Self-Confidence, Overconfidence and Prenatal Testosterone Exposure: Evidence from the Lab*

Patricio S. Dalton † Sayantan Ghosal ‡

February 9, 2014

Abstract

This paper examines whether the degree of confidence and overconfidence in one's ability is determined biologically. In particular, we study whether foetal testosterone exposure correlates with an incentive-compatible measure of confidence within an experimental setting. We find that men (rather than women) who were exposed to high testosterone levels in their mother's womb are less likely to overestimate their actual performance, which in turn helps them to gain higher monetary rewards. Men exposed to low prenatal testosterone levels, instead, set unrealistically high expectations which results in self-defeating behavior. These results from the lab are able to reconcile hitherto disconnected evidence from the field, by providing a link between traders' overconfidence bias, long-term financial returns and prenatal testosterone exposure.

JEL Classification: C91, D03, D87.

Keywords: 2D:4D, testosterone, neuroeconomics, expectations, overconfidence, self-confidence, goals.

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

†Corresponding author. Tilbug University, CentER, CAGE. Warandelaan 2, 5037 AB, Tilburg, The Netherlands. E-mail: p.s.dalton@uvt.nl. phone: +31134664032

‡Adam Smith Business School, University of Glasgow.

1 Introduction

Self-confidence and overconfidence play a crucial role in people's decisions and welfare. While positive thinking can enhance motivation and improve performance, being overly confident - i.e. believing one is better than one actually is - can be self-defeating (Benabou and Tirole, 2002). Indeed, overconfidence bias has been used to explain phenomena such as business failures (Camerer and Lovallo, 1999), stock market bubbles and excessively frequent trading (Barber and Odean, 2001; Grinblatt and Keloharju, 2009).

An important question that arises is what determines the level of self-confidence and overconfidence. It is known that nurture does play a role. Mastering own experiences and observing successful experiences of similar others can influence people's confidence (Bandura 1997). Does nature play any role too?

We address this question by examining whether prenatal testosterone exposure determines people's confidence and overconfidence about their own ability to perform a rather unfamiliar and challenging task.[1] We found that, ceteris paribus, male subjects exposed to low prenatal testosterone levels were more likely to overestimate their actual performance. Such overestimation, rather than being a rational strategy to increase motivation and hence performance, showed to be self-defeating. Overconfident participants gained significantly less earnings than participants who were rather conservative in their expectations.

As a marker for the strength of prenatal testosterone exposure we used the ratio of the length of the index finger to the length of the ring finger (2D:4D). We followed the vast literature started by Manning et al. (1998) which shows that individuals with conditions associated with very high prenatal testosterone levels exhibit significantly smaller 2D:4D (Brown et al., 2002).[2]

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

[1]Prenatal testosterone exposure has been shown to have important organizing effects on brain development, several psychological traits and behavior (see Tobet and Baum, 1987)

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

[3]The incentive-compatible scheme of payments we used was also implemented by Mobius and Rosenblant (2006) to measure self-confidence in a lab setting. Next section describes in detail the mechanism.

confidence.

This paper contributes to the literature in several ways. First, overconfidence is "perhaps the most robust finding in the psychology of judgment" (De Bondt and Thaler, 1995, p. 389). Here we provide evidence that it is - at least partially - biologically determined.

Second, our results unify two well-known empirical findings in the literature of economics and finance. On the one hand, Barber and Odean (2001) find that overconfident traders earn lower returns than more conservative traders. On the other hand, Coates et al. (2009) show that male traders who earn higher long term returns and remain longer time on business have been exposed to high prenatal testosterone levels (i.e. lower 2D:4D). Hence, our results reconcile these two pieces of independent evidence, providing a plausible explanation of Coates' et al. (2009) findings based on Barber and Odean's (2001) evidence. Namely, financial traders with higher prenatal testosterone exposure have higher returns in the long-run and stay longer on business because they are less likely to suffer from overconfidence bias.

Third, Benabou and Tirole's (2002) seminal model predicts that overconfidence can harm welfare but individuals may nevertheless display it. Our paper does not only provide empirical evidence for this theoretical finding, but it also suggests a biological origin.

