Children's rationality, risk attitudes and misbehavior

Marco Castillo*, Jeffrey L. Jordan**, Ragan Petrie*

*Interdisciplinary Center for Economic Science (ICES), George Mason University, Fairfax, VA 22030 USA
Castillo: mcastil8@gmu.edu, Petrie: rpetrie1@gmu.edu

**Agricultural and Applied Economics, University of Georgia, Griffin, GA 30223 USA
Jordan: jjordan@uga.edu

July 2015

JEL codes: C91, D81, J13, J18
Keywords: rational behavior, risk preferences, children, economic experiments, misbehavior

Abstract: We investigate the relationship between rational behavior, risk attitudes and the field choices of children by conducting economic experiments with 1,275 8th graders. We find a propensity for children to act irrationally independent of the theory of choice used as a benchmark. However, the hypothesis that children's behavior is simply random choice is rejected. They respond to their underlying preferences, albeit in a distorted manner. In addition, measured risk preferences that account for decision error predict future field behavior. Children who are more risk seeking are more likely to have disciplinary referrals one and two years after the experiment, even controlling for rationality, family background, scholarly achievement and past misbehavior. The magnitude of the effect of risk attitudes on field behavior is comparable to that of sex, race and family structure.

1. Introduction

According to the Center for Disease Control (CDC), 29,551 individuals aged 15-24 died in 2010. Of those deaths, 20.7 percent were due to accidental discharge of firearms, accidental poisoning, exposure to noxious substances, and assault. By comparison, the frequency of these causes of death in the rest of the population is 1.5 percent. As reported by childrenstats.gov, the percent of 8th graders who admit using drugs is 7.7, and of the 47.4 percent who say they have had sex, three-quarters did not use contraceptives and 40 percent did not use a condom. While many of these behaviors may be viewed as imprudent, some could be considered rational if, for example, the returns to illegal activities are high or a child is willing to take risks. We investigate this type of behavior more generally by asking two questions: (1) Are children's choices over risky outcomes rational, as defined by economic theory? and (2) Do risk preferences predict field behavior? We address these by collecting experimental and field behavior on 1,275 8th graders.

Several reasons motivate studying the nature of children's preferences and their relationship with field behavior. First, children make many decisions on their own and these decisions have consequences for future economic outcomes. Recent studies from labor economics show that misbehavior during childhood and adolescence have long term consequences on earnings (Segal, 2013; Heckman et al., 2006). Second, economic theory (Freeman, 1999) suggests that children's misbehavior must respond to both the relative costs to human capital accumulation and their preferences. Third, measurement error may be important when relating experimental estimates of preferences to outcomes, and misbehavior can be used to test the extent of the error and a methodology to correct for it.

We attempt to improve our understanding of risk preferences, rationality and misbehavior with several methods of collecting data. Information on the child's household environment is gathered with a survey, past and future misbehavior come from school records and risk preferences are measured using incentivized economic experiments. Measuring preferences with experiments is advantageous because this allows us to observe the behavior of children from various backgrounds over identical choice sets. Similarly, data on household environment are potentially informative of the barriers some children might face in attempting to accumulate human capital.[1] For instance, children in single parent households might face lower costs to misbehaving. An important component of our study is a brief survey on family structure completed by children participating in the experiment. This drastically reduces the loss of information due to non-response from mailed surveys to parents and allows us to control for variables that might be correlated with behavior in the experiment and the field. To control for the influence of cognitive and non-cognitive abilities (Heckman et al., 2006), we use data from school records on standardized test scores and disciplinary referrals prior to the experiment.[2]

We examine field behavior by looking at the effect of children's risk preferences on disciplinary referrals measured in the future. Field behavior is likely to be influenced by many factors that are difficult to measure. The controls mentioned above reduce the potential of an omitted variable problem.

Before relating measures of risk preferences to field behavior, however, it is crucial to determine what these experimental measures of risk do reveal. It would be inappropriate to interpret behavior in experiments as a measure of preferences if the behavior is not rational and merely reflects random choice. Even if children possess well-behaved preferences over risky prospects, they might be distorted in experiments by execution mistakes or inattention. This measurement error could make it difficult to detect a relationship between preferences and field behavior even if one exists. To overcome these problems, our experimental design is a series of lottery decisions that are structured specifically to measure rational behavior. The generated data permit

[1]While personality might also be important in explaining behavior, we do not have these measures. Instead, we control for household environment and past misbehavior.

[2]Achievement test are correlated with personality traits as well (see Borghans et al. (2011)). Niederle and Vesterlund (2010) suggest that they might also be related to competitive attitudes. Disciplinary referrals and test scores are themselves imperfect measures of cognitive and non-cognitive abilities (Heckman et al., 2006). They might then be imperfect controls for their influence on future behavior. A more complete approach would take into consideration the measurement error in risk preferences as well as skills (Cunha et al., 2010). We examine measurement error in the former but not the latter.

us to identify whether children's choices can be rationalized by theories of behavior under risk and to what extent this behavior is noisy. For instance, our experiments allow us to test if a person is consistent with expected utility theory, but also with less restrictive models of behavior such as Neilson (1992a), Diecidue et al. (2004), and cumulative prospect theory.

We have several findings. First, children's decisions systematically deviate from the theories of decision making under risk we consider, including expected utility, as well as, non-expected utility theory. While we find evidence of departures from these theories, we reject the hypothesis that children's decisions are simply random. For instance, the proportion of children adhering to expected utility theory is almost three times larger than what purely random behavior would predict. Similarly, the proportion of children adhering to a version of cumulative prospect theory is 40 percent larger than random choice. Using Harless and Camerer (1994)'s approach, we find that, once we allow for errors in decision making and the parsimony of alternative theories, expected utility theory performs no worse than other models (Vuong, 1989).

Using this finding, we then introduce a new estimate of risk preferences that removes some of the measurement error due to decision mistakes. Assuming an expected utility model of decision making, we structurally estimate the conditional probability that a child has a certain set of preferences given the pattern of errors in the population and the actual choices made by the child.[3] This approach provides a measure of preferences that is cleaned of decision error and we use to test for rational behavior and correlate with misbehavior.

Second, we find some heterogeneity in rational behavior and risk preferences, although the patterns are not particularly strong or robust across the various theories of decision making we consider. Children who are older within the 8th grade cohort and whose fathers only finished high school are more likely to choose at random, and children who are older and with lower standardized reading tests are less likely to behave risk aversely.

[3]We test other models as well but none is stable in the specification that controls for covariates.

Third, regarding children's field behavior, we find that boys, black children, children living with a single or no parent, and children eligible for free or reduced price school meals have more disciplinary referrals. The opposite is true for children who are first borns and have higher standardized math scores. Our estimates are robust to the inclusion of past disciplinary referrals. These results are consistent with previous research on children's behavior (Bertrand and Pan, 2013; Segal, 2013; Currie and Tekin, 2006) and demonstrates that our sample is not atypical.

Fourth, as predicted by economic theory, a child's propensity to choose safe lotteries is negatively correlated with future disciplinary referrals. This holds even controlling for a measure of rationality, past disciplinary referrals, children's scholastic performance and family background. This is important because it shows that the risk measure captures other traits of the child. The misbehavior being explained occurred one and two years after risk preferences were measured, so our results show a relationship between a preference measure and future behavior.[4] In addition, our results show that experimental measures can be clouded by measurement error.[5] Both an aggregate measure of lottery choices and each separate lottery decision are not significantly correlated with field behavior, but our cleaned measure of risk preferences is. Because this is an estimate, we test if our results disappear when we account for sampling error by using bootstrap methods. They do not. In sum, our results show that experimental measures, properly estimated, can be externally valid, correlate with behavior up to two years in the future and partly explain heterogeneity in behavior.

Very little is known about the rationality of children's decisions under risk. Harbaugh et al. (2002) examine risk attitudes of children and test for deviations from expected utility, but they do not test for rational behavior.[6] Our study is unique in

[4]We note that precedence does not necessarily resolve the reverse causality problem since current behavior might depend on expected future behavior and incentives. We did not collect information on children's expectations.

[5]See (Kimball et al., 2006, 2008)

[6]They find that children appear to underweight low-probability events and overweight high-probability ones.

that it tests for the rational behavior of children in decisions under risk. Consistent with studies using adults (Harless and Camerer, 1994) and a similar instrument, we find that a significant proportion of behavior can be attributed to error (see also Jacobson and Petrie (2009); Ashraf et al. (2006)). We find, however, that the behavior of children can be rationalized by standard theory coupled with decision mistakes.

A main contribution of our paper is to show the importance of measurement error in the study of children's risk preferences and its ability to correlate with future field behavior.[7] Our results suggest that measurement error might obscure the relationship between risk preferences and field behavior and this might explain the weak correlation detected in previous studies (Sutter et al., 2013).[8] There is a small but growing body of results showing significant correlation of behavior in experiments with future field behavior (Buser et al., 2014; Castillo et al., 2011), and our findings add further evidence.

Finally, our results are important for policy design. Incentive schemes in schools designed to promote investment in human capital are likely to have a differential impact across children if the decision to participate reflects a child's underlying preferences. If misbehavior is positively correlated with impatience and risk taking behavior, it is unclear if those who might benefit the most from an incentive would be more or less likely to take advantage of it. If they are less likely, then incentives tied to a target level of performance could merely be a transfer to children that would have invested in human capital anyway in the absence of the incentives. The recent experience providing children monetary payments to improve scholarly performance (Angrist and Lavy, 2009; Fryer, 2011) provides a cautionary tale. Preferences may interact in important ways with offered incentives and would then have an impact on the success of policies.

[7] There is growing evidence that measurement error is a problem in estimating risk preferences (Kimball et al., 2006, 2008; von Gaudecker et al., 2011). Beauchamp et al. (2011) shows that once measurement error is accounted for risk preferences predict field behavior.

