.533***	0.125**	0.521***	0.0995	0.578***
	(0.048)	(0.047)	(0.053)	(0.051)	(0.083)	(0.096)
Impulsive (I)	0.295***	0.760***	0.272***	0.728***	0.331***	0.789***
	(0.036)	(0.037)	(0.047)	(0.062)	(0.049)	(0.045)
Other (O)	0.237***	0.582***	0.130*	0.415***	0.307***	0.665***
	(0.071)	(0.087)	(0.078)	(0.101)	(0.107)	(0.123)
Obs.	8,064	8,064	4,152	4,152	3,912	3,912
P-val. diff. R - I	0.003***	0.000***	0.042***	0.012***	0.016***	0.052*
P-val. diff. R - O	0.176	0.626	0.955	0.369	0.162	0.582
P-val. diff. I - O	0.441	0.068*	0.124	0.009***	0.835	0.357

Table 6: CRT and social preferences: Fehr and Schmidt (1999)'s structural estimation. Maximum likelihood estimates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The estimates in Table 6 show that cognitive ability, measured using the CRT groups, explains inequality aversion, thus confirming the results in Cabrales *et al.* (2010). When we test pairwise differences in the estimates between CRT groups we find that impulsive subjects have higher distributional concerns than reflective ones, as the p-values at the bottom of the Table 6 show. In addition, we find that impulsive subjects are also weakly more guilty than the residual group, and this is mostly driven by males' behavior.

6 Is CRT another rationality test ?

In this section we study whether CRT scores and groups are related with measures of "consistency" associated with subjects' behavior in the experiments, as well as alternative proxies of subjects' cognitive ability. As for the former, our indicator of consistency is related with the lottery choices in MPL experiments. As for the latter, we consider two additional measures of cognitive ability: *educational achievement* and *financial literacy.* Even though these two measures may depend on many factors, we follow Frederick (2005)'s intuition that certain aspects of cognitive ability, such as reading comprehension and mathematical skills, may aid performance in CRT and are likely to correlate with educational achievement and financial literacy, too.

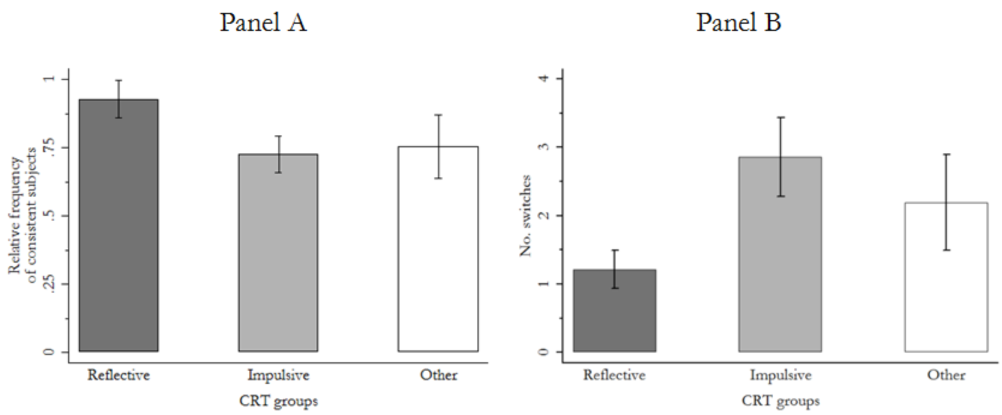
6.1 Rationality in lotteries

In this section we test whether cognitive reflection explains subjects' *consistency* across choices over lotteries by using the MPL data in projects 4 and 8. In this respect, a "consistent" subject is defined as one whose choices satisfy these conditions:

1. She should always choose Lottery B (A) in Decision 1 (21) in the sequence. This condition is due to first-order stochastic dominance.