Finally, the paper contributes to an emerging literature in economics studying the relationship between 2D:4D and economic preferences and outcomes. For example, 2D:4D has been shown to be correlated with social preferences (van den Bergh and Dewitte, 2006; Millet and Dewitte, 2009; Buser, 2012) risk preferences (Apicella et al., 2008; Sapienza et al., 2008) and economic outcomes such as cooperation in prisoner's dilemma (Sanchez-Pages and Turiegano, 2010), traders' profits (Coates et al., 2009) or career choices (Sapienza et al., 2008). However, to our knowledge, this is the first paper investigating the link between 2D:4D and confidence and overconfidence.[4]

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

2 Method

We designed an experiment to measure the three variables of interest: (ex-ante) self-confidence, ex-post overconfidence and the second to fourth digit ratio (2D:4D). Through emails and leaflets, we recruited two hundred fifty-five undergraduate and graduate students from the University of Warwick. We conducted twelve sessions with approximately twenty students each. Each session lasted sixty minutes. The average payment was £ 14 including a show up fee of £ 5. In each session, the sequence of the experiment was as

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

follows. Once each subject read and signed the consent form, the experimenter would read out loud the experimental instructions, which included a description of the task and the monetary payments.[5] Participants were informed that they had twenty minutes to complete the same task and that they would be payed 100 points (equivalent to £ 1) per completed task. Subjects were given one minute of practice time to get familiar with the task and after that, we elicited their self-confidence in the following way.[6] We asked them to predict the number of tasks they expected to successfully complete in the twenty minutes of performance time. The answer to that question constituted our measure of self-confidence. In Section 2.1 below we describe the incentive-compatible mechanism of self-confidence elicitation. Once the subjects reported their prediction, they started performing the task for twenty minutes. When they finished, they were asked to fill in a questionnaire, they were payed and their right hands were scanned. Below we describe in more detail the manner in which self-confidence, overconfidence and the 2D:4D were measured.

2.1 Confidence, Overconfidence and Incentives Scheme

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

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

Although these scales have been proved to be useful in many settings, they were not appropriate for the purpose of this paper for the following reasons. First, we required a unidimensional and easily interpretable measure of how confident the person was about his/her capacity to perform an unfamiliar task in the lab. These scales are rather multidimensional and general. Second, this paper also aimed at measuring overconfidence, so we needed to be able to evaluate how far were expectations from actual performance. The existing psy-

[5]See Appendix A for the instructions and appendices B and C for a snapshot of the screen the subjects saw.

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

[7]Perceived self-efficacy is a very different concept to self-esteem. While perceived self-efficacy is a judgment of capability, self-esteem is a judgment of worth (Bandura, 1977, pg. 309).

chological scales are simply not developed to measure this construct. Finally, we needed to capture the true expectations of own performance and at the same time, we wanted to ensure that subjects performed up to their maximum capacity during performance time. To achieve that, subjects needed to be provided with the right material incentives. In the absence of incentives, they may have answered to conform to the experimenter's expectations, they may have not put enough care to think about the answer or they may have not put enough effort when performing.

In light of the above, we applied the following incentive scheme. Subjects were asked to solve a practice task for one minute. Once the practice period was over, their self-confidence C was measured by asking them to report how many tasks they expected to solve during the 20-minute period. The subject received a piece rate of 100 points per solved task, P, minus 40 points for each task that he mispredicted when estimating future performance:

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

The misprediction penalty provided the subjects with an incentive to truthfully report the median of their perceived performance distribution. Note that this scheme implies that the effective piece rate of performance was 140 points for each successfully completed task as long as they stay below their estimate and 60 points for each successfully completed task thereafter. Hence, truthful elicitation of self-confidence was bound to somewhat distort incentives during the performance period. For this reason, we chose a generous exchange rate from points to money (0.01 £ per point) to ensure that even 60 points represented a salient reward and the subject had high enough incentives to continue putting effort. Moreover, once the subject reached his estimate, it meant that he figured out the way to solve the task, and the effort put thereafter was bound to be less costly.

Recall that above and beyond confidence, we were interested in measuring the degree of overconfidence. Moore and Healy (2008) defines overconfidence as the overestimation of one's actual performance and we apply this definition for this paper.[8]

Like self-confidence, the degree of overconfidence is usually measured through answers to surveys or experimental questionnaires in a non-incentivised way. For the same reasons exposed above, we used an incentive compatible measure of overconfidence. A person was considered to be overconfident when he/she expected to perform better than his/her actual performance. This measure pins down overconfidence in an incentive compatible way because subjects had material incentives to both, announce their expectations as accurately as possible and perform as well as possible.

[8] Overconfidence has also been defined in the literature as the overplacement of one's performance relative to others and as the overestimation of the precision in one's knowledge (Moore and Healy, 2008).

