Dec, 2025
GRINS DISCUSSION PAPER SERIES DP N° 105/2026 ISSN 3035-5576

GRINS

Exploring Choice Errors in Children

DP N° 105/2026

Authors:
Carla Guerriero, Valentino Dardanoni, Marco Mariotti, Paola Manzini, Daniele Caliari

Dec, 2025

GRINS DISCUSSION PAPER SERIES DP N° 105/2026 ISSN 3035-5576

Exploring Choice Errors in Children

Carla Guerriero, Valentino Dardanoni, Marco Mariotti, Paola Manzini, Daniele Caliari

KEYWORDS

correlation neglect | bounded rationality | violations of first order stochastic dominance

JEL CODE

D01, D90

ACKNOWLEDGEMENTS

Acknowledgments: We thank Uta Bolt, Dorothea Kuebler, Lorenzo Neri, Elia Sartori, Nicola Persico and Ivan Soraperra for insightful comments. Valentino Dardanoni and Carla Guerriero acknowledge funding from the European Union's Next Generation EU program under the GRINS - Growing Resilient, Inclusive, and Sustain- able Project (GRINS PE00000018– CUP E63C22002140007). Valentino Dardanoni also acknowledges support from PRIN 2022 (Prot.20222Z3CR7) . Carla Guerriero is additionally supported by PRIN 2022 (Prot.2022KL4J4J). The study was approved by the Ethical Committee of the University Federico II of Naples: ref. 335/17/ESPA5+ICF2. Paola Manzini acknowledges support from the British Academy under grant SRG20\201008 for the work reported with adult women, ethically approved by the University of Sussex n. ER/PM392/1 (and subsequently confirmed by the University of Bristol).

CITE THIS WORK

Author(s): Carla Guerriero, Valentino Dardanoni, Marco Mariotti, Paola Manzini, Daniele Caliari. Title: Exploring Choice Errors in Children. Publication Date: 2025.

Mistakes in decision-making can have serious consequences; while adult choice behaviour is well studied, much less is known about how children make decisions.
We study experimentally how children's ability to avoid choice errors develops over time, focusing on both riskless and risky decisions among primary school children.
Unsurprisingly, inconsistent or erroneous choices in younger children abound. However, by ages 10–11, some display error rates comparable to adults.
We also identify a framing effect—correlation neglect—previously only studied in adults.
Our findings are consistent with a limited stochastic consideration choice model, reflecting the bounded nature of children's decision-making.

Exploring Choice Errors in Children

Daniele Caliari
Wissenschaftszentrum, Berlin

Valentino Dardanoni
University of Palermo

Carla Guerriero
University of Naples - Federico II

Paola Manzini
University of Bristol

Marco Mariotti
Queen Mary University of London

This version: December 2025

Abstract

Mistakes in decision making can have serious consequences; while adult choice behaviour is well studied, much less is known about how children make decisions. We study experimentally how children's ability to avoid choice errors develops over time, focusing on both riskless and risky decisions among primary school children. Unsurprisingly, inconsistent or erroneous choices in younger children abound. However, by ages 10–11, some display error rates comparable to adults. We also identify a framing effect—correlation neglect—previously only studied in adults. Our findings are consistent with a limited stochastic consideration choice model, reflecting the bounded nature of children's decision-making.

Keywords: correlation neglect, bounded rationality, violations of first order stochastic dominance.
JEL Codes: D01; D90

Acknowledgments: We thank Uta Bolt, Dorothea Kuebler, Lorenzo Neri, Elia Sartori, Nicola Persico and Ivan Soraperra for insightful comments. Valentino Dardanoni and Carla Guerriero acknowledge funding from the European Union's Next Generation EU program under the GRINS - Growing Resilient, Inclusive, and Sustainable Project (GRINS PE00000018– CUP E63C22002140007). Valentino Dardanoni also acknowledges support from PRIN 2022 (Prot.20222Z3CR7) . Carla Guerriero is additionally supported by PRIN 2022 (Prot.2022KL4J4J). The study was approved by the Ethical Committee of the University Federico II of Naples: ref. 335/17/ESPA5+ICF2. Paola Manzini acknowledges support from the British Academy under grant SRG20\201008 for the work reported with adult women, ethically approved by the University of Sussex n. ER/PM392/1 (and subsequently confirmed by the University of Bristol).

1 Introduction

Rationality is a cornerstone assumption in economics. Yet increasing evidence suggests that people's response to incentives may be influenced by various cognitive biases, and economic policies designed for rational agents may fall short of their intended outcomes. Most of he experimental economics literature providing plentiful evidence of rationality failures concerns adults; in this paper, instead, we focus on children.

We investigate failures of rationality as *objective* choice errors that can make individuals worse off: understanding their extent is important in general. In the case of children it is particularly critical, for three key reasons. First, studying rationality in childhood may inform our understanding of adult behaviour by uncovering if and how (bounded) rationality evolves with age/schooling or persists into adulthood. Second, there is increasing recognition of children's relevance as autonomous economic actors.[1] Third, there is mounting evidence that the development of cognitive skills in childhood predicts economically relevant adult outcomes.[2] Despite this, evidence on children's choice behaviour remains underdeveloped. In this paper we examine choice errors in primary school age children. We use our experimental design to explore the interplay between different classes of errors: *(i)* intransitive choice, *(ii)* menu effects, *(iii)* violations of first order stochastic dominance and *(iv)* correlation neglect. These rationality violations are problematic: cyclical choices can turn a decision maker into a money pump;[3] menu effects imply that the decision maker can be manipulated into choosing what the menu designer prefers; and failures of (stochastic) dominance lead to choosing inferior outcomes.

These errors are objective because they are independent of preference heterogeneity, i.e. different likelihoods of making errors lead to different choices even when preferences are identical. Therefore, heterogeneity in errors must be distinguished from heterogeneity in preferences as a driver or choice. To this end, we introduce a structural model of choice that allows separate identification of these two forms of heterogeneity (Manzini & Mariotti, 2014; Dardanoni et al., 2023). and provide a measure of the development of rationality which can be estimated. The model posits that each item in a menu is considered not necessarily with certainty, but with some probability. The latter, interpreted as a consideration parameter, serves as a proxy for error avoidance: when preferences are stable, a fully rational agent is represented as having full consideration (i.e. consideration parameter close to unity). At the other extreme,

[1] For instance, Calvert (2008) found that already in 2004, US children under 12 years of age spent 25 billion dollars and might influence additional $200 billion purchases per year - the direct influence on household spending is also increasing, see e.g. Dauphin et al. (2011).

[2] As an example, Gill & Prowse (2024) find that greater creativity (i.e. the ability to produce novel ideas to solve problems) in childhood predicts higher levels of educational attainment, higher occupational categories and higher earnings when adults.

[3] A "money pump" arises when an agent with cyclical preferences is exploited by being enticed into a sequence of trades against money that make him poorer by taking him back to the original allocation. E.g. if the agent prefers item x_1 to item x_2, item x_2 to item x_3, and item x_3 to item x_1, then he can be enticed into buying $x_i - 1$ when he is holding x_i (modulo 3), ending up in the starting position every third trade, but with less money.

an agent choosing each alternative with equal probability is represented as not considering anything (consideration parameter close to zero), and choosing randomly.

We generate choice data by running an experiment in the field with primary school-age children across four Italian state schools, selected from catchment areas with different socioeconomic status. We use this novel dataset to estimate both preference and error-making heterogeneity as well as the rationality parameter, tracking how it evolves with schooling.[4] Our simple settings span two (ubiquitous) domains: riskless and risky choice. In riskless choice, we ask children to pick their favourite one from a menu of options. Here, violations of standard utility maximising behavior occur due to the failure of either of the two constituent components of the Weak Axiom of Revealed Preferences (WARP): *choice acyclicity* (i.e. binary choice must be transitive) and *menu independence* (i.e. a 'winning' option in binary comparisons – Condorcet winner – must be chosen in the union of the binary sets). To test both these requirements, unlike previous literature (Harbaugh et al., 2001; Brocas et al., 2019b), we elicit choices from all the subsets drawn from the same universal set of items (i.e. the full choice function). We then design a new elicitation method to study errors in choice under risk operationalised as consistency with first-order stochastic dominance (FOSD), according to which a lottery that pays more than another regardless of the outcome should always be preferred. With this design, we can also investigate a specific type of framing effect consistent with correlation neglect (i.e. overvaluing the probability of the union of two events by discounting their common source/correlation).

Our contribution. The novelty of our contribution can be summarised as follows:

1. unlike previous children studies, in riskless choice we elicit the *full choice function*. This way we can study how the two types of rationality failures that can arise (menu effects and intransitive choice) interact, which isn't possible within the traditional settings of either choice from budgets or binary choices;

2. in risky choice we design and implement a new simple experimental setup to test for violations of first order stochastic dominance, and document an important choice 'anomaly', correlation neglect, as yet not studied in children;

3. we study the interrelation between the four types of rationality violations;

4. we develop and estimate a structural model of children's choice behaviour, and single out a parameter that quantifies and tracks error avoidance as children grow.

[4]Cognitive psychology focuses on the various developmental stages of 'executive functions', i.e. the various cognitive processes driving directed action, such as inhibition, working memory, and cognitive flexibility (see e.g. Diamond (2013) on executive functions in general, and Ferguson et al. (2021) on the development of executive functions). Our more modest ambition is to identify and measure an index of (a specific type of) rationality in choice, one of the most fundamental economic contexts.

In broad terms, we find that errors are pervasive, with 50% of the children failing the standard tenets of rationality in economics and doing so across domains, but that errors in the older children (10-11 years old) are already comparable to those seen in adults.[5] Error rates fall as age increases and as socioeconomic status (proxied by the school's location) improves. While the extant literature on cognitive skills generally concludes that these persist throughout life (see e.g. Attanasio et al. (2022)) in our sample we find that the rationality gap (i.e. difference in error rates) between children from more and less privileged backgrounds reduces with years of schooling. On the other hand, results from so-called 'fluid intelligence' (namely children's performances in Raven matrices) are less clear cut as we find that performance between children with different socio-economic backgrounds diverges with age.[6] Diving deeper, in the riskless domain, we find that while menu effects are more prevalent than intransitivities in *levels*, violations of transitivity are more prevalent than menu effects once we control for the number of choice patterns that can generate such rationality violations. More importantly, intransitive binary choice is predictive, in the sense of statistic association, of menu effects. Consequently dismissing cyclical binary choice as a 'tremble' risks overlooking the associated choice inconsistencies in larger sets. Turning to the risky domain, we ask children to choose from outcome equivalent lotteries in two frames. In one, a lottery dominates transparently the other ones, while in the other frame the presentation makes this dominance less obvious, and FOSD violations can take the form of correlation neglect. We find that correlation neglect is significant and persistent (though as noted error rates decrease with age), and it is associated with FOSD violations observed in the transparent frame. Moreover, it is independent of menu effects in the riskless choice tasks, and so are FOSD violations: menu effects, framing effects and violations of first order stochastic dominance are different, independent form of mistakes. Finally, we show that our structural model can accommodate the choice patterns related to correlation neglect we observe in both younger and older children, and provide an interpretation through the lenses of the theory of consideration sets.

[5]In both riskless and risky choice, comparisons with the adult population are partly based on existing results in the literature (Harbaugh et al., 2001; Brocas et al., 2019b; Rees-Jones et al., 2024), partly on a replication of our experiments with a subpopulation of adults (middle-aged women) - more details in section 4.3 and Appendix E.

[6]This evidence aligns with the existing literature on early educational intervention (Heckman, 2006), which shows that the window of opportunity to intervene is relatively narrow, and on the short-term effect of moving to wealthier neighbourhoods (Fryer Jr & Katz, 2013). Neighbourhood effects have been the subject of an enormous amount of literature in economics, sociology, and psychology. Comprehensive reviews are Leventhal & Brooks-Gunn (2000); Durlauf (2004); Chyn & Katz (2021). Our data do not allow us to investigate how these gaps persist over subsequent years of compulsory schooling or even more in the long-term. This is a point worth exploring in the future, but outside the scope of this paper, and it would contribute to existing evidence on long-term (educational) gaps (see e.g. Oreopoulos (2003), Deming (2009), Chetty et al. (2016)). We are also agnostic about the nurture-nature roles (see Muslimova et al. (2020), Houmark et al. (2024); for some recent contributions) in determining our measures of rationality.

1.1 Related literature

The cognitive psychology literature on child development is extensive - as in broad lines the approach is different from economics, we refer the reader to standard references, as for instance Siegler et al. (2024). We only note here that our structural model is broadly aligned with the information processing approaches to child development modelling that focus on the role of attentional control in children's decision making (see e.g. Field & Lester (2010) and Steinbeis (2023)). Turning to the economics literature, Our paper fits the recent small but expanding literature on experiments with children. Recent comprehensive reviews are Sutter et al. (2019) and List et al. (2023) - the papers closest to our study are Brocas et al. (2019b) and Harbaugh et al. (2001).[7] The latter investigates rationality in children intended as consistency in choice out of discretised budget sets. A total of 74 children of different ages (either around 7 or around 11 years old) were presented with choices from either three or seven bundles, each consisting of bags of crisps and fruit juice cartons. This setup made it possible to investigate violations of the Generalised Axiom of Revealed Preferences (GARP): in broad terms, GARP rules out preference reversals in chains of choices, or equivalently violations of transitivity. As noted in Brocas et al. (2019b), GARP is an indirect test of transitivity, as children are not asked to make direct binary comparisons and thereby reveal their base preference. Binary choices are indeed the object of study in Brocas et al. (2019b) (in riskless, risky, and social/group choices), with 134 children in a similar age group to ours (from kindergarten to 5th grade, ages about 5–11) from an affluent bilingual private school.[8] However, these do not allow checking for the presence of menu effects, as we do (while we do not cover the social preference dimension). In our contribution we go beyond these important papers by looking at *both* intransitivity of choice and menu effects in a larger and more socioeconomic diverse children population, finding a non obvious and so far unknown interplay between violations of these two constituent parts of rationality; this is only possible by eliciting the whole choice function from respondents.

As we also investigate violations of first-order stochastic dominance (FOSD) and the correlation neglect phenomenon,[9] our study is also related to the experimental literature on these topics, absent from the extant work on children. Virtually all the literature documenting FOSD violations relies on covert implementation, with questions framed so that the FOSD relation

[7] This literature has mainly focused on the evolution of strategic thinking (Czermak et al., 2016; Brocas & Carrillo, 2020a,b, 2021, 2022; Fe et al., 2022) and the theory of mind (Sher et al., 2014; Charness et al., 2019; Fe et al., 2022), while less on individual decision-making. There are several interesting papers in economics that look at children's learning, as for instance Apesteguia et al. (2018) and Barash et al. (2019); exploring how children learn (or not) is beyond the scope of this paper.

[8] Other papers study risk preferences in children; Castillo et al. (2018) too has children choosing from binary sets, with subjects being older children (average age 13.78) than our experimental cohort. A recent study of risk preferences in children is Piovesan & Willadsen (2021), which also focuses on binary choices in the lottery domains and enriches the analysis by also investigating personality traits. Both these studies focus on risk preferences, while our focus is on choice errors.

[9] Our setup is such that more sophisticated forms of misperception of correlation do not apply. For instance, our experimental subjects violate the Weak Monotonicity axiom in Ellis & Piccione (2017).

isn't apparent - two notable exceptions are Charness et al. (2007) and Agranov & Ortoleva (2017).[10] Interestingly, both papers find violations of FOSD in university students to be low, and much lower than in our findings for children. The literature on correlation neglect is arguably less mature but developing rapidly, possibly as it crops up in a wide range of settings[11] and is particularly problematic in view of the large amount of correlated information present on all pervasive social media. The experimental literature on correlation neglect (see e.g. Eyster & Weizsacker (2016), Enke & Zimmermann (2019), Rees-Jones et al. (2024), Vitaku (2024)), is also focused entirely on the adult population, which is unsurprising given the intrinsic complexity of the notion of correlation. The simplest setup, which is closest to ours, is used in Rees-Jones et al. (2024), who consider the case of school choice. They simulate in the lab the case of applicants to university places. Entry is gained on the basis of an applicant's 'grade' (drawn from a uniform distribution). Before knowing their grade, subjects have to rank universities they wish to be matched to in order of preference, with more desirable universities (i.e. providing higher payoff in case of a match) having higher entry requirements than less desirable ones. When admittance is gained on the basis of a single priority score (as for instance in the case of admission based on GPA), if the top two schools differ only slightly in entry requirements, risk-neutral and risk-averse applicants should select the most and least selective schools, yet Rees-Jones et al. (2024) document that a large percentage of experimental subjects ignore correlation and apply to the top two schools. Indeed, roughly half of their participants display correlation neglect. This is similar to the choice behaviour we observe in our older children, as we show below.

2 Experimental design

As noted, in our experiment we wished to explore four types of choice mistakes: cyclical choice and menu effects in riskless choice; failure of stochastic dominance, and correlation neglect in risky choice. We describe our approach to testing each of these below.

2.1 Errors in riskless choice.

We start with the simplest possible setup for deterministic choice: pick one from a menu of (equally valued) items. In such a setting, rationality takes the form of a particular type of consistency, which is equivalent to each subject being able to rank all items from best to worst

[10] In Li (2017) terminology, our presentation is such that the best strategy is 'obviously dominant'.

[11] These span as diverse context as health (Agarwal et al. (2023) find that professional radiologists using AI assistive technology underperform as they fail to recognise the dependence between their own information and AI predictions), job search (e.g. Huang et al. (2025) find that prospective job applicants to the Chinese civil service overweight multiple negative signals regarding the job even when they evidently originate from a common source), and voting (e.g. Ortoleva & Snowberg (2015) find that correlation neglect generates overconfidence in voters, pushing them towards more extreme parties).

and consistently pick the best available item from each choice problem; that is, subjects choose the preference-maximising item. Whenever this fails, an error is produced. Error-free choice patterns are those that satisfy the **Weak Axiom of Revealed Preferences (WARP)**. This is a consistency property requiring that if an alternative, call it x, is picked out of a menu, it can never be the case that one of the unchosen items is picked in another menu that also contains x. This implies that all alternatives can be ranked from best to worst, so when choices satisfy WARP, choice behaviour can be described as if a preference is being maximised.

When WARP fails, failures of rationality can be attributed to one of *only two* possible forms of non-rational behaviour: **cyclical choice** or (a particular type of) **menu effects** (as showed in Manzini & Mariotti (2007)). Cyclical choice arises whenever a choice pattern shows that there are at least three items, say x, y and z, such that a subject has picked x in the choice between x and y, y in the choice between y and z, and z in the choice between x and z. This aspect of rationality therefore requires that **No Binary Cycles** (NBC) are observed, which is the term we will use to refer to this property, following Manzini & Mariotti (2007). Menu effects, instead, consist in a pattern of choice such that an alternative is not chosen in a menu where it is the Condorcet winner. So for instance x is picked in the choice between x and y, and also in the choice between x and z, but not from a menu that includes all three alternatives, x y and z. The rationality requirement that is violated in this case is called **Always Chosen** (AC).