2. She should switch from Option B to Option A only once in the sequence. This is due to monotonicity and transitivity.

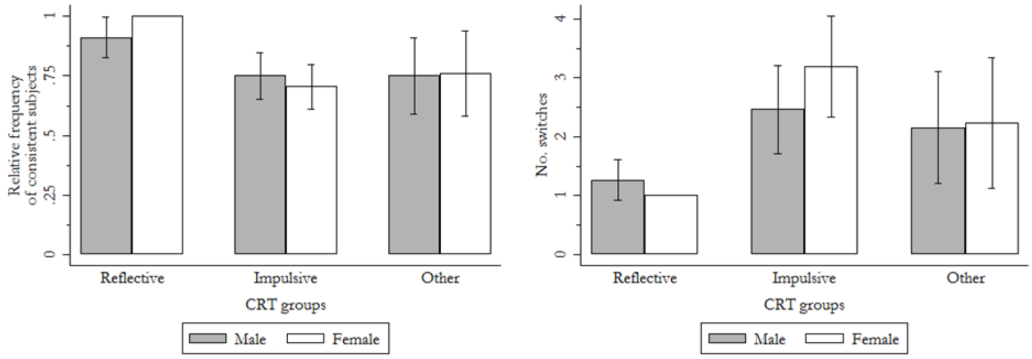
This joint condition yields a dummy equal to 1 for "consistent subjects", i.e., those who satisfy conditions 1-2 for all the choices in the sequence and 0 otherwise. We also use another proxy for consistency. By counting the number of switches observed for any given individual, we define "inconsistency" as growing with the number of switches.



Panel A Panel B

Figure 9: Consistent subjects and number of switches in lottery choices and 95% confidence intervals

Panel A in Figure 9 shows the relative frequency of consistent subjects in the different CRT groups. About 90% of reflective subjects are consistent, while only around 75% are in the other two groups. The 95% confidence intervals in Figure 9 show that reflective subjects are significantly more consistent than any of the other groups, while the difference between the two subgroups who are not reflective is not. As for Panel B, it shows that the number of switches for the impulsive group and all others is three (two) times greater than that of the reflective (residual) group, respectively. This difference is significant, although the difference between the two subgroups that are not reflective is not.



Panel A Panel B

Figure 10: Consistent subjects and number of switches in lottery choices by gender and 95% confidence intervals

As for Figure 10, it shows that there are no significant gender differences in consistency within each CRT group.[12] In these data, there are only 11 females in the reflective group and they all behave consistently. This is why the variance for reflective females displayed in Figure 10 is zero.

6.2 Grades and financial literacy

Extensive evidence documents that educational achievement, which we measure using subjects' grades at university from 0 to 100 (GPA), is positively correlated with labor market outcomes (Heckman *et al.*, 2006). Similarly, financial literacy has been shown to correlate with stockholding (Christelis *et al.*, 2010) and is an increasingly important objective in high school curricula (Mandell and Klein, 2009).

We measure financial literacy by asking 3 questions on subjects' general knowledge of financial markets. Consistently with Frederick (2005), the ordered logit

[12]In Table 4 in Appendix B we implemented Mann-Whitney-Wilcoxon tests for pairwise comparisons between all CRT groups for the full sample and disaggregated by gender. Finally, in Table 5 we performed the same tests for gender differences. The results are in line with those in Figure 9 and 10.

Dependent variable: number of correct answers in the CRT

	(1)	(2)	(3)	(4)
GPA	0.019*	0.022**	0.010	0.009
	(0.010)	(0.010)	(0.023)	(0.023)
Female		-1.141***		-1.447***
		(0.200)		(0.473)
Financial Literacy			0.573**	0.312
			(0.240)	(0.265)
Observations	432	432	96	96

Table 7: CRT, GPA and financial literacy. Ordered Logit estimates. Robust standard errors in parentheses. * p<0.1; ** p<0.05; *** p<0.01

estimates in Table 7 show that GPA is a significant predictor of CRT scores. Financial literacy is also positively and significantly correlated with CRT. However, after controlling for gender the effect is no longer significant. It seems that the aggregate correlation between CRT and financial literacy is driven by the fact that women in our sample have lower financial literacy.[13]

7 Discussion

Overall, our results show a strong gender component in CRT performance, with females scoring significantly lower than males and falling in the impulsive group significantly more often. With regards to other individual characteristics, we found significant, although much smaller, correlations between CRT and 2D:4D, personality traits and family education. Overall, our results are in line with the evidence on cognitive reflection and 2D:4D in Bosch-Domènech *et al.* (2014) and with the non-experimental evidence on personality traits in Borghans *et al.* (2008a).

