Laboratory Measure of Cheating Predicts School Misconduct*

Alain Cohn and Michel André Maréchal

May 2017

Abstract

Laboratory experiments provide insights into the drivers of cheating behaviour, but it is unclear to what extent cheating in the lab generalizes to the field. We conducted an experiment with middle and high school students to test whether a common laboratory measure of cheating predicts three types of school misconduct: (i) disruptiveness in class, (ii) homework non-completion, and (iii) absenteeism. We find that students who cheat in the experimental task are more likely to misbehave at school, suggesting that experimental measures of cheating generalize to rule violating behaviour in naturally occurring environments.

JEL classification: C93, K42

Keywords: Cheating, Honesty, Rule violation, Experiment, External validity, School Misbehaviour.

---

* Cohn: Booth School of Business, University of Chicago, 5807 S Woodlawn Ave, Chicago, IL 60637, U.S., email: alain.cohn(at)chicagobooth.edu. Maréchal: University of Zurich, Department of Economics, Bluemlisalpstrasse 10, CH-8006 Zurich, Switzerland, email: michel.marechal(at)econ.uzh.ch. We thank Jonas Addor and Joël Farronato for excellent research assistance and David Tannenbaum, John List, Kjell Salvanes (editor), and two anonymous referees for helpful comments. Financial support from the Gottlieb Duttweiler Institute is gratefully acknowledged. Correspondence to: Cohn and Maréchal.

1 Introduction

Cheating, misconduct and other forms of rule violating behaviour are pervasive problems in many important areas of social and economic life. Examples range from scandals in the business world (e.g., Volkswagen's recent emission fraud or interest and exchange rate manipulations in the financial industry) to rigged sport competitions (Duggan et al. 2002), rampant corruption in developing countries (Pande and Olken 2012; Banerjee et al. 2013), and student and teacher cheating (Jacob and Levitt 2003; Levitt and Lin 2015).

Given the prevalence and cost of dishonesty to society, a rapidly growing literature has emerged with the aim to provide a better understanding of the determinants of lying, cheating, and stealing (see Ariely 2012; Irlenbusch and Villeval 2015, and Shalvi et al. 2015 for recent reviews). Due to its clandestine nature, dishonest behaviour is typically difficult to measure reliably using observational field data (Zitzewitz 2012). As a consequence, the majority of empirical findings originates from controlled laboratory environments.[1]

A widely used experimental paradigm to measure cheating is to instruct subjects to perform a simple task of chance (e.g., flipping coins or rolling dice) and asking them to report their outcomes. Because the actual outcomes are not observed by the experimenter and only certain outcomes are rewarded, subjects face the temptation to increase their earnings by misreporting their outcomes without any risk of getting caught (e.g., Bucciol and Piovesan 2011; Shalvi et al. 2011; Fischbacher and Föllmi-Heusi 2013; Cohn et al. 2014; Abeler et al. 2016).[2] Although cheating cannot be detected at the individual level, researchers can measure cheating at the group level as the true distribution of the underlying random process is known. Moreover, because higher earnings are less likely

[1]See Pierce and Balasubramanian (2015) for a survey of the literature on dishonest behaviour that uses observational data and field experiments.

[2]Another common approach are interactive sender-receiver games where senders can increase their earnings by sending deceptive messages to the receiver (e.g., Gneezy 2005; Sutter 2009).

to be the result of chance, earnings claimed by individual subjects can serve as a proxy for their cheating behaviour. While this paradigm has been used extensively to study the determinants of dishonesty and rule violating behaviour, the extent to which the insights gained from the lab can be extrapolated to naturally occurring environments remains unclear. Common objections to the generalizability of lab experiments are that subjects make low-stakes decisions in artificial environments and that they know their behaviour is being recorded and analysed (Levitt and List 2007; Falk and Heckman 2009).

In this paper we investigate whether cheating in the lab predicts rule violating behaviour in the field. To this end, we matched a common laboratory measure of cheating with teacher evaluations of students' misbehaviour in school. We experimentally measured cheating by asking the students to toss ten coins in private and report their outcomes. Students only received financial rewards when reporting "heads," and thus had a financial incentive to misreport their outcomes for unsuccessful coin flips. Our measures of school misbehaviour are based on the US National Education Longitudinal Survey. Specifically, we asked teachers to assess their students along three dimensions: disruptiveness in class, non-completion of homework, and absenteeism. These measures of school misconduct are important as they have been shown to reliably predict future educational achievement and labour market outcomes (Segal 2013; Autor et al. 2015).[3] We expect the laboratory measure of cheating to be predictive of school misconduct because both cheating and school misconduct require people to break rules.