In our experiment each child had to pick one item out of each of the 11 possible subsets of a grand set of four alternatives. There were two sets of such tasks (pens and pencils, see Figure 1), so in total 22 choices were elicited.[12] The questions were presented in random order, with a distractor task between each choice.[13] After both sets of 11 questions, we elicited preference intensities for each item on a Likert scale from 1 to 5.[14] Lastly, subjects answered eight Raven's Coloured Progressive Matrices (RCPM), presented in increasing difficulty order.[15] At the end of this stage, one of the choice screens was drawn at random for each set of choices, and the choice on that screen was assigned as the prize (hence, each child won a pencil and a pen). Screenshots for all the tasks are reported in Appendix F. We will refer to this deterministic choice task as the pencils/pens task.

[12]This part of the experiment was administered via tablet computers connected to a local server. The code is available here: *https://github.com/DCaliari/RationalityInChildren.git*

[13]We randomized whether the pupil would face the questions regarding pencils or pens first, and within each set of questions, we also randomised their order and the positions of the alternatives on the screen (children also answered one extra question with four notepads meant for a different exercise, hence excluded from the analysis). The distractor task consisted of a short nature video (about 10 seconds).

[14]This task is similar to the "ranking task" (Brocas et al. (2019b)), where we allow for indifferences. This way we are able to devise a preference intensity index, see Online Appendix D.5 for details.

[15]These matrices were chosen from the set of RCPMs after piloting to guarantee that the answers were better than random for 1st-grade pupils and different from perfect accuracy for 5th-grade pupils.

Figure 1: Sets of pencils and pens.

2.2 Errors in risky choice - testing for failures of first-order stochastic dominance and correlation neglect.

Secondly, we look at the simplest possible setup to investigate children's ability to avoid mistakes when outcomes are stochastic, that is, they take the form of lotteries. Lotteries are arguably harder to evaluate than deterministic prizes; to make the notion of a risky prospect accessible to children, we introduce a novel design in the form of a coin drop game that induces a uniform distribution over the outcomes.[16] More precisely, a token ('coin') is dropped from the top of a sloping wooden board with pins placed in such a way that the coin has an equal probability of ending up in each of the eight pockets at the bottom. A lottery takes the form of a strip of eight squares, either white or yellow, placed below the pockets, see Figure 2.

Figure 2: The coin drop game: the FOSD test lotteries are on the left, the correlation neglect test lotteries are on the right.

A win occurs only if the coin falls in a pocket above a yellow square for the chosen strip/lottery - hence the setup provides an arguably clear visualisation of how a random event maps to outcomes. The lotteries we used to investigate FOSD violations are those in the left panel in Figure

[16]The pin placement follows Akitaya et al. (2018).

2, which have either the five, six, or seven leftmost squares in yellow and the remaining in white. The top lottery is the least likely to result in a win, while the bottom lottery is the most likely to result in a win (in this setup, we have state-wise dominance between the lotteries to make it more easily recognisable). For reasons that will become more apparent in a moment, we refer to this set of three lotteries as 'independent' and to the task as 'independent task/coin drop game'. Children recorded their choice on a sheet picturing of the three lotteries, as in the left panel in Figure 3. The '1' and '0' reinforced that 1 prize (a sheet of stickers) would be won in correspondence with the yellow squares, and no prize otherwise.[17]

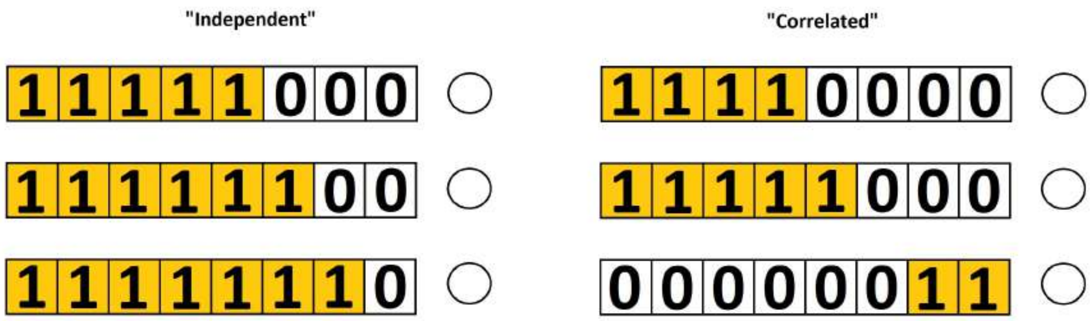

Figure 3: The answer sheets in the two tasks.

The right panels in Figures 2 and 3 display the lotteries that we used to test for correlation neglect. Correlation neglect is the failure to account for the correlation of two events. Suppose, for instance, that university admission relies on the final mark received by an applicant in a state exam and that applications have to be made ahead of the results being known (we borrow this example from Rees-Jones et al. (2024)). Suppose also that each applicant can only apply to two universities. Then there is little point in only applying to elite universities (with stringent entry requirements), since missing the grade for one likely means also missing the grade for the other: unless an applicant is very risk loving, such applicant's behaviour would be disregarding the correlation between the two events (admission to either university depends on the same mark).

We implemented a test for correlation neglect in children using the lotteries depicted in the right panel in Figure 2, which we refer to as 'correlated'. Now children had to pick **two** out of three lotteries (see Figure 3), with a win realising if **any** of the yellow squares **across the two** chosen lotteries corresponded to the pocket with the dropped coin. The strip/lottery at the top had the four leftmost squares in yellow and the other four in white; the middle one had the five

[17]The answers were collected by pen and paper, and the overall experiment was performed with the help of specifically trained research assistants. After the instructions (see appendix F), children recorded their choices on answer sheets (see Figure 3; actual answer sheets did not have the "independent" and "correlated" labels).

leftmost squares in yellow and the three rightmost ones in white; and the bottom one had the two rightmost squares in yellow and the rest in white. Picking the two lotteries with the four or five leftmost squares in yellow attests correlation neglect, for conditional on picking the lottery with the highest probability of a win (the one with five yellow squares), adding the second best (with four yellow squares) brought no additional benefit.

Each pair of lotteries in the correlated task maps into the same probability of winning as a corresponding lottery in the independent task: the top two lotteries in the correlated set produce a win in 5 out of 8 cases, as does the top independent lottery. The bottom two lotteries in the correlated task produce a win in 7 out of 8 cases, as does the bottom lottery in the independent task; and the extreme lotteries in the correlated task produce a win in 6 out of 8 cases, as does the middle lottery in the independent task. Put differently, the two lottery choice problems provide different frames for essentially the same possible outcomes. Hence, we should expect rational agents to choose the same FOSD dominant lottery in both tasks (the bottom lottery in the independent task and the bottom two lotteries in the correlated task).

We ran the independent task prior to the correlated task; being a transparent test of first-order stochastic dominance, the independent task also doubles as a comprehension test. Finally, all three lotteries were played out, and prizes were assigned accordingly.

3 Data

Our sample consists of 676 children enrolled in four primary state schools in different districts of a large city in Italy, Naples.[18] Our sample comprises classes across all years for which we obtained consent to conduct the experiments from the school president, the school board, and the teachers.[19] The schools differ in catchment areas, providing a good degree of socio-economic variability. Our analysis focuses on the two schools at the two extremes of the sampled socioeconomic spectrum, which we denote as **School L** and **School H** (for lowest and highest socioeconomic status), where we collected the majority of our data. School L is located in one of the census areas in Naples with the lowest per capita income, with 37% unemployment rate and only 32% of adults having a high school or bachelor's degree. School L is located in a

[18]Our complete dataset comprises 732 primary school children. Due to COVID-19, the data collection for the pencils/pen task began in May 2022 (8%) and was completed in October-November 2022 (92%). Only 27 pupils repeated this part both in May 2022 and October-November 2022. In our main dataset, we consider only the data collected in October-November 2022 but re-run the analysis considering the first choices of the pupils who repeated the experiment, confirming all results (see Online Appendix I).

[19]Compulsory education in Italy covers all children up to age 16. It comprises five stages, which are nursery, primary school (normally 6-11 years old), lower secondary school (normally 11-14 years old), upper secondary school (normally 14 to 19 years old), and university. Written consent was secured from the parents, and oral consent was obtained from the children. The children were given a brief explanation of the activities and were informed that they could withdraw from the study at any time, though none chose to do so. The experiment was administered by Carla Guerriero, along with a team of trained interviewers (aged 30 to 46), all educated to degree level and with prior experience of conducting scientific research with children. See Appendix F for the script.

privileged part of Pozzuoli, a town within the wider metropolitan area around Naples. We use data from the remaining two Schools, also located in Pozzuoli, in a comprehensive series of robustness checks (see Online Appendices).

As a proxy for socioeconomic status, Figure 4 displays self-reported parental educational achievements in our sample, with the distribution of educational attainment by parents in School H dominating the one for School L. The pens/pencil task was administered in October-November 2022, while the coin drop tasks were administered in March-April 2023. Due to the intrinsic complexity of the lottery tasks, we administered it only to children in years 3-5 inclusive.

	sample size	**% male**	**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**	**parental ed. coverage**
School L	255	49.02	28.24	11.37	16.08	21.57	22.75	77.25
School 2	64	53.13	20.31	21.88	0.00	34.38	23.44	84.38
School 3	113	53.10	17.70	18.58	15.04	19.47	29.20	81.42
School H	244	50	14.75	20.08	20.90	23.36	20.90	84.84
total	676	50.44	20.86	16.72	16.12	23.08	23.22	81.36
L+H	499	49.50	21.64	15.63	18.44	22.44	21.84	80.96

Table 1: Sample size, gender, grades, and survey coverage.

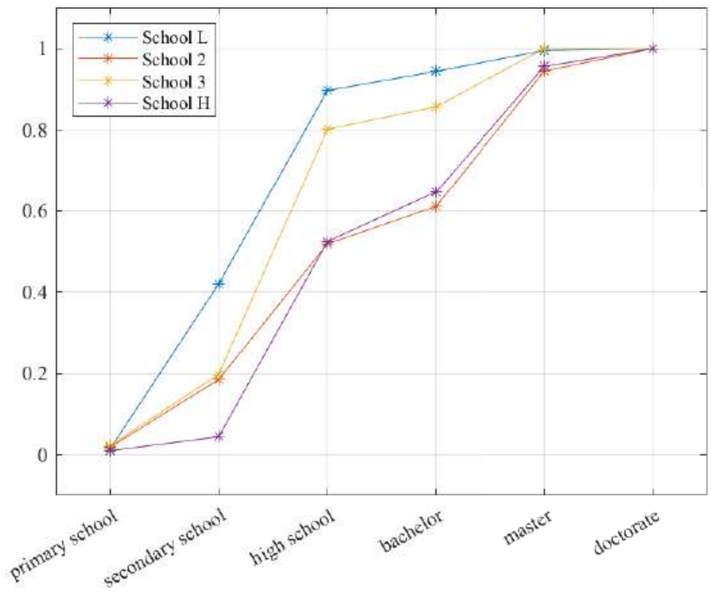

Figure 4: Cumulative distributions of parental education

4 Results

Roadmap. This section is organized as follows. First, in Section 4.1, we present the aggregate evidence on the extent and interplay of the four categories of errors. In Section 4.2, we introduce a model of stochastic consideration that yields an index of rationality, operationalized by the consideration parameter. Ultimately, our aim is to provide estimates for each grade within schools to compare children's development along the socio-economic dimension (Section 4.3). To do so, we turn to showing the predictions of the stochastic consideration set model for the choices in our experiments. We begin with the coin drop games, as each task involves choices from a single menu and with a unique preference ordering, and then move to the pencils/pens choice task, where we expect heterogeneity not just in cognition but also in preferences. In what follows, we try to keep formalism to a minimum - the choice shares in the various tasks and any calculations supporting hypotheses are in Appendix B, while the structural estimation is described in Appendix C. Finally, we conclude by comparing our results on the development of economic rationality with those on the development of fluid intelligence, as proxied by scores on a Raven's matrices task (Section 4.4).

4.1 Aggregate evidence

As discussed, errors in the deterministic choice domain take the form of violations of the No Binary Cycles (NBC) property and/or of the Always Chosen (AC) property; errors in the lottery choice domain take the form of violations of First Order Stochastic Dominance and/or Correlation Neglect.

Choices in the four main tasks are summarised in Figure 5. The first two panels from the left refer to the pencils and pens tasks; the responses of each child are allocated to one cell in the relevant table, depending on which axiom, if any, is violated by that child's choices. The rightmost table presents the joint distribution of choices in the independent and correlated lottery task, where entries '5', '6' and '7' denote the number winning outcomes characterizing the lotteries. Colour intensity maps the relative weight of each cell.

We uncover the following empirical regularities:

(I) **deterministic choice**:

(a) cyclical choice is highly associated to menu effects, in the sense that children who exhibit cyclical choices are more likely to display menu effects (Chi squared test rejects independence, $p < 0.001$); moreover, while menu effects are much more prevalent than cycles in terms of raw data, there are *proportionally* more cycles than menu effects; moreover .

(b) rationality violations in the pen and pencils tasks are highly associated (Chi squared test rejects independence, $p < 0.001$).

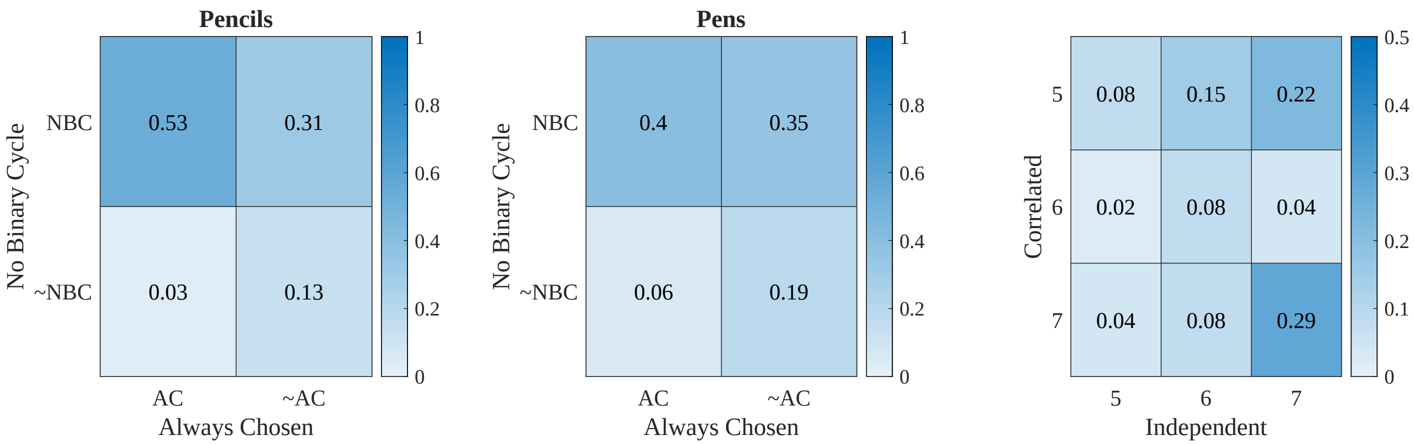

Figure 5: Axiom violations (NBC and AC) and framing effects

(II) **lottery choice**: violations of correlation neglect are very substantial, and larger than violations of first order stochastic dominance. Here too the violations are not independent (Chi squared test rejects independence, $p < 0.001$).

(III) **menu effects vs framing effects**: from (I) and (II) above we know these are significant; what is more, violations in the deterministic choice tasks are independent of violations in the lotteries tasks (Chi-square test has p-value of 0.882 and 0.359 for the independent and correlated lotteries, respectively). In other words, the framing effects observed in the lottery tasks are not associated to the menu effects observed in the deterministic tasks, and a child making rational choices in one set of tasks may make irrational choices in the other set of tasks.

Regarding point (I), the two properties that make up WARP are not independent in our sample: an NBC violator is more likely than not to also violate AC, while an NBC rational subject is more likely than not to also be AC rational. This is the more remarkable when noting that the presence of a cycle in binary choice makes it technically *less likely for menu effects to arise*: this is because a cycle entails fewer alternatives chosen over all others (in a choice cycle each alternative is dominated by at least one other), hence fewer opportunities to violate the axiom.[20]

Our assertion that there are *proportionally* more cycles than violations of AC requires further explanation. In comparing the success rates of axioms we must take into account the number of cases that are theoretically compatible with that axiom. We do so is by computing Selten's measure of predictive success: for a given property P, this measure is defined as the difference

[20] More formally, a conditional statement holds vacuously whenever the premise is false. In our context, without cycles the maximum number of possible AC violations is 5. With cycles, depending on the specific configuration the maximum number of violations is between 2 and 4.

between the proportion a of observations agreeing with the property under investigation and the relative proportion r of cases that are *theoretically compatible* with the property (see Selten (1991)). A property that cannot be disproved (so that $r = 1$) has a Selten measure of zero, since all observations satisfy the property, so that $a = 1$ and $s_P = a - r = 0$, where s_p stands for Selten's measure of predictive success for property P. In our experiment $s_{AC} = 0.515$ and $s_{NBC} = 0.463$ for the pencils task and $s_{AC} = 0.515$ and $s_{NBC} = 0.463$ for the pens task, whence our conclusion that AC performs better than NBC.[21]

Turning to (II), while the options presented in the two coin drop tasks were payoff equivalent, the observed marginal distributions are quite different in the two cases, with the choice distribution in the correlated task being bimodal. Fewer than half of the children make the same choice in both tasks, and fewer than a third of children make the optimal choice in both task. Around 40% of the children who picked the dominant lottery in the independent task display correlation neglect. Choosing the dominated lottery is much more likely in the correlated task than in the independent task, while choosing the dominant lottery is much more likely in the independent task than in the correlated task.

The evidence in support of (III) is reported in Table 2. Whether or not a subject is prone to menu effects bears no relation to either failure of first order stochastic dominance or to the presence of correlation neglect. Because choices in the Independent and Correlated task were equivalent, correlation neglect is a particular instance of framing effects. Therefore in particular menu effects and framing effects are independent of one another, and whether or not deterministic choices are subject to menu effects does not change the relative probabilities of picking one particular lottery. Put it differently, menu effects and framing effects are different and independent forms of error, at least in our context.

		Pencils task (n=272)						**Pens task (n=272)**					
		Independent			**Correlated**			**Independent**			**Correlated**		
		5	6	7	5	6	7	5	6	7	5	6	7
AC	yes	9.5	20.8	38.0	29.4	8.8	30.1	7.7	18.6	28.8	25.4	7.7	22.0
rational	no	5.1	10.2	16.4	14.7	5.5	11.4	6.9	12.4	25.5	18.8	6.6	19.5

Table 2: Menu and Framing effects

Before drilling down into our results, we introduce a theoretical framework that can account for the experimental evidence.