In addition, we have studied whether cognitive reflection explains risk aversion and social preferences and we find that it does. Our structural estimates overall obtained using RLP data show that reflective subjects tend to be less risk averse than impulsive ones, particularly for females. In contrast, MPL data show no significant difference by CRT group or gender, in line with the criticism to risk-elicitation methods in Andersson *et al.* (2013).[14] As for social preferences, impulsive subjects are more envious and guilty than reflective ones, and impulsive males are more

[13] After performing Mann-Whitney test for gender differences, we find that financial literacy is significantly lower for females (z=3.588, p-value = 0.0003)

[14] See also Charness *et al.* (2013); Filippin and Crosetto (2014) for a discussion of the relative advantages and disadvantages of different risk elicitation protocols.

guilty than the residual group, while females are not. This evidence complements the findings in Di Cagno *et al.* (2014) who employ the Dictator Game data of Project 2 and find that, once again, impulsive subjects are those whose behaviour markedly differs from that of the other two groups (again, in the direction of inequity aversion).

Finally, we have studied the correlation between cognitive reflection and alternative proxies of cognitive ability. Here we have found that reflective subjects are more likely to satisfy basic consistency conditions in their lottery choice of lotteries, in contrast with the other two groups, which are, instead, equally likely to violate such conditions. In line with Frederick (2005), we have also found that academic performance (GPA) is positively correlated with CRT. Similar considerations hold for financial literacy, which is also correlated with CRT. However, in this case, the effect seems to be uniquely driven by the underlying gender difference. Additional experimental sessions seem required to increase the low sample size and obtain more robust evidence with respect to this result.

We conclude by recalling that this paper exploits the richness of our dataset only partially, with particular reference to our behavioral data, in that it focuses on individual decision tasks (mainly related with risk and social preferences). The link between cognitive reflection and behavior in strategic environments is being studied elsewhere (take, for example, Projects 1, 2, 4 6 or 7). For example, Ponti and Carbone (2009) find a negative correlation between CRT scores and the level of noise of subjects' play in an experimental model of informational cascades, while Ponti *et al.* (2014b), within the realm of a simple principal-agent model with moral hazard, show that reflective principals offer higher wages, which, in turn, yield higher effort levels and profits. By the same token, reflective agents put more effort, which also results in higher expected profits in the experiment. Moving to a rather different behavioral domain, Ferrara *et al.* (2014) find that, for reflective subjects, sleep deprivation makes it more likely to choose riskier lotteries and induce a more altruistic behavior. By contrast, Albano *et al.* (2014), in an experimental procurement auction, do not detect significant differences across CRT groups in both winning frequencies or in expected profits. A more detailed study to relate such a dispersed evidence is currently under way.

Appendix

Appendix A

Personality trait	Definition
Openness	Being open to new ideas and intellectually curious, imaginative, nonconforming, unconventional an autonomous
Neuroticism	Tendency to experience psychological distress, exhibit poor emotional adjustment and experience negative affects, such as anxiety, insecurity and hostility
Agreeableness	Tendency to be compassionate, cooperative, trusting, compliant, caring and gently
Conscientiousness	Tendency to show control and self-discipline, is comprised on two related facets: achievement and dependability
Extraversion	Pronounced engagement with outside world, it represents the tendency to be sociable, assertive, active and experience positive affects such as energy and zeal