2.2 2D:4D and other Measures

At the end of the experiment, we scanned the right hand of each subject, we measured the length of their second and fourth finger, and calculated their ratio (2D:4D ratio).[9] Finger length was measured by two independent research assistants using a digital caliper. All data analysis was done using the average of the two independent measures of ratios.[10]

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

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

2.3 The Task

For our experiment, we chose a computerized puzzle which consisted of a modified version of the so-called "Tower of Hanoi"(ToH) puzzle. The standard ToH consists of three rods, and a number of disks of different sizes which can slide onto any rod.[11] The puzzle starts with the disks in a neat stack in ascending order of size on one rod, the smallest at the top, thus making a conical shape. The objective of the puzzle is to move the entire stack to another rod, obeying the following rules: (a) only one disk can be moved at a time, (b) each move consists of taking the upper disk from one of the rods and sliding it onto another rod, on top of the other disks that may already be present on that rod and (c) no disk may be placed on top of a smaller disk. We used a slightly modified version of the original ToH to increase difficulty. In our case, instead of having disks of different sizes, there were disks of different colors. The rule was to always preserve the original order of colours of the disks (pink, green, blue, turquoise, brown). For example, brown could be moved on top of any other disks, but green could only be moved on top of the pink, etc.[12]

[9] 2D:4D was determined from right-hand measurements only, because right-hand digit ratios have been shown previously to display more robust sex differences and are thus thought to be more sensitive to prenatal androgens.

[10] Both independent measures displayed a high repeatability (intraclass correlation 0.875). The results if we used the two measurements separately are qualitatively the same.

[11] The standard ToH has been extensively studied by cognitive psychologists but very rarely used in economics (McDaniel and Rutström, 2001).

[12] A screenshot of the computerized puzzle can be seen in Appendix C.

We chose this puzzle for several reasons. First, the rules of the task are easy to understand, which reduces the possibility of noise. Second, the task has a unique solution (involving thirty one moves) which is computed by backward induction. Third, it is quite unfamiliar to subjects and it constitutes a Eureka-type of problem (Cooper and Kagel, 2005): it appears to be challenging at first glance, but simple to solve once the algorithm is figured out. This is a desirable property for a self-confidence and overconfidence measure, since it allowed us to elicit expectations within a setting in which people had imperfect knowledge of their own abilities.[13] In fact, in our experiment, only five subjects managed to solve the task in the practice time, but all eventually made it during the performance time.

3 Data

Two hundred and fifty five students from Warwick University participated in the study. The sample was proportionally balanced by gender. Five subjects who solved the task in the practice time were excluded from all the analysis. We decided to exclude them because their prediction of expected performance would not involve any level of uncertainty about their capacity to perform. Further, we excluded one outlier with an overconfidence level forty times higher than the mean and two subjects who did not report their gender. Therefore, the final sample we analyze consisted of two hundred and forty nine subjects.

Table 1 shows the summary statistics of our experimental measure of self-confidence. On average, subjects expected to solve about ten ToHs in twenty minutes, with a standard deviation of about six. As Figure 1 shows, the frequency distribution of confidence in our data is quite disperse and rather skewed to the right, with a median at eight, a mode at five, a minimum at zero and a maximum at thirty. Finally, although this paper is not about gender differences, it is worth noticing that in average men expected to perform forty percent better than women ($P < 0.01$).[14]

We also looked at other variables that we expected to be positively correlated with our measure of self-confidence (see Table 2). As expected, we observed a significant positive correlation with Schwarzer and Jerusalem's (1995) general measure of perceived self-efficacy ($P < 0.01$). Likewise, self-confidence was positively correlated with some proxies of the ability to solve the task such as being enrolled in a mathematical oriented degree ($P < 0.01$) and being familiar with the task ($P < 0.10$). We also looked at its correlation with risk aversion, since one could expect that risk averse subjects set lower expectations. However we don't find evidence of a link between these two variables.

Table 3 and Figure 2 describe the data on overconfidence. Recall that those subjects whose expectations were higher (respectively lower) than their actual performance are clas-

[13] Imperfect knowledge of own ability is one of the key assumptions made by Benabou and Tirole (2002) to model self-confidence.

[14] This and all the tests reported hereafter are two sided.

sified as overconfident (respectively underconfident). As it can be seen in Table 3, the sample is equally divided between these two groups of subjects, with only 7 percent of the subjects performing exactly the way they expected to perform. Interestingly, the number of overconfident (hence underconfident) subjects is equal for men and women.

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

4 Results

4.1 Self-confidence and Prenatal Testosterone Exposure

In Table 5 we report the results of a linear regression analysis examining the relation between our measure of self-confidence and the digit ratio.[15] Self-confidence was significatively positively correlated with the digit ratio, suggesting that high self-confidence was associated with low prenatal testosterone exposure. When data were analyzed separately for men and women, we found that the effect was entirely driven by men. Also, as expected, men exhibited significantly higher self-confidence than women ($P < 0.01$).