[21] In our setup of 4 alternatives where choices are elicited from six sets with two alternatives, four sets with 3 alternatives and one grand set of four alternatives, we have $2^6 \cdot 3^4 \cdot 4 = 20,736$ possible choice combinations, i.e. choice functions. Of these, $40 \cdot 3^4 \cdot 4$ are the choice functions that exhibit binary cycles, so that $\frac{2^6 \cdot 3^4 \cdot 4 - 40 \cdot 3^4 \cdot 4}{2^6 \cdot 3^4 \cdot 4} = \frac{3}{8} = 37.5\%$ of the possible choice functions satisfy the No Binary Cycles property, while 960 of all possible choice functions, or 4.6% of them, satisfy Always Chosen. Since in the pencils task we observe 83.8% of choice functions satisfying NBC and just over 56% satisfying AC, Selten's measures of predictive success are $s_{NBC} = 0.838 - 0.375 = 0.463$ and $s_{AC} = 0.561 - 0.046 = 0.515$.

4.2 Interpreting the data through a model of limited consideration

Choices displaying menu and/or framing effects cannot be explained via the maximisation of a utility function, no matter how 'rich'. The reason is straightforward: if object a is assigned a higher utility than object b, a utility maximising agent will never pick b over a when both are available. The only way for such choice behaviour to arise is if the alternative with higher utility is somehow out of the picture, for instance due to bounded rationality in the form of limited consideration.

Because they distinguish 'base rationality' (i.e. preferences) from consideration (i.e. focussing on a subset of the options available), consideration set models are well suited to investigate the forms of non fully rational behaviour that we have uncovered. The model we present below enables us to:

1. identify and disentangle heterogeneity in taste from heterogeneity in cognition (seen as limited consideration);
2. identify and estimate an index of rationality, the consideration parameter.
3. tie correlation neglect and limited consideration in the risky setting where preference heterogeneity collapses into FOSD.

Our experimental subjects were tasked with selecting a single option from each choice menu - to model these *discrete choices* we use Horan (2019)'s variant of the Stochastic Consideration Set (SCS) model introduced in Manzini & Mariotti (2014) which is suitable for experiments with so called 'forced choice' (i.e. skipping tasks is not allowed). In the SCS model a decision maker is endowed with a standard rational preference $\succ$, which ranks from better to worst all available alternatives. This agent does pick the best alternative according to her preference, but may fail to consider some alternatives. In other words, the decision maker picks the most preferred alternative *among those considered*. Letting $\gamma_a \in (0,1]$ denote the probability that alternative a is considered, the probability that an agent with preference $\succ$ chooses alternative a from menu A , which we denote by $p^{\succ}(a, A)$, is:

$$p^{\succ}(a, A) = \frac{\gamma_a \prod_{b \in A : b \succ a} (1 - \gamma_b)}{1 - \prod_{b \in A} (1 - \gamma_b)} \tag{1}$$

where at the numerator the first term is the probability that a is considered, and the second term is the probability that no alternative better than a is considered (if considered, any alternative better than a would be picked over a). At the denominator, we have the probability that some alternative is considered (where $\prod_{b \in A} (1 - \gamma_b)$ is the probability that no alternative is considered in set A).

In this model the consideration parameter is a **proxy for rationality**, intended as (absence of) limited consideration. A rational agent considers all alternatives: in the language of our model, this sets $\gamma_a = 1$ for all a. It is immediate from (1) that $p^{\succ}(a, A) = 1$ if and only if a is the preference maximising alternative, with $p^{\succ}(a, A) = 0$ otherwise. At the other extreme, in the limit as the consideration parameter approaches 0, $p^{\succ}(a, A) = \frac{1}{|A|}$, where $|A|$ is the number of alternatives in set A (i.e., its cardinality - the agent picks an alternative at random, with uniform probability).

4.2.1 The coin drop game - independent lotteries task

The choice menu in this task consisted of lotteries differing only in the probabilities of obtaining a prize, ordered naturally by the overall winning probability of each lottery. We use ℓ_n^I to denote the lottery strip with n winning outcomes in the Independent lotteries task, and L^I to label the set of independent lotteries (that is $L^I = \{\ell_5^I, \ell_6^I, \ell_7^I\}$); γ_n^I denotes the consideration probability for the lottery with n winning outcome.

A rational agent has perfect consideration, that is $\gamma_n^I = 1$ for all n. Such agent makes no mistake, and picks the dominant lottery ℓ_7^I with probability 1. Hence as a preliminary test of rationality we have:

Implication 0. *For a rational agent $p\left(\ell_7^I, L^I\right) = 1$ and $p\left(\ell_6^I, L^I\right) = 0 = p\left(\ell_5^I, L^I\right)$.*

Ideally we would like to be able to identify the three consideration parameters, γ_5, γ_6 and γ_7; however under the constraint that choice probabilities must add up to one, we have only two independent choice probabilities in three unknowns, which results in under-identification. However under the assumption that consideration is weakly increasing with the number of winning outcomes, the following holds:

Implication 1. *If the model holds and consideration is monotonically non decreasing in n, then*

$$p\left(\ell_7^I, L^I\right) > p\left(\ell_6^I, L^I\right) > p\left(\ell_5^I, L^I\right) \tag{2}$$

We can achieve identification of the consideration parameter is we assume that it is lottery independent: given the observed probabilities of choice of each lotteries, any of these equations suffices to identify the unobserved consideration parameter γ^I. Then the following condition is necessary for the model, hence a testable prediction:

Implication 2. *If the model holds with $\gamma_n^I = \gamma^I$ for all n, then:*

$$\frac{p\left(\ell_7^I, L^I\right)}{p\left(\ell_6^I, L^I\right)} = \frac{p\left(\ell_6^I, L^I\right)}{p\left(\ell_5^I, L^I\right)} \tag{3}$$

4.2.2 The coin drop game - correlated lotteries task

In the correlated lotteries task subjects pick *two* lotteries. Now the position of the yellow squares matters: so if $L^C = \{\ell_2, \ell_4, \ell_5\}$ is the set of lotteries, the pair $\{\ell_4, \ell_5\}$ is equivalent to a single lottery in which a win occurs 5 times out of eight, which we denote by ℓ_5^C; the pair $\{\ell_4, \ell_2\}$ is equivalent to a lottery that wins 6 times out of eight, labelled ℓ_6^C; and the pair $\{\ell_5, \ell_2\}$ is equivalent to a single lottery that wins 7 times out of eight, labelled ℓ_7^C.

In the absence of framing effects we would expect participants to choose in exactly the same way in the two tasks:

Implication 0′. *In the absence of framing effects* $p\left(\ell_n^C, L^C\right) = p\left(\ell_n^I, L^I\right)$ *for all* n.

However if each lottery in a pair is appraised separately from the other, a pair of lotteries will be selected only if both "component" lotteries are considered and either these are the best combination, or no lottery forming a better combination is considered.[22]

An agent immune to correlation neglect would have $p\left(\ell_7^C, L^C\right) = 1$, while someone succumbing to correlation neglect entirely would have $p\left(\ell_5^C, L^C\right) = 1$; this suggests a simple correlation neglect index ι defined as

$$\iota = \frac{p\left(\ell_7^C, L^C\right) - p\left(\ell_5^C, L^C\right)}{p\left(\ell_7^C, L^C\right) + p\left(\ell_5^C, L^C\right)} \in [-1, 1] \tag{4}$$

In addition, we are able to test the following hypotheses, which are necessary conditions for the model to hold:

Implication 1′. *If the model holds and consideration is monotonically non decreasing in* n*, then lottery* $\ell_6^C = \{\ell_2, \ell_4\}$ *is the least likely to be chosen, that is:*

$$\frac{p\left(\ell_5^C, L^C\right)}{p\left(\ell_6^C, L^C\right)} > 1 \textit{ and } \frac{p\left(\ell_7^C, L^C\right)}{p\left(\ell_6^C, L^C\right)} > 1 \tag{5}$$

Implication 2′. *If the model holds with* $\gamma_n^C = \gamma$ *the probability of choosing* $\ell_6^C = \{\ell_4, \ell_2\}$ *is the same as the probability of choosing* $\ell_5^C = \{\ell_4, \ell_5\}$*, and both are smaller than the probability of choosing* $\ell_7^C = \{\ell_5, \ell_2\}$*, that is:*

$$\frac{p\left(\ell_5^C, L^C\right)}{p\left(\ell_6^C, L^C\right)} = 1 \textit{ and } \frac{p\left(\ell_7^C, L^C\right)}{p\left(\ell_i^C, L^C\right)} > 1, i = 5, 6 \tag{6}$$

4.2.3 Stochastic consideration and choice from multiple sets

The pencils/pens task involved for each subject the elicitation of choices from a collection of sets: for both pencils and pens each subject picked an item from each of 11 possible choice sets.

[22] This differs from the setup in Ellis & Freeman (2024), where the focus is on misunderstanding the integration of separate problems, the usual issue addressed in choice bracketing.

In this context it would be unreasonable to assume that preferences are the same. Because the underlying model is one of individual decision-making, it isn't possible to test the axioms with our data. We can, however, test some of the assumptions behind our estimates. We use $\pi(\succ)$ to denote the probability that a decision maker has preference $\succ$ (her *type*). We postulate that subjects select the preferred alternative across those considered, according to model (1). Each combination of 11 choices constitutes a complete *choice function*, i.e. a map C assigning a single item to each choice menu; hence for each experimental subject we collect two complete choice functions, one for pencils and one for the pens.

We wish to estimate the distribution of preferences and the consideration parameters for each subject from our data set, the frequency distribution of these observed choice functions. Such *Mixture Choice Function*, which we denote by μ, combines heterogeneity in both tastes and consideration (Dardanoni et al., 2023).

Under the assumption that preferences $\succ$ are independent of consideration, and that choices at each menu are independent, the probability $\mu(C)$ of observing a given subject making choices according to C can be expressed as

$$\mu(C) = \sum_{\succ \in P} \pi(\succ) \prod_{A \in \mathcal{A}} p^{\succ}(C(A), A)$$

where $p^{\succ}(C(A), A)$ is a shorthand for the probability that item $C(A)$ is selected in choice set A according to equation (1). So in addition to the consideration parameters γ_a for the various choice objects (pens and pencils) we also wish to estimate $\pi(\succ)$ for each of the possible 24 preference ordering. This is only possible however is the model is identified, which turns out to be the case:

Proposition 1. *Let $\gamma_a = \gamma$ for all $a \in A$. Then the distribution of preference types and the (type-dependent) consideration parameters in the Conditional Stochastic Consideration Set model are generically identified from our mixture choice data.*

Here we outline the logic for our proof, relegating details to Appendix D. It relies on arguments in Allman et al. (2009) (henceforth AMR), who prove the generic identifiability of a general class of latent class models (of which ours is a special case) when parameters are unrestricted. We cannot invoke their result directly because imposing a choice model means restricting the range in which parameters can change, which could result in a singularity. This point applies to one specific step of the proof in AMR, which requires showing the generic invertibility of some matrices. We prove such invertibility for our model, which then ensures that the rest of the proof by AMR also holds in our case. Then since the theoretical model is identified from the collection of binary and ternary sets (see Horan (2019)), we know that we can use the type conditional probabilities identified in the previous "step" to retrieve the model's "deep parameters" (preference and consideration parameters) for each type from our dataset/mixture choice function. Once we have established identification, we can estimate our model using the maximum likelihood method (see Appendix C for details).

The model makes the arguably strong assumption that rationality is independent of preferences. Exploiting variation in preferences along the gender dimension, which we expected given our choice of the animal and color characteristics of pencils/pens, we can test this assumption. Namely, if the independence hypothesis holds, we should expect the consideration parameter to be invariant with respect to gender, regardless of preference heterogeneity. If independence fails, preferences and the consideration parameter become entangled, and, in the presence of preference heterogeneity, differences across gender may arise. We can therefore formulate the following:

Implication 3. *If the model holds, gender differences in preferences should not imply differences in the consideration parameter.*

4.3 The development of rationality

4.3.1 Errors in the Pencils task

In the main text, we focus on the pencils task while relegating the pens task to Online Appendix E, given the highly overlapping results. In light of Proposition 1, we estimated a single consideration parameter in various subpopulations based on grade, school, and gender. Our results and plotted in Figures 6 and 7.

Figure 6 reports, on the left, the distribution of the preference types at the school level and, on the right, the same distribution by gender. The heading 'ABCD' is a shorthand for the preference ranking A first, then B, then C, then D. The letters in the table are the initials of the eraser on top of each pencil (Duck, Ladybird, Frog, and Shark).

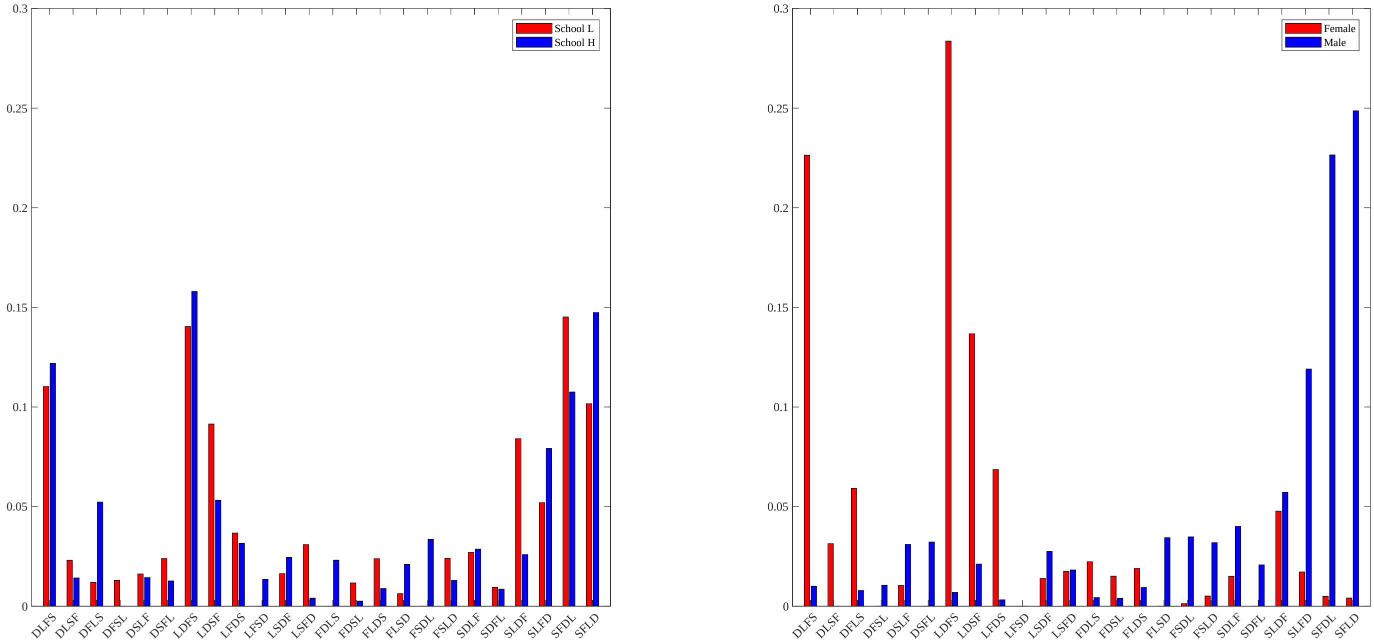

Figure 6: Distribution of the preference types by school and gender (pencils task).

Figure 7 plots the estimated consideration parameters, which, as discussed, act as a proxy for the ability to avoid errors.

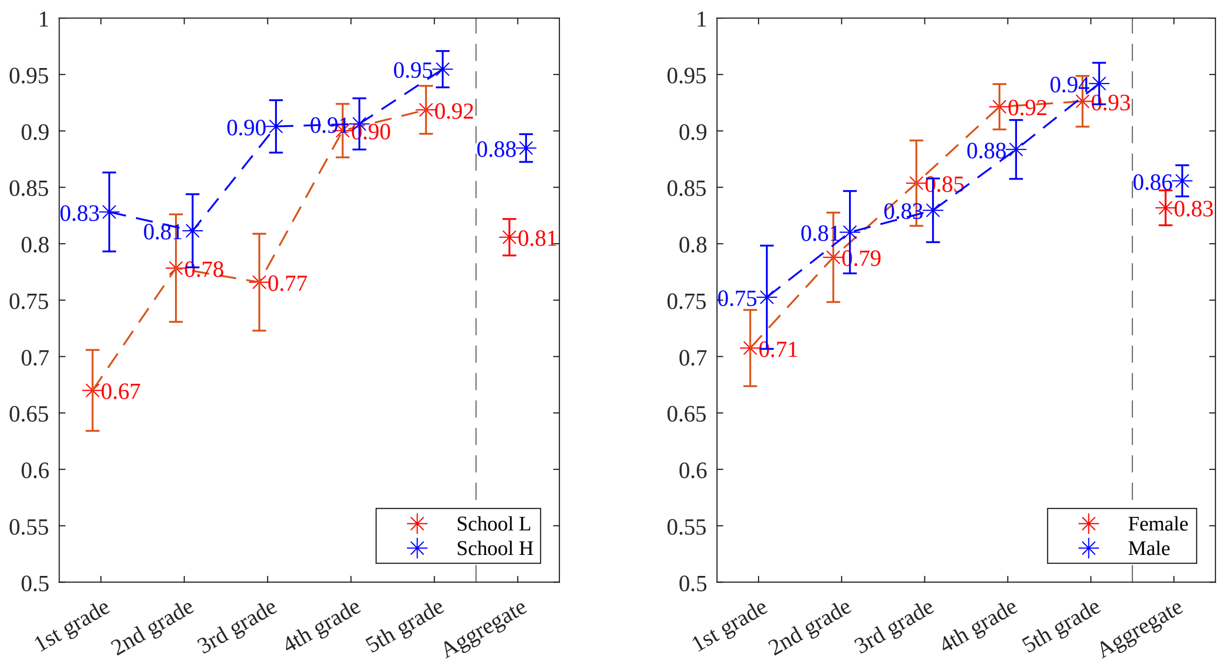

Figure 7: Development of the consideration parameter by school and gender (pencils task)

Figures 6 and 7 highlight the main strength of our methodology: preferences and consideration parameters are disentangled, giving us interesting insights into the heterogeneity of children's choices across the socio-economic and gender dimensions. As expected, preferences are different along the gender dimension (chi-square test, p-value < 0.01), with boys preferring the Shark and girls preferring the Duck and Ladybird, while they are remarkably similar along the socio-economic dimension (chi-square test, p-value > 0.1). On the other hand, the development of the consideration parameter exhibits opposite patterns: it differs along the socio-economic dimension but overlaps along the gender dimension (t-tests show no significant differences at any age or in the aggregate, p-values > 0.1), confirming Implication 3. Our estimates show that the ability to avoid errors is much more pronounced in School H than in School L, with this difference being statistically significant in the aggregate (t-test, p-value < 0.01). The distribution of the estimated consideration parameter across grades in School L first-order stochastically dominates the corresponding distribution in School H - however, there is a composition effect: while the difference in the consideration parameter between youngest and oldest children (1st and 5th grade) is considerable in both schools (p-value < 0.01), by the time children reach fifth grade the differences between schools that are significant in the 1st grade (p-value < 0.01) are no more statistically significant (p-value > 0.1).