Table A1: Big 5 personality traits

	Female				Male			
	Mean			Kruskal-Wallis	Mean			Kruskal-Wallis
	Reflective	Impulsive	Other	p-value	Reflective	Impulsive	Other	p-value
Left hand 2D:4D	0.981	0.986	0.993	0.668	0.965	0.976	0.970	0.050
Rightt hand 2D:4D	0.975	0.984	0.989	0.494	0.964	0.972	0.971	0.366
Neuroticism	0.538	0.548	0.506	0.612	0.394	0.459	0.418	0.035
Extraversion	0.601	0.576	0.617	0.497	0.574	0.645	0.553	0.000
Openness	0.773	0.682	0.686	0.007	0.706	0.714	0.677	0.008
Agreeableness	0.727	0.679	0.685	0.324	0.681	0.692	0.650	0.001
Conscientiousness	0.731	0.688	0.702	0.382	0.672	0.688	0.661	0.148
Family education (1+ parent uni. degree)	0.446	0.394	0.473	0.295	0.521	0.370	0.488	0.002
N. languages >2	0.414	0.430	0.471	0.885	0.453	0.297	0.286	0.080

Table A2: Means of individuals' characteristics and p-values of Kruskal-Wallis test of differences among CRT groups

	Full sample			Female			Male		
	Reflective -		Impulsive -	Reflective -		Impulsive -	Reflective -		Impulsive -
	Impulsive	Other	Other	Impulsive	Other	Other	Impulsive	Other	Other
Left hand 2D:4D	0.011	0.014	0.366	0.830	0.415	0.417	0.022	0.080	0.649
Right hand 2D:4D	0.025	0.073	0.792	0.339	0.208	0.653	0.144	0.575	0.843
Neuroticism	0.002	0.069	0.485	0.997	0.381	0.352	0.015	0.051	0.893
Extraversion	0.321	0.574	0.071	0.486	0.877	0.287	0.025	0.210	0.000
Openness	0.070	0.005	0.031	0.002	0.022	0.917	0.808	0.014	0.002
Agreeableness	0.573	0.023	0.009	0.134	0.291	0.789	0.721	0.004	0.000
Conscientiousness	0.981	0.252	0.271	0.160	0.434	0.730	0.413	0.187	0.062
Family education (1+ parent uni. degree)	0.001	0.815	0.014	0.342	0.701	0.160	0.001	0.607	0.058
N. languages >2	0.214	0.508	0.781	0.876	0.654	0.664	0.033	0.134	0.907

Table A3: Mann-Whitney-Wilcoxon p-values of differences in means of individuals' characteristics among CRT groups

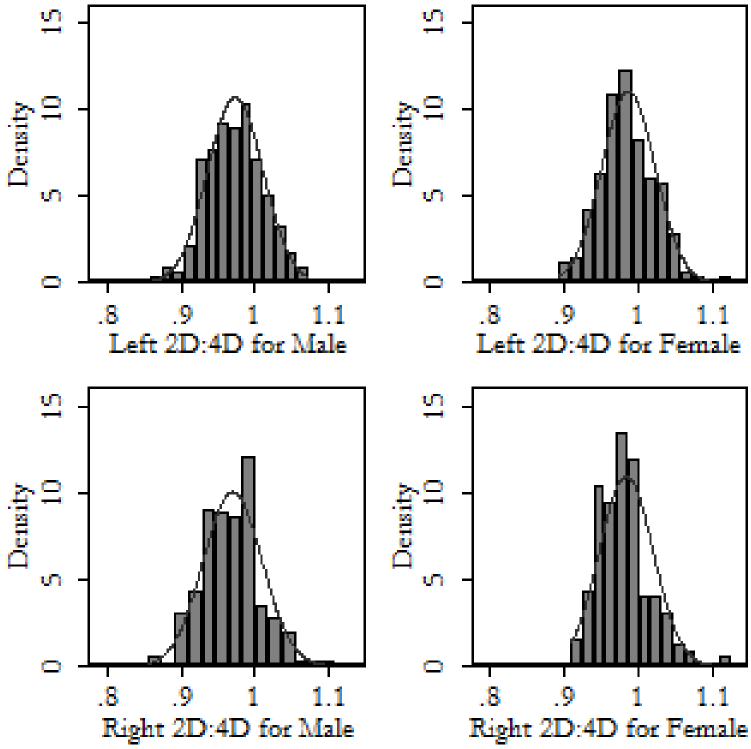


Figure A1: Second to fourth digit ratio (2D:4D) histogram by gender

Appendix B

Relative frequency of consistent subjects

	Reflective	Impulsive	Other
Reflective	.	.	.
Impulsive	0.002***	.	.
Other	0.013***	0.672	.