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

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

4.2 Overconfidence and Prenatal Testosterone Exposure

Table 6 reports results on the relation between our measure of overconfidence and digit ratio. Recall that overconfidence is defined as expectations minus actual performance, so this variable takes positive values when the person is overconfident, and is increasing in the degree of overconfidence. When we regressed this measure on digit ratio, we found that they were significatively positive correlated, suggesting that high overconfidence was associated with low prenatal testosterone exposure (Table 6). After controlling for possible confounding variables, like previous experience with the task, risk aversion and self-efficacy, the association between prenatal testosterone exposure and overconfidence became even stronger. (Table 6, columns III and IV). Again, we found this effect only in men. Also, as expected, we found that the higher the degree of previous expertise with the task and the higher the self-efficacy, the lower the overconfidence.[16]

4.3 Overconfidence and Experimental Earnings

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

As pointed out by Benabou and Tirole (2002), the answer is not straightforward. On the one hand, setting high expectations can improve earnings by motivating higher effort and hence improving performance. On the other hand, setting excessively high expectations can only increase the cost of not reaching them. Thus, whether overconfidence is in the end a good or a bad strategy is an empirical question. We examined this question by regressing an overconfidence dummy on the final experimental earnings (see Table 8). Our regressions confirm that being overconfident was on average a bad strategy in our experiment. Non-overconfident subjects who set their expectations below their actual potential ended up winning on average eight to nine British pounds more than overconfident subjects.[17] These results are true for both, men and women, and controlling for a series of possible confounders. The magnitude of the cost of overconfidence on earnings was very high: it more than doubled the cost of not having previous experience with the task. Interestingly, the 2D:4D ratio did not affect earnings directly, but trough its effect on self-confidence.

The subjects who performed better in the lab seemed to have pursued a strategy that the psychologists know as "defensive pessimism": setting low expectations in uncertain situations to harness anxiety and thus perform better. This strategy was also discussed in

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

[17] Note that given that we created the dummies *Exceeded and Correct Expectations*, the benchmark variable for comparisons is *Unreached Expectations*.

the economic model of Benabou and Tirole (2002). In their theory, "defensive pessimism" comes as a result from assuming that ability is a substitute rather than a complement of effort in generating future pay-offs. This gives the person an incentive to discount or repress signals of high ability, as these would increase the temptation to "coast" or "slack off." In other words, considering the possibility of failure may motivate higher effort to avoid that possibility, and it is a rational strategy to follow inasmuch it increases performance. This is, indeed, what we observe in our experimental data: overconfident subjects gained substantially lower earnings than subjects who set more modestly their expectations. Overconfidence was self-defeating.

5 Conclusion

This is the first paper examining the biological determinants of self-confidence and overconfidence. We provide evidence that men with higher prenatal testosterone exposure (i.e. low 2D:4D ratio) are less likely to set unrealistically high expectations about their own performance. Importantly, we also show that such bias has normative implications: overconfidence was detrimental for individuals' earnings.

The evidence in this paper can be understood as a plausible explanation of why male financial traders with higher prenatal testosterone exposure remain longer on business or have higher long term profits (Coates et al., 2009). According to our findings, these traders may be less likely to suffer from overconfidence bias, and this helps them to be more successful in the long run. This interpretation is consistent with the empirical findings of Barber and Odean (2001), who show that overconfidence is negatively correlated with traders financial returns.[18]

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

[18]The other alternative explanations to Coates' et al. (2009) findings rely on risk preferences or preferences for competition. However, there is no unambiguous empirical evidence on the link between these two preferences and 2D:4D. Moreover, we found no significant relationship of 2D:4D and risk aversion in our data (neither for men nor for women), which is also the findings of Sapienza et al. (2008) and Apicella et al. (2008). Likewise, Pearson and Schipper (2011) found no correlation between 2D:4D and competitive behavior in markets.

References

[1] Apicella, C. L., A. Dreber, B. Campbell, P. B. Gray, M. Hoffman, M. and A. C. Little, 2008. Testosterone and Financial Risk Preferences. Evolution and Human Behavior, 29(6), 384-390.

[2] Bandura, A., 1977. Self-efficacy: Toward a Unifying Theory of Behavioral Change. Psychological Review, 84, 191-215.

[3] Bandura, A., 1986. Social Foundations of Thought and Action. Englewood Cliffs, NJ.

[4] Barber, B. M. and T. Odean, 2001. Boys will be boys: Gender, Overconfidence, and Common Stock Investment. Quarterly Journal of Economics, 116(1), 261-292.

[5] Benabou R, and J. Tirole, 2002. Self-Confidence and Personal Motivation. Quarterly Journal of Economics, 117(3), 871-915.