We found a positive and significant correlation between the laboratory measure of cheating and students' misbehaviour in school. This relationship remains strong after adding controls for age, gender, nationality, school level, parental education and cognitive ability. Our estimates indicate that the difference in school misbehaviour between students who claimed ten coins (presumably cheaters) and those who claimed five coins

[3] Disruptive and noncompliant behaviour in school also seem to matter for students' current academic performance as we found negative and significant correlations between students' self-reported grade point average (GPA) and the three measures of school misbehaviour (disruptiveness: $p = 0.001$, homework: $p < 0.001$, absenteeism: $p = 0.002$, Spearman tests).

(presumably honest individuals) is, on average, 0.53 standard deviations. For comparison, we observe the same gap in school misbehaviour between students whose cognitive abilities (i.e., crystallized intelligence) differ by 2.7 standard deviations. Together, these results suggest that the cheating paradigm from the lab provides an externally valid measure of rule violating behaviour in the field.

Our paper contributes to several strands of the literature. First, a growing number of studies combines lab and field data from the *same* subjects to examine the external validity of laboratory measures of behaviour.[4] For example, Karlan (2005) found that second-mover behaviour in a trust game correlates with the likelihood of loan repayment among participants of a microcredit program in Peru.[5] Using experimental measures of present bias, Sutter et al. (2013) show that more impatient children and adolescents are more likely to buy alcohol and cigarettes, are more likely to be obese, and are less likely to save money.[6] Our findings suggest that cheating in the lab provides a reliable indicator of rule violating behaviour in the field. Only a few studies analysed the relationship between rule violation in the lab and the field. Hanna and Wang (forthcoming) examined cheating in a sample of government nurses in India. They found that nurses who cheated more in a dice task also tended to show up at work less often. Cohn et al. (2015*b*) conducted a coin tossing experiment with inmates from a maximum-security prison. They found a positive correlation between claimed earnings from the coin tosses and misconduct in prison (e.g., illegal drug possession or aggression against guards and other inmates). However, the latter studies used rather unusual participants drawn from the extreme ends of the honesty distribution. Recently, Dai et al. (forthcoming) reported a die-rolling experiment with public transport passengers showing that the proportion of fully dishonest participants is higher among those who did not hold a valid ticket. It is

[4]See Camerer (2015) for an overview of experimental studies linking behaviour in the lab and field.

[5]Benz and Meier (2008), Carpenter and Myers (2010), Fehr and Leibbrandt (2011), Burks et al. (2015), and Cohn et al. (2015*a*) provide further evidence for positive associations between lab and field measures of prosociality.

[6]Meier and Sprenger (2010) show that experimentally elicited present bias is a reliable predictor of credit card borrowing.

reassuring that these papers provide evidence that is consistent with our study despite using different methods and subject pools.[7]

Second, our paper also speaks to a growing literature on school misconduct as manifestations of non-cognitive skills.[8] For example, Segal (2013) shows that students misbehaving in eighth grade are almost three times less likely to finish high school and have almost 10 percent lower earnings as adults relative to non-disruptive students. Bertrand and Pan (2013) found that behavioural problems in school are more prevalent among boys, especially if they grow up in single-mother households. This finding may explain the widening gender gap in academic achievement in the United States and other developed countries (Goldin et al. 2006; Becker et al. 2010; Fortin et al. 2015). We find that male students cheat significantly more and that this gender difference in the coin tossing task explains about one-fifth of the gender gap in school misbehaviour. Our paper also links to an emerging literature on the relationship between economic preferences and non-cognitive skills (Becker et al. 2012; Almlund et al. 2011). The identified relationship between cheating behaviour and school misconduct raises the possibility that intrinsic honesty and expressions of non-cognitive skills at school share a common underlying mechanism.

2 Design

We conducted a paper-and-pencil experiment with 162 students from eight classes in two Swiss public schools—one middle and one high school. Students were between 12 and 20 years old, and 43 percent of them were female. They were informed that their data

[7] List (2009) analysed a subsample of 17 sellers from open air markets for which he observed lab and field behaviour. He found that sellers who breached collusive agreements in contextualized lab experiments were also more likely to do so in the field. More recently, Potters and Stoop (2016) and Kröll and Rustagi (2017) find that subjects who cheat in the lab are also less likely to report "accidental" overpayments and are more likely to adulterate milk with water, respectively.

[8] Externalizing behaviour and misconduct in school are typically seen as expressions of non-cognitive skills and relate to personality traits such as agreeableness and conscientiousness (see Ehrler et al. 1999; Almlund et al. 2011).

will be treated confidentially and that we will not reveal their data to others, including their teachers and school authorities. The experiment took place in the classrooms in absence of the teachers. We set up a mobile laboratory and installed partition walls to shield subjects from sight and therefore ensure privacy (see Figure A.1 in the online appendix).[9] Although participation was voluntary, all students gave their consent to participate in the study. We ran the experiment simultaneously in all four classes at each school to avoid cross-talk between subjects.