[8] Previous research has examined the correlation between preference measures and contemporaneous outcomes. For example, several studies show that survey measures of risk preferences are correlated with field behavior in adults (Dohmen et al., 2011; Kimball et al., 2008; Jaeger et al., 2010; Bonin et al., 2007; Burks et al., 2009).

The paper is organized as follows. Section 2 describes the risk instrument used in the experiments and how it can detect irrational behavior. The section also describes the experimental implementation. Section 3 presents results, including summary statistics, a description of rationality of behavior, an estimation of risk preferences, the relationship between theories of rationality and individual characteristics, and the correlation between rationality, risk preferences and disciplinary referrals. Section 4 concludes.

2. Experimental Design and Implementation

In this section, we describe the basic experimental design and implementation. We also show how the design can be used to measure risk preferences and test for adherance to alternative models of behavior under risk.

2.1. Design

The experimental design, based on (Chew and Waller, 1986), requires three ordered payoffs, $x_L < x_M < x_H$, two alternative lotteries, $A = (x_M, 1)$ and $B = (x_H, \alpha; x_L, 1-\alpha)$, and a parameter $0 < \beta < 1$ which specifies how the A and B lotteries and payoffs are combined into new lotteries used in the experiment.[9] Subjects make five decisions, and each decision consists of choosing one out of two lotteries. The lotteries used in the five decisions are constructed from the three payoffs, the A and B lotteries and β, and are devised such that we can test for Allais' paradox (the common consequence effect) and the common ratio property of expected utility.[10] The five-decision design has the advantage of being simple for the children and gives

[9]Lotteries A and B are referred to as pair O. Chew and Waller (1986) called their design HILO, from high (H), intermediate (I) and low (L) outcomes plus lottery O.

[10]Allais' paradox is tested by comparing what we call decisions D1 and D2, and the common ratio property is tested by comparing choices in decision D3 and D2.

us data that allow for a robust test of rationality.[11] A simpler design of choosing one lottery from many options (e.g. Binswanger, 1981) or a multiple price list of binary lottery choices (e.g. Holt and Laury, 2002) would not provide rich enough data to distinguish between various models of behavior.

We have two designs, the on-border design and the off-border design, which only differ in the possibility of receiving a certain payoff. We will first discuss the set-up of the on-border design. The off-border design is very similiar and discussed later in this section. In the on-border design, $x_L = \$0$, $x_M = \$30$, $x_H = \$40$, $\alpha = 0.8$ and $\beta = 0.25$.[12] The off-border design shares the same ordered payoffs and β.

Figure 1 presents, in the Marshack-Machina (MM) triangle (Machina, 1987), the lotteries used in both of our designs, Figure 2 shows the decision sheets in the on-border design as presented to subjects, and Figure 3 shows the decision sheets for the off-border design.[13] For the on-border design, in Figure 1, each decision is represented by two solid dots connected by a line. The dots represent the two lottery options for that decision. The five decisions which subjects make are labelled D1, D2, D3, D4 and D5 in the figure. By construction, along each line, the bottom left lottery is the safer (S) option, and the upper right lottery is the riskier (R) option. In Figures 2 and 3, for each decision, option A is the safer lottery and option B is the riskier lottery.

An expected utility maximizer would choose either all safe or all risky options since his/her indifference curves are linear and parallel (Mas-Colell et al., 1995). The possible choice patterns over the five decisions for an expected utility maximizer would be $SSSSS$ or $RRRRR$. Other choice patterns are possible if we consider

[11] An alternative design could have been Choi et al. (2007), however, this would have been much more difficult to implement in our setting since it requires more decisions and computer terminals. Our design uses only paper and pencil, and more importantly, it allows us to examine a larger set of theories of behavior (see Polisson and Quah (2013)). This is important since we would like to distinguish between systematic deviations from theory and mere noisy behavior.

[12] The five decisions used in the experiment are constructed as follows. D1 compares $\beta A \oplus (1-\beta)x_M$ and $\beta B \oplus (1-\beta)x_M$. D2 compares $\beta A \oplus (1-\beta)x_L$ and $\beta B \oplus (1-\beta)x_L$. D3 compares A and B. D4 compares $\beta B \oplus (1-\beta)x_M$ and B. D5 compares $\beta A \oplus (1-\beta)x_H$ and $\beta B \oplus (1-\beta)x_H$.

[13] The complete decision sheets used by subjects in the experiment are included in Appendix B.

alternative models of behavior. Linear, but not parallel, indifference curves as in (Dekel, 1986; Gul, 1991; Neilson, 1992a) require only that choices satisfy expected utility along each separate line. This property is called betweenness. Models of preferences for certainty like that of (Diecidue et al., 2004) and (Bleichrodt and Schmidt, 2002) predict that behavior will follow expected utility theory except for lottery A which gives \$30 with certainty. Other models, like (Neilson, 1992b) and cumulative prospect theory, produce richer patterns of behavior. Table A1 in Appendix A details the choice patterns predicted by the six theories of behavior under uncertainty we consider, which include expected and non-expected utility theories.[14]

Our on-border design has the advantage of being simple and compact. However, as noted by Sopher and Gigliotti (1993) and Conslik (1989), this design might overstate departures from expected utility if choices are different when certain options are available than when none are available. That is, the on-border design might reject the null hypothesis of expected utility too often. To test the robustness of the results, we created our off-border design. This design only has pairs of lotteries which do not include a certain option. These are based on lotteries $A^\star = (x_H, 0.16; x_M, 0.8; x_L, 0.04)$ and $B^\star = (x_H, 0.6; x_M, 0.2, x_L, 0.16)$.[15]

None of the lotteries in the off-border design involve complete certainty, and all are within the borders of the Marshack-Machina triangle. These lotteries are presented in Figure 1 as open dots. Notice that the two options available for each decision in either the on-border or off-border designs lie on the same line. The on-border decisions (solid dots) may include options where a certain payoff is available, and the off-border decisions (open dots) have no options with certain payoffs.

[14]These theories are betweenness, Neilson (1992b), Diecidue et al. (2004), cumulative prospect theory with convex weighting function, cumulative prospect theory with concave weighting function, and expected utility theory.

[15]The five new decisions are created in a similar manner to the on-border design, using the following rule: $A_i^\star = \beta A^\star \oplus (1-\beta)x_i$ and $B_i^\star = \beta B^\star \oplus (1-\beta)x_i$ for $i = L, M, H$. In particular, D1* compares $\beta A^\star \oplus (1-\beta)x_M$ and $\beta B^\star \oplus (1-\beta)x_M$. D2* compares $\beta A^\star \oplus (1-\beta)x_L$ and $\beta B^\star \oplus (1-\beta)x_L$. D3* compares $A^\star$ and $B^\star$. D4* compares $\beta B^\star \oplus (1-\beta)x_M$ and $B^\star$. D5* compares $\beta A^\star \oplus (1-\beta)x_H$ and $\beta B^\star \oplus (1-\beta)x_H$. These lotteries are constructed using the same value of β as the original lotteries.

2.2. Implementation and Data Collection

In each experimental session, subjects are assigned a unique identification code. This code is private, and subjects do not know the identification codes of other subjects. Instructions are given orally by reading from a script, and a bingo cage with numbered balls is used to represent probabilities (e.g. out of 20 balls, those numbered 1-15 pay $30 and those numbered 16-20 pay $40).[16] During the instructions, several written examples are used to make sure subjects understand probabilities, to show that payoffs are tied to probabilities and to make clear the nature of the decision task. To maximize comprehension, subjects write down the answers to the examples, and any questions are addressed.

Upon completion of the instructions and questions, subjects make their choices privately by circling option A or option B on the decision sheet for each of the five decisions.[17] Only one decision is shown on each page, and subjects are free to move back and forth between pages when making their choices. After completing their choices, a subject puts her decision sheets in an envelope with her identification code written discretely on the bottom and the envelopes of all subjects are collected. Subjects then complete a short survey about family structure, parent's education and number of siblings.[18] This survey is put in a separate envelope with the identification code and also collected.

Once all the envelopes are collected, one of the decisions is chosen at random for payment by taking index cards with the decision numbers written on them, shuffling them in front of the subjects, presenting them "face down," and asking a subject to choose one card. The number on the selected card is the decision number to be paid for the three subjects in each session who are randomly chosen to receive

[16]The decision sheet the subject sees has numbers next to various payments. The number on the ball that comes out of the bingo cage determines which payment is realized.

[17]Subjects made an additional lottery decision that is not analyzed in this paper. The additional decision varied across classrooms and is therefore controlled for with classroom fixed effects in the regression results. The decision sheets the subjects used to make their choices in the on-border and off-border designs are in Appendix B.

[18]The survey is included in Appendix B.

payment. Upon selecting a decision to be paid, the bingo cage is turned to select a ball numbered between 1 and 20. The selected ball determines the payoff corresponding to the subject's choice (A or B) for that decision.[19] For example, if decision 1 (D1) in the on-border design was chosen for payment, and the number chosen from the bingo cage was 18, the subject would earn $30 if she chose option A and $40 if she chose option B.

After determining the decision to be paid and the amount to be paid for choosing option A or option B, all the envelopes with the decision sheets are shuffled in front of the subjects, and three envelopes are chosen for payment. This is done in each session. The identification codes of those chosen to receive payment are written on the blackboard. Because identification codes are kept private by each subject, no other subject knows which subjects have been chosen to receive payment. Those who are chosen to receive payment are paid with a Wal-Mart gift card within a few days of the experiment.[20] The subjects who are chosen to be paid go privately to the principal's office to pick up their gift cards, and their names and payment are kept confidential. Subjects know all of these procedures before making their decisions.