To conclude, we aim to decompose the index of rationality in cyclical choice and menu effects. We note that the consideration parameter closely tracks these rationality violations reported in Figure 8. Specifically, we report the proportion of children who satisfy NBC and, among these, the proportion of those who also satisfy AC. The interpretation is as follows. First, we ask how many children can correctly express their preferences in the simplest case, i.e., binary comparisons. Second, we see whether they are immune to menu effects. The evidence shows that the vast majority of children satisfy NBC; however, there are differences between schools in the first and third grades. Further, menu effects among these children substantially decrease with age in both schools. This analysis shows that the difference between schools, which are concentrated mainly in the first and third grades, appears to be driven by a higher proportion of children who violate NBC in School L, whereas those who satisfy NBC suffer from menu effects similarly in both schools.

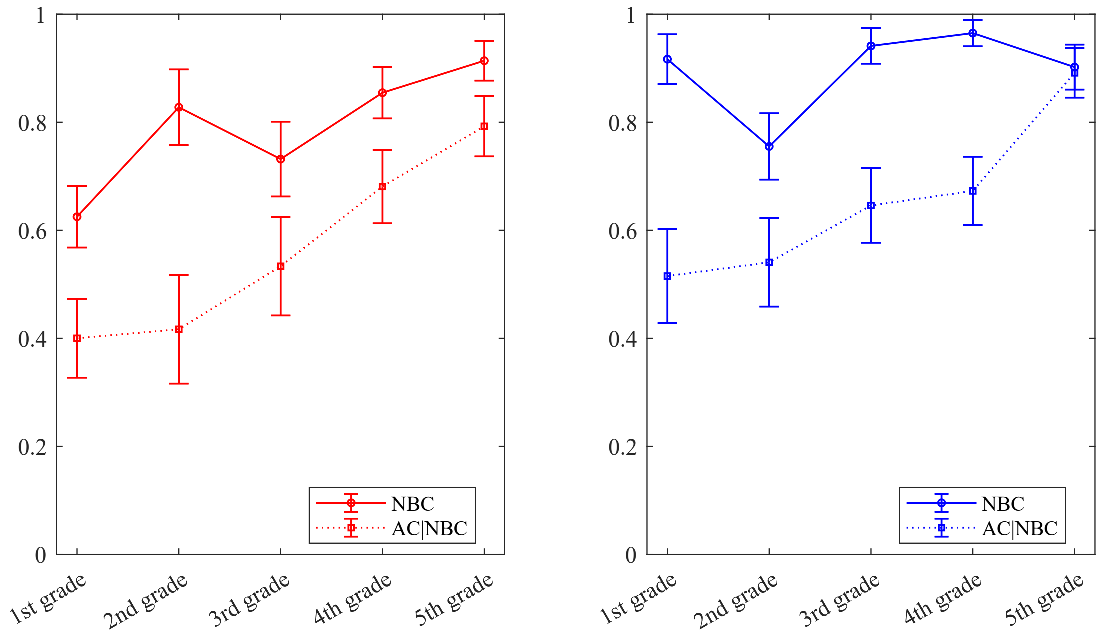

Figure 8: Development of the components of rationality in School L (left) and School H (right).

4.3.2 Errors in the coin drop games

The Independent Coin Drop task Because the Independent coin drop task is also a transparent test of first order stochastic dominance, it constitutes a stark test of rationality. Very clearly, Implication **0** is disproved: the overall fraction of children that picked the dominant lottery is only 55.3%. Distinguishing by school and grade, the ability to avoid errors improves with age (supporting our focus on older children in the lotteries task) and is larger in School H than in School L. Despite the large number of errors, in all cases our Implication **1**, i.e. $p\left(\ell_7^I, L^I\right) > p\left(\ell_6^I, L^I\right) > p\left(\ell_5^I, L^I\right)$, which relies on the weak requirement of a monotone relation between the consideration parameter and the number of winning states, is confirmed. With very few exceptions, such transparent tests of first-order stochastic dominance are rare; after all, it is deemed a basic tenet of rationality. As mentioned in the introduction, Charness et al. (2007) is one of the very few exceptions. As part of a wider experiment carried out with university students, they found that in a choice between two lotteries paying the same prize ($2) but with different probabilities ($\frac{5}{9}$ in one lottery and $\frac{6}{9}$ in the other), about 9% of their subjects chose the dominated lottery.

In our experiment, we can see that, focusing on the oldest children (5th grade), 54% of respondents in School L and 66% of respondents in School H picked the dominant lottery. While different from the error rates in the student population of Charness et al. (2007), this is comparable to what we observed when we administered the questions as in Figure 3 as part of a separate incentivised online experiment to a sample of adult female participants (average self-declared age: 58 years old).[23] Surprisingly to us, in this non-student sample, the error rate is even higher than in our school children: only 19 out of 50 (38%) respondents selected the dominant lottery.

As discussed in section 4.2.1 the consideration parameter is uniquely identified only under the assumption that it does not vary across lotteries. Table 3 reports the results of the (likelihood ratio) test of Implication **2**: as we can see the implication isn't rejected, but there is considerable variation both across schools and within schools, revealing substantial composition effects.

School L - Independent lotteries task

Grade	3rd	4th	5th	Aggregate
LRT (p-value)	**0.0714**	**0.2667**	**0.3782**	**0.9035**
Observed frequencies				
of choosing ℓ_5	0.33	0.17	0.11	0.20
of choosing ℓ_6	0.20	0.39	0.34	0.32
of choosing ℓ_7	0.47	0.44	0.55	0.48
N	36	46	44	126

School H - Independent lotteries task

Grade	3rd	4th	5th	Aggregate
LRT (p-value)	**0.9674**	**0.1246**	**0.4859**	**0.2193**
Observed frequencies				
of choosing ℓ_5	0.13	0.10	0.06	0.10
of choosing ℓ_6	0.27	0.38	0.27	0.31
of choosing ℓ_7	0.60	0.52	0.67	0.59
N	48	50	48	146

Table 3: Likelihood ratio tests ($\gamma_n^I = \gamma$), independent lottery task

Based on Table 3, we estimate lottery independent consideration parameters, which we plot in Figure 9. They different significantly in the aggregate between schools (0.36 vs 0.57, p-value < 0.05), and in the 3rd grade (0.19 vs 0.55, p-value < 0.05), while as in the riskless choice, differences in the 5th grade are not statistically significant (0.50 vs 0.66, p-value > 0.1). The development is limited in School H (from 0.55 to 0.66, p-value > 0.1), while more significant in School L (from 0.19 to 0.50, p-value < 0.1).

[23] We describe this experiment in detail, including results, in appendix E.

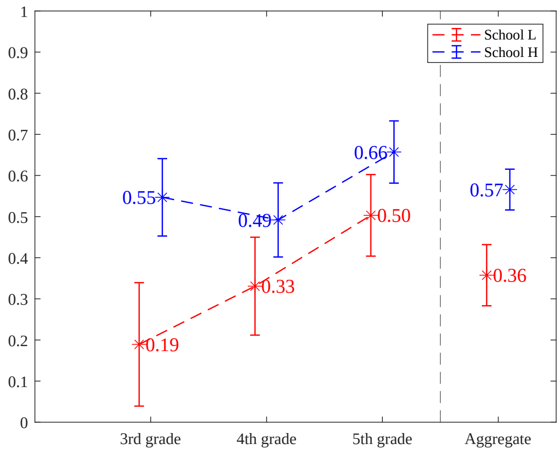

Figure 9: Estimated consideration parameter in the Independent Coin drop task

The Correlated Coin Drop task The substantial amount of correlation neglect we saw in the aggregate data persists when disaggregating by age and school, and Implication **0′** fails in all cases (chi-square tests have p-values < 0.01), with one exception: the majority of oldest children (5th Grade) choose the dominant lottery, and for school H children who fail to do so distribute evenly between the two choice errors (choosing ℓ_5 and choosing ℓ_6), a behaviour already quite close to that documented in Rees-Jones et al. (2024).[24] In this latter sub-sample, Implication **0′** is not rejected (chi-square test, p-value = 0.098). The limited consideration model is compatible with correlation neglect, hence we now turn to Implications **1′** and **2′**.

Table 4 reports the results of the likelihood ratio test for Implication **2′**, which required the choice probabilities for lotteries $\ell_5^C = \{\ell_4, \ell_5\}$ and $\ell_6^C = \{\ell_4, \ell_2\}$ to be the same, and both lower than the probability of choosing $\ell_7^C = \{\ell_5, \ell_2\}$. This Implication is generally rejected - notably though we are unable to reject it for the case of the oldest group of children (5th grade) in School H.[25], with the development of $\boldsymbol{\iota}$ depicted in Figure 10.

[24] Rees-Jones et al. (2024) find that averaging over a number of scenarios with a similar structure to our correlated Coin Drop Task, 43.5% of their adult participants (university students) display correlation neglect, with just over a third (36.2%) making the rational choice (see Figure 2 in Rees-Jones et al. (2024)).

[25] In light of this observation, for this sub-sample we can estimate a unique consideration parameter. Interestingly, we find that $\gamma = 0.66$ in the Independent task and $\gamma = 0.67$ in the Correlated task.

School L - Correlated lotteries Task

Grade	3rd	4th	5th	Aggregate
LRT (p-value)	**0.0000**	**0.0423**	**0.0138**	**0.0000**
Observed frequencies				
of choosing ℓ_5	0.67	0.48	0.37	0.49
of choosing ℓ_6	0.11	0.26	0.11	0.17
of choosing ℓ_7	0.22	0.26	0.52	0.34
$\boldsymbol{\iota}$	**-0.5**	**-0.29**	**0.18**	**-0.18**
N	36	46	44	126

School H - Correlated lotteries Task

Grade	3rd	4th	5th	Aggregate
LRT (p-value)	**0.0000**	**0.0007**	**0.8185**	**0.0000**
Observed frequencies				
of choosing ℓ_5	0.54	0.44	0.21	0.40
of choosing ℓ_6	0.08	0.1	0.19	0.12
of choosing ℓ_7	0.38	0.46	0.60	0.48
$\boldsymbol{\iota}$	**-0.18**	**0.02**	**0.49**	**0.09**
N	48	50	48	146

Table 4: Likelihood ratio test ($\gamma_n^C = \gamma$) and index of correlation neglect.

Implication **1′** requiring $\ell_6^C = \{\ell_2, \ell_4\}$ to be the least likely to be chosen, is confirmed in all cases, similarly to the Independent task, providing again support for the assumption that the consideration parameter is non decreasing in the number of states in which a prize is won.

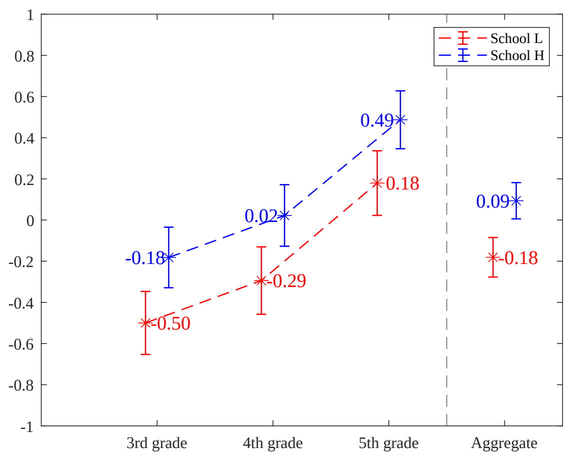

Figure 10: Index of correlation neglect ι

In summary, the limited consideration model can account both for the 'correlation neglect' lottery ℓ_5^C being chosen with larger frequency than its independent version ℓ_5^I, and also for lottery $\ell_6^C = \{\ell_2, \ell_4\}$ being the least likely to be chosen, as long as consideration is monotonically non decreasing in the prize. Finally, we plot the correlation neglect proxy, ι, in our sub-samples. T-tests comparing the indexes report significant differences in the aggregate between schools (p-value < 0.05) and in both schools between 3rd and 5th Grade (p-values < 0.01), signalling a clear development in rationality by age. On the other hand, when controlling for grades, differences across schools are never statistically significant.

It could be argued that because γ is bounded by 1, the above is not a good test of convergence. To address this point, we turn to the odds ratios[26] of choosing lotteries ℓ_5, ℓ_6, and ℓ_7, by school and age in the Independent and Correlated tasks. These are constructed from Tables 3 and 4, and reported in the two panels in Table 5.

From inspection of the left panel, we can see that in the Independent task, the odds ratios of choosing the dominant lottery in grade 3 vs grade 5 are the same across schools. This indicates that the gradient of development in rationality is the same independently of the schools, despite the fact that children in School H choose more rationally across all ages (from Table 3).

Independent Task, Grade 5 to Grade 3

Odds ratios of choosing ℓ_7

	Odds Ratio
School H	1.31
School L	1.34

Correlated to Independent Task

Odds ratios of choosing ℓ_5

	3rd	4th	5th
School H	8.27	7.07	3.95
School L	4.00	4.35	4.46

Table 5: Estimated odds ratios for schools across grades and conditions

On the other hand, the right panel supports the conclusion that children in School H are more susceptible to framing than children in School L. The right panel reports the odds ratios of choosing the dominated lottery ℓ_5 in the Correlated vs the Independent task. Because the same children choose under both frames, the distribution of the consideration probabilities is the same in the two tasks. The odds ratios show that children in School H are much more likely than children in School L to choose ℓ_5 in the Correlated task than in the Independent task in the lower grades, while the ratios converge by the time they reach fifth grade implying a greater development in School H w.r.t. correlation neglect.

4.4 Raven scores and rationality

Our analysis highlights that children's ability to make consistent choices and avoid mistakes improves with years of schooling, regardless of school location, and that the gap between children attending schools in neighbourhoods of different socioeconomic status narrows. This isn't

[26]Recall that the odds of an event are the ratio of the probability q_E of the event occurring vs the probability of the event not occurring, that is $\frac{q_E}{1-q_E}$. The odds ratio, as the name suggests, is the ratio of two odds, typically across different groups, that is $\frac{q_E(1-q_E^{'})}{(1-q_E)q_E^{'}}$ if $q_E^{'}$ denotes the probability of the event in the second group. An odds ratio of 1 is taken to indicate independence.

the case for performance in the Raven tests. Figure 11 depicts the Raven score by grade, normalised to the [0,1] interval.

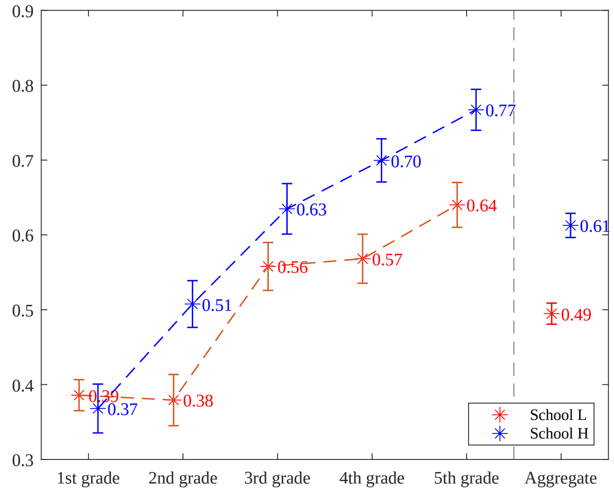

Figure 11: Normalised Raven's scores by grades and schools.

In the 1st grade, children's scores are statistically the same across schools, while the Raven score for 5th-grade children in School H is 20% higher than in School L, with this difference being statistically significant (p-value < 0.01). This seems to suggest a divergent trend of development in the two schools. We confirm this hypothesis by simply looking at the interaction term between grade and a school dummy in a linear regression (see Online Appendix D.4). We find that the slope of development of the children in School H is about 35% higher than that of the children in School L (p-value < 0.05).[27]

5 Concluding remarks

In this paper, we have described a novel experiment exploring economic rationality, intended as the ability to avoid errors, in primary school-age children, and how this ability develops with age. The errors we study constitute violations of two cornerstones of economic modelling, namely preference maximisation and dominance, hence their understanding is of obvious consequence. To explore these violations, our experiment elicits the full choice function over deterministic choices and studies choice behaviour over lotteries using a new simple design to present lotteries to children. We provide the first evidence of correlation neglect in children

[27] This evidence seems to confirm findings in psychology (see e.g. Von Stumm & Plomin (2015); Burneo-Garcés et al. (2019). Further Gneezy et al. (2019) find that performance in the Raven test might depend on the fact that with low stakes some students may not feel incentivised enough to put in sufficient effort. On the other hand there is evidence that whether the presentation is as a test or as a game matters Croizet & Dutrévis (2004)). Since in our experiment we presented all tasks as games, we don't think our results should be particularly adversely affected by the incentive structure.

and how it develops with age. Importantly, we introduce an index that tracks error avoidance ability as children develop. With plentiful evidence of choice errors in adults, it comes as no surprise that choice errors abound in children - what is striking though is that already by age 10-11, children overall display error rates which are close to those observed in adults. Interestingly, we find no correlation between violations in the riskless and risky task; however, we also observe that menu effects and framing effects reduce with age. Socioeconomic background seems to impact the ability to avoid errors depending on the task. While we find that the gap tends to close over time in our riskless task, this is not evident in the risky task, and the gap even widens in the fluid intelligence task. Finally, whether these results are due to schooling, parental influence, peer effects, or growing is not possible to ascertain from our data and warrants further investigation.

References

Agarwal, N., Moehring, A., Rajpurkar, P., & Salz, T. (2023). Combining human expertise with artificial intelligence: Experimental evidence from radiology. Tech. rep., National Bureau of Economic Research.

Agranov, M., & Ortoleva, P. (2017). Stochastic choice and preferences for randomization. *Journal of Political Economy*, *125*(1), 40–68.

Akitaya, H. A., Demaine, E. D., Demaine, M. L., Hesterberg, A., Hurtado, F., Ku, J. S., & Lynch, J. (2018). Pachinko. *Computational Geometry*, *68*, 226–242.

Allman, E. S., Matias, C., & Rhodes, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *Annals of Statistics*, *37*(6A), 3099–3132.