Number of switches

	Reflective	Impulsive	Other
Reflective	.	.	.
Impulsive	0.005**	.	.
Other	0.108	0.326	.

(a) Full sample

	Reflective	Impulsive	Other
Reflective	.	.	.
Impulsive	0.031**	.	.
Other	0.060*	1.000	.

	Reflective	Impulsive	Other
Reflective	.	.	.
Impulsive	0.033**	.	.
Other	0.230	0.608	.

(b) Male

	Reflective	Impulsive	Other
Reflective	.	.	.
Impulsive	0.039**	.	.
Other	0.088*	0.591	.

	Reflective	Impulsive	Other
Reflective	.	.	.
Impulsive	0.113	.	.
Other	0.306	0.408	.

(c) Female

Table 4: P-values of Mann-Whitney-Wilcoxon tests of relative frequency of consistent subjects and number of switches for pairs of CRT groups *p-value<0.1, **p-value<0.05, ***p-value<0.01

	Relative frequency of consistent subjects	Number of switches
Reflective	0.314	0.613
Impulsive	0.511	0.511
Other	0.932	0.947

Table 5: P-values of Mann-Whitney-Wilcoxon tests of gender differences in the relative frequency of consistent subjects and number of switches by CRT group. *p-value<0.1, **p-value<0.05, ***p-value<0.01

References

Albano, G., Di Paolo, R., Paolillo, A. and Ponti, G. (2014). *Absolute vs. Relative Scoring in Experimental Procurement.* mimeo, LUISS Guido Carli Roma.

Andersen, S., Harrison, G. W., Lau, M. I. and Rutström, E. E. (2008). Eliciting risk and time preferences. *Econometrica*, **76** (3), 583–618.

Andersson, O., Tyran, J.-R., Wengstrm, E. and Holm, H. J. (2013). *Risk Aversion Relates to Cognitive Ability: Fact or Fiction?* Working Papers 2013:9, Lund University, Department of Economics.

Apicella, C. L., Dreber, A., Campbell, B., Gray, P. B., Hoffman, M. and Little, A. C. (2008). Testosterone and financial risk preferences. *Evolution and Human Behavior*, **29** (6), 384–390.

Barrick, M. R. and Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, **44** (1), 1–26.

Ben-Ner, A., Kong, F. and Putterman, L. (2004). Share and share alike? Gender-pairing, personality, and cognitive ability as determinants of giving. *Journal of Economic Psychology*, **25** (5), 581–589.

Benet-Martinez, V. and John, O. P. (1998). Los cinco grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the big five in spanish and english. *Journal of personality and social psychology*, **75** (3), 729.

Benjamin, D. J., Brown, S. A. and Shapiro, J. M. (2013). Who is behavioral? cognitive ability and anomalous preferences. *Journal of the European Economic Association*, **11** (6), 1231–1255.

— and Shapiro, J. M. (2005). Does cognitive ability reduce psychological bias? *Harvard University.*

Bergman, O., Ellingsen, T., Johannesson, M. and Svensson, C. (2010). Anchoring and cognitive ability. *Economics Letters*, **107** (1), 66–68.

Borghans, L., Duckworth, A. L., Heckman, J. J. and ter Weel, B. (2008a). The Economics and Psychology of Personality Traits. *Journal of Human Resources*, **43** (4).

—, Golsteyn, B. H. H., Heckman, J. J. and Meijers, H. (2009). Gender Differences in Risk Aversion and Ambiguity Aversion. *Journal of the European Economic Association*, **7** (2-3), 649–658.

—, Meijers, H. and Weel, B. T. (2008b). The Role Of Noncognitive Skills In Explaining Cognitive Test Scores. *Economic Inquiry*, **46** (1), 2–12.

Bosch-Domènech, A., Brañas-Garza, P. and Espín, A. M. (2014). Can exposure to prenatal sex hormones (2d: 4d) predict cognitive reflection? *Psychoneuroendocrinology*, **43**, 1–10.