All experiments were conducted by the authors. In total, 1,275 8th-grade students participated. One hundred and twenty two students were randomly chosen to be paid, and the average payment was $34.55. The experiments were conducted during the school day in the home-room classrooms of the entire 8th-grade cohort of a county in Georgia, USA during the 2008-2009 and 2011-2012 school years. All experimental sessions in a particular school were completed sequentially during one morning or afternoon to minimize the chance that students spoke to one another about the experiment across sessions within a school.[21] The experiment took 30 minutes to

[19]In the on-border design, the bingo cage had 20 numbered balls to make the decision task simpler. In the off-border design, because we had probabilities of less than 5%, there were 100 numbered balls.

[20]In order to minimize using up classroom time, we paid subjects a day or two later. We chose to pay with a Wal-Mart gift card because it minimizes potential problems associated with giving children cash and it can be transformed into many goods that children desire, so it is very similar to cash. Also, the school administration did not want us to use cash.

[21]Spillover across sessions was extremely unlikely. The children were kept in their home rooms

complete, and the participation rate was very high, with about 95% of each 8th-grade cohort completing the experiment.[22] Subject characteristics are presented in more detail in the next section.

The experimental data provide a measure of risk preferences, and the survey data collected from the subjects after the experiment provide information on family background and structure. In addition, the school district supplied information from the student's school records on gender, race, standardized math and reading scores, whether the student qualified for free or reduced price school meals and the number of disciplinary referrals the student received in 7th, 8th and 9th grades. A disciplinary referral happens when an offending student is sent to the administrative office (by a teacher, administrator or bus driver) and the behavior is entered into the student's official record (i.e. reprimand, detention, suspension, etc.). This measure does not include referrals to the office that do not result in a recorded entry in the student's record. Referrals in 9th grade are typically given for more serious infractions. All the data, from the school and the experiment, were anonymized to protect students' privacy.

3. Results

In this section we summarize the survey and school record data and describe the experimental results and evidence of deviations from rational behavior. An estimate of risk preferences is presented, as well as the correlation between rational behavior and individual characteristics and the relationship between rationality, risk and field behavior.

while the experiments were conducted and were not allowed to interact with children seated in other classrooms either waiting to participate in the experiment or having completed the experiment. The children did not know the exact nature of the experiment prior to our arrival in the room to conduct the session.

[22]Non-participation was primarily due to absence on the day the experiments were implemented. We had only a handful of students across the three times the experiments were implemented who declined to participate.

3.1. Description of the sample

Summary statistics of our sample are shown in Table 1. About one half of the subjects are male, and 44% are given a racial classification by the school district as black and 48% as white. The average age is 13.8 years, and 34% of children report living with one or no parent. These numbers hide the fact that 49% of black children report living with one or no parent while only 22% of white children do. About two thirds of children qualify for free or reduced price school meals. The average number of disciplinary referrals is 1.56 in the 7th grade, 2.06 in the 8th grade and 1.33 in the 9th grade. While the number of referrals are highly correlated across 7th and 8th grades ($\rho = 0.5778$, p-value < 0.0001), they are statistically significantly different (t-test = -3.9976, p-value = 0.0001). Figure 4 shows that the distribution of referrals for 7th, 8th and 9th grades is skewed towards zero. For any grade year, between 65 and 78 percent of children receive no referrals.

3.2. Are choices rational or random?

The distributions of choices for each of the five decisions in the on-the-border and off-the-border designs are presented in Table 2. The modal choice is risk averse. In total, safe options (S) are chosen about three-fifths of the time across both designs. Looking at the distribution of choices, it is not uniform in either design, as would be predicted by expected utility theory. For example, the proportion of times the safe option is chosen varies from 51 to 69 percent of the time across the five decisions. We also observe that, consistent with previous experiments, when the option of $30 for sure is available, as in D1 and D3 of the on-the-border design, subjects are more likely to choose it. For instance, S is chosen 69 percent of the time in D1 and 63 percent of the time in D3. Subjects are also more likely to switch to R in D5 in both designs.[23] The common consequence effect can be evaluated using D1 and D2. The pattern typically found in previous experiments is for subjects to choose S

[23] D5 is represented on the upper left hand corner of the MM triangle in Figure 1.

more frequently than R in D1 compared to D2. Table 2 shows that this pattern is present in the on-the-border design, but the opposite pattern (from R to S) emerges in the off-the-border design. The difference in the distribution of choices between the on-border and off-border designs is significant.[24]

Harbaugh et al. (2001) show that children as young as eight years old satisfy the Generalized Axiom of Revealed Preferences (GARP) when choosing between two goods and no risk. Moreover, they show that this behavior is statistically significantly different from random choice. Our design allows us to examine if this holds when decisions are made under uncertainty. We turn now to examining whether decisions are rational or random in the face of risk.[25]

Each row in Table 3 compares the predicted frequency of patterns of behavior if children chose at random and the observed frequency of such patterns in the data for the six theories of decision making under uncertainty we consider. These theories include one that requires behavior satisfies betweenness, two that allow for different utilities over certain and uncertain payoffs, Neilson's (1992) model and Diecidue et al's (2004) model, cumulative prospect theory (CPT) with convex weighting function, CPT with concave weighting function and expected utility. While the list of models we consider is not exhaustive, some of the theories deviate significantly from expected utility and therefore increase the chance of detecting whether choices different from expected utility are systematic or not.[26]

[24] A χ^2 test of the equality of distributions across the two designs yields a p-value of 0.000.

[25] Our measure of rationality is valid independent of the underlying distribution of preferences. Individual risk attitudes affect the distribution of patterns of behavior, but not whether a person acts rationally. While the experimental design might not be well calibrated to capture individual differences in risk attitudes, our design is still powerful enough to detect deviations from rational behavior. For example, under expected utility theory, the calibration of the experiment might affect whether a subject chose always S or always R, but not that he only chose one of these two patterns. The distribution of underlying preferences, however, does affect the power of the test of alternative theories of behavior under risk. That is, the calibration might affect the relative, but not the absolute, performance of a particular model in rationalizing the data.

[26] Following the convention in the literature, we compare patterns of behavior to random choice. Other patterns of "irrational" behavior could manifest, such as choosing all A's. If so, it would be difficult to distinguish this pattern from adherence to expected utility. For this reason, we examine five additional theories that produce further patterns of behavior. Also, we might expect to see this

The first row shows that a child choosing at random would satisfy betweenness 25 percent of the time while behavior consistent with betweenness occurred 38.6 percent of the time in the on-the-border design and 38.9 percent in the off-the-border design. The z-scores testing differences in means corresponding to these comparisons are 7.98 and 8.05 and show that behavior in the experiment is significantly different from random choice when assuming preferences satisfy betweenness.[27] Indeed, for five of the six theories we examine, the patterns of choices made by children are significantly different from random.[28]

Additional evidence confirms that children's choices are not just noise. The scale reliability coefficient (Cronbach's alpha) between the answers to the five lottery questions is 0.31. This coefficient would be about 0.08 if the data were generated at random. Similarly, the largest eigenvalue corresponding to the five decisions using factor analysis is 0.508, while the largest eigenvalue, if the data were generated at random, would be 0.076 (Horn, 1965). We conclude that the choices over the five lottery decisions contain more information that random choice would predict.

3.3. Estimation of risk preferences

We have seen that the decisions of children are noisy but not random. In this section, we examine which of the six theories considered best explain choice patterns. In so doing, we describe a method to estimate risk preferences that accounts for noise. This cleaned estimate is used in section 3.5 to correlate with field behavior.

In comparing across the six theories, we would like to account for the fact that some theories predict more patterns of behavior than others. To do this, we follow Harless and Camerer (1994)'s approach to modelling behavior in the MM triangle.

---

type of "irrationality" when the decision problem is more complex, e.g. in the off-border design. However, the proportion of children choosing all A's is not significantly different in the off-border and on-border designs (13.2% in the on-border and 12.0% in in the off border, p-value=0.528).

[27]For an observed and predicted frequency x and y the z-score is calculate as $z = \frac{x-y}{\sqrt{\frac{y(1-y)}{n}}}$ where n is the sample size.

[28]The only theory for which observed behavior is not significantly different than random choice is cumulative prospect theory with concave weighting function.

The basic approach is to model observed patterns of behavior as the result of adherence to a theory with probability $(1-\omega)$ or random choice with probability ω. For instance, suppose pattern $SSSSS$ is chosen with probability p and pattern $RRRRR$ is chosen with probability $1-p$, then if a person makes a mistake with probability ω then the probability of observing pattern $SSSSS$ according to this theory of behavior is $p(1-\omega)^5 + (1-p)\omega^5$. That is, $SSSSS$ could have been chosen without making errors or because a person switched from the desired pattern $RRRRR$ by making errors.

Since for a given p and ω this approach allows us to calculate the likelihood of observing any pattern of behavior, we can use maximum likelihood methods to estimate the parameters of this model or any theory that predicts a subset of all possible patterns. The number of parameters of a theory so defined is then equal to the number of patterns of behavior predicted by the theory.[29] Note that this approach treats all mistakes as equal regardless of the absolute difference in the expected value of the lotteries considered. This should not be an issue in our case, however, because we do not find evidence that subjects are less likely to make mistakes the larger the difference in the expected value of the lotteries.[30] Table A2 in Appendix A presents estimates for the theories we consider. All theories have large levels of error, and these error rates are on the higher side of those reported for adults in Harless and Camerer (1994).

An alternative approach to the one described above is the framework of random utility theory. For instance, we could assume that subjects either choose according to a constant absolute risk aversion utility function with some probability or choose completely at random otherwise. Our approach uses the data more thoroughly.

[29] Because probabilities add up to one, there are the number of patterns minus one parameters. The noise parameter adds one more, which makes the total number of parameters equal to the number of patterns.

[30] We test if decision errors are sorted according to the difference in expected value between lotteries by allowing decision errors to decrease with the difference in expected value between lotteries. We cannot reject the hypothesis that errors are not sorted in this manner ($\chi^2(1) = 1.13$, p-value = 0.2881). The test restricts decision errors to be the same across all participants, but allows the propensity to choose risk aversely to depend on covariates.