[6] Brown, W. M., M. Hines, B. A. Fane and S. M. Breedlove, 2002. Masculinized Finger Length Patterns in Human Males and Females with Congenital Adrenal Hyperplasia. Hormones and Behavior. 42(4), 380-386.

[7] Buser, T., 2012. Digit Ratios, the Menstrual Cycle and Social Preferences. Games and Economic Behavior, 76, 457470.

[8] Camerer, C. and D. Lovallo, 1999. Overconfidence and Excess Entry: An Experimental Approach. American Economic Review, 89(1), 306-318.

[9] Coates, J. M., M. Gurnell and A. Rustichini, 2009. Second-to-fourth Digit Ratio Predicts Success Among High-frequency Financial Traders. Proceedings of the National Academy of Sciences, 106(2), 623-628.

[10] Cooper, D. J. and J. H. Kagel, 2005. Are two Heads Better Than One? Team Versus Individual Play in Signaling Games. American Economic Review, 95(3), 477-509.

[11] De Bondt, W. and R. H. Thaler. 1995. Financial Decision-making in Markets and Firms: a Behavioral Perspective, in R. A. Jarrow, V. Maksimovic and W. T. Ziemba, eds, Finance, Handbooks in Operations Research and Management Science, vol. 9, pp. 385410. Amsterdam: North Holland.

[12] Eckel, C. and P. Grossman, 2002. Sex Differences and Statistical Stereotyping in Attitudes Toward Financial Risk. Evolution and Human Behavior, 23(4), 281-295.

[13] Grinblatt, M. and M. Keloharju, 2009. Sensation Seeking, Overconfidence, and Trading Activity, The Journal of Finance, 64(2), 549-578.

[14] Lutchmaya, S., S. Baron-Cohen, P. Raggatt, R. Knickmeyer and J. T. Manning, 2004. 2nd to 4th Digit Ratios, Fetal Testosterone and Estradiol. Early Human Development, 77(1), 23-28.

[15] Manning, J. T., D. Scutt, J. Wilson and D. I. Lewis-Jones, 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.

[16] Manning, John T., Barley, L., Walton, J., Lewis-Jones, D.I., Trivers, R.L., Singh, D., Thornhill, R., Rohde, P., Bereczkei, T., Henzi, P., Soler, M., Szwed, A., 2000. The 2nd:4th Digit Ratio, Sexual Dimorphism, Population Differences, and Reproductive Success: Evidence for Sexually Antagonistic Genes?. Evolution and Human Behavior, 21(3), 163-183.

[17] Manning, J. T. and R. P. Taylor, 2001. Second to Fourth Digit Ratio and Male Ability in Sport: Implications for Sexual Selection in Humans. Evolution and Human Behavior, 22(1), 6169.

[18] Millet, K. and S. Dewitte, 2009. The Presence of Aggression Cues Inverts the Relation between Digit Ratio (2D:4D) and Prosocial Behaviour in a Dictator Game. British Journal of Psychology, 100(1), 151162.

[19] McDaniel, T. M. and E. E. Rutstrm, 2001. Decision Making Costs and Problem Solving Performance. Experimental Economics, 4(2), 145-161.

[20] Mobius M. M. and T. S. Rosenblat, 2006. Why Beauty Matters. American Economic Review, 96(1), 222-235.

[21] Moore, D. A. and P. J. Healy, 2008. The Trouble with Overconfidence, Psychological Review, 115(2), 502.

[22] Pearson, M. and B. C. Schipper, 2012. The Visible Hand: Finger ratio (2D: 4D) and competitive bidding. Experimental Economics, 15(3), 510-529.

[23] Robinson, S.J. and J. T. Manning, 2000. The Ratio of 2nd to 4th Digit Length and Male Homosexuality. Evolution and Human Behavior, 21 (5), 333345.

[24] Sanchez-Pages, S. and E. Turiegano, 2010. Testosterone, Facial Symmetry and Cooperation in the Prisoners Dilemma. Physiology and Behavior, 99(3), 355361.

[25] Sapienza, P., L. Zingales, L. and D. Maestripieri, 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.

[26] Schwarzer, R. and M. Jerusalem, 1995. Generalized Self-Efficacy Scale. In J. Weinman, S. Wright, and M. Johnston. Measures in Health Psychology: A users Portfolio. Causal and Control Beliefs, pp. 35-37. Windsor, UK: NFER-NELSON.

[27] Schwarzer, R. and A. Born, 1997. Optimistic self-beliefs: Assessment of General Perceived Self-efficacy in Thirteen Cultures. World Psychology, 3(1-2), 177-190.