In the first part of the experiment, we asked subjects some basic socio-demographic questions such as age, gender, nationality, and parental education (see Table 1 for descriptive statistics). In part two, we measured their cognitive ability using two short tests from Dohmen et al. (2010): the word fluency test and the symbol-digit correspondence test.[10] Both tests are related to working memory and processing speed, which is often part of the reason children thrive or struggle in school, but they measure distinct concepts of reasoning capability (Carroll 1993).[11] The word fluency test measures "crystallized intelligence" (ability to solve problems using knowledge and experience) by asking subjects to list as many different animals as possible within 90 seconds. Subjects received one point for each correct and unique animal named. The symbol-digit correspondence test measures "fluid intelligence" (innate ability to solve problems) and consists of decoding sets of unfamiliar symbols into single digits as fast as possible within 90 seconds. For each set, subjects had to write down the correct numbers under a grid of nine symbols using a predefined mapping between symbols and digits. Subjects scored one point for each correct symbol-digit pair.

The last part of the experiment comprised the coin tossing task—our laboratory

[9] We took these measures to mitigate potential confidentiality concerns. Such concerns could, in principle, lead to an overestimation of the relationship between our laboratory measure of cheating and misbehavior at school if the most well-behaved students in the class were more worried about data privacy. We thank an anonymous referee for pointing this out.

[10] The two cognitive ability tests are based on submodules of the Wechsler Adult Intelligence Scale (WAIS)—one of the most frequently used intelligence tests.

[11] Test scores are positively correlated in our sample (Spearman's rho $= 0.423, p < 0.001$).

Table 1: Descriptive statistics

Variable	Mean	SD
Age	14.938	2.015
Female	0.432	0.497
Swiss nationality	0.673	0.471
High school	0.488	0.501
Parental education	0.364	0.483
Crystallized intelligence	20.401	7.471
Fluid intelligence	43.370	10.331
Grade point average (self-reported)	4.632	0.535
Absenteeism	0.981	1.522
Disruptiveness	0.981	1.530
Homework non-completion	1.815	1.991
School misbehavior index	1.259	1.355

This table reports descriptive statistics. Age is measured in years. Female, Swiss nationality, High school, and Parental education are dummy variables. Parental education equals to one if at least one parent has a university degree. Crystallized and Fluid intelligence are based on the scores from the word fluency test and the symbol-digit correspondence test, respectively. Grade point average is the self-reported grade point average on a scale from 1 (worst) to 6 (best). Disruptiveness, Homework non-completion and Absenteeism are three measures of school misconduct, based on the teachers' assessments on a scale from "never misbehaves" (= 0) to "always misbehaves" (= 6). School misbehaviour index is the average of the three items of school misconduct. The number of observations is 162, except for age (N=161) because one subject did not state his age.

measure of cheating. Subjects first opened an envelope containing ten coins, each worth 0.5 Swiss francs (about US $0.55). Then, they were instructed to toss each coin in private and report their outcomes on paper. For every coin toss for which subjects reported the outcome "heads" they were allowed to keep the coin; they had to put the coin back into the envelope otherwise. Participants thus faced a financial incentive to cheat by misreporting the outcomes of their coin flips without any risk of getting caught.[12] The stakes were considerable as the maximum possible payoff in this task corresponds roughly to half the amount students of similar age receive in pocket money every week (e.g., see `www.budgetberatung.ch`). After completing the coin tossing task, subjects were asked to put their envelope with the remaining coins into a container.

[12]Nine subjects reported a lower number of heads than the number of coins they actually took out of the envelope. For our analysis, we use the number of coins taken as the outcome variable. Our results remain the same if we use the reported number of heads instead (see Table A.1 in the online appendix).