Estimates are based on the population data and the individual decisions of a subject and provide a more precise measure of individual preferences.[31] The random utility model would allow for this if we assume that individual preferences are characterized by a distribution rather than a singleton. Our data, however, are not rich enough to satisfy this assumption.[32] Instead, the approach outlined above addresses the issue of idiosyncratic heterogeneity in preferences, independent of covariates, subject to the data generated by our experimental design.

Using our approach to modeling decisions over risk, we now turn to examine which theory of decision making is better at describing the data. Table 4 presents the pairwise comparisons of the six theories using the (Vuong, 1989) test for model selection. The test compares the likelihood that the data have been produced by a particular theory and corrects for the number of parameters (or patterns of behavior) of the theory.[33] The numbers in the table show the test of the theory in a row against the theory in a column. Positive numbers mean that the theory in the row is better at explaining the data than the theory in the column and negative numbers mean the opposite.

We see that, with the exception of CPT with convex weighting, betweenness does better than other theories of behavior in rationalizing the data. This holds because the loss in patterns of behavior explained by betweenness is compensated by having

[31] That is, our approach allows an estimation of the conditional probability an individual has risk averse preferences given his specific choices in the experiment and not just on his individual characteristics. This approach assumes the existence of idiosyncratic differences in preferences and uses individual decisions to best guess what these are.

[32] Such an approach is proposed by von Gaudecker et al. (2011) and ours is a discrete version of it. von Gaudecker et al. (2011) assume that subjects either evaluate lotteries according to a parametrically specified expected utility function or choose at random. To model heterogeneity across individuals, and in the spirit of mixed logit models, they characterize preference parameters as drawn from a continuous parametric distribution function. For any given set of parameters, an estimate of an individual's expected utility parameters can be constructed based on observed choices and the individual's characteristics. We have attempted this approach in our data but could not obtain estimates of the model either using a constant relative risk aversion or a constant absolute risk aversion utility function. This suggests that our experimental design is not rich enough to estimate underlying preferences using this approach.

[33] We note that these tests are joint hypotheses of the theory and the particular model of random errors and therefore the results should be understood as such.

a more parsimonious representation of preferences. Overall, though, the table shows that it is difficult to come up with a reasonable ranking of the theories. Expected utility theory performs only worse than betweenness, but no strong ranking pattern across the six theories emerges.

In summary, parametric and non-parametric approaches suggest an important amount of noise in the data. Behavior, however, is not random. As in experiments with adults (Harless and Camerer, 1994), expected utility theory does not do worse than alternatives theories at explaining the data once parsimony is taken into account.

3.4. Rationality and individual characteristics

Rationality may be correlated with individual characteristics. This is examined in Table 5 which presents a linear probability model of a variable that equals 1 if a subject's choice pattern is consistent with a choice pattern predicted by a particular theory and equals 0 otherwise on individual characteristics. For instance, column 7 in Table 5 presents the linear probability model corresponding to expected utility theory. For expected utility theory, the dependent variable equals 1 if a subject always chose S or always R, yielding a choice pattern of $SSSSS$ or $RRRRR$. For the other columns, the dependent variable equals 1 if the choice pattern is one of those listed in Table A1 and 0 otherwise. For reference, in the first column of Table 5, an ordinary least squares regression is presented of the number of safe choices made across the five lotteries. This is a measure of the individual level of risk aversion. All the estimations include fixed-effects at the classroom level to account for unobserved heterogeneity.

The first column in Table 5 shows that younger children within the cohort and children who perform better in standardized reading tests are more likely to choose safe options.[34] The remaining columns look at consistency with predicted patterns

[34]Note that the effect of age in any of our results is not due to older children (e.g. those > 14 years old) who may have been held back and are repeating 8th grade. The effect is solely due to natural age variation within a grade cohort. A potential reason for age effects is hormonal changes,

of behavior for each theory and show that older children within the cohort tend to be less likely to choose consistently, but this is not significant for all theories considered. Children whose father's highest level of education is high school are less likely to act rationally, but again, this is not robust. We also examine an alternative measure of rationality: the number of decisions that need to be changed to make a choice pattern consistent with a particular behavioral theory. For instance, an individual choosing a pattern predicted by theory will have a cost of zero while an individual who would conform to theory by changing at most one decision will have a cost of one. In general, we find that this alternative measure reproduces the same results as Table 5, but the correlation between individual characteristics and costs to rationality is even weaker.[35]

So, while some individual characteristics are correlated with making rational decisions, no consistent relationships exist across various definitions of rationality.

3.5. Rationality, risk attitudes and field behavior

We now turn to the relationship between risk attitudes, rationality and field behavior. We focus on disciplinary referrals for field behavior as these are acts that have been shown to predict economic outcomes later in life, such as education achievement and lower wages (Bowles et al., 2001; Heckman et al., 2006; Lang and Ruud, 1986; Segal, 2013), as well as high school drop-out rates (Alexander et al., 1997; Rumberger, 1995). Disciplinary referrals are therefore a good benchmark to evaluate the influence of preferences on behavior and to test the ability of experiments to uncover them.[36] Following Freeman (1999), we expect disciplinary referrals to be negatively correlated with risk aversion.

however, we do not have a direct way to test this hypothesis.

[35]Table A5 in Appendix A shows these results. In addition, Table A3 shows that, for the case of expected utility, the propensity to deviate from rational behavior is correlated with age, father's education and math scores.

[36]Time preference may also be important, however, we do not have a measure of this to include in the regressions. Instead, we include relevant covariates. The time preference experimental data used in Castillo et al. (2011) are from previous school cohorts.

Tables 6-8 present negative binomial regressions of discipline referrals on three different measures of risk preferences, and all include a measure of rationality as defined by expected utility theory and covariates. We use a negative binomial specification because disciplinary referrals are a count variable and many children have no referrals (see Figure 3 for the distribution of referrals across 7th, 8th and 9th grades). Our control for rationality is a dummy variable indicating choices consistent with expected utility because, as shown in Section 3.2, it is as good at explaining the data as other theories and has been used frequently in the literature. All reported results in this section hold using any of the alternative theories to define rationality.[37] The regressions include fixed effects at the classroom level to account for any variation due to experimental implementation or selection of students into classrooms.

We look at disciplinary referrals at the end of the 8th grade, 9th grade and the sum of both of them. 8th grade referrals occurred in middle school, and 9th grade referrals occurred in high school. Having referrals for the same child from two different schools provides a stress test on our results because these are two different sets of school administrators deciding whether to record the discipline infraction on the student's permanent record. These referrals occurred up to 2 years after the experiments were conducted. All the regressions also control for disciplinary referrals in the 7th grade. While our results are qualitatively similar whether this variable is included or not, including past misbehavior diminishes the omitted variable problem due to unobservable conditions that might influence an individual's behavior in the field and in the experiment.

We first examine whether our first risk measure, a simple aggregate measure of risk preferences based on the total number of safe choices across the five decisions, can explain disciplinary referrals. Table 6 shows these results for referrals in 8th, 9th and 8th and 9th grades combined.[38]

Consistent with previous research on the behavior of children (Bertrand and

[37] These results are reported in Tables A6-A14 in Appendix A.

[38] Tables A6-A8 in Appendix A show the same results for all six of the theories we consider. These tables show that the results reported in the paper are robust to our use of expected utility to define rationality.

Pan, 2013), we find that male and black children are more likely to have disciplinary referrals.[39] Children who qualify for free or reduced price school meals are also more likely to have disciplinary referrals. The opposite is true for children living in a two-parent household, first-born children, children with no older siblings living in the household, and children who perform better in standardized math tests, although some of these results are not robust across years. Past disciplinary referrals are a strong predictor of future disciplinary referrals. This is consistent with disciplinary referrals being the product of unobserved personality traits, conditions faced by the child or expectations. In the absence of additional individual information on these factors, we cannot distinguish between these hypotheses.

Looking at the effect of risk preferences on referrals, Table 6 shows that, using our first measure of risk, the more safe choices a child makes, the fewer disciplinary referrals he receives. However, this is only significant when 8th and 9th grade referrals are combined. Table 7 shows the results using our second measure, for each decision, whether the riskier option was chosen. Only one of the decisions, D1, is significantly correlated with referrals and only in 8th grade and 8th and 9th grades combined.[40] Taken together, our first two measures suggest that risk preferences are not significantly correlated with disciplinary referrals.

For our third measure of risk preferences, we turn to the approach developed in section 3.3 to construct a new measure that is free of measurement error. Following Harless and Camerer (1994), we estimate a structural model that predicts that a person adheres to expected utility with probability $1-\varepsilon$ and reverses her preferences (makes a mistake) with probability ε. Since, in our experiment, expected utility allows only two choice patterns ($SSSSS$ or $RRRRR$) we allow this probability and the probability of making a mistake (ε) to depend on the characteristics of the individual. Using the entire sample, we then estimate these parameters and use them to calculate the posterior probability that a person's preferences are consistent with all

[39] In our data, boys have 2.59 disciplinary referrals and girls have 1.51 disciplinary referrals (p-value < 0.000). Black children have 2.99 disciplinary referrals on average and non-black children have 1.32 disciplinary referrals (p-value < 0.000).