[28] Tobet, S. A. and M. J. Baum, 1987. Role for Prenatal Estrogen in the Development of Masculine Sexual Behavior in the Male Ferret. Hormones and Behavior, 21(4), 419-429.

[29] van den Bergh, B. and Dewitte, S., 2006. Digit Ratio (2D:4D) Moderates the Impact of Sexual cues on Mens decisions in ultimatum Games. Proceedings of the Royal Society of Biological Sciences B, 273 (1597), 20912095.

Tables

Table 1: Self-Confidence: Summary Statistics

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

Table 2: Self-confidence: Pair-wise Correlations

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

Notes: *** significant at 1%, ** significant at 5%, * significant at 10%.

Table 3: Predicted and Actual Performance

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

Table 4: 2D/4D: Summary Statistics

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

Table 5: OLS Regressions of 2D:4D on Self-Confidence

	Both Genders				Women				Men			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
Average digit ratio	26.749**	30.298**	30.513**	29.594**	12.32	13.01	13.07	13.10	36.734**	39.409*	39.201*	38.099*
	(11.11)	(11.68)	(11.06)	(10.80)	(16.45)	(15.40)	(15.60)	(15.69)	(16.71)	(18.43)	(18.40)	(17.86)
Gender: Male = 1	3.931***	3.540***	3.572***	3.310***	.	.	.	.	.	.	.	.
	(0.96)	(0.96)	(0.94)	(0.88)	.	.	.	.	.	.	.	.
Familiarity with task		0.994	1.006	0.918		1.74	1.737	1.511		(0.19)	(0.17)	(0.12)
		(1.19)	(1.20)	(1.19)		-2.118	-2.198	-2.181		(1.06)	(0.94)	(0.95)
Math degree		1.676*	1.573	1.612		-0.332	-0.335	-0.282		3.231**	3.134**	3.166**
		(0.93)	(1.04)	(1.07)		-1.954	-1.92	-1.976		(1.19)	(1.24)	(1.22)
Risk Aversion			(0.16)	(0.19)			-0.005	-0.059			(0.38)	(0.37)
			(0.28)	(0.29)			-0.369	-0.393			(0.43)	(0.42)
Self-efficacy				0.180*				0.160*				0.12
				(0.09)				-0.088				(0.24)
Observations	247	245	244	244	128	128	128	128	119	117	116	116

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

Table 6: OLS Regression of 2D:4D on Expectations minus Actual Performance

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

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

Table 7: Ordered Logit Regression of 2D:4D on Under/Over-confidence

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

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

Table 8: OLS Regression of Under/Over-confidence on Actual Earnings

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

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

Figures

Figure 1: Self-Confidence measure: Frequency

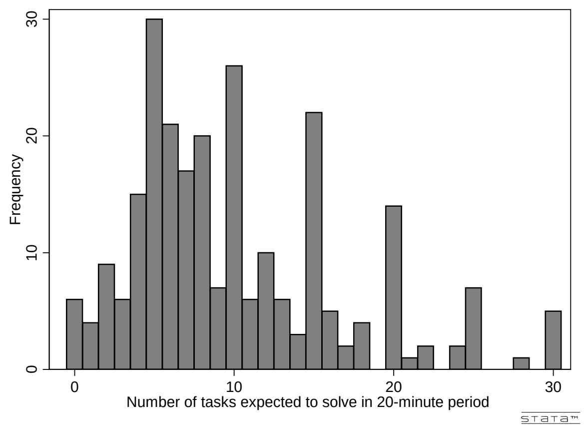
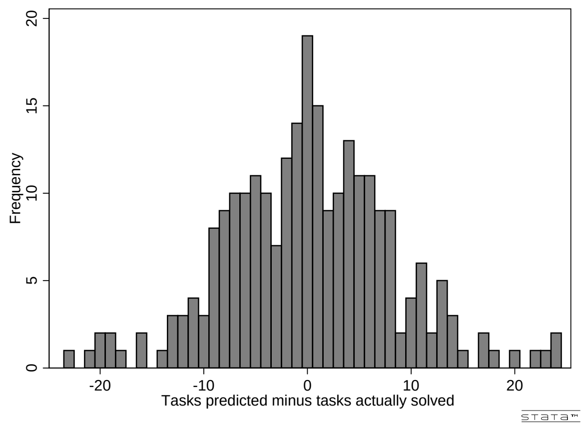


Figure 2: Prediction minus Actual Performance



Appendix

A Session Instructions

[As Subjects (Ss) arrive, the experimenter welcomes them, hands them an ID card and invites them to sit in the computer desk corresponding to the ID. Tell them to wait and not to login in the computer. Once everyone is sat, the experimenter asks the students to read the information sheet and consent forms and sign up if they agree. The experimenter reads the following:]

Pre-instructions

Welcome to this research on individual decision making. My name is xxx. These are my assistants xxx and xxx who are going to help me in this research project today. As you were informed in the recruitment process, you will be asked to perform multiple rounds of a task and fill in a questionnaire. Your total estimated participation time for today is 60 minutes. Before we begin, please read the information sheet and consent form that are placed on your desk, and put your signature on the consent form if you agree to participate. Once you are done, raise your hand and one of our assistants will come to your desk to collect the form.