Alós-Ferrer, C., & Garagnani, M. (2021). Choice consistency and strength of preference. *Economics Letters*, *198*, 109672.

Andreoni, J., Di Girolamo, A., List, J. A., Mackevicius, C., & Samek, A. (2020). Risk preferences of children and adolescents in relation to gender, cognitive skills, soft skills, and executive functions. *Journal of economic behavior & organization*, *179*, 729–742.

Apesteguia, J., Huck, S., Oechssler, J., Weidenholzer, E., & Weidenholzer, S. (2018). Imitation of peers in children and adults. *Games*, *9*(1), 11.

Arcidiacono, P., & Jones, J. B. (2003). Finite mixture distributions, sequential likelihood and the em algorithm. *Econometrica*, *71*(3), 933–946.

Attanasio, O., Cattan, S., & Meghir, C. (2022). Early childhood development, human capital, and poverty. *Annual Review of Economics*, *14*, 853–892.

Barash, J., Brocas, I., Carrillo, J. D., & Kodaverdian, N. (2019). Heuristic to bayesian: The evolution of reasoning from childhood to adulthood. *Journal of Economic Behavior & Organization*, *159*, 305–322.

Brocas, I., & Carrillo, J. D. (2020a). The evolution of choice and learning in the two-person beauty contest game from kindergarten to adulthood. *Games and Economic Behavior*, *120*, 132–143.

Brocas, I., & Carrillo, J. D. (2020b). Iterative dominance in young children: Experimental evidence in simple two-person games. *Journal of Economic Behavior & Organization*, *179*, 623–637.

Brocas, I., & Carrillo, J. D. (2021). Steps of reasoning in children and adolescents. *Journal of Political Economy*, *129*(7), 2067–2111.

Brocas, I., & Carrillo, J. D. (2022). The development of randomization and deceptive behavior in mixed strategy games. *Quantitative Economics*, *13*(2), 825–862.

Brocas, I., Carrillo, J. D., Combs, T. D., & Kodaverdian, N. (2019a). Consistency in simple vs. complex choices by younger and older adults. *Journal of Economic Behavior & Organization*, *157*, 580–601.

Brocas, I., Carrillo, J. D., Combs, T. D., & Kodaverdian, N. (2019b). The development of consistent decision-making across economic domains. *Games and Economic Behavior*, *116*, 217–240.

Burneo-Garcés, C., Cruz-Quintana, F., Pérez-García, M., Fernández-Alcántara, M., Fasfous, A., & Pérez-Marfil, M. N. (2019). Interaction between socioeconomic status and cognitive development in children aged 7, 9, and 11 years: a cross-sectional study. *Developmental neuropsychology*, *44*(1), 1–16.

Calvert, S. L. (2008). Children as consumers: Advertising and marketing. *The future of children*, (pp. 205–234).

Caron, R., & Traynor, T. (2005). The zero set of a polynomial. *WSMR Report*, (pp. 05–02).

Castillo, M., Jordan, J. L., & Petrie, R. (2018). Children's rationality, risk attitudes and field behavior. *European economic review*, *102*, 62–81.

Charness, G., Karni, E., & Levin, D. (2007). Individual and group decision making under risk: An experimental study of bayesian updating and violations of first-order stochastic dominance. *Journal of Risk and uncertainty*, *35*, 129–148.

Charness, G., List, J. A., Rustichini, A., Samek, A., & Van De Ven, J. (2019). Theory of mind among disadvantaged children: Evidence from a field experiment. *Journal of Economic Behavior & Organization*, *166*, 174–194.

Chetty, R., Hendren, N., & Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, *106*(4), 855–902.

Chyn, E., & Katz, L. F. (2021). Neighborhoods matter: Assessing the evidence for place effects. *Journal of Economic Perspectives*, *35*(4), 197–222.

Croizet, J.-C., & Dutrévis, M. (2004). Socioeconomic status and intelligence: Why test scores do not equal merit. *Journal of Poverty*, *8*(3), 91–107.

Czermak, S., Feri, F., Glätzle-Rützler, D., & Sutter, M. (2016). How strategic are children and adolescents? experimental evidence from normal-form games. *Journal of Economic Behavior & Organization*, *128*, 265–285.

Dardanoni, V., Manzini, P., Mariotti, M., Petri, H., & Tyson, C. J. (2023). Mixture choice data: revealing preferences and cognition. *Journal of Political Economy*, *131*(3), 687–715.

Dauphin, A., El Lahga, A.-R., Fortin, B., & Lacroix, G. (2011). Are children decision-makers within the household? *The Economic Journal*, *121*(553), 871–903.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from head start. *American Economic Journal: Applied Economics*, *1*(3), 111–134.

Diamond, A. (2013). Executive functions. *Annual review of psychology*, *64*(1), 135–168.

Durlauf, S. N. (2004). Neighborhood effects. *Handbook of regional and urban economics*, *4*, 2173–2242.

Ellis, A., & Freeman, D. J. (2024). Revealing choice bracketing. *American Economic Review*, *114*(9), 2668–2700.

Ellis, A., & Piccione, M. (2017). Correlation misperception in choice. *American Economic Review*, *107*(4), 1264–1292.

Enke, B., & Zimmermann, F. (2019). Correlation neglect in belief formation. *The Review of Economic Studies*, *86*(1), 313–332.

Eyster, E., & Weizsacker, G. (2016). Correlation neglect in portfolio choice: Lab evidence. *Available at SSRN 2914526*.

Fe, E., Gill, D., & Prowse, V. (2022). Cognitive skills, strategic sophistication, and life outcomes. *Journal of Political Economy*, *130*(10), 2643–2704.

Ferguson, H. J., Brunsdon, V. E., & Bradford, E. E. (2021). The developmental trajectories of executive function from adolescence to old age. *Scientific reports*, *11*(1), 1382.

Field, A. P., & Lester, K. J. (2010). Is there room for 'development'in developmental models of information processing biases to threat in children and adolescents? *Clinical child and family psychology review*, *13*(4), 315–332.

Fryer Jr, R. G., & Katz, L. F. (2013). Achieving escape velocity: Neighborhood and school interventions to reduce persistent inequality. *American Economic Review*, *103*(3), 232–237.

Gill, D., & Prowse, V. (2024). The creativity premium: Exploring the link between childhood creativity and life outcomes. *Journal of Political Economy Microeconomics*, *2*(3), 495–526.

Gneezy, U., List, J. A., Livingston, J. A., Qin, X., Sadoff, S., & Xu, Y. (2019). Measuring success in education: The role of effort on the test itself. *American Economic Review: Insights*, *1*(3), 291–308.

Harbaugh, W. T., Krause, K., & Berry, T. R. (2001). Garp for kids: On the development of rational choice behavior. *American Economic Review*, *91*(5), 1539–1545.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900–1902.

Horan, S. (2019). Random consideration and choice: A case study of "default" options. *Mathematical Social Sciences*, *102*, 73–84.

Houmark, M. A., Ronda, V., & Rosholm, M. (2024). The nurture of nature and the nature of nurture: How genes and investments interact in the formation of skills. *American Economic Review*, *114*(2), 385–425.

Huang, Y., Jiang, Y., & Lu, Z. (2025). Correlation neglect on social media: Effects on civil service applications in china. *The Economic Journal*, (p. ueaf047).

Ingleton, A. W. (1956). The rank of circulant matrices. *Journal of the London Mathematical Society*, *s1-31*(4), 445–460.

Kass, R. E., & Steffey, D. (1989). Approximate bayesian inference in conditionally independent hierarchical models (parametric empirical bayes models). *Journal of the American Statistical Association*, *84*(407), 717–726.

Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: the effects of neighborhood residence on child and adolescent outcomes. *Psychological bulletin*, *126*(2), 309.

Li, S. (2017). Obviously strategy-proof mechanisms. *American Economic Review*, *107*(11), 3257–3287.

List, J. A., Petrie, R., & Samek, A. (2023). How experiments with children inform economics. *Journal of Economic Literature*, *61*(2), 504–564.

Manzini, P., & Mariotti, M. (2007). Sequentially rationalizable choice. *American Economic Review*, *97*(5), 1824–1839.

Manzini, P., & Mariotti, M. (2014). Stochastic choice and consideration sets. *Econometrica*, *82*(3), 1153–1176.

Muslimova, D., van Kippersluis, H., Rietveld, C. A., von Hinke, S., & Meddens, S. F. W. (2020). Nature-nurture interplay in educational attainment. *arXiv preprint arXiv:2012.05021*.

Oreopoulos, P. (2003). The long-run consequences of living in a poor neighborhood. *The quarterly journal of economics*, *118*(4), 1533–1575.

Ortoleva, P., & Snowberg, E. (2015). Overconfidence in political behavior. *American Economic Review*, *105*(2), 504–535.

Piovesan, M., & Willadsen, H. (2021). Risk preferences and personality traits in children and adolescents. *Journal of Economic Behavior & Organization*, *186*, 523–532.

Rees-Jones, A., Shorrer, R., & Tergiman, C. J. (2024). Correlation neglect in student-to-school matching. *American Economic Journal - Microeconomics*, *16*(3), 1–42.

Selten, R. (1991). Properties of a measure of predictive success. *Mathematical social sciences*, *21*(2), 153–167.

Sher, I., Koenig, M., & Rustichini, A. (2014). Children's strategic theory of mind. *Proceedings of the National Academy of Sciences*, *111*(37), 13307–13312.

Siegler, R. S., Saffran, J., Gershoff, E., & Graham, S. (2024). *How Children Develop*. Macmillan.

Steinbeis, N. (2023). A rational account of cognitive control development in childhood. *Annual Review of Developmental Psychology*, *5*(1), 217–238.

Sutter, M., Zoller, C., & Glätzle-Rützler, D. (2019). Economic behavior of children and adolescents–a first survey of experimental economics results. *European Economic Review*, *111*, 98–121.

Vitaku, V. (2024). Correlation neglect in risky choice. *Available at SSRN 4985359*.

Von Stumm, S., & Plomin, R. (2015). Socioeconomic status and the growth of intelligence from infancy through adolescence. *Intelligence*, *48*, 30–36.

Appendix

A Theoretical distribution of violations of NBC and AC

In Section **??**, we reported the Selten measure referred to No Binary Cycle and Always Chosen. This is based on the theoretical distribution of violations of these axioms which is derived based on the following:

1. number of **choice functions that satisfy NBC and fail AC**: since NBC holds, the restriction of each choice function in this class to the subdomain of binary sets maximises a linear order. Of the 64 possible choice functions, only 24 maximise a linear order. Fixing the choice in the binary sets, the choice in the remaining sets (of cardinality three and four) will be rational in only one configuration, corresponding to the maximisation of a linear order. It follows that the number of choice functions that satisfy NBC and fail AC is $24 \cdot \left(3^4 \cdot 4 - 1\right) = 24 \cdot 323 = 7,752$. This is a proportion $\frac{7,752}{20,736} = 37.38\%$ of all choice functions.

2. number of **choice functions that fail NBC and satisfy AC**: cycles are present in each of the remaining 40 choice functions restricted to the subdomain of binary sets. We have to distinguish based on the cardinality of the sets where the premise of AC does not apply because of the cycles. There are 8 cases with one free set of cardinality 3; 10 cases with

one free set of cardinality 3 and the grand set; and 22 sets with two free sets of cardinality 3 and the grand set. The total is

$$8 \cdot 3 + 10 \cdot 3 \cdot 4 + 22 \cdot 3^2 \cdot 4 = 12\,(2 + 10 + 66) = 936$$

which is a proportion is a proportion $\frac{936}{20{,}736} = 4.51\%$ of all choice functions.

3. number of **choice functions that satisfy both NBC and AC**: these are the rational choice functions, which are each rationalised by the maximisation of one of the 24 possible linear ordering of the universe of four alternative. Hence there are 24 such choice functions, a proportion $\frac{24}{20{,}736} = 0.12\%$ of all possible choice functions.

4. number of **choice functions that fail both axioms** are the remaining $20,736 - 24 - 936 - 7,752 = 12,024$ corresponding to 57.99% of all choice functions.

Selten's measure of predictive success can be obtained by subtracting the theoretical figures from the observed distribution.

B Choice probabilities: Stochastic Consideration Set model

B.1 Lotteries task

For the independent lotteries task, equation 1 prescribes the following the following choice probabilities:

$$p\left(\ell_5^I, L^I\right) = \frac{\gamma_5^I\left(1-\gamma_6^I\right)\left(1-\gamma_7^I\right)}{1-\left(1-\gamma_5^I\right)\left(1-\gamma_6^I\right)\left(1-\gamma_7^I\right)} \tag{7}$$

$$p\left(\ell_6^I, L^I\right) = \frac{\gamma_6^I\left(1-\gamma_7^I\right)}{1-\left(1-\gamma_5^I\right)\left(1-\gamma_6^I\right)\left(1-\gamma_7^I\right)} \tag{8}$$

$$p\left(\ell_7^I, L^I\right) = \frac{\gamma_7^I}{1-\left(1-\gamma_5^I\right)\left(1-\gamma_6^I\right)\left(1-\gamma_7^I\right)} \tag{9}$$

If we set $\gamma_n^I = \gamma^I$ for all n in equations (7)-(9) it follows that $\frac{p\left(\ell_7^I, L^I\right)}{p\left(\ell_6^I, L^I\right)} = \frac{1}{1-\gamma} = \frac{p\left(\ell_6^I, L^I\right)}{p\left(\ell_5^I, L^I\right)}$, justifying Implication 2.

For the correlated lotteries task, We let L^C denote the set of correlated lotteries, that is $L^C = \{\ell_2, \ell_4, \ell_5\}$, and denote each *pair* of lotteries with the overall winning probability, that is $\ell_5^C = \{\ell_4, \ell_5\}$, $\ell_6^C = \{\ell_4, \ell_2\}$ and $\ell_7^C = \{\ell_5, \ell_2\}$. With γ_n^C denoting the consideration probability for lottery with n winning entries in this treatment, under the assumption of independent consideration the choice probabilities are:

$$p\left(\ell_5^C, L^C\right) = \frac{\gamma_4^C\gamma_5^C\left(1-\gamma_2^C\right)}{\gamma_2^C\gamma_4^C+\gamma_2^C\gamma_5^C+\gamma_4^C\gamma_5^C-2\gamma_2^C\gamma_4^C\gamma_5^C} \tag{10}$$

$$p\left(\ell_6^C, L^C\right) = \frac{\gamma_2^C\gamma_4^C\left(1-\gamma_5^C\right)}{\gamma_2^C\gamma_4^C+\gamma_2^C\gamma_5^C+\gamma_4^C\gamma_5^C-2\gamma_2^C\gamma_4^C\gamma_5^C} \tag{11}$$

$$p\left(\ell_7^C, L^C\right) = \frac{\gamma_2^C\gamma_5^C}{\gamma_2^C\gamma_4^C+\gamma_2^C\gamma_5^C+\gamma_4^C\gamma_5^C-2\gamma_2^C\gamma_4^C\gamma_5^C} \tag{12}$$

where the first two lines derive from observing the two lotteries that combine suboptimally without observing the third, while the third line is the probability of considering the two lotteries that combine optimally, regardless of whether or not the third is also considered.[28]

Implication **1**′ holds since on the one hand $\gamma_5^C \geq \gamma_2^C \implies \frac{\gamma_5^C}{1-\gamma_5^C} > \frac{\gamma_2^C}{1-\gamma_2^C}$ so that $p\left(\ell_5^C, L^C\right) > p\left(\ell_6^C, L^C\right)$; and on the other hand $\gamma_5^C \geq \gamma_4^C \implies \frac{\gamma_5^C}{1-\gamma_5^C} > \gamma_4^C$, so that $p\left(\ell_7^C, L^C\right) > p\left(\ell_6^C, L^C\right)$.

B.2 Pencils/pens task

Let C denote a generic choice function, and let $\mathcal{C}$ denote the collection of all possible choice functions. As mentioned in the main text, a Mixture Choice Function μ is a probability distribution over $\mathcal{C}$; i.e., a $\mu : \mathcal{C} \to [0,1]$ such that $\sum_{c\in\mathcal{C}} \mu(C) = 1$.

Under the assumption that choices at each menu are independent, the probability $\mu(C)$ of observing a given subject making choices according to C can be expressed as the product of all the individual choices at each menu, that is

$$\mu(C) = \prod_{A\in\mathcal{A}} p\left(C(A), A\right)$$

In our setup we observe the choice function of a population of experimental subjects. We capture heterogeneity by postulating that each subject is of a specific (preference) type $\succ$ from a collection P of admissible types, and we postulate that their choice behaviour is driven by some hypothesised model. Let $p^{\succ}(a, A)$ denote the type conditional probability that a subject of type $\succ$ selects a from menu A; then the probability of any given choice function C can be written as the mixture

$$\mu(C) = \sum_{\succ\in P} \pi(\succ) \prod_{A\in\mathcal{A}} p^{\succ}\left(C(A), A\right)$$

[28] The probability of not choosing anything is the probability of considering at most one lottery, given by

$$1-\left(\gamma_2^C\left(1-\gamma_4^C\right)\left(1-\gamma_5^C\right)+\gamma_4^C\left(1-\gamma_2^C\right)\left(1-\gamma_5^C\right)+\gamma_5^C\left(1-\gamma_4^C\right)\left(1-\gamma_2^C\right)+\left(1-\gamma_2^C\right)\left(1-\gamma_4^C\right)\left(1-\gamma_5^C\right)\right)$$

which reduces to the expression at the denominator.

More in general the mixture choice function can be expressed as

$$\boldsymbol{\mu} = \boldsymbol{Z}\boldsymbol{\pi}, \tag{13}$$

where $\boldsymbol{\mu} = \left[\mu\left(C_1\right), \mu\left(C_2\right), ..., \mu\left(C_{|\mathcal{C}|}\right)\right]'$, is the vector form of μ, $\mathbf{Z}$ is the matrix of type conditional probabilities according to the postulated model describing choice behaviour (with each column inputing the demand by a different preference type $\succ$), and $\boldsymbol{\pi} = \left[\pi\left(\succ_1\right), \mu\left(\succ_2\right), ..., \mu\left(\succ_{|P|}\right)\right]'$ is the vector form of the (unknown) type distribution. The matrix of type conditional probabilities in equation (13) can be written more extensively as

$$\mathbf{Z} = \begin{bmatrix} \prod_{A\in\mathcal{A}} p^{\succ_1}\left(C_1\left(A\right), A\right) & \prod_{A\in\mathcal{A}} p^{\succ_2}\left(C_1\left(A\right), A\right) & \cdots & \prod_{A\in\mathcal{A}} p^{\succ_{|P|}}\left(C_1\left(A\right), A\right) \\ \prod_{A\in\mathcal{A}} p^{\succ_1}\left(C_2\left(A\right), A\right) & \prod_{A\in\mathcal{A}} p^{\succ_2}\left(C_2\left(A\right), A\right) & \cdots & \prod_{A\in\mathcal{A}} p^{\succ_{|P|}}\left(C_2\left(A\right), A\right) \\ \vdots & \vdots & \ddots & \vdots \\ \prod_{A\in\mathcal{A}} p^{\succ_1}\left(C_{|\mathcal{C}|}\left(A\right), A\right) & \prod_{A\in\mathcal{A}} p^{\succ_2}\left(C_{|\mathcal{C}|}\left(A\right), A\right) & \cdots & \prod_{A\in\mathcal{A}} p^{\succ_{|P|}}\left(C_{|\mathcal{C}|}\left(A\right), A\right) \end{bmatrix}$$

The specific form that the type conditional probabilities take depend on the model postulated. In our case, it is the Stochastic Consideration Set model (1).