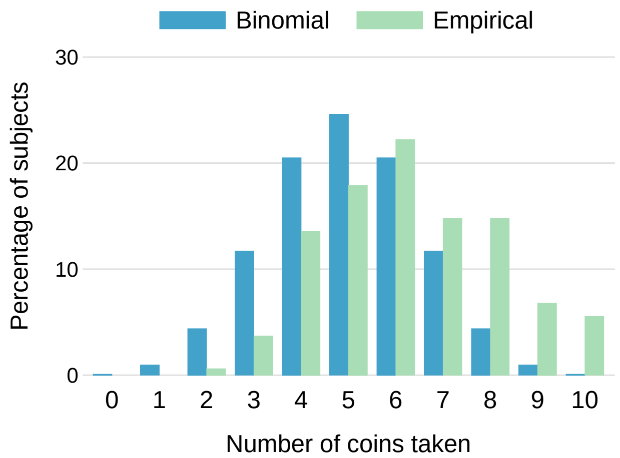
Teachers were asked to assess their students along three dimensions: disruptiveness in class, non-completion of homework, and absenteeism. For each item the teachers evaluated the students on a scale from "never misbehaves" (= 0) to "always misbehaves" (= 6). These measures of school misbehaviour were inspired by the US National Educational Longitudinal Survey—a study that followed a nationally representative sample of more than 20,000 students over several years. We chose these measures of school misbehaviour as they have been shown to reliably predict future educational achievement and labour market outcomes (Segal 2013; Autor et al. 2015). Because the three items are strongly correlated (Cronbach's $\alpha = 0.718$) we created an index of school misbehaviour using the unweighted average of all three items. Our regression analysis uses the school misbehaviour index to reduce the influence of measurement error, but we also report the results using the three measures of misbehaviour separately (see Table A.2 in the online appendix). We matched teachers' evaluations with the experimental data using identification codes to preserve subjects' anonymity.

3 Results

The results indicate that a significant proportion of the subjects cheated by inflating their number of successful coin tosses. Figure 1 shows that the empirical distribution of coins taken is shifted towards a higher number relative to the honest benchmark provided by the binomial distribution. The outcomes ten, nine, and eight coins are significantly overrepresented ($p < 0.001$ for all three outcomes, binomial tests), whereas two, three, four, and five coins are significantly underrepresented ($p = 0.011$, $p < 0.001$, $p = 0.032$, and $p = 0.055$, binomial tests). On average, the students took 62.8% of the coins in the envelopes (95% confidence interval: 60.0%, 65.7%).[13] Assuming that none of the participants cheated to his or her disadvantage we estimate that 25.7% of the coins were

[13]If we use reported outcomes instead, the percentage of heads is 61.6% (95% confidence interval: 58.9%, 64.3%).

Figure 1: Students' behaviour in the coin tossing task



The figure indicates that a significant proportion of students cheated in the coin tossing task. The empirical distribution of coins taken (green bars) is shifted towards higher earnings relative to the binomial distribution implied by fully honest behaviour (blue bars).

misreported.[14]

We also analysed individual determinants of cheating using multivariate regression analysis. Higher earnings are less likely to be the result of chance. Thus, we use the number of coins each subject took as a proxy for cheating in the regression analysis. Column (1) of Table 2 indicates that female students behaved more honestly than male students as they took significantly less coins ($p < 0.000$, t-test).[15] Moreover, we found that high school students took significantly less coins than those from middle school after controlling for age ($p = 0.011$, t-test), which could be explained by less deviant students selecting into higher education. Earnings in the coin tossing task and the two measures of cognitive ability are negatively correlated. However, the correlations do not reach

[14]The calculation of percentage of misreported coin tosses is straightforward if we assume that none of the participants cheated to his or her disadvantage (see Houser et al. 2012). Let h be the percentage of coins taken from the envelopes and m be the percentage of misreported coin tosses. For any given coin toss, a participant who cheats keeps it with a probability of 1. By contrast, a participant who is truthful keeps each coin with a probability of 0.5. Thus, the percentage of coins taken from the envelope is $h = m * 1 + (1 - m) * 0.5 = 0.5 * (1 + m)$. Solving the equation yields the percentage of misreported coin tosses $m = 2 * h - 1$.

[15]Dreber and Johannesson (2008) document a similar gender difference in dishonest behaviour.

Table 2: Determinants of Behaviour in the Coin Task and School Misbehaviour

	(1)	(2)	(3)	(4)
Dependent variable	# of coins	School Misbehaviour Index		
# of coins taken		0.150**	0.145**	
		(0.015)	(0.015)	
Age	-0.038	0.489***	0.472***	0.467***
	(0.731)	(0.006)	(0.007)	(0.008)
Female	-1.061***	-0.621**	-0.663**	-0.817**
	(0.000)	(0.015)	(0.010)	(0.011)
Swiss nationality	-0.411	0.143	0.186	0.127
	(0.140)	(0.226)	(0.206)	(0.440)
High school	-1.018*	-1.360***	-1.033*	-1.181**
	(0.050)	(0.008)	(0.054)	(0.030)
Parental education	-0.120	0.462	0.532	0.514
	(0.657)	(0.190)	(0.154)	(0.163)
Crystallized intelligence	-0.080		-0.267**	-0.279**
	(0.599)		(0.044)	(0.035)
Fluid intelligence	-0.041		0.040	0.034
	(0.744)		(0.771)	(0.808)
Constant	8.145***	-6.321***	-6.239***	-5.056**
	(0.000)	(0.004)	(0.005)	(0.020)
Observations	161	161	161	161
R^2	0.226	0.310	0.333	0.303