[40] Complete estimates are shown in Tables A9-A11 in Appendix A.

safe choices given choice patterns in the experiment and individual characteristics.[41] These estimates using the entire population allow us to assess how much an individual decision is due to the strength of preferences or mistakes. We refer to this new measure of risk attitudes as $Pr(AAAAA|Choice)$ in the regression table.[42]

Table 8 shows the results of using this third measure. Now, we see that future disciplinary referrals are strongly and negatively correlated with risk aversion, even controlling for demographics, household characteristics and rationality. Children who are more risk averse are less likely to receive disciplinary referrals up to two years in the future. The effect is large, consistent with theory and holds in both 8th and 9th grades when the child is in the last year of middle school and the first year of high school. The average difference in the likelihood of being risk averse between those who chose always S and those who chose always R is 0.5. This implies that the effect of risk preferences on disciplinary referrals in 9th grade ($0.5\times$-0.65 = -0.325) is half the effect of being black and almost three times the effect of being a boy. For the sum of referrals in 8th and 9th grades, the effect of choosing always safe options versus risky is the same as the effect of being a boy or three quarters the effect of being black. Rationality, however, does not explain referrals.[43] Finally, the

[41]The results of these estimations are shown in Table A3 in Appendix A and are largely consistent with previous research. For example, boys are less likely to choose all safe options, but this is not significant. Those with higher math scores are less likely to choose all safe and have lower error rates. The off-border design (partially proxied by the dummy variable "Experiment run in 2011" yields higher error rates.

[42]We tried alternative ways of modeling $Pr(AAAAA|Choice)$, including the five other models discussed in the paper, but could not obtain stable results. To obtain consistent estimates of the effect $Pr(AAAAA|Choice)$ on field behavior, it is necessary to include all the covariates considered in Table 8 in the estimation and calculation of the posteriors of $P(AAAAA|Choice)$ (Kimball et al., 2008). Using this approach, only the model of expected utility provided stable results.

[43]Results using alternative definitions of rationality can be seen in Tables A12-A14 in Appendix A. These results show that, with the exception of Diecidue et al. (2004), none of the theories of decision making under uncertainty explain misbehavior. It is interesting to note that Diecidue et al. (2004) allow for certainty baises (people value certain payoff more than uncertain ones) and uncertainty biases (people value uncertain payoff more than certain ones). It is the presence of uncertainty biased subjects that explains the positive correlation between rationality and disciplinary referrals. These confirm that whether risk attitudes are measure by the concavity of the utility function or through a distortion of expected utility, it is the propensity to take risky actions that is correlated with disciplinary referrals.

negative and significant effect of risk aversion on misbehavior is robust to alternative specifications.[44]

This third measure of risk aversion is an estimate and therefore is likely to be affected by sampling error. Table A4 in Appendix A presents the distribution of the estimated effect of risk aversion on disciplinary referrals in the 8th grade, 9th grade and 8th and 9th grades combined based on 1,000 bootstrap replications of the estimates.[45] As expected, sampling error is important, nonetheless, the estimates of risk aversion are still significantly negatively correlated with future disciplinary referrals in 9th grade and 8th and 9th grades combined. So, even taking into account sampling error, our result that more risk averse children are less likely to have disciplinary referrals up to two years in the future is still robust.

In sum, the results in this section show that repeated measures of risk preferences are advantageous as they allow us to construct estimates that are purged of measurement error. Eliminating these errors with estimations reveal a significant relationship between preferences and field behavior that was not apparent in aggregate and one-response measures constructed from the raw data.

4. Conclusions

We set out to investigate the relationship between children's rationality, risk preferences and field behavior. Our study is motivated by economic theory which suggests that misbehavior is, ceteris paribus, correlated with the willingness to take risks. One of the main findings is that children who are more risk averse are less

[44] If instead of the number of disciplinary referrals, we use as the dependent variable a binary variable that equals 1 if the child has had any disciplinary referrals > 0 and equals 0 if the child has had no referrals and rerun all the specifications in Table 8, we get the same results. Risk averse children have a lower probability of having at least one disciplinary referral. The effect is roughly 11 percent in 8th grade and 16 percent in 9th grade and 8th and 9th grades combined. The effect is significant at the 5-percent level for 9th grade and 8th and 9th grades combined and at the 15-percent level for 8th grade.

[45] In particular, we re-estimate the model 1,000 times using a bootstrapped sample and used those estimates to re-estimate the posterior probability that a subject was risk averse.

likely to receive disciplinary referrals up to two years after the experiment. That is, risk attitudes not only significantly correlate with behavior in the field but also with behavior in the future.[46] More importantly, our estimates show that risk preferences have an effect on behavior separate from rationality, cognitive abilities, household environment and past behavior. Risk preferences are therefore important for understanding heterogeneity of behavior in the field and could interact with the response to policy by children.

A main contribution of our paper is a direct test of the rationality of children in risky environments under alternatives models of behavior. Our data suggest that children's behavior is noisier than that of adults, but otherwise is informative. That is, children are not always rational, as defined by various theories, but their choices are not random either. These results suggest that similar issues are likely to emerge in less educated populations as well and that the study of the effect of preferences in other context may require appropriately addressing measurement problems.

Our paper also contributes to the discussion on the external validity of experimental measures of preferences by showing that the relationship between lab data and field behavior is obscured by error. Simple aggregate measures of risk preferences do not correlate with future field behavior, but an estimate of risk preferences that controls for error does. In addition, we control for past field behavior when examining the relationship between future field behavior and the estimated risk measure. This allows us to examine the correlation of risk preferences on future misbehavior separate from the child's past.

If preferences largely determine responses to incentives, policies designed to foster investment in human capital among children are likely to have heterogeneous effects in the population. In the extreme, if those more likely to take risks are also less likely to take advantage of these type of incentives, rewards to good behavior might just be a transfer to children that would have invested optimally in the absence of

[46]Buser et al. (2014) find that willingness to compete predicts career choice among Dutch teenagers, and Castillo et al. (2011) show that impatience among children predicts disciplinary referrals.

the incentive. Our findings might lend some insight into the limited effectiveness of incentives that reward achieving a certain level of performance on standardized tests. Optimal policies might ultimately require differential treatment across individuals. Our results certainly suggest that designs of incentives for children to accumulate human and non-human capital may need to take into consideration heterogeneity of preferences.

5. References

Alexander, K.L., Entwisle, D.R., Horsey, C.S., 1997. From first grade forward: Early foundations of high school dropout. Sociology of Education 70, 87–107.

Angrist, J., Lavy, V., 2009. The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial. American Economic Review 99, 1384–1414.

Ashraf, N., Karlan, D., Yin, W., 2006. Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines. Quarterly Journal of Economics 121, 673–697.

Beauchamp, J., Cesarini, D., Johannesson, M., 2011. The psychometric properties of measures of economic risk preferences. Working paper.

Bertrand, M., Pan, J., 2013. The trouble with boys: social influences and the gender gap in disruptive behavior. American Economic Journal-Applied Economics 5, 32–64.

Bleichrodt, H., Schmidt, U., 2002. A context-dependent model of the gambling effect. Management Science 48, 802–812.

Bonin, H., Dohmen, T., Falk, A., Huffman, D., Sunde, U., 2007. Cross-sectional earnings risk and occupational sorting: The role of risk attitudes. Labour Economics 14, 926–937. Workshop on Education and Risk, Bonn, GERMANY, APR, 2006.

Borghans, L., Golsteyn, B.H.H., Heckman, J., Humphries, J.E., 2011. Identification problems in personality psychology. Personality and Individual Differences 51, 315–320.

Bowles, S., Gintis, H., Osborne, M., 2001. The determinants of earnings: A behavioral approach. Journal of Economic Literature 39, 1137–1176.

Burks, S.V., Carpenter, J.P., Goette, L., Rustichini, A., 2009. Cognitive skills affect economic preferences, strategic behavior, and job attachment. Proceedings of the National Academy of Sciences of the United States of America 106, 7745–7750.

Buser, T., Niederle, M., Hessel, O., 2014. Gender, Competitiveness and Career Choices. Quarterly Journal of Economics 129, 1409–1447.

Castillo, M., Ferraro, P.J., Jordan, J.L., Petrie, R., 2011. The today and tomorrow of kids: time preferences and educational outcomes of children. Journal of Public Economics 95, 1377–1385.

Chew, S.H., Waller, W.S., 1986. Empirical tests of weighted utility theory. Journal of Mathematical Psychology 1, 55–72.

Choi, S., Fisman, R., Gale, D., Kariv, S., 2007. Consistency and heterogeneity of individual behavior under uncertainty. American Economic Review 97, 1921–1938.

Conslik, J., 1989. 3 variants on the Allais example. American Economic Review 79, 392–407.

Cunha, F., Heckman, J.J., Schennach, S.M., 2010. Estimating the Technology of Cognitive and Noncognitive Skill Formation. Econometrica 78, 883–931.

Currie, J., Tekin, E., 2006. Does child abuse cause crime? NBER Working Paper Series 12171.

Dekel, E., 1986. An axiomatic characterization of preferences under uncertainty - weakening the independence axiom. Journal of Economic Theory 40, 304–318.

Diecidue, E., Schmidt, U., Wakker, P., 2004. The utility of gambling reconsidered. Journal of Risk and Uncertainty 29, 241–259.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G., 2011. Individual risk attitudes: measurement, determinants, and behavioral consequences. Journal of the European Economic Association 9, 522–550.

Freeman, R.B., 1999. The Economics of Crime. Handbooks in Economics, vol. 5. Amsterdam; New York and Oxford: Elsevier Science, North-Holland.

Fryer, Jr., R.G., 2011. Financial Incentives and Student Achievement: Evidence from Randomized Trials. Quarterly Journal of Economics 126, 1755–1798.

von Gaudecker, H.M., van Soest, A., Wengstrom, E., 2011. Heterogeneity in risky choice behavior in a broad population. American Economic Review 101, 664–694.

Gul, F., 1991. A theory of disappointment aversion. Econometrica 59, 667–686.

Harbaugh, W.T., Krause, K., Berry, T.R., 2001. GARP for kids: On the development of rational choice behavior. American Economic Review 91, 1539–1545.

Harbaugh, W.T., Krause, K., Vesterlund, L., 2002. Risk Attitudes of Children and Adults: Choices Over Small and Large Probability Gains and Losses. Experimental Economics 5, 53–84.