C Estimation: Finite Mixture Model

We estimate our model (type distribution $\pi(\succ)$ and consideration parameter γ) from the deterministic choice dataset by maximum likelihood.[29]

In our i.i.d realised sample $(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_I)$, each $\mathbf{y}_i$ is a multinomial random vector of the choices made by the i-th subject at all choice menus, that is $\mathbf{y}_i = \left(y_{iA_1}, y_{iA_2}, ... y_{iA_{|\mathcal{A}|}}\right)$.[30] The individual likelihood that observation $\mathbf{y}_i$ could have been generated by type $\succ$ is given by:

$$\mathcal{L}_i\left(\succ; \boldsymbol{y}_i\right) = p\left(\boldsymbol{y}_i\right) = \prod_{A\in\mathcal{A}} p\left(y_{iA}\right) \tag{14}$$

where the last equality follows from the independence assumption. Given that types are unobservable for the researcher, each individual likelihood can be written as a finite mixture

$$\mathcal{L}_i\left(\succ; \boldsymbol{y}_i\right) = \sum_{\succ\in P} \pi\left(\succ\right) \left[\prod_{A\in\mathcal{A}} p^{\succ}\left(y_{iA}\right)\right]. \tag{15}$$

[29] The estimation follows an iterated two-step procedure as in Arcidiacono & Jones (2003). At each step, we proceed to maximise the likelihood with respect to one unknown conditional on the estimated values of the other.

[30] Each subject i makes a single choice from each of the 11 choice menus in each of the pen or pencils domains (6 binary choices, 4 ternary choices, and the choice from the grand set), hence with $2^6 \cdot 3^4 \cdot 4 = 20,736$ possible choice functions, $\mathbf{y}_i$ will be a $2^6 \cdot 3^4 \cdot 4 \times 1$ vector having 1 in the element corresponding to the observed choice function, and zeros elsewhere. More in general, if there are n options, subject i is of type $\succ$, and choices are independent conditional on type, the $2^{\binom{n}{2}} \times 3^{\binom{n}{3}} \times \cdots \times n$)-dimensional vector $\mathbf{y}_i$ is distributed as a single draw from a multinomial distribution with probability given by the corresponding entry in the $\succ$-th column of $\mathbf{Z}$.

When the model is identified (that is, we can recover uniquely the type distribution and the consideration parameter from the available data), estimation can be obtained by maximising the log-likelihood of the whole sample subject to the model restrictions, namely that $p^{\succ}$ is as in equation (1):

$$\max\ \log \mathcal{L}\left(\pi, \gamma; (\succ; \boldsymbol{y})\right) = \sum_{i=1}^{|I|} \log \left(\sum_{\succ \in P} \pi\left(\succ\right) \prod_{A \in \mathcal{A}} p^{\succ}\left(y_i\right) \right). \tag{16}$$

For the lotteries task, we can test the implications of our model. Let γ be a triple of consideration parameters (or a unique one) and Y_n the number of subjects choosing lottery ℓ_n. The maximum-likelihood problem to estimate γ is:

$$\max_{\gamma} \log \mathcal{L}\left(\gamma; \boldsymbol{Y}\right) = \sum_{\ell_n \in A} Y_n \log(p(\ell_n, A)) \tag{17}$$

D Proof of Proposition 1

Outline and background

Recall from the main text that our objective is to identify the type distribution and the type conditional probabilities from knowledge of the mixture choice function from equation (13). Formally, our observable, $\boldsymbol{\mu}$ is a tensor. AMR (Allman et al., 2009) seminal paper contains results on tensor decomposition that lead to the identification of the unknown type conditional probabilities. These generic identification results however assume that the unknowns can take any value in the parameter space; in our case the theoretical model restricts these parameters, hence we need to prove that AMR's results holds in our case, too. To this effect we have to modify part of one step in the proof of AMR's main theorem. In our setup this requires partitioning the collection of choice occasions (i.e. the 11 choice menus) into three subcollections such that:

1. each matrix of type conditional probabilities restricted to each of these subcollection is invertible; and
2. the products of the cardinality of each set in the subcollections are related in a specific way (guaranteeing that there are enough degrees of freedom for identification).

Regarding the first point, we construct three sub-collections $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$ comprising, respectively, the grand set only (collection $\mathcal{A}_1$), all thr binary sets (collection $\mathcal{A}_2$), and all ternary sets (collection $\mathcal{A}_3$). Let $\kappa_i = \prod_{A \in \mathcal{A}_i} |A|$ for all $i = 1, 2, 3$. Then with our construction we have $\kappa_1 = 4$, $\kappa_2 = 2^6 = 64$ and $\kappa_3 = 3^4 = 81$.

Corollary 3 in AMR states that r types can be identified, up to label swapping, as long as $\sum_{i=1}^{3} \min\left\{\kappa_i, r\right\} \geq 2r + 2$. In our case, with 24 unknown types (i.e. the 24 strict preference orders over four alternatives), this requires

$$\min\{\kappa_1, 24\} + \min\{\kappa_2, 24\} + \min\{\kappa_3, 24\} \geq 48 + 2 \Rightarrow 4 + 24 + 24 \geq 50$$

Hence the condition for the identification of the 24 types holds. For each collection, we construct the matrix of type conditional choice probabilities restricted to that collection and show that each has full Kruskal rank, where the Kruskal rank of a matrix is the largest k such that any set of k columns from such matrix are linearly independent. To prove that the Kruskal rank is full, from each such matrix we extract a 24×24 minor that has a non-zero determinant. Since the determinant is a polynomial function, and since a polynomial function is either identically zero or non-zero almost everywhere (see Caron & Traynor (2005)), by exhibiting one parameter value that sets the determinant different from zero, we show that the matrix is generically invertible. In turn, by appealing to ARM, we are able to establish the generic identifiability of the model.

Proof

Let $\mathcal{A}_1 = \{X\}$, $\mathcal{A}_2 = \{A \subset X : |A| = 2\}$, and $\mathcal{A}_3 = \{A \subset X : |A| = 3\}$, and correspondingly construct the restriction of each type conditional mixture choice function to the collection $\mathcal{A}_i$, with $i = 1, 2, 3$. We now show that the matrix of type conditional choice probabilities restricted to each collection has full Kruskal rank. Consider collection $\mathcal{A}_1$ first. Here the restricted type conditional matrix has dimension 4×24, with each row corresponding to one possible choice $C(X)$ from the grand set and each row corresponding to a different preference type. Hence the Kruskal rank is at most 4. Relabel the four alternatives as x_i with $i = 1, .., 4$. Then we can select a minor consisting of the four types $\succ_j$ each with $x_i \succ_j x_{i+1}$ modulo 4 and with $j = 1, .., 4$ and $x_j \succ_j x$ for all $x \neq x_j$. This way, we can arrange orders to obtain the circulant minor

$$\begin{bmatrix} a & ab & ab^2 & ab^3 \\ ab & ab^2 & ab^3 & a \\ ab^2 & ab^3 & a & ab \\ ab^3 & a & ab & ab^2 \end{bmatrix} \tag{18}$$

where we have used the shorthand notation

$$a = \frac{\gamma}{1 - (1 - \gamma)^4}; b = 1 - \gamma$$

Exploiting the properties of the rank of circulant matrices, that establish that the rank is given by the difference between the number of rows and the degree of the matrix (see e.g. Ingleton

(1956)), it follows that minor (18) has full rank, hence full Kruskal rank.[31]

Next, consider $\mathcal{A}_2$, which collects all binary sets. The matrix of type conditional probabilities has dimensions 64×24. Extract a 24×24 minor selecting, for each type, the unique row corresponding to the choice functions selecting the top alternatives in all 6 sets according to that type. Note that the type conditional probability for choice functions restricted to this collection when they maximise an order is

$$\left(\frac{\gamma}{1-(1-\gamma)^2}\right)^6$$

which is therefore the largest value achievable in a given row. By collecting this term outside of the matrix, we can permute rows so as to obtain a square minor with 1 on the main diagonal, with all other values less than 1. Now setting $(1-\gamma) = \frac{1}{24}$ the minor is a diagonally dominant matrix[32], which is invertible. Hence, the determinant is non-zero, and since the determinant of this minor is a polynomial function, it is different from zero almost everywhere. We conclude that the minor is generically invertible, and as a consequence has full Kruskal rank. For the third and final collection we proceed similarly, only this time for each type we select the row such that the corresponding choice function restricted to the subcollection, selects, for each preference type, the top alternative from the first three sets, and the bottom alternative from the fourth set. The type conditional probability in correspondence of the matching preference type is

$$\left(\frac{\gamma}{1-(1-\gamma)^3}\right)^4 (1-\gamma)^2$$

with all other terms being smaller. Hence we can collect this term, and again permute this minor so as to obtain a matrix with 1 on the main diagonal. Again setting $(1-\gamma) = \frac{1}{24}$ the minor is a diagonally dominant matrix, which is invertible, and we can proceed as above.

E Lottery tasks with adult women

While, as mentioned in the main text, there are existing transparent tests of violations of first order stochastic dominance and correlation neglect, these experiments are typically carried out with undergraduate students. In order to have a more generic comparison group, we admin-

[31]For a circulant matrix with first row $[a_0, a_1, ...a_{n-1}]$ the associated polynomial is the function $\sum_{i=0}^{n-1} a_i x^i$. The degree of the circulant matrix is the number of common zeros between the associated polynomial and 1—x^n. The associated polynomial of matrix (18) is

$$\frac{\gamma}{1-(1-\gamma)^4}\left(1+(1-\gamma)x+(1-\gamma)^2x^2+(1-\gamma)^3x^3\right)$$

Since the only real root is $-\frac{1}{1-\gamma} \neq \pm 1$, the degree of this circulant minor is zero, hence the minor has full rank.

[32]A square matrix is diagonally dominant if, for each row, the aboslute value of the element on the diagonal is greater or equal to the sum of the absolute values of the other elements in that row

istered the lottery tasks to a group of adult women, since we had the opportunity to insert those questions as part of a separate choice experiment.[33] Our subject pool consisted of 50 female participants recruited through a private crafting Facebook group (average self-declared age was 58). The full experiment was for a separate project and focused on the elicitation of partial choice functions. In view of the transparency of the FOSD test we employed a between subject design, with half of the respondents answering the independent lotteries question and half answering the correlated lotteries question. In addition, both groups answered the same set of choice questions eliciting a partial choice function out of a grand set of 5 alternatives,[34] depicted in Figure 13, choosing one item from each of sixteen screens (comprising the grand set, all sets of cardinality four, and all binary sets). Hence, for these women, we can estimate the consideration parameter.[35] The appearance of the lottery questions was essentially the same as in the experiment with children, as can be seen from the screenshots in Figure 12.

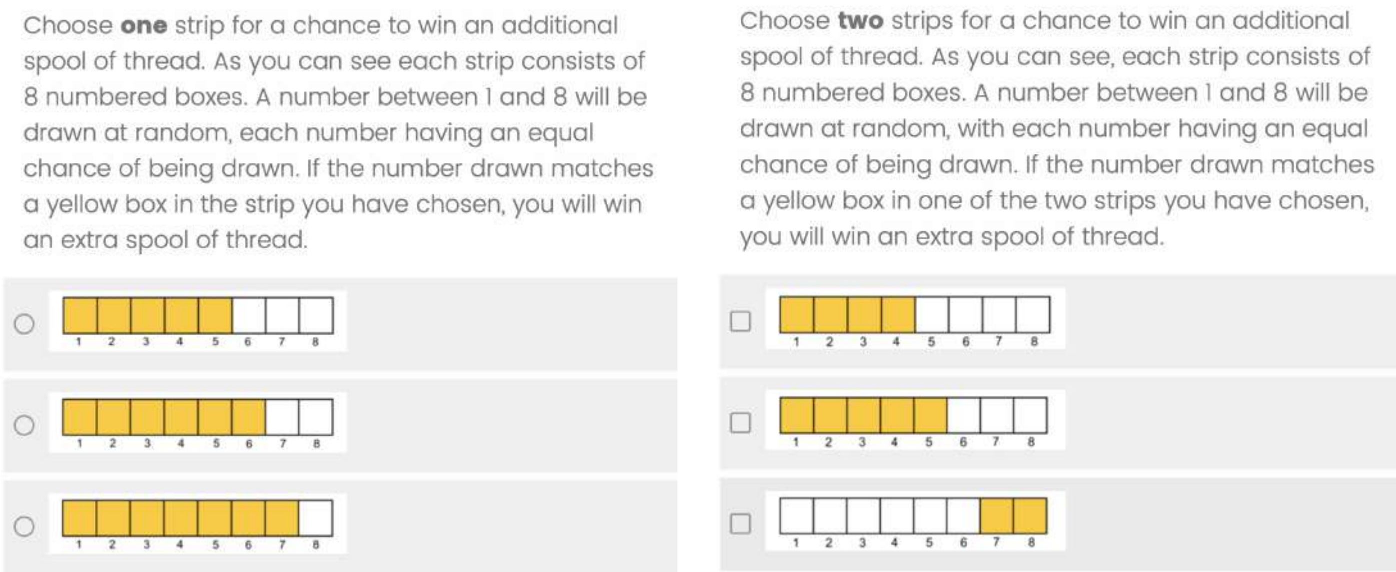

Figure 12: Lottery questions, adult women

[33]Examples of the comparisons of how performance in various tests of rationality vaaries across the ages either include comparisons between children and undergraduate students (Harbaugh et al. (2001); Brocas et al. (2019b)), and or comparisons between younger (18-34) and older (59-89) adults. In the first case by the time children are in 5th grade their performance is already close to that of undergraduate students, while when comparing different adult groups on consistency measure, significant differences emerge, with older adults being less consistent (see (Brocas et al., 2019a)). We combine these approaches by comparing children to an adult subpopulation.

[34]These were high-value items, instructional DVDs on bobbin lace retailing for more than ¤100.

[35]See appendix E.3 for proof that the model is identified in this partial domain.

Figure 13: The grand set of items, adult women

E.1 Errors in riskless choice.

Among the 100 women who completed the survey, 91 answered all the choice questions regarding the five instructional DVDs. The proportion of fully rational adult women is only 18%, lower than in the case of children (see Figure **??**). We should note however that these older women answered 16 choice questions (rather than the 11 administered to children): with a fixed consideration parameters, the probability of being fully rational is by necessity lower as the number of choice sets increases. Leveraging our identification results (see section 4.2.3 and Appendix E.3), we estimate a unique consideration parameter for the population of old women. The result is 0.74, which is in line with the results from the pupils of the 3rd grade in School L, while lower than any results from the pupils in School H (see table **??**).

We also administered three Raven matrices to these older adults. The correlation between individual consideration parameters and cognitive abilities is only 0.08 (p-value > 0.1), and focusing only on the 18 adult women with a perfect Raven score, the estimated consideration parameter is still only 0.8, substantially lower than even fourth-grade pupils.

Interestingly, we find a negative correlation between the individually estimated consideration parameters and age, -0.18 (p-value < 0.1) which, focusing only on the adult women with perfect Raven scores, becomes even stronger: -0.5 (p-value < 0.05). This result suggests, speculatively, that the dimension of rationality captured by our consideration parameter is related to cognitive skills that increase quickly with age among the pupils, while they decrease in later stages of life.

E.2 Errors in risky choice.

The results from the lottery choices are summarized in tables 6 and 7. We also estimated the unique consideration parameter in the independent lottery task (the likelihood ratio test does not reject the hypothesis of a unique consideration parameter), which is equal to 0.24, similar to the result of School L children in third grade (see Figure 9). The adult women, instead, performed better in the correlated task, at least compared to children. The index of correlation neglect (0.14) is in line with that of School L children in fifth grade, and substantially lower than those in School H (see table 4). With a high proportion of adult women choosing ℓ_5^C, the likelihood-ratio test rejects the hypothesis of a unique consideration parameter.

Independent lotteries Task	
LRT (p-value)	**0.1548**
Observed frequencies	
of choosing ℓ_5	0.20
of choosing ℓ_6	0.42
of choosing ℓ_7	0.38
N	50

Table 6: Independent lotteries task.

Correlated lotteries Task	
LRT (p-value)	**0.0470**
Observed frequencies	
of choosing ℓ_5	0.36
of choosing ℓ_6	0.16
of choosing ℓ_7	0.48
$\boldsymbol{\iota}$	**0.14**
N	50

Table 7: Correlated lotteries task.

E.3 Identification in the partial domain of the adult women experiment

A necessary condition for our identification from mixture choice data (Proposition 1) is that the constituent model is identified. To establish the result in this case too, with choices were elicited from sets of two, four, and five alternatives, we extend Horan (2019)'s proof (which requires choices over all binary and ternary sets in addition to the grand seet) to show that the model is identified in this partial domain.