Brañas-Garza, P., Garca-Muoz, T. and Gonzlez, R. H. (2012). Cognitive effort in the beauty contest game. *Journal of Economic Behavior & Organization*, **83** (2), 254 – 260.

— and Rustichini, A. (2011). Organizing effects of testosterone and economic behavior: Not just risk taking. *PloS one*, **6** (12), e29842.

Cabrales, A., Miniaci, R., Piovesan, M. and Ponti, G. (2010). Social preferences and strategic uncertainty: An experiment on markets and contracts. *American Economic Review*, **100** (5), 2261–78.

Charness, G., Gneezy, U. and Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization*, **87**, 43–51.

Chen, C.-C., Chiu, I.-M., Smith, J. and Yamada, T. (2013). Too smart to be selfish? Measures of cognitive ability, social preferences, and consistency. *Journal of Economic Behavior & Organization*, **90** (C), 112–122.

Christelis, D., Jappelli, T. and Padula, M. (2010). Cognitive abilities and portfolio choice. *European Economic Review*, **54** (1), 18–38.

Coates, J. M., Gurnell, M. and Rustichini, A. (2009). Second-to-fourth digit ratio predicts success among high-frequency financial traders. *Proceedings of the National Academy of Sciences*, **106** (2), 623–628.

Cueva, C., Iturbe-Ormaetxe, I., Ponti, G. and Tomás, J. (2014). *An Experimental Study on the Disposition Effect with Competitive and Taxes Scheme*. Tech. rep., Universidad de Alicante.

Daly, M., Harmon, C. P. and Delaney, L. (2009). Psychological and biological foundations of time preference. *Journal of the European Economic Association*, **7** (2-3), 659–669.

Del Pozo, X., Galliera, A., Ponti, G. and Sikora, I. (2013). *Social Preferences, Risk Preferences and the Hexagon Condition*. Tech. rep., Universidad de Alicante.

Di Cagno, D., G.W., H., R., M. and Ponti, G. (2014). *Social Preferences over Utilities*. Tech. rep., LUISS Guido Carli Roma.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual review of psychology*, **41** (1), 417–440.

Dohmen, T., Falk, A., Huffman, D. and Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, **100** (3), 1238–1260.

Donkers, B., Melenberg, B. and Van Soest, A. (2001). Estimating risk attitudes using lotteries: A large sample approach. *Journal of Risk and Uncertainty*, **22** (2), 165–195.

Evans, J. S. B. T. (1984). Heuristic and analytic processes in reasoning*. *British Journal of Psychology*, **75** (4), 451–468.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, **114** (3), 817–868.

Ferrara, M., Bottasso, A., Tempesta, D., Carrieri, M., Gennaro, L. and Ponti, G. (2014). *Differential effects of sleep deprivation on risk attitude and inequality aversion: Evidence from an economic experiment*. mimeo, LUISS GUIDO Carli Roma.

Filippin, A. and Crosetto, P. (2014). *A Reconsideration of Gender Differences in Risk Attitudes*. IZA Discussion Papers 8184, Institute for the Study of Labor (IZA).

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, **19** (4), 25–42.

Gill, D. and Prowse, V. (2014). *Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis*. Economics Series Working Papers 712, University of Oxford, Department of Economics.

Grimm, V. and Mengel, F. (2012). An experiment on learning in a multiple games environment. *Journal of Economic Theory*, **147** (6), 2220–2259.

Harrison, G. W. and McDaniel, T. (2008). Voting games and computational complexity. *Oxford Economic Papers*, p. gpm045.

Hauge, K. E., Brekke, K. A., Johansson, L.-O., Johansson-Stenman, O. and Svedsäter, H. (2009). Are social preferences skin deep? dictators under cognitive load. *rapport nr.: Working Papers in Economics 371*.

Hawkins, K. A., Faraone, S. V., Pepple, J. R., Seidman, L. J. and Tsuang, M. T. (1990). Waisr validation of the wonderlic personnel test as a brief intelligence measure in a psychiatric sample. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, **2** (2), 198.