Harless, D.W., Camerer, C.F., 1994. The predictive utility of generalized expected utility theories. Econometrica 62, 1251–1289.

Heckman, J.J., Stixrud, J., Urzua, S., 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. Journal of Labor Economics 24, 411–482.

Horn, J.L., 1965. A rationale and test for the number of factors in factor analysis. Psychometrika 30, 179–185.

Jacobson, S., Petrie, R., 2009. Learning from mistakes: What do inconsistent choices over risk tell us? Journal of risk and uncertainty 38, 143–158.

Jaeger, D.A., Dohmen, T., Falk, A., Huffman, D., Sunde, U., Bonin, H., 2010. Direct evidence on risk attitudes and migration. Review of Economics and Statistics 92, 684–689.

Kimball, M., Sahm, C., Shapiro, M., 2006. Preferences in the PSID: Individual Imputations and Family Covariation. American Economic Review 99, 363–368.

Kimball, M.S., Sahm, C.R., Shapiro, M.D., 2008. Imputing Risk Tolerance From Survey Responses. Journal of the American Statistical Association 103, 1028–1038.

Lang, K., Ruud, P.A., 1986. Returns to schooling, implicit discount rates and black-white wage differentials. Review of Economics and Statistics 68, 41–47.

Machina, M.J., 1987. Choice under uncertainty - problems solved and unsolved. Journal of Economic Perspectives 1, 121–154.

Mas-Colell, A., Whinston, M.D., Green, J.R., 1995. Microeconomic theory. Oxfor University Press.

Neilson, W.S., 1992a. A mixed fan hypothesis and its implications for behavior toward risk. Journal of Economic Behavior and Organizations 19, 197–211.

Neilson, W.S., 1992b. Some mixed results on boundary effects. Economics Letters 39, 275–278.

Niederle, M., Vesterlund, L., 2010. Explaining the Gender Gap in Math Test Scores: The Role of Competition. Journal of Economic Perspectives 24, 129–144.

Polisson, M., Quah, J.K.H., 2013. Revealed preference tests under risk and uncertainty. University of Leicester Working Paper No/ 13/24.

Rumberger, R.W., 1995. Dropping out of middle school - a multilevel analysis of students and schools. American Educational Research Journal 32, 583–625.

Segal, C., 2013. Misbehavior, education, and labor market outcomes. Journal of the European Economic Association 11, 743–779.

Sopher, B., Gigliotti, G., 1993. A Test of generalized expected utility theory. Theory and Decision 35, 75–106.

Sutter, M., Kocher, M.G., Glaetzle-Ruetzler, D., Trautmann, S.T., 2013. Impatience and uncertainty: experimental decisions predict adolescents' field behavior. American Economic Review 103, 510–531.

Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57, 307–333.

6. Figures and Tables

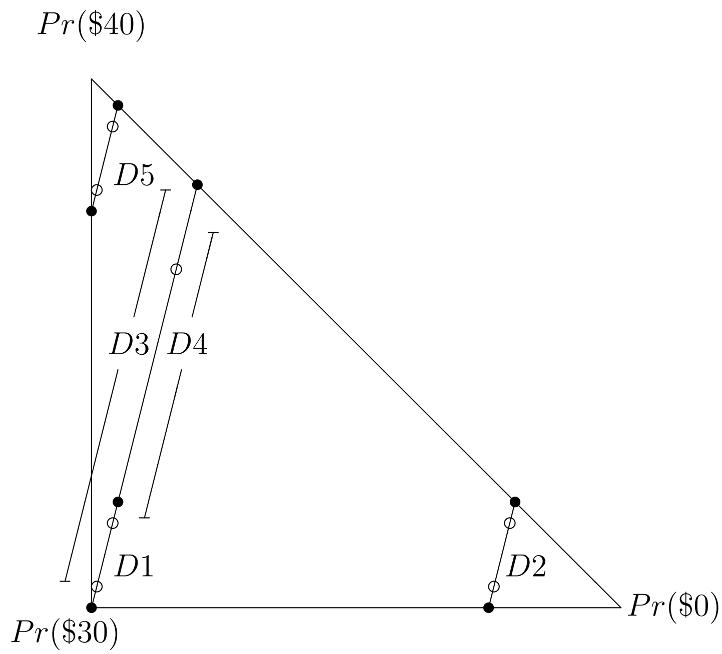

Figure 1: Experimental design represented in the Marschak-Machina triangle

The lines connect the two lottery options available for each of the five decisions (D1, D2, D3, D4, D5). Solid dots represent options for the on-border design, and open dots are for the off-border design. The dot located on the lower left along the line is the safer option (S) and the one on the upper right is the riskier option (R).

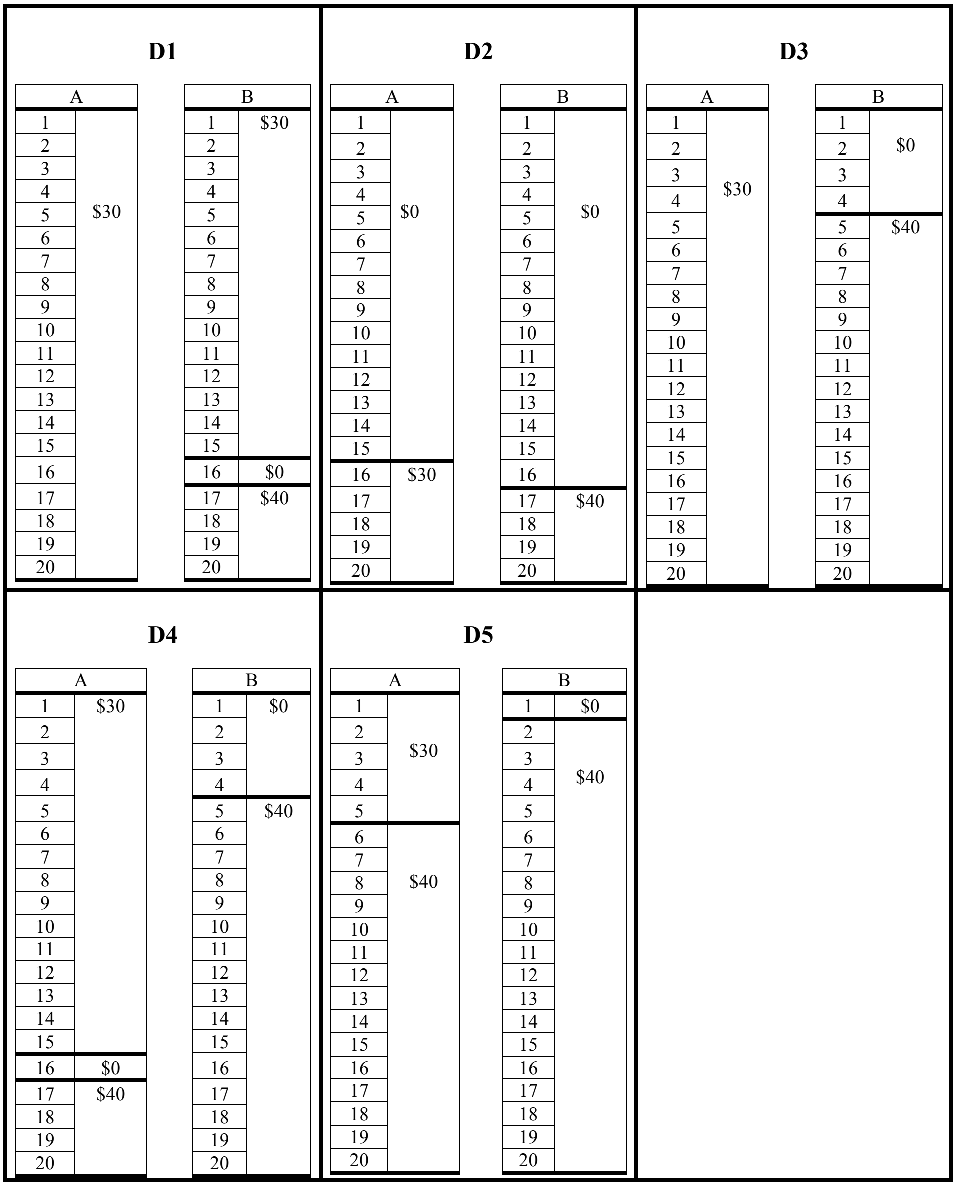

Figure 2. Decisions in the on-border design as represented to subjects

For each of the five decisions (D1, D2, D3, D4, D5), subjects chose either A or B . A is the safer option (S) and B is the riskier option (R). For each option, payoffs are associated with numbers, from 1-20. These represent the probability of receiving that payment. Payment is determined by choosing one numbered ball (numbered from 1-20) from a bingo cage.