This table reports OLS coefficient estimates. p-values are reported in parenthesis. In column (1), we regress the number of coins taken in the coin tossing task on a set of individual characteristics and two measures of cognitive ability. Age is measured in years. Female, Swiss nationality, High school, and Parental education are dummy variables. Parental education equals to one if at least one parent has a university degree. Crystallized and Fluid intelligence are based on the scores from the word fluency test and the symbol-digit correspondence test, respectively. Both cognitive ability measures are normalized to have a mean of zero and a standard deviation of one. In columns 2 to 4, the dependent variable is the School misbehaviour index, which is constructed by averaging the three items of school misconduct, including disruptiveness in class, failure to complete homework, and absenteeism (all measured on a scale from "never misbehaves" (= 0) to "always misbehaves" (= 6)). Because the models in columns 2 to 4 use teacher evaluations, we computed p-values that are robust to clustering at the class level. To account for the low number of clusters we applied the wild cluster bootstrap procedure (Cameron et al. 2008) using Webb's (2013) 6-point distribution of weights (see online appendix for a description of the procedure). The number of observations is 161 instead of 162 because one subject did not state his age. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

statistical significance, neither for crystallized nor for fluid intelligence ($p = 0.599$ and $p = 0.744$, t-tests).

We next examined whether our experimental measure of cheating is related to school misconduct. Panels (a) to (c) in Figure 2 illustrate the average scores for the three measures of school misconduct for subjects who took more than five coins (i.e., subjects who presumably cheated) and those who took five coins or less.[16] Together, the three panels highlight that behaviour in the coin tossing task is positively associated with each measure of school misbehaviour. Subjects who took more than five coins score 0.5 points (or 72 percent) higher on disruptiveness in class, 0.9 points (or 69 percent) higher on non-completion of homework, and 0.4 points (or 61 percent) higher on absenteeism relative to the other subjects. Using the raw data, we find statistically significant correlations between the number of coins taken and disruptiveness and homework non-completion ($p = 0.003$ and $p = 0.020$), but the correlation with absenteeism fails to reach statistical significance ($p = 0.136$, Spearman tests).

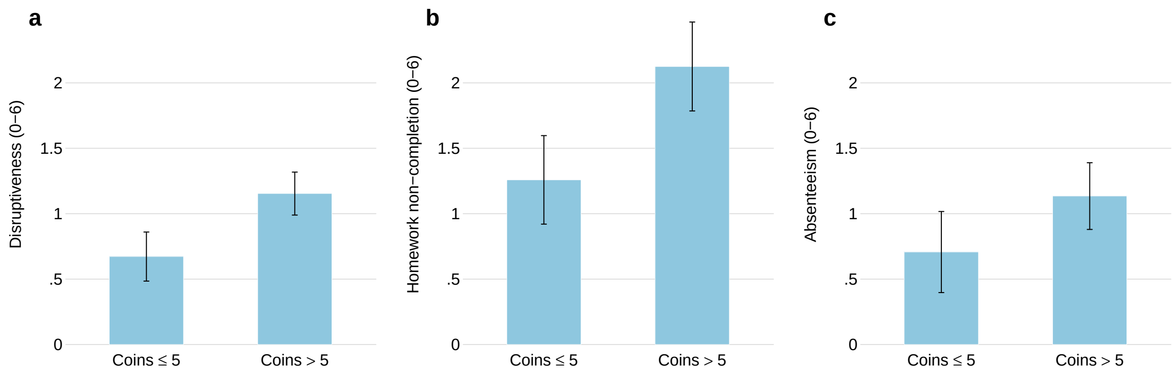
We additionally estimated regression models to control for factors that might jointly influence cheating and school misbehaviour. In the regression analysis, we use the school misbehaviour index, which is the average score of all three individual measures of school misbehaviour (see Figure A.3 in the online appendix for a graph depicting the distribution of the school misbehaviour index). Our main results are similar if we analyse each measure of school misbehaviour separately (see Table A.2 in the online appendix).

Column (2) of Table 2 confirms that behaviour in the coin tossing task is significantly related to school misbehaviour when controlling for age, gender, nationality, education level, and parental education. A higher number of coins taken is associated with increased behavioural problems in school ($p = 0.015$).[17] Interestingly, in addition to pocketing a

[16] Five coins corresponds to the median number of claims. Alternatively, Figure A.2 in the online appendix illustrates that there is a monotonic relationship for all three measures of school misbehaviour when the data is split by tertiles of coins taken.

[17] We computed p-values that are robust to clustering at the class level. To account for the low number of clusters we applied the wild cluster bootstrap procedure (Cameron et al. 2008) using Webb's (2013) 6-point distribution of weights (see online appendix for a description of the procedure).