Heckman, J. J. and LaFontaine, P. A. (2010). The american high school graduation rate: Trends and levels. *The review of economics and statistics*, **92** (2), 244–262.

— and Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the ged testing program. *American Economic Review*, pp. 145–149.

—, Stixrud, J. and Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, **24** (3), 411–482.

Hey, J. D. and Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica: Journal of the Econometric Society*, pp. 1291–1326.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American economic review*, **92** (5), 1644–1655.

Hoppe, E. I. and Kusterer, D. J. (2010). *Behavioral biases and cognitive reflection.* Cologne Graduate School Working Paper Series 01-03, Cologne Graduate School in Management, Economics and Social Sciences.

Jang, K. L., Livesley, W. J. and Vemon, P. A. (1996). Heritability of the big five personality dimensions and their facets: a twin study. *Journal of personality*, **64** (3), 577–592.

John, O. P., Naumann, L. P. and Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research*, **3**, 114–158.

— and Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, **2** (1999), 102–138.

Judge, T. A., Higgins, C. A., Thoresen, C. J. and Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel psychology*, **52** (3), 621–652.

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.

Kirby, K. N., Winston, G. C. and Santiesteban, M. (2005). Impatience and grades: Delay-discount rates correlate negatively with college {GPA}. *Learning and Individual Differences*, **15** (3), 213 – 222.

Loehlin, J. C. (1992). *Genes and environment in personality development.* Sage Publications, Inc.

—, McCrae, R. R., Costa Jr, P. T. and John, O. P. (1998). Heritabilities of common and measure-specific components of the big five personality factors. *Journal of Research in Personality*, **32** (4), 431–453.

Loewenstein, G. F., Thompson, L. and Bazerman, M. H. (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social psychology*, **57** (3), 426.

Mandell, L. and Klein, L. S. (2009). The impact of financial literacy education on subsequent financial behavior. *Journal of Financial Counseling and Planning*, **20** (1), 15–24.

Manning, J. T., Scutt, D., Wilson, J. and Lewis-Jones, D. I. (1998). The ratio of 2nd to 4th digit length: a predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. *Human reproduction*, **13** (11), 3000–3004.

Morsanyi, K., Busdraghi, C. and Primi, C. (2014). Mathematical anxiety is linked to reduced cognitive reflection: a potential road from discomfort in the mathematics classroom to susceptibility to biases. *Behavioral and Brain Functions*, **10** (1), 31.

Oechssler, J., Roider, A. and Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization*, **72** (1), 147–152.

Pearson, M. and Schipper, B. C. (2012). The visible hand: Finger ratio (2d: 4d) and competitive bidding. *Experimental Economics*, **15** (3), 510–529.

Ponti, G. and Carbone, E. (2009). Positional learning with noise. *Research in Economics*, **63** (4), 225–241.

—, Rodriguez-Lara, I. and Di Cagno, D. (2014a). *Doing it Now or Later with Payoff Externalities: Experiental Evidence on Social Time Preference*. mimeo, LUISS GUIDO Carli Roma.

—, Sartarelli, M., Sykora, I. and Zhukova, V. (2014b). *The Price of Ent repr eneurship. Evidence from the Lab*. mimeo, Universidad de Alicante mimeo.

Sapienza, P., Zingales, L. and Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proceedings of the National Academy of Sciences*, **106** (36), 15268–15273.

Toplak, M., West, R. and Stanovich, K. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, **39** (7), 1275–1289.

Wilcox, N. T. (2011). Stochastically more risk averse:a contextual theory of stochastic discrete choice under risk. *Journal of Econometrics*, **162** (1), 89–104.

Zhao, H. and Seibert, S. E. (2006). The big five personality dimensions and entrepreneurial status: a meta-analytical review. *Journal of Applied Psychology*, **91** (2), 259.

serie ad

Ivie

Guardia Civil, 22 - Esc. 2, 1º
46020 Valencia - Spain
Phone: +34 963 190 050
Fax: +34 963 190 055

Website: www.ivie.es
E-mail: publicaciones@ivie.es