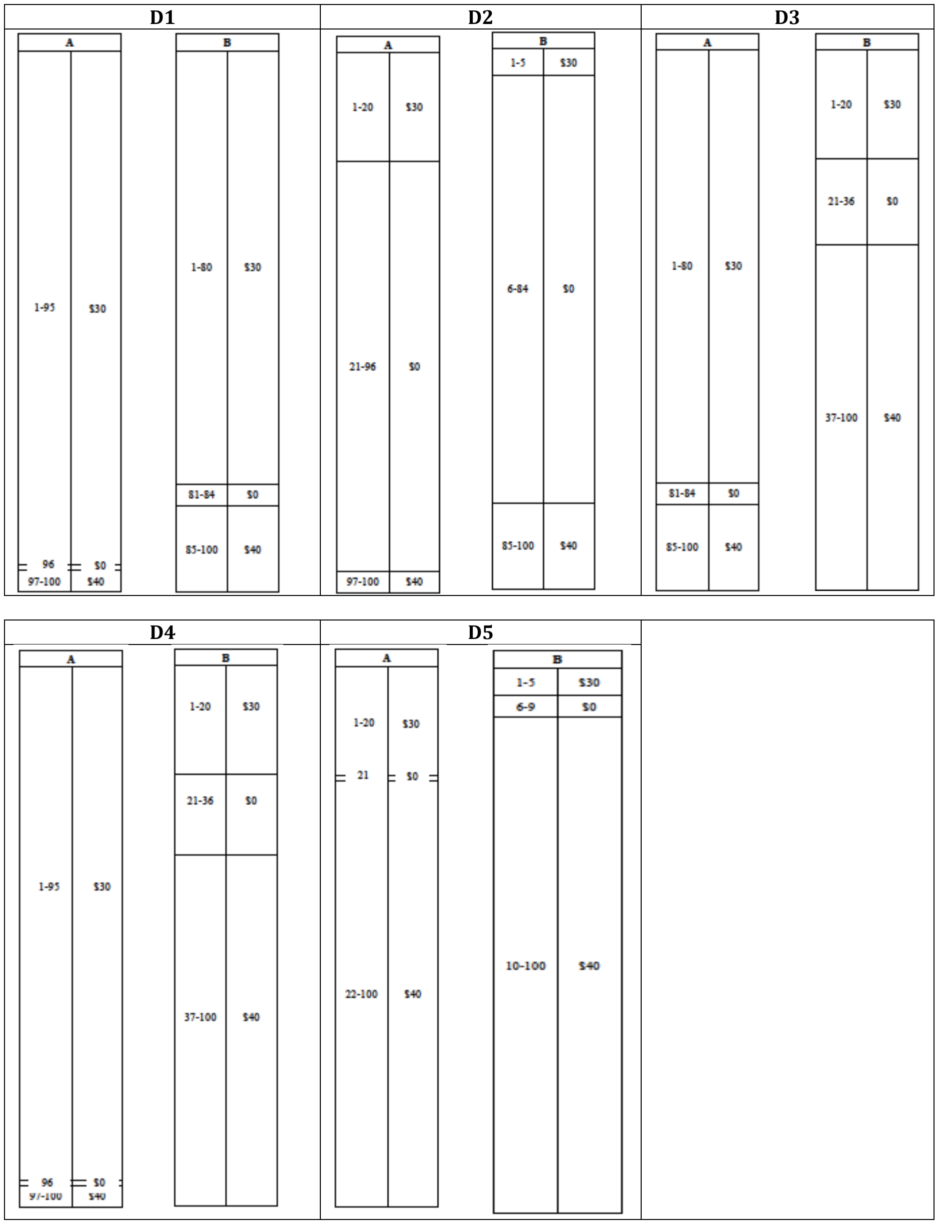

Figure 3. Decisions in the off-border design as represented to subjects

For each of the five decisions (D1, D2, D3, D4, D5), subjects chose either A or B . A is the safer option (S) and B is the riskier option (R). For each option, payoffs are associated with numbers, from 1-100. These represent the probability of receiving that payment. Payment is determined by choosing one numbered ball (numbered from 1-100) from a bingo cage.

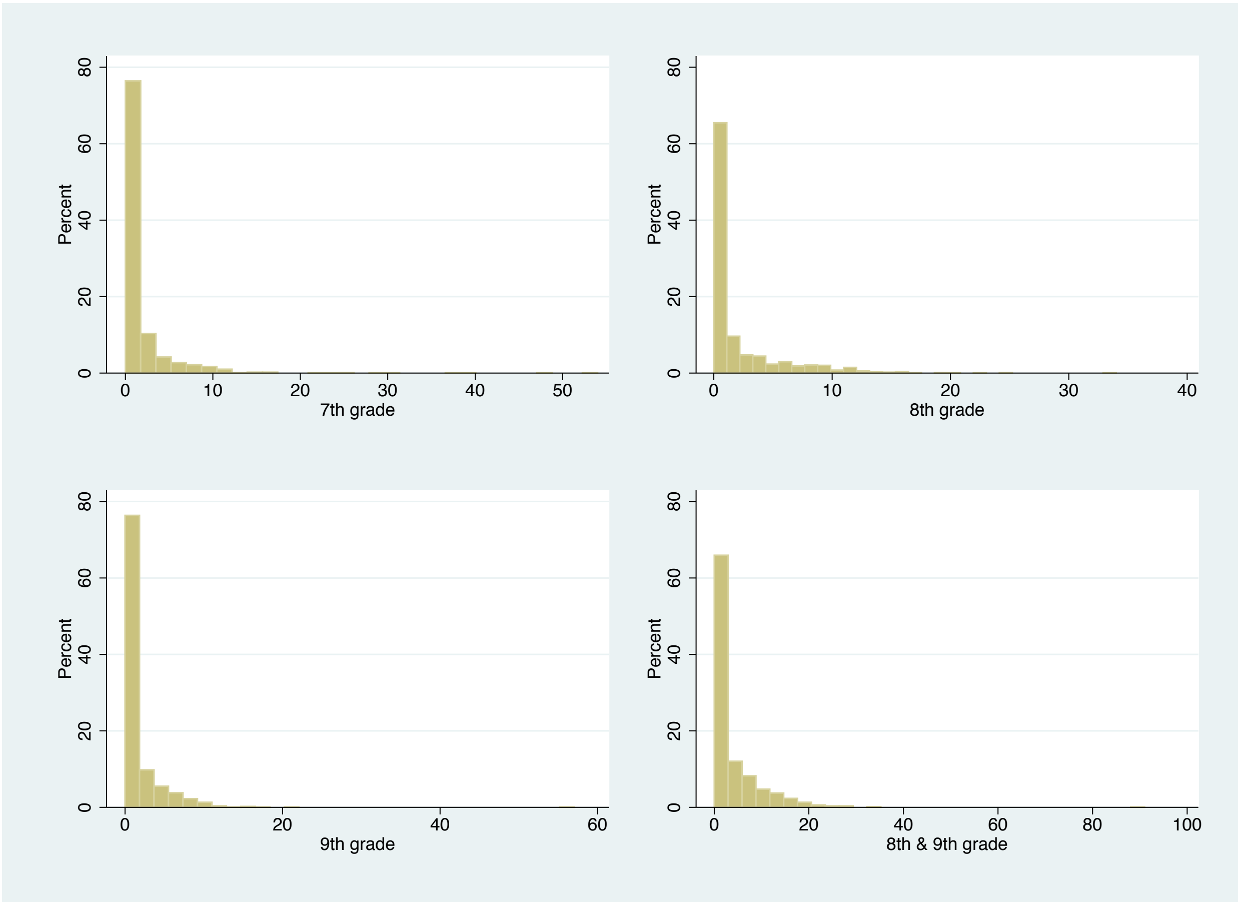

Figure 4. Distribution of Disciplinary referrals

For 7th, 8th, 9th and 8th/9th grades.

Table 1. Descriptive statistics

	Mean	S.D.
Total (n)	1,275	
Male (n)	648	
Black (n)	563	
White (n)	614	
Age in years	13.78	0.64
Two-parent household (%)	66.00	47.39
Mother finished high school (no college) (%)	39.02	48.80
Father finished high school (no college) (%)	35.35	47.83
Mother has college degree (%)	32.56	46.88
Father has college degree (%)	23.26	42.26
First born (%)	37.80	48.51
No older sibling in household (%)	48.63	50.00
Math score (8th grade)	812.51	32.31
Reading score (8th grade)	831.27	23.00
Free and reduced price lunch (%)	64.84	47.77
# of Disciplinary referrals (7th grade)	1.56	3.92
# of Disciplinary referrals (8th grade)	2.06	3.55
# of Disciplinary referrals (9th grade)	1.33	3.00

Some subjects are missing data on age, test scores (because they left the school system between the experiment and testing), or household characteristics (because they failed to answer a post-experiment survey question).

Table 2. Distribution of lottery decisions, percent in parentheses

	On-border lotteries		Off-border lotteries	
	Safe	Risky	Safe	Risky
D1	447 (68.98)	201 (31.02)	379 (60.45)	248 (39.55)
D2	384 (59.26)	264 (40.74)	420 (66.99)	207 (33.01)
D3	411 (63.43)	237 (36.57)	410 (65.39)	217 (34.61)
D4	394 (60.80)	254 (39.20)	374 (59.65)	253 (40.35)
D5	332 (51.23)	316 (48.77)	289 (46.09)	338 (53.91)
Total	1,968 (60.74)	1,272 (39.26)	1,872 (59.71)	1,263 (40.29)

Test of equality of distributions: $\chi^2(31) = 65.7085$, p-value = 0.000

Table 3. Observed behavior versus random behavior, in percent

	On-the-border design			Off-the-border design		
Theory (# of patterns)	Observed	Random	z-score	Observed	Random	z-score
Betweenness (8)	38.58	25.00	7.98	38.92	25.00	8.05
Neilson (1992) (5;2)	26.70	15.63	7.76	13.56	6.25	7.56
Diecidue et al. (2004) (6;2)	27.62	18.75	5.79	13.56	6.25	7.56
CPT w/convex weights (13)	61.42	40.63	10.78	52.47	40.63	12.48
CPT w/concave weights (13)	40.43	40.63	-0.10	44.50	40.63	1.97
Expected utility (2)	16.51	6.25	10.79	13.56	6.25	7.56

If the number of predicted patterns differ in the on-border and off-border designs, they are listed as (x;y). The z-score is the test statistic comparing observed to random behavior for the on-border or off-border designs.