Figure 2: Behaviour in the Coin Task and School Misconduct



The figure shows that, relative to those who took five coins or less (i.e. five coins corresponds to the median), students who took more than five coins (i.e., those who presumably cheated to a greater extend) disrupt the class to a larger degree (a), fail to do their homework more often (b), and are more frequently absent from school (c). Error bars indicate the standard error of the mean (adjusted for clustering at the class level).

lower number of coins, female and high school students also misbehave less frequently ($p = 0.015$ and $p < 0.008$, respectively). The model reported in column (3) additionally controls for cognitive ability to address potential issues of third variables that correlate with both school misbehaviour and dishonesty.[18] We find that crystallized intelligence is negatively associated with school misbehaviour ($p = 0.044$), but fluid intelligence is not ($p = 0.771$). While differences in cognitive ability explain some variation in disruptive and noncompliant behaviour, the predictive power of the coin tossing task for school misbehaviour remains high after controlling for key background characteristics as well as cognitive ability ($p = 0.015$). The coefficient estimate implies that the difference in school misbehaviour between students who took ten coins (presumably cheaters) and those who took five coins (presumably honest individuals) is more than 0.7 points (or 0.53 standard deviations) on average. For comparison, it would require students to differ by 2.7 standard deviations in cognitive ability (i.e., crystallized intelligence) to produce the same difference in school misbehaviour. The difference in school misbehaviour be-

[18]For example, Ruffle and Tobol (2017) and Deckers et al. (2016) found negative associations between cognitive ability and immoral behaviour.

tween presumable cheaters and honest students is also larger than the widely discussed gender gap in misbehaviour (e.g., Bertrand and Pan 2013). In column (4) of Table 2 we removed our laboratory measure of cheating from the regression model and found that the gender coefficient increases from -0.663 to -0.817. This suggests that gender differences in experimentally elicited rule violating behaviour explain almost one-fifth of the gender gap in school misbehaviour.[19]

4 Conclusion

In this paper, we examined whether a common laboratory measure of cheating is a reliable predictor of rule violating behaviour in the field. We present evidence on the link between rule violating behaviour in the lab and field using middle and high school students. We combined experimental data from an incentivized coin tossing task with measures of disruptive and noncompliant behaviour at school. Our main result is that students who presumably cheated more in the coin tossing task also misbehave more often at school. The relationship holds when controlling for students' socioeconomic background and cognitive ability.

Our findings contribute to the active debate about the generalizability of laboratory experiments, i.e., whether data obtained in the lab can be extrapolated to naturally occurring environments (Levitt and List 2007; Falk and Heckman 2009). We find a significant relationship between lab and field measures of rule violating behaviour despite differences across the two settings, including the context of the choice situation and the degree of scrutiny—factors which have been argued to make inferences from lab to field environments difficult. Our findings concur with very recent results from studies that document positive correlations between lab and field measures of dishonesty (Potters and Stoop 2016, Dai et al. forthcoming, and Kröll and Rustagi 2017). The fact that these

[19] We found very similar results using the pooled Blinder-Oaxaca decomposition method—a technique that was initially developed for studying gender gaps in labour market earnings (Blinder 1973; Oaxaca 1973).

correlations emerge from independent studies that use different methods and subject pools is reassuring for the usefulness of laboratory measures of behaviour, especially cheating behaviour, as cheating has been conjectured to be more context-sensitive than other types of behaviour, such as cooperativeness and consumption choices (Abeler et al. 2014).

In a broader sense, our paper also adds to a nascent literature on the relationship between economic preferences and non-cognitive skills (Becker et al. 2012; Almlund et al. 2011). Our results raise the possibility that intrinsic honesty and expressions of non-cognitive skills at school share a common underlying mechanism.

References

Abeler, Johannes, Anke Becker, and Armin Falk (2014): Representative Evidence on Lying Costs, *Journal of Public Economics*, Vol. 113, No. C, pp. 96–104.

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond (2016): Preferences for Truth-Telling, *University of Oxford Working Paper*.

Almlund, Mathilde, Angela Lee Duckworth, James J Heckman, and Tim D Kautz (2011): Personality Psychology and Economics, in: Stephen Machin Eric A. Hanushek and Ludger Woessmann (Eds.), Handbook of the Economics of Education, Vol. 4, pp. 1–181, Amsterdam: Elsevier.

Ariely, Dan (2012): The (Honest) Truth about Dishonesty: How we Lie to Everyone–Especially Ourselves, New York: Harpercollins.

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman (2015): Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes, *Northwestern University Working Paper Series*.

Banerjee, Abhijit, Rema N. Hanna, and Sendhil Mullainathan (2013): Corruption, in: Robert Gibbons and John Roberts (Eds.), Handbook of Organizational Economics, pp. 1109–1147, Princenton: Princeton University Press.