[The Consent form notes that if they stay in the room, they are agreeing to participate. If Ss refse to participate, then pay the show-up fee and send them on their way. Once all the consent forms have been collected the experimenter reads the following instruction at loud.]

Thank you for agreeing to participate.

[the experimenter activates the login page]

Now you will see the login page on your screen. Please use the computer ID card you have received from us to login. Once you are logged in, you will see a set of instructions. Once you finish reading the instructions, you will have to click on the tab ”play the game” at the top of the page. From now on you will advance the session through your own input on the screen and you will not receive any further oral instruction from me. So please follow the instructions on the screen very carefully. If you have any questions at any point please raise your hands and one of us will come to you to assist you privately.

Instructions

Today's session is part of a research project at the University of Warwick, Economics Department. You will receive £ 5 as a participation fee as well as additional earnings depending on your individual performance during this session. Your total earnings will be paid to you privately in cash at the end of this session. From now on, you are requested not to communicate with the other participants. If you have any questions, please raise your hand and one of us will answer your questions privately. Please do not ask them out aloud.

You will be asked to perform multiple repetitions of a task that will appear on your computer screen.

In the first part of this session, you will have the opportunity to practice this task in order to get familiar with it. You will be given 1 minute of practice time.

After your practice time is over, we will ask you to complete this task as many times as you can in 20 minutes. Before you do so, we will ask you to predict, as accurately as you can, the number of times you think you can successfully complete the task in 20 minutes.

Your final compensation will be determined by your participation fee £ 5 plus additional earnings based on your total points, where every 100 points are worth £ 1.

The total points you earn will be equal to your performance points minus a prediction error penalty, where:

Performance points: 100 points for each time that you successfully complete the task in the 20 minutes allotted.

Prediction error penalty: A penalty of 40 points each for each completion above or below your prediction. For instance, if your actual performance exceeds – or falls short – of your predicted number of completions by say 5 times, your prediction penalty will be 200 (or 5 x 40) points in either case. Therefore, the best way to maximize your earnings is to predict your performance as accurately as you can.

Clearly, your prediction may not be exactly right. If, while performing the task, you reach your predicted number and still have time left during the 20 minutes allocated, remember that each additional completion will still earn you a net of 60 additional points (100 performance points for the additional completion - 40 penalty points for being above your prediction).

You must fully complete the task each time before you can start a new one. Our central computer will record your total number of completions in the time allowed and will use this number, together with your initial prediction, to compute the total amount due to you. This amount will be paid in full at the end of the session.