Table 4. Pairwise comparison of theories using Vuong (1996) test (N = 1,275)

	Betweenness	Diecidue et al. (2004)	Neilson (1992)	CPT w/convex weights	CPT w/concave weights
Diecidue et al. (2004)	-2.541**				
	(0.011)				
Neilson (1992)	-2.389**	0.634			
	(0.017)	(0.526)			
CPT w/convex weights	-1.456	1.434	1.276		
	(0.146)	(0.152)	(0.202)		
CPT w/concave weights	-3.340***	-0.599	-0.800	-1.809*	
	(0.001)	(0.549)	(0.424)	(0.070)	
Expected utility	-1.985**	1.199	0.873	-0.827	1.400
	(0.047)	(0.231)	(0.383)	(0.408)	(0.162)

Reported numbers in the table are the test statistic comparing the row theory to the column theory. Positive numbers mean the row theory is better at explaining the data than the column theory. p-value in parentheses. *** p<0.01, ** p<0.05, * p<0.10

Table 5. Linear probability model of rationality as defined by various theories

Variables	Number of safe decisions	Betweenness	Neilson (1992)	Preference for certainty	CPT convex w.	CPT concave w.	Expected utility
Male	-0.112	-0.009	-0.038	0.007	-0.001	-0.000	-0.001
	[0.077]	[0.031]	[0.031]	[0.025]	[0.025]	[0.031]	[0.022]
Black	-0.040	-0.061*	-0.058	-0.012	-0.026	0.031	-0.025
	[0.090]	[0.036]	[0.036]	[0.029]	[0.029]	[0.036]	[0.026]
Age in years	-0.230***	-0.041	-0.023	-0.048*	-0.056**	-0.032	-0.057**
	[0.078]	[0.031]	[0.031]	[0.025]	[0.025]	[0.031]	[0.022]
Two-parent household	0.043	-0.045	-0.007	-0.013	-0.021	0.001	-0.012
	[0.088]	[0.035]	[0.035]	[0.028]	[0.028]	[0.035]	[0.025]
Mother's highest education is high school	0.043	0.008	0.073*	0.031	-0.003	0.077*	0.023
	[0.105]	[0.042]	[0.042]	[0.034]	[0.034]	[0.043]	[0.030]
Father's highest education is high school	-0.042	-0.045	-0.073*	-0.062**	-0.054*	-0.056	-0.064**
	[0.098]	[0.039]	[0.039]	[0.031]	[0.031]	[0.040]	[0.028]
Mother's highest education is college	-0.021	0.036	0.091*	0.009	-0.023	0.051	0.011
	[0.118]	[0.047]	[0.047]	[0.037]	[0.038]	[0.048]	[0.034]
Father's highest education is college	0.150	-0.020	0.000	-0.011	0.000	-0.024	-0.025
	[0.116]	[0.046]	[0.046]	[0.037]	[0.037]	[0.047]	[0.033]
First born	-0.029	0.035	0.040	0.042	0.031	-0.009	0.028
	[0.092]	[0.037]	[0.037]	[0.029]	[0.030]	[0.037]	[0.026]
No older sibling in household	0.052	-0.046	-0.011	-0.016	-0.012	0.047	-0.007
	[0.091]	[0.036]	[0.036]	[0.029]	[0.029]	[0.037]	[0.026]
Math score (8th grade)	-0.001	0.001	0.000	0.000	0.000	-0.000	0.000
	[0.002]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Reading score (8th grade)	0.005**	0.001	0.000	0.000	0.001	-0.000	0.001
	[0.002]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Free/reduced price meal	0.025	-0.043	0.024	-0.056*	-0.041	0.018	-0.037
	[0.096]	[0.038]	[0.038]	[0.031]	[0.031]	[0.039]	[0.028]
Constant	2.703	-0.761	0.175	0.332	0.349	1.328	-0.017
	[2.335]	[0.932]	[0.932]	[0.743]	[0.750]	[0.943]	[0.672]
Observations	1,065	1,065	1,065	1,065	1,065	1,065	1,065
R-squared	0.024	0.021	0.015	0.018	0.018	0.009	0.024
Number of classrooms	65	65	65	65	65	65	65

Dependent variable in columns 2-7 equals 1 if the observed decision pattern is predicted by the theory. Fixed effects at the classroom level. Robust standard errors in brackets. Off-border design is controlled for with classroom fixed effects. *** p<0.01, ** p<0.05, * p<0.10.

Table 6. Fixed effects negative binomial regression on number of disciplinary referrals using number of safe decisions as risk measure

	8th grade	9th grade	8th & 9th grade
Number of safe decisions	-0.050	-0.067	-0.057*
	[0.034]	[0.044]	[0.032]
Consistent with Expected Utility	0.146	0.238	0.161
	[0.124]	[0.158]	[0.117]
Disciplinary referrals (7th grade)	0.060***	0.054***	0.055***
	[0.005]	[0.006]	[0.004]
Male	0.449***	0.143	0.288***
	[0.079]	[0.101]	[0.074]
Black	0.259***	0.586***	0.355***
	[0.093]	[0.121]	[0.088]
Age in years	-0.134*	-0.085	-0.134*
	[0.077]	[0.099]	[0.073]
Two-parent household	-0.193**	0.022	-0.146*
	[0.084]	[0.110]	[0.080]
Mother's highest education	-0.041	0.007	-0.015
is high school	[0.103]	[0.132]	[0.097]
Father's highest education	-0.048	-0.112	-0.074
is high school	[0.098]	[0.126]	[0.093]
Mother's highest education	0.026	0.079	0.023
is college	[0.119]	[0.153]	[0.113]
Father's highest education	-0.211*	-0.267*	-0.158
is college	[0.123]	[0.160]	[0.115]
First born	-0.210**	-0.230*	-0.217**
	[0.097]	[0.123]	[0.090]
No older siblings in household	-0.174*	-0.046	-0.132
	[0.093]	[0.119]	[0.087]
Math score (8th grade)	-0.010***	-0.009***	-0.011***
	[0.002]	[0.003]	[0.002]
Reading score (8th grade)	-0.003	-0.004	-0.003
	[0.003]	[0.003]	[0.002]
Free/reduced price meal	0.277***	0.328**	0.261***
	[0.106]	[0.139]	[0.099]
Constant	11.599***	10.462***	12.144***
	[2.504]	[3.191]	[2.338]
Observations	1,060	1,055	1,060
Number of classrooms	62	60	62
Log-Likelihood	-1453	-1145	-1833

Fixed effects at the classroom level. Robust standard errors in brackets. Off-border design is controlled for with classroom fixed effects. *** p<0.01, ** p<0.05, * p<0.10.

Table 7. Fixed effects negative binomial regression on number of disciplinary referrals using each individual decision separately as risk measure. Numbers in table are coefficients on each decision.

	8th grade	9th grade	8th & 9th grade
Risk taking in D1	0.156*	0.082	0.133*
	[0.083]	[0.108]	[0.079]
Risk taking in D2	0.018	0.132	0.044
	[0.083]	[0.107]	[0.079]
Risk taking in D3	0.030	0.013	0.040
	[0.084]	[0.109]	[0.079]
Risk taking in D4	0.098	0.144	0.120
	[0.081]	[0.105]	[0.077]
Risk taking in D5	-0.017	0.010	-0.017
	[0.084]	[0.109]	[0.079]
Covariates inlcuded?	yes	yes	yes
Observations	1,060	1,055	1,060
Number of classrooms	62	60	62

Full regression results are reported in Tables A9-A11 in Appendix A. Fixed effects at the classroom level. Robust standard errors in brackets. Off-border design is controlled for with classroom fixed effects. *** p<0.01, ** p<0.05, * p<0.10.

Table 8. Fixed effects negative binomial regression on disciplinary referrals using estimated risk measure

	8th grade	9th grade	8th & 9th grade
$Pr(AAAAA\|Choice)^{+}$	-0.464**	-0.651**	-0.584***
	[0.231]	[0.299]	[0.213]
Consistent with Expected Utility	0.098	0.172	0.104
	[0.118]	[0.150]	[0.111]
Disciplinary referrals 7th	0.061***	0.055***	0.056***
	[0.005]	[0.006]	[0.004]
Male	0.439***	0.128	0.274***
	[0.079]	[0.102]	[0.074]
Black	0.300***	0.648***	0.408***
	[0.096]	[0.125]	[0.090]
Age in years	-0.134*	-0.080	-0.133*
	[0.077]	[0.099]	[0.073]
Two-parent household	-0.194**	0.024	-0.145*
	[0.083]	[0.109]	[0.079]
Mother's highest education	-0.021	0.037	0.009
is high school	[0.104]	[0.134]	[0.097]
Father's highest education	-0.012	-0.064	-0.030
is high school	[0.099]	[0.128]	[0.094]
Mother's highest education	0.042	0.103	0.042
is college	[0.119]	[0.154]	[0.113]
Father's highest education	-0.194	-0.242	-0.135
is college	[0.124]	[0.161]	[0.116]
First born	-0.257**	-0.290**	-0.273***
	[0.100]	[0.127]	[0.093]
No older siblings in household	-0.152	-0.023	-0.106
	[0.093]	[0.120]	[0.087]
Math score (8th grade)	-0.011***	-0.011***	-0.012***
	[0.002]	[0.003]	[0.002]
Reading score (8th grade)	-0.002	-0.003	-0.002
	[0.003]	[0.003]	[0.002]
Free/reduced price meal	0.281***	0.335**	0.264***
	[0.106]	[0.139]	[0.099]
Constant	12.199***	11.023***	12.809***
	[2.510]	[3.194]	[2.342]
Observations	1,060	1,055	1,060
Number of classrooms	62	60	62
Log-Likelihood	-1452	-1143	-1831

$^{+}Pr(Choice) = Pr(Choice|\alpha,\epsilon)\alpha + Pr(Choice|1-\alpha,\epsilon)(1-\alpha)$, $\alpha = Pr(AAAAA)$. $Pr(AAAAA|Choice) = \frac{Pr(Choice|\alpha,\epsilon)\alpha}{Pr(Choice)}$.

Fixed effects at the classroom level. Robust standard errors in brackets. Off-border design is controlled for with classroom fixed effects. *** p<0.01, ** p<0.05, * p<0.10.