Proposition 2. *Given a stochastic choice function p, the consideration parameter γ and the preference $\succ$ are uniquely identified by choices from sets of two, four, and five alternatives.*

Proof. Let $X = \{a, b, c, d, e\}$. Recall from Horan (2019) the following definitions: for any $x, y, z \in X$:

$$L_{[x:y]}(A) = \frac{p(x, A \cup \{a, b\})}{p(y, A \cup \{a, b\})}$$

$$R_{[x:y]}(A; z) = \frac{L_{[x:y]}(A)}{L_{[x:y]}(A \cup \{z\})} = \frac{\frac{p(x, A\cup\{x,y\})}{p(y, A\cup\{x,y\})}}{\frac{p(x, A\cup\{x,y,z\})}{p(y, A\cup\{x,y,z\})}}$$

To identify the **preference** we can follow exactly Horan (2019), that is $x \succ y$ iff either $R_{[x:y]}(A; i) < 1$, or $R_{[x:i]}(A; y) < 1$, or $R_{[i:y]}(A; x) < 1$ for some 'intermediate' alternative i. Hence setting $A = X \setminus \{x, y, i\}$ we need all the sets of cardinality 4 and the grand set (of cardinality 5). To identify the consideration parameters, relabel all alternatives from best to worst so that $a \succ b \succ c \succ d \succ e$. Let $A = \{c, d\}$. Then:

$$R_{[a:e]}(A; b) = \frac{L_{[a:e]}(A)}{L_{[a:e]}(A \cup \{b\})} = \frac{\frac{p(a,\{a,c,d,e\})}{p(e,\{a,c,d,e\})}}{\frac{p(a,X)}{p(e,X)}} = \frac{\frac{\gamma_a}{\gamma_e(1-\gamma_a)(1-\gamma_c)(1-\gamma_c)}}{\frac{\gamma_a}{\gamma_e(1-\gamma_a)(1-\gamma_b)(1-\gamma_c)(1-\gamma_c)}} = \frac{1}{\frac{1}{1-\gamma_b}} = 1 - \gamma_b$$

More in general, for any $i \neq a, e$:

$$\gamma_i = 1 - \frac{\frac{p(a,X\setminus i)}{p(e,X\setminus i)}}{\frac{p(a,X)}{p(e,X)}}$$

which follows from

$$R_{[a:e]}\left(A;i\right) = \frac{\frac{p(a,X\setminus i)}{p(e,X\setminus i)}}{\frac{p(a,X)}{p(e,X)}} = 1 - \gamma_i$$

To identify the consideration parameters for the two extreme alternatives we use one of the values found for the consideration parameters of any intermediate alternative i to obtain γ_e from $L_{[i:e]}\left(\emptyset\right)$ and γ_a from $L_{[a:i]}\left(\emptyset\right)$, as follows:

$$L_{[i:e]}\left(\emptyset\right) = \frac{p\left(i,\{i,e\}\right)}{p\left(e,\{i,e\}\right)} = \frac{\gamma_i}{\gamma_e\left(1-\gamma_i\right)} \Rightarrow \gamma_e = \frac{p\left(e,\{i,e\}\right)}{p\left(i,\{i,e\}\right)}\frac{\gamma_i}{1-\gamma_i}$$

$$L_{[i:a]}\left(\emptyset\right) = \frac{p\left(i,\{a,i\}\right)}{p\left(a,\{a,i\}\right)} = \frac{\gamma_i\left(1-\gamma_a\right)}{\gamma_a} \Rightarrow \gamma_a = \frac{\gamma_i}{p\left(i,\{a,i\}\right)+\gamma_i p\left(a,\{a,i\}\right)}$$

□

E.4 Extending the proof of Proposition 1 w.r.t. to the experiment in appendix E

This proof follows closely the one for Proposition 1 (see appendix D) - now our three collections $\mathcal{A}_1$,$\mathcal{A}_2$ and $\mathcal{A}_3$ include, respectively, the grand set (this time 5 alternatives), the collection of all binary sets (as before), and the collection of all quaternary sets. These enable us to identify all the $5! = 120$ types.[36]

More formally, let $\mathcal{A}_1 = \{X\}$, $\mathcal{A}_2 = \{A \subset X : |A| = 2\}$, and $\mathcal{A}_3 = \{A \subset X : |A| = 4\}$, and correspondingly construct the restriction of each type conditional mixture choice function to the collection $\mathcal{A}_i$, with $i = 1,2,3$. Similarly to the previous proof, for collection $\mathcal{A}_1$ first, the restriction means that the matrix of type conditional probabilities restricted to the choices from the grand set has dimension 5×120, hence the Kruskal rank is at most 5. Relabel the five alternatives as x_i with $i = 1,..,5$. Then we can select a minor consisting of the five types $\succ_j$ each with $x_i \succ_j x_{i+1}$ modulo 5 and with $j = 1,..,5$ and $x_j \succ_j x$ for all $x \neq x_j$. This way we can

[36]Now $\kappa_1 = 5$, $\kappa_2 = 2^{10} = 1024$ and $\kappa_3 = 4^5 = 1024$, hence we can identify the 120 preference types since $5 + 120 + 120 \geq 2 \cdot 120 + 2$ types.

arrange orders to obtain the circulant minor

$$\begin{bmatrix} a & ab & ab^2 & ab^3 & ab^4 \\ ab & ab^2 & ab^3 & ab^4 & a \\ ab^2 & ab^3 & ab^4 & a & ab \\ ab^3 & ab^4 & a & ab & ab^2 \\ ab^4 & a & ab & ab^2 & ab^3 \end{bmatrix} \tag{19}$$

where $a = \frac{\gamma}{1-(1-\gamma)^5}$ and $b = 1-\gamma$. As seen before, minor (19) has a full rank, hence full Kruskal rank.

The reasoning for collection $\mathcal{A}_2$ is as in the proof of Proposition 1, where now we construct a 120×120 minor selecting, for each type, the unique row corresponding to the choice functions selecting the top alternatives in all 10 sets according to that type. The type conditional probability for choice functions restricted to this collection when they maximise an order now becomes

$$\left(\frac{\gamma}{1-(1-\gamma)^2}\right)^{10}$$

which is therefore the largest value achievable in a given row. Again, by collecting this term outside of the matrix, we can permute rows so as to obtain a square minor with 1 on the main diagonal, with all other values less than 1. Now setting $(1-\gamma) = \frac{1}{120}$ the minor is an invertible diagonally dominant matrix. In turn, this establishes that the minor is generically invertible and, hence has full Kruskal rank. For the third and final collection, we proceed in now familiar fashion, selecting for each preference type the choice function that picks the top alternative from the first four sets, and the bottom alternative from the fifth set. The type conditional probability of the corresponding preference type is

$$\left(\frac{\gamma}{1-(1-\gamma)^4}\right)^5 (1-\gamma)^3$$

with all other terms being smaller. Hence we can collect this term, and again permute this minor so as to obtain a matrix with 1 on the main diagonal. Again setting $(1-\gamma) = \frac{1}{120}$ the minor is a diagonally dominant matrix, which is invertible, and we can proceed as above.

$$L_{[i:a]}(\emptyset) = \frac{p(i,\{a,i\})}{p(a,\{a,i\})} = \frac{\gamma_i(1-\gamma_a)}{\gamma_a} \Rightarrow \gamma_a = \frac{\gamma_i}{p(i,\{a,i\}) + \gamma_i p(a,\{a,i\})}$$

F Screenshots and Instructions

Figure 14: Screenshot for the problem in which all the pencils were available.

Figure 15: Screenshot for the problem in which all the pens were available.

Figure 16: Screenshot for the Likert-scale task for pencils.

Figure 17: Screenshot for the Likert-scale task for pens.

Figure 18: Screenshot for one of the Raven question.

We report below the script followed by the interviewers who conducted the experiments under Carla Guerriero's direct supervision. Given the young age of our subjects, the interviewers followed the script as guideline in instructing the children rather than reading them word by word. The following pages show the original Italian version, followed by an English translation.

Figure 19: Instructions used to train the interviewers, pencils and pens task (Italian original and English translation).

ITALIAN VERSION

(ESPERIMENTO 1)

Presentazione ai bambini.

1. In questo gioco dovrete scegliere tra un gruppo di alternative (si mostrano le alternative). Ci sono quattro tipi di matite, quattro tipi di penne, e quattro block notes. **[Mostrare le alternative mette tutti i bambini sullo stesso piano a livello di conoscenza delle alternative.]**
2. Vi verranno fatte delle domande, ed in ognuna dovete scegliere l'alternativa che vi piace di piu' senza pensare alle scelte che avete fatto prima.
3. Più scegliete l'alternativa che preferite e più è probabile che la riceverete. Alla fine del gioco riceverete una sola matita, una sola penna e un solo block notes, solo uno per tipo. Quindi mi raccomando se qualcosa vi piace molto dovete sceglierlo quando appare.
4. Quando avete finite di scegliere dovrete dire quanto vi piacciono le alternative. Con una pagina stampata si fa vedere come dire che le cose piacciono o non piacciono.
5. E infine avrete un gioco, in cui dovete completare una figura con I pezzi mancanti come i puzzle.

ENGLISH VERSION

(EXPERIMENT 1)

Presentation to the children.

1. In this game, you'll have to choose from a group of options (the options are shown). There are four types of pencils, four types of pens, and four types of notepads. **[Displaying all the alternatives puts all children on an equal footing in terms of knowing what the available options are.]**
2. You will be asked some questions, and for each one, you should choose the option you like the most without thinking about the choices you have made before.
3. The more you pick the option you prefer, the more likely it is you will get it. At the end of the game, you will receive just one pencil, one pen, and one notepad – only one of each. So if you really like something, make sure to choose it when it appears.
4. When you have finished making your choices, you'll be asked how much you like each of the options. **[Show a printout of the preference intensity display]**
5. Lastly, there will be a game where you have to complete a picture using the missing pieces, like a puzzle.

Figure 20: Instructions used to train the interviewers, lotteries tasks (Italian original).

ITALIAN VERSION

(ESPERIMENTO 2)

PACHINKO:

- Ciao a tutti.
- Questo gioco si chiama **Pachinko**. E' come un flipper. E' costruito in modo che se una pallina viene inserita finisce lo stesso numero di volte negli otto cestini finali.
- Come vedete sotto il Pachinko, ci sono dei posti dove e' possibile mettere delle strisce colorate (**si mostrano le strisce colorate**). Il gioco riguardera' scegliere delle strisce da mettere sotto il Pachinko. E scegliendo le strisce potrete vincere dei premi. Ora vi spiego cosa vogliono dire queste strisce.
- Le strisce hanno due colori diversi: bianco e giallo. Con il giallo vincete uno e con il bianco non vincete niente. I numeri sono scritti sopra i colori cosi che non possiamo sbagliarci.
- Come fate a vincere? Allora, immaginiamo di mettere una striscia colorata sotto il Pachinko. Se la pallina finisce in un cestino che sta sopra il giallo vincete una matita, se finisce sopra il bianco non vincete nulla [**si fa un esempio**].
- Il gioco avra' due domande. In entrambe le domande potrete vincere dei premi.
- Il gioco e' molto semplice, ma nella seconda domanda dovrete scegliere due strisce, non una, delle tre che vedete sul foglio. Fate attenzione. Quindi come funziona in questo caso? **Il giallo vince sul bianco.** Quindi se mettete queste due strisce, (**si mostra un esempio con due strisce diverse da quelle nelle domande**) come vedete la pallina e' finita sia sopra il colore Giallo-1 e Bianco-0, quindi vincete un premio.
- Una volta fatte le vostre scelte, i fogli verranno raccolti e poi una pallina verra' lanciata e cosi vedremo cosa avrete vinto.

GIOCO DELLE CARTE

- Prima di procedure all'estrazione faremo un altro gioco.
- Ora giochiamo ad un altro gioco. In questo gioco ci sono otto carte. Come vedete ci sono sette carte che raffigurano delle pecore e una carta che raffigura un lupo (**si mostrano le carte**).
- Importante: se viene girata la carta con il lupo si perde.
- Per vincere a questo gioco dovrete decidere quante carte verranno girate. Quanto potete vincere a questo gioco? **Potete vincere 1 matita per ogni pecora che girate.**
- Quindi, se vengono girate cinque carte e sono tutte pecore vincete 5 matite. Ma se tra le cinque carte c'e' un lupo allora vincerete 0 matite. Piu' carte girate piu' potete vincere, ma piu' probabile e' che appaia un lupo.
- In questo gioco, ognuno di voi ha un foglio con numeri da 0 a 8. Questi numeri indicano il numero di carte che volete girare.
- Ognuno di voi deve decidere **uno e un solo** numero di carte da scoprire.
- Una volta che avrete deciso, i fogli verranno raccolti e la maestra girera' una carta alla volta fino a trovare il lupo. E ognuno di voi vincera' se avete scelto un numero di carte minore a quello che ha girato la maestra quando ha trovato il lupo.

PROTOCOLLO:

L'esperimento va nel seguente ordine:

1. Si svolge il Pachinko.
   a. Si distribuiscono i fogli delle risposte.
   b. Si spiega il gioco.
   c. I bambini fanno le scelte e poi i fogli vengono raccolti.
2. Si svolge il gioco delle carte.
   a. Si distribuisce il foglio delle risposte.
   b. Si spiega li gioco.
   c. I bambini decidono quante carte girare.
3. Si inserisce la pallina nel Pachinko e si riporta il risultato.
4. Si girano le carte e si riporta il risultato.
5. Si „pagano" i bambini.

Figure 21: Instructions used to train the interviewers, lotteries tasks (English translation).

ENGLISH VERSION

(EXPERIMENT 2)

PACHINKO

- Hello everyone.
- This game is called Pachinko. It is similar to pinball. It is designed so that if a ball is dropped, it has the same chance of landing in each of the eight final baskets.
- As you can see, under the Pachinko there is space to put coloured strips **[the coloured strips are displayed]**. The game is about choosing strips to place under the Pachinko. By choosing the strips, you can win prizes. Let me explain what these strips mean.
- The strips have two different colours: white and yellow. With yellow you win one **[item]**, while with white you win nothing. The numbers are written on the colours so we cannot go wrong.
- How do you win? Let's imagine placing a coloured strip under the Pachinko. If the ball lands in a pocket on top of yellow, you win a pencil; if it lands on top of white, you win nothing **[example]**.
- There will be two questions. You can win prizes with both questions.
- The game is very simple, but in the second question you will have to choose **two** strips, not just one, from the three you see on the sheet. Pay attention. So how does it work in this case? Yellow beats white. So if you place these two strips **[an example is shown with two strips different from those in the actual questions],** you can see that the ball has landed on both the Yellow-1 and the White-0, so you win a prize.
- Once you have made your choices, the sheets will be collected, and then a ball will be released, and we will see what you have won.

CARD GAME

- Before the draw we will play another game.
- Now we are going to play a different game. In this game there are eight cards. As you can see, seven cards show sheep, and one card shows a wolf **[display the cards].**
- Important: if the card with the wolf is flipped, you lose.
- To win in this game, you must decide how many cards to turn. How much can you win in this game? You can win 1 pencil for each card that reveals a sheep.
- So, if five cards are turned and they are all sheep, you win 5 pencils. But if among those five cards there is a wolf, you win 0 pencils. The more cards you flip, the more you can win, but the more likely it is that the wolf appears.
- In this game, each of you has a sheet with numbers from 0 to 8. These numbers represent the number of cards you want to turn.
- Each of you must choose one and only one number of cards to turn.
- Once you've decided, the sheets will be collected, and the teacher will turn the cards one by one until the wolf is found. And each of you will win if the number of cards you chose is **lower** than the number of cards the teacher turned before finding the wolf.

PROTOCOL [for the experimenters]:

The experiment proceeds in the following order:

1. Run the Pachinko game.
    a. Answer sheets are handed out.
    b. The game is explained.
    c. Children make their choices, and the sheets are collected.
2. Run the card game.
    a. Answer sheets are handed out.
    b. The game is explained.
    c. Children decide how many cards to turn.
3. The marble is released into the Pachinko, and the result is recorded.
4. The cards are turned, and the result recorded.
5. The children are "paid".

Online Appendix

A NBC and AC by grade and school

In Section **??**, Figure 5, we have presented data about violations of No Binary Cycle and Always Chosen. Here, we report the same evidence by grade, highlighting the development of the relationship between NBC and AC, and by school, highlighting the differences across socio-economic background.

	Grade 1 (n=108)			Grade 2 (n=78)			Grade 3 (n=92)		
	AC rational yes	AC rational no	Total	AC rational yes	AC rational no	Total	AC rational yes	AC rational no	Total
NBC yes	32.4	39.8	72.2	38.4	39.8	78.2	51.1	33.7	84.8
NBC no	4.6	23.2	27.8	5.1	16.7	21.8	2.2	13.0	15.2
Total	37.0	63.0	100.0	43.5	56.5	100.0	53.3	46.7	100.0

	Grade 4 (n=112)			Grade 5 (n=105)		
	AC rational yes	AC rational no	Total	AC rational yes	AC rational no	Total
NBC yes	61.6	29.5	91.1	76.1	14.7	90.8
NBC no	2.7	6.2	8.9	1.8	7.4	9.2
Total	64.3	35.7	100.0	77.9	22.1	100.0

	School L (n=255)			School H (n=244)		
	AC rational yes	AC rational no	Total	AC rational yes	AC rational no	Total
NBC yes	46.3	31.7	78.0	59.8	29.9	89.7
NBC no	4.7	17.3	22.0	1.7	8.6	10.3
Total	51.0	49.0	100.0	61.5	38.5	100.0

Table 8: Pencils Experiment: Proportion of AC and NBC rationality violations by School

	Grade 1 (n=108)			Grade 2 (n=78)			Grade 3 (n=92)		
	AC rational yes	AC rational no	Total	AC rational yes	AC rational no	Total	AC rational yes	AC rational no	Total
NBC yes	23.1	39.8	62.9	29.5	35.9	65.4	41.3	37.0	78.3
NBC no	6.5	30.6	37.1	7.7	26.9	34.6	4.3	17.4	21.7
Total	29.6	70.4	100.0	37.2	62.8	100.0	45.6	54.4	100.0

	Grade 4 (n=112)			Grade 5 (n=105)			Aggregate		
	AC rational yes	AC rational no	Total	AC rational yes	AC rational no	Total	AC rational yes	AC rational no	Total
NBC yes	49.1	31.2	80.3	53.2	33.9	87.1	39.9	35.4	75.3
NBC no	6.3	13.4	19.7	4.6	8.3	12.9	5.8	18.9	24.7
Total	55.4	44.6	100.0	57.8	42.2	100.0	45.7	54.3	100.0

	School L (n=255)			School H (n=244)		
	AC rational yes	AC rational no	Total	AC rational yes	AC rational no	Total
NBC yes	32.9	38.0	70.9	47.1	32.8	79.9
NBC no	6.3	22.8	29.1	5.3	14.8	20.1
Total	39.2	60.8	100.0	52.4	47.6	100.0

Table 9: Pens Experiment: Proportion of AC and NBC rationality violations by School