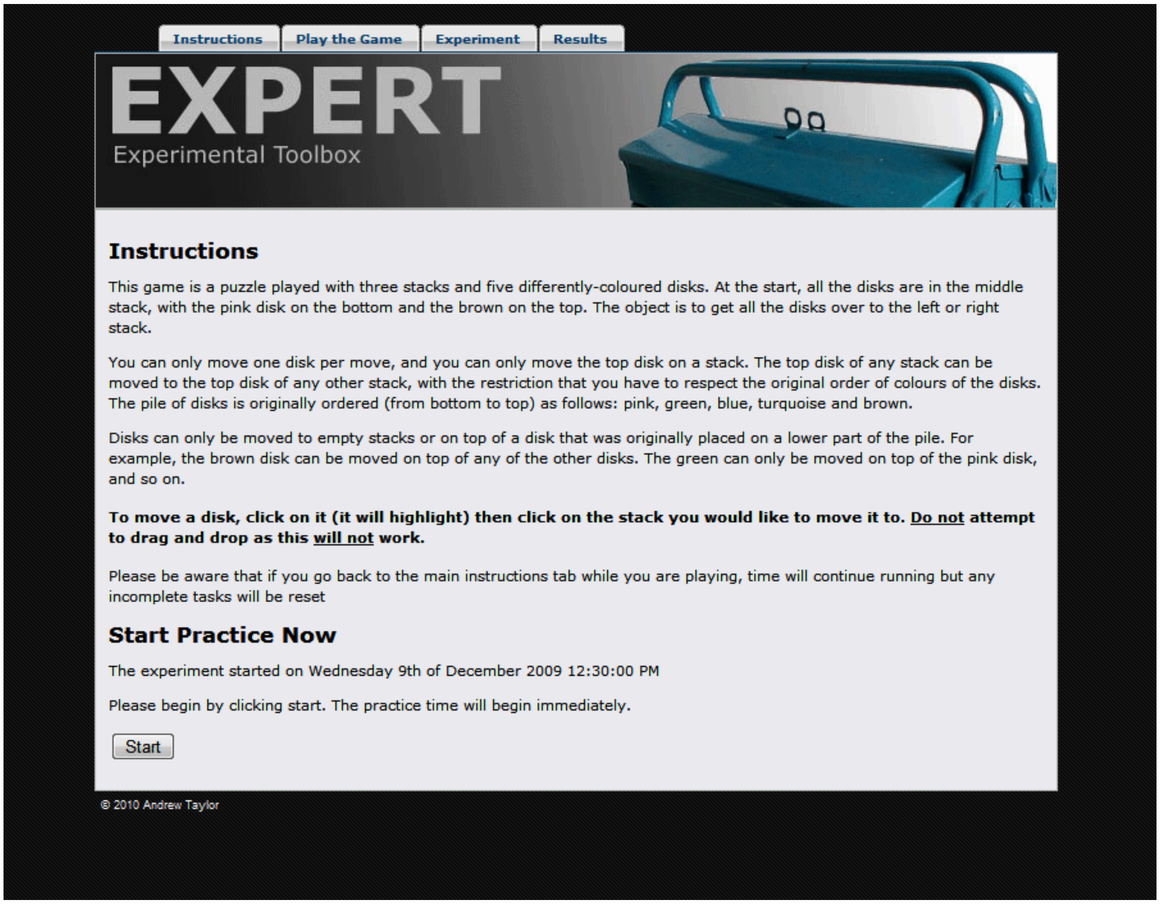
If you have any doubts please raise your hand now. One of us will come to you and answer your question privately. Please don't ask at loud.

[After all concerns have been addressed privately, the experimenter continues reading the instructions:]

After the assigned 20 minutes are over, you will be requested to complete a questionnaire. Once you have completed it, please wait until you are called upon to collect your final payment at the front desk.

Now you can start with the session. Please click "Play the Game" tab at the top of this page and follow the instructions on the screen.

B Puzzle Instructions Screen


Instructions | Play the Game | Experiment | Results

EXPERT
Experimental Toolbox

Instructions

This game is a puzzle played with three stacks and five differently-coloured disks. At the start, all the disks are in the middle stack, with the pink disk on the bottom and the brown on the top. The object is to get all the disks over to the left or right stack.

You can only move one disk per move, and you can only move the top disk on a stack. The top disk of any stack can be moved to the top disk of any other stack, with the restriction that you have to respect the original order of colours of the disks. The pile of disks is originally ordered (from bottom to top) as follows: pink, green, blue, turquoise and brown.

Disks can only be moved to empty stacks or on top of a disk that was originally placed on a lower part of the pile. For example, the brown disk can be moved on top of any of the other disks. The green can only be moved on top of the pink disk, and so on.

To move a disk, click on it (it will highlight) then click on the stack you would like to move it to. Do not attempt to drag and drop as this will not work.

Please be aware that if you go back to the main instructions tab while you are playing, time will continue running but any incomplete tasks will be reset

Start Practice Now

The experiment started on Wednesday 9th of December 2009 12:30:00 PM

Please begin by clicking start. The practice time will begin immediately.

Start

© 2010 Andrew Taylor

C Puzzle Practice Screen

D Self-Efficacy Scale (Schwarzer and Jerusalem, 1995)

For each of the following ten statements indicate how true you think each statement is for you. (1 = not at all true, 2 = hardly true, 3 = moderately true, 4 = exactly true) (write your answer in the space left at the beginning of each statement)

1) .... I can always manage to solve difficult problems if I try hard enough.

2) .... If someone opposes me, I can find the means and ways to get what I want.

3) .... It is easy for me to stick to my aims and accomplish my goals.

4) .... I am confident that I could deal efficiently with unexpected events.

5) .... Thanks to my resourcefulness, I know how to handle unforeseen situations.

6) .... I can solve most problems if I invest the necessary effort.

7) .... I can remain calm when facing difficulties because I can rely on my coping abilities.

8) .... When I am confronted with a problem, I can usually find several solutions.

9) .... If I am in trouble, I can usually think of a solution.

10)........ I can usually handle whatever comes my way.

Scoring: Responses are made on a 4-point scale. Sum up the responses to all 10 items to yield the final composite score with a range from 10 to 40.