Becker, Anke, Thomas Deckers, Thomas Dohmen, Armin Falk, Fabian Kosse, et al. (2012): The Relationship Between Economic Preferences and Psychological Personality Measures, *Annual Review of Economics*, Vol. 4, No. 1, pp. 453–478.

Becker, Gary S, William HJ Hubbard, Kevin M Murphy, et al. (2010): The Market for College Graduates and the Eorldwide Boom in Higher Education of Women, *American Economic Review*, Vol. 100, No. 2, pp. 229–233.

Benz, Matthias and Stephan Meier (2008): Do People Behave in Experiments as in the Field? Evidence from Donations, *Experimental Economics*, Vol. 11, No. 3, pp. 268–281.

Bertrand, Marianne and Jessica Pan (2013): The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior, *American Economic Journal. Applied Economics*, Vol. 5, No. 1, pp. 32–64.

Blinder, Alan S (1973): Wage Discrimination: Reduced Form and Structural Estimates, *Journal of Human Resources*, Vol. 8, No. 4, pp. 436–455.

Bucciol, Alessandro and Marco Piovesan (2011): Luck or Cheating? A Field Experiment on Honesty with Children, *Journal of Economic Psychology*, Vol. 32, No. 1, pp. 73–78.

Burks, Stephen V, Daniele Nosenzo, Jon Anderson, Matthew Bombyk, Derek Ganzhorn, Lorenz Goette, Aldo Rustichini, et al. (2015): Lab Measures of Other-Regarding Preferences Can Predict Some Related on-the-Job Behavior: Evidence from a Large Scale Field Experiment, University of Nottingham Working Paper CeDEx 2015-21.

Camerer, Colin F (2015): The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Reply to Levitt and List, in: G Fréchette and A Schotter (Eds.), The Methods of Modern Experimental Economics, 207-248, Oxford University Press Oxford, UK.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller (2008): Bootstrap-based Improvements for Inference with Clustered Errors, *Review of Economics and Statistics*, Vol. 90, No. 3, pp. 414–427.

Carpenter, Jeffrey and Caitlin Knowles Myers (2010): Why Volunteer? Evidence on the Role of Altruism, Image, and Incentives, *Journal of Public Economics*, Vol. 94, No. 11, pp. 911–920.

Carroll, John B (1993): Human Cognitive Abilities: A Survey of Factor-analytic Studies, Cambridge University Press.

Cohn, Alain, Ernst Fehr, and Lorenz Goette (2015*a*): Fair Wages and Effort Provision: Combining Evidence from a Choice Experiment and a Field Experiment, *Management Science*, Vol. 61, No. 8, pp. 1777–1794.

Cohn, Alain, Ernst Fehr, and Michel A. Maréchal (2014): Business Culture and Dishonesty in the Banking Industry, *Nature*, Vol. 516, No. 7529, pp. 86–89.

Cohn, Alain, Michel A. Maréchal, and Thomas Noll (2015*b*): Bad Boys: How Criminal Identity Salience Affects Rule Violation, *Review of Economic Studies*, Vol. 82, No. 4, pp. 1289–1308.

Dai, Zhixin, Fabio Galeotti, and Marie Claire Villeval (forthcoming): Cheating in the Lab Predicts Fraud in the Field: An Experiment in Public Transportation, *Management Science*.

Deckers, Thomas, Falk Armin, Fabian Kosse, and Nora Szech (2016): Homo Moralis: Personal Characteristics, Institutions, and Moral Decision-making, IZA Discussion paper No. 9768.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2010): Are Risk Aversion and Impatience Related to Cognitive Ability?, *American Economic Review*, Vol. 100, No. 3, pp. 1238–60.

Dreber, Anna and Magnus Johannesson (2008): Gender Differences in Deception, *Economics Letters*, Vol. 99, No. 1, pp. 197–199.

Duggan, Mark, Steven D Levitt, et al. (2002): Winning Isn't Everything: Corruption in Sumo Wrestling, *American Economic Review*, Vol. 92, No. 5, pp. 1594–1605.

Ehrler, David J, J Gary Evans, and Ron L McGhee (1999): Extending Big-Five Theory into Childhood: A Preliminary Investigation into the Relationship Between Big-Five Personality Traits and Behavior Problems in Children, *Psychology in the Schools*, Vol. 36, No. 6, pp. 451–458.

Falk, Armin and James J Heckman (2009): Lab experiments are a major source of knowledge in the social sciences, *Science*, Vol. 326, No. 5952, pp. 535–538.

Fehr, Ernst and Andreas Leibbrandt (2011): A Field Study on Cooperativeness and Impatience in the Tragedy of the Commons, *Journal of Public Economics*, Vol. 95, No. 9, pp. 1144–1155.

Fischbacher, Urs and Franziska Föllmi-Heusi (2013): Lies in Disguise - An Experimental Study on Cheating, *Journal of the European Economic Association*, Vol. 11, No. 3, pp. 525–547.

Fortin, Nicole M, Philip Oreopoulos, and Shelley Phipps (2015): Leaving Boys Behind: Gender Disparities in High Academic Achievement, *Journal of Human Resources*, Vol. 50, No. 3, pp. 549–579.

Gneezy, Uri (2005): Deception: The Role of Consequences, *American Economic Review*, Vol. 95, No. 1, pp. 384–394.

Goldin, Claudia, Lawrence F Katz, and Ilyana Kuziemko (2006): The Homecoming of American College Women: The Reversal of the College Gender Gap, *Journal of Economic Perspectives*, Vol. 20, No. 4, pp. 133–156.

Hanna, Rema and Shing-Yi Wang (forthcoming): Dishonesty and Selection into Public Service, *American Economic Journal: Economic Policy*.

Houser, Daniel, Stefan Vetter, and Joachim Winter (2012): Fairness and Cheating, *European Economic Review*, Vol. 56, No. 8, pp. 1645–1655.

Irlenbusch, Bernd and Marie Claire Villeval (2015): Behavioral ethics: how psychology influenced economics and how economics might inform psychology?, *Current Opinion in Psychology*, Vol. 6, pp. 87–92.

Jacob, Brian A and Steven D Levitt (2003): Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating, *Quarterly Journal of Economics*, Vol. 118, No. 3, pp. 843–877.

Karlan, Dean S. (2005): Using Experimental Economics to Measure Social Capital and Predict Financial Decisions, *American Economic Review*, Vol. 95, No. 5, pp. 1688–1699.

Kröll, Markus and Devesh Rustagi (2017): Got Milk? Dishonesty, Market Power, and Cheating in Informal Milk Markets in India, Working paper.

Levitt, Steven D and Ming-Jen Lin (2015): Catching Cheating Students, NBER Working Paper No. 21628.

Levitt, Steven D and John A List (2007): What do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?, *Journal of Economic Perspectives*, Vol. 21, No. 2, pp. 153–174.

List, John A (2009): The economics of open air markets, NBER Working Paper No. 15420.

Meier, Stephan and Charles Sprenger (2010): Present-biased Preferences and Credit Card Borrowing, *American Economic Journal: Applied Economics*, Vol. 2, No. 1, pp. 193–210.

Oaxaca, Ronald (1973): Male-Female Wage Differentials in Urban Labor Markets, *International Economic Review*, Vol. 14, No. 3, pp. 693–709.

Pande, R and B Olken (2012): Corruption in Developing Countries, *Annual Review of Economics*, Vol. 4, No. 1, pp. 479–509.

Pierce, Lamar and Parasuram Balasubramanian (2015): Behavioral Field Evidence on Psychological and Social Factors in Dishonesty and Misconduct, *Current Opinion in Psychology*, Vol. 6, pp. 70–76.

Potters, Jan and Jan Stoop (2016): Do Cheaters in the Lab also Cheat in the Field?, *European Economic Review*, Vol. 87, No. C, pp. 26–33.

Ruffle, Bradley J. and Yossef Tobol (2017): Clever Enough to Tell the Truth, *Experimental Economics*, Vol. 20, No. 1, pp. 130–155.

Segal, Carmit (2013): Misbehavior, Education, and Labor Market Outcomes, *Journal of the European Economic Association*, Vol. 11, No. 4, pp. 743–779.

Shalvi, Shaul, Jason Dana, Michel J.J. Handgraaf, and Carsten DeDreu (2011): Justified Ethicality: Observing Desired Counterfactuals Modifies Ethical Perceptions and Behavior, *Organizational Behavior and Human Decision Processes*, Vol. 115, No. 2, pp. 181–190.

Shalvi, Shaul, Francesca Gino, Rachel Barkan, and Shahar Ayal (2015): Self-Serving Justifications Doing Wrong and Feeling Moral, *Current Directions in Psychological Science*, Vol. 24, No. 2, pp. 125–130.

Sutter, Matthias (2009): Deception Through Telling the Truth?! Experimental Evidence From Individuals and Teams, *Economic Journal*, Vol. 119, No. 534, pp. 47–60.

Sutter, Matthias, Martin Georg Kocher, Stefan T Trautmann, et al. (2013): Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior, *American Economic Review*, Vol. 103, No. 1, pp. 510–31.

Webb, D., Matthew (2013): Reworking Wild Bootstrap Based Inference for Clustered Errors, Queen's Economics Department Working Paper No. 1315.

Zitzewitz, Eric (2012): Forensic Economics, *Journal of Economic Literature*, Vol. 50, No. 3, pp. 731–769.