B Joint distribution choices in the lottery tasks by grade and school

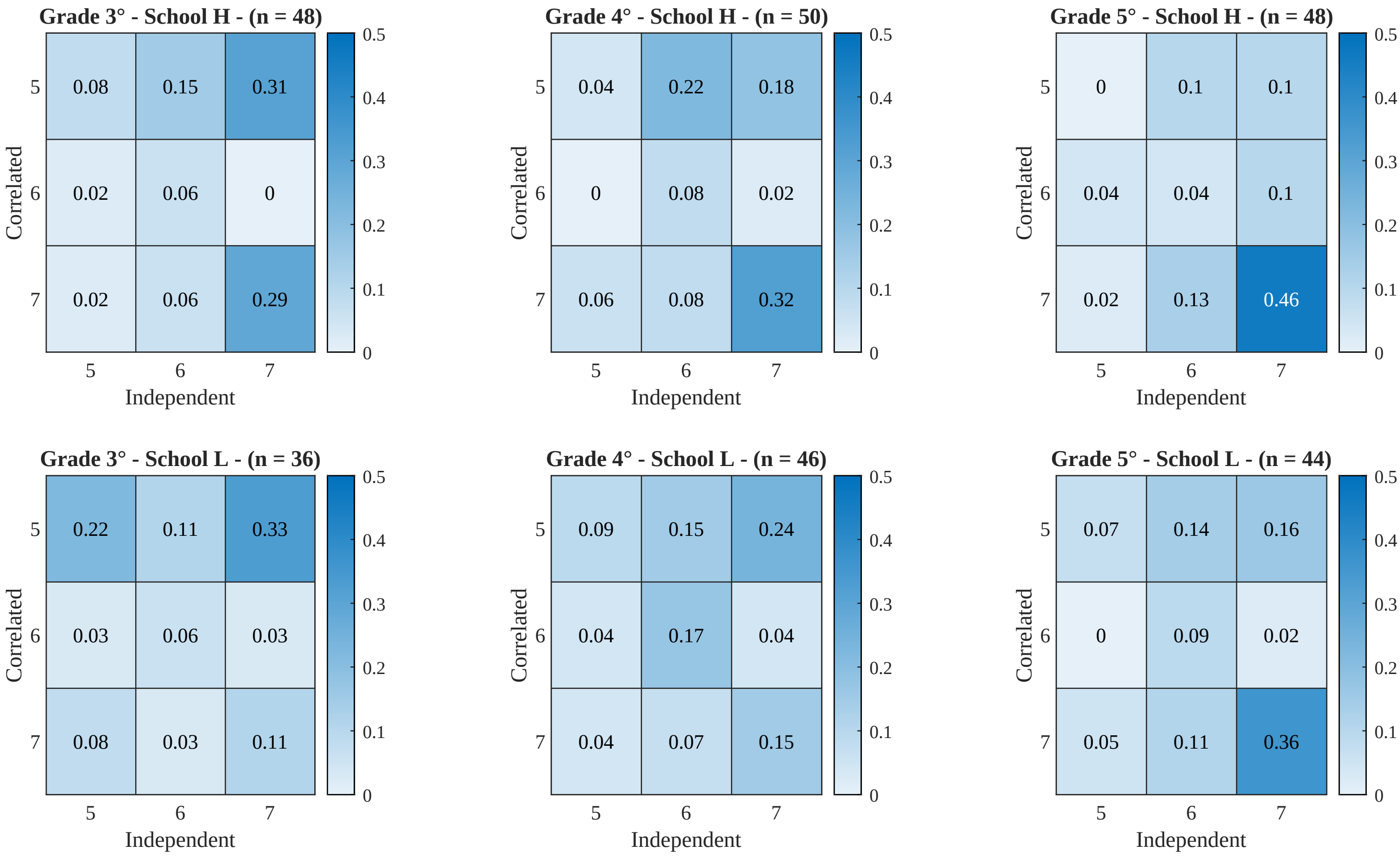

Figure 22: Joint distribution of the 'risky choices' by school and grade

C Errors in the Pens task

In this appendix, we report the results of the pens task - these broadly replicate those seen in section 4.3.1 of the main text for the pencils task.

Comparisons of the distributions of preference types reflect similar patterns to those in the pencils task (Figure 23). The differences are more marked in the gender(p-value = 0.043) than in the school dimension (chi-square test, p-value = 0.081). Recall that while pencils differed both in colour and in the animal shaped erasers, pens only differed in colour. Our estimates of the consideration parameters are reported in Table 10 and and broadly in line with what already observed in the pencil task: again the consideration/rationality parameter iincreases with grade in both schools (differences between 1st and 5th grade are large and significant, t-tests, p-values < 0.01). Further, differences across schools are larger in the first grade (t-test, p-value < 0.05) than in the fifth grade (p-value > 0.1), while in the aggregate, we find a significant difference in the ability to avoid errors across schools (p-value < 0.01). As for the pencil taks, the development of the consideration parameter does not differ across gender, with differences that are never statistically significant at any age or in the aggregate.

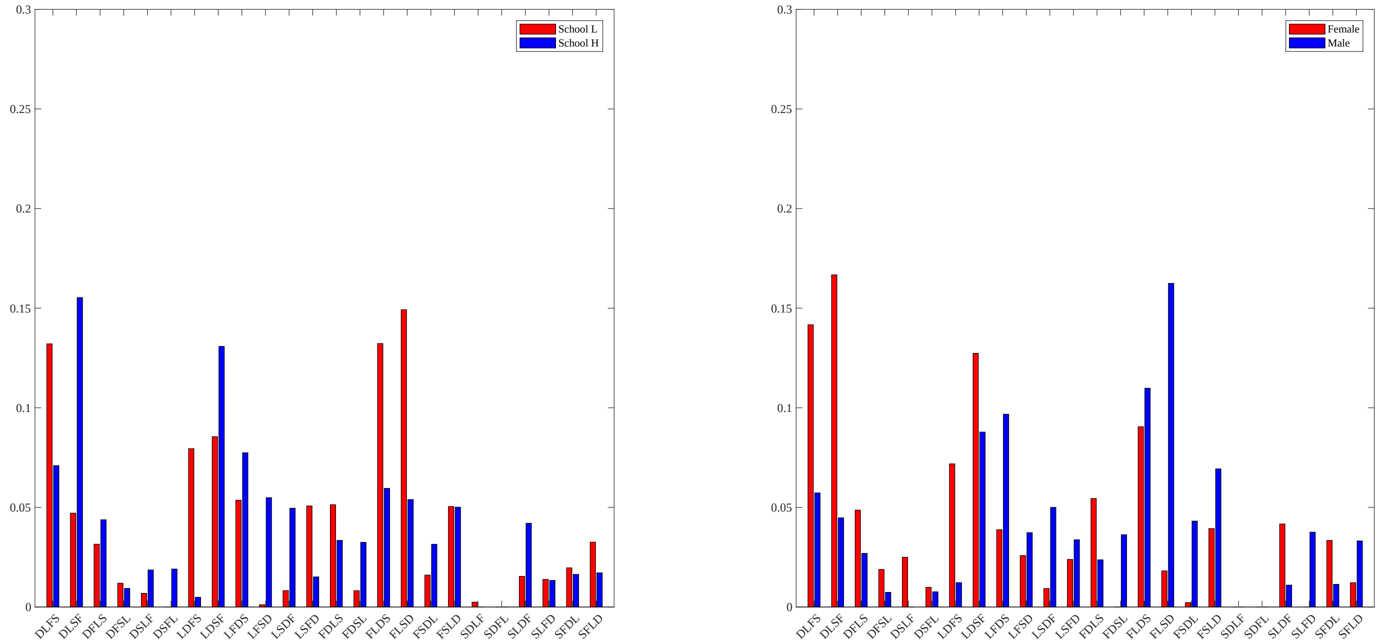

Figure 23: Distribution of the preference types by school and gender (pens task).

	School L	School H		Female	Male	
1st Grade	0.63	0.77		0.65	0.73	1st Grade
2nd Grade	0.72	0.77		0.74	0.77	2nd Grade
3rd Grade	0.72	0.85		0.79	0.80	3rd Grade
4th Grade	0.88	0.87		0.89	0.86	4th Grade
5th Grade	0.86	0.90		0.85	0.91	5th Grade
Schools	0.76	0.84		0.78	0.82	Gender
	Socio-economic			Gender		

Table 10: Estimated values of γ (pens task)

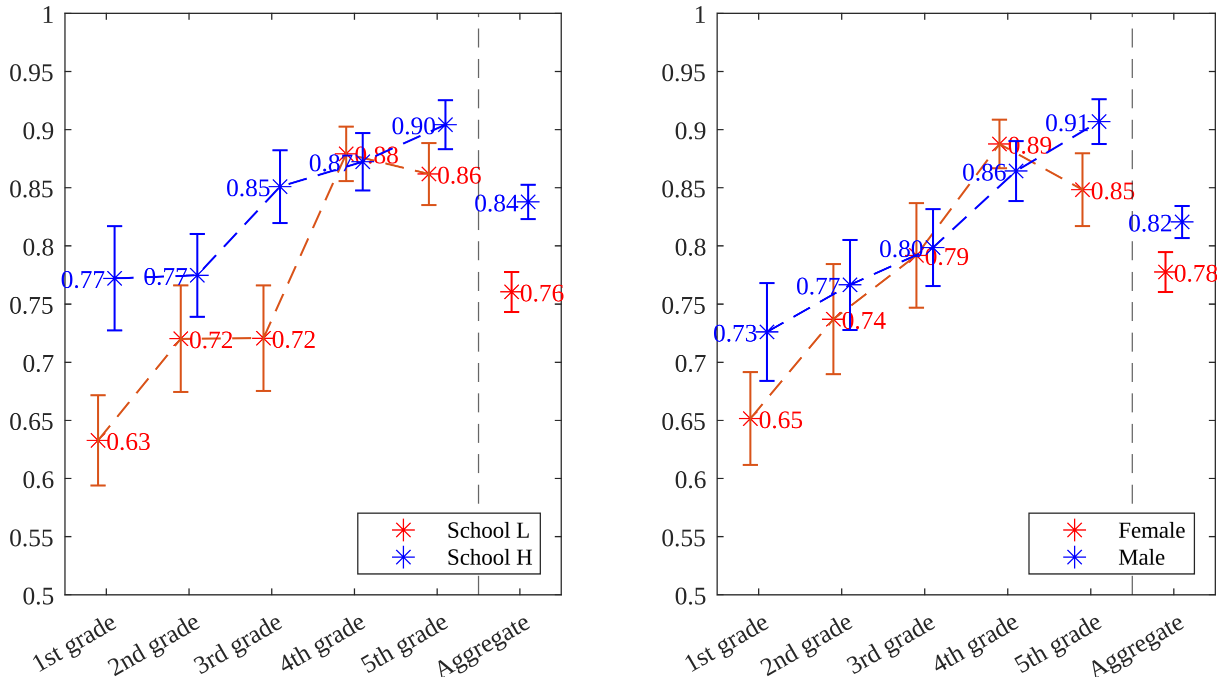

Figure 24: Development of the consideration parameter by school and gender

D Relevant factors associated with the consideration parameter

This appendix contains some results on how the covariates we have available might influence the consideration parameter γ in the various tasks, focusing on individual-level estimates.[37]

D.1 Riskless choice

For the analysis in this section, we estimate a consideration parameter $\hat{\gamma}_{i,s}$ for each participant i enrolled in school s. Since $\hat{\gamma}_{i,s}$ is an estimated variable at the individual level, we will account for the resulting heteroskedasticity using robust standard errors clustered at the level of the experimental session.

We run an OLS regression for $\hat{\gamma}_{i,s}$ on the available independent variables collected in the vector $\mathbf{X}_{i,s}$, and using a vector $\mathbf{Z}_{i,s}$ of control variables.[38] The independent variables are the Raven scores of each child, the highest educational attainment by either parent and the grade year the child is in, while the controls include gender, and preference intensity, measured on

[37] While Proposition 1 establishes identification in the pencils/pens task, individual estimates may be noisy due to the small sample size. In a separate online appendix, we replicate our results using an empirical Bayesian approach (known as "approximate" (Kass & Steffey, 1989)) within our latent class model to reduce this concern. The results are robust everywhere, and if anything, correlations are greater.

[38] Since $\hat{\gamma}_{i,s}$ takes values between 0 and 1, in a separate online appendix, we replicate the results using a beta regression with logistic transformation. The results are robust.

a Likert scale as mentioned in section 2.1.[39] We also include a school fixed effect, denoted by ρ_s is the school fixed effect. The distinction between variables in $\mathbf{X}_{i,s}$ and $\mathbf{Z}_{i,s}$ is readily seen. Our objective is to explore potential determinants of choice errors conditional on pupils' preferences. Both gender and preference intensity allow us to control for any residual effect of the pupil's preferences that are not captured by the model.[40] The equation we estimate is:

$$\hat{\gamma}_{i,s} = \alpha + \beta_1 \mathbf{X}_{i,s} + \beta_2 \mathbf{Z}_{i,s} + \rho_s + \varepsilon_{i,s} \tag{20}$$

Results are presented in Table 16 where the covariates are coded such that the constant represents the consideration parameter of a pupil who is male, has both parents with the lowest educational attainment, has scored the lowest in the Raven test, is attending the 1st grade in **School L**, and is indifferent between all alternatives. Specifications that contain parental education have a smaller sample size due to frictions in our survey, as reported in Table 1.

[39] See the online appendix for descriptive statistics on preference intensity and its calculation.

[40] Our theoretical analysis allows us to disentangle preference heterogeneity from the consideration parameter. Our results rely on preferences entering our model ordinally. This does not account for different degrees of preference intensity (Alós-Ferrer & Garagnani, 2021). Children, however, may develop their preference intensity differently between schools. Confirming our results after controlling for preference intensity provides an important robustness check for our structural analysis.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	0.207***			0.0992**	0.190***			0.0912*
	(0.0408)			(0.0478)	(0.0404)			(0.0493)
Second Grade		0.0736**		0.0701*		0.0569*		0.0633
		(0.0328)		(0.0381)		(0.0329)		(0.0394)
Third Grade		0.0952***		0.0661*		0.0834***		0.0623*
		(0.0344)		(0.0391)		(0.0276)		(0.0365)
Fourth Grade		0.143***		0.115***		0.134***		0.114***
		(0.0265)		(0.0327)		(0.0254)		(0.0325)
Fifth Grade		0.189***		0.150***		0.181***		0.149***
		(0.0249)		(0.0314)		(0.0224)		(0.0316)
Parental education			0.0265**	0.0226**			0.0171	0.0169*
			(0.0102)	(0.00867)			(0.0106)	(0.00912)
Preference Intensity	0.181***	0.186***	0.193***	0.183***	0.172***	0.172***	0.180***	0.174***
	(0.0411)	(0.0386)	(0.0520)	(0.0428)	(0.0427)	(0.0399)	(0.0514)	(0.0432)
Gender	-0.0151	-0.00290	-0.0149	0.00508	-0.0160	-0.00482	-0.0165	0.00334
	(0.0162)	(0.0158)	(0.0175)	(0.0156)	(0.0158)	(0.0153)	(0.0172)	(0.0155)
School Fixed Effect					0.0394*	0.0522***	0.0453	0.0295
					(0.0224)	(0.0169)	(0.0283)	(0.0219)
Constant	0.675***	0.680***	0.699***	0.569***	0.669***	0.670***	0.713***	0.584***
	(0.0366)	(0.0279)	(0.0451)	(0.0428)	(0.0366)	(0.0240)	(0.0442)	(0.0459)
Observations	499	499	404	404	499	499	404	404
R-squared	0.120	0.158	0.078	0.184	0.129	0.174	0.087	0.188

Table 11: Consideration, age, parental education, and cognitive abilities (pencils task).

The dependent variable is the estimated consideration parameter in the pencils task. The covariates are coded such that the constant represents the consideration parameter of a pupil who is male, has both parents with the lowest educational attainment, has scored the lowest in the Raven test, is attending the 1st grade in **School L**, and is indifferent between all alternatives. Parental education is coded from 1 to 6 following the categories in Figure 4. The coefficient of Raven's scores is the difference between the highest and lowest-scoring pupils. Since the pupils answered eight matrices, the coefficient can be divided by eight to obtain the effect of each correct answer. The regression models are estimated in Stata (Robust clustered standard errors at the experimental session level). *** <0.01, ** <0.05, *<0.1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	0.176***			0.115**	0.200***			0.104*
	(0.0434)			(0.0506)	(0.0407)			(0.0525)
Second Grade		0.0670**		0.0639*		0.0491		0.0556
		(0.0327)		(0.0373)		(0.0310)		(0.0377)
Third Grade		0.0989***		0.0655		0.0856***		0.0607
		(0.0358)		(0.0406)		(0.0278)		(0.0365)
Fourth Grade		0.153***		0.123***		0.143***		0.122***
		(0.0269)		(0.0338)		(0.0257)		(0.0333)
Fifth Grade		0.191***		0.144***		0.182***		0.142***
		(0.0245)		(0.0316)		(0.0215)		(0.0313)
Parental education			0.0279**	0.0230**			0.0164	0.0157
			(0.0104)	(0.00870)			(0.0110)	(0.00945)
Preference Intensity	0.196***	0.125**	0.111*	0.123**	0.109*	0.110**	0.100	0.115**
	(0.0487)	(0.0481)	(0.0608)	(0.0458)	(0.0557)	(0.0495)	(0.0605)	(0.0467)
Gender	-0.0331**	-0.000984	-0.0127	0.00675	-0.0140	-0.00327	-0.0148	0.00442
	(0.0161)	(0.0157)	(0.0175)	(0.0151)	(0.0158)	(0.0151)	(0.0170)	(0.0149)
School Fixed Effect					0.0427*	0.0569***	0.0549*	0.0377
					(0.0231)	(0.0170)	(0.0296)	(0.0227)
Constant	0.670***	0.710***	0.735***	0.592***	0.694***	0.698***	0.748***	0.609***
	(0.0344)	(0.0262)	(0.0434)	(0.0377)	(0.0344)	(0.0213)	(0.0431)	(0.0412)
Observations	499	499	404	404	499	499	404	404
R-squared	0.094	0.131	0.044	0.160	0.103	0.149	0.058	0.166

Table 12: Consideration, age, parental education, and cognitive abilities (pens task).

The dependent variable is the estimated consideration parameter in the pens task. The covariates are coded such that the constant represents the consideration parameter of a pupil who is male, has both parents with the lowest educational attainment, has scored the lowest in the Raven test, is attending the 1st grade in **School L**, and is indifferent between all alternatives. Parental education is coded from 1 to 6 following the categories in Figure 4. The coefficient of Raven's scores is the difference between the highest and lowest-scoring pupils. Since the pupils answered 8 matrices, the coefficient can be divided by eight to obtain the effect of each correct answer. The regression models are estimated in Stata (Robust clustered standard errors at the experimental session level). *** <0.01, ** <0.05, *<0.1.

Results confirm that grade, a proxy for age, strongly influences the consideration parameter. Performance in the Raven test and parental education are also large and significant determinants of consideration. Specifications (1)-(3) and (5)-(7) report the unconditional correlations with and without school fixed effect, and conditional on preference intensity, which is, as expected, significantly correlated with the consideration parameter; and gender, which instead does not affect the consideration parameter. Since our independent variables are highly correlated with each other, i.e. the correlation between Raven's scores and Grade is +0.48 ($p < 0.01$), while between Raven's scores and Parental education is +0.18 ($p < 0.01$), in specifications (4) and (8), we show the conditional correlations to better identify their individual roles. In the pencils task (similarly in the pens task), the analysis of individual grades (specification 8) shows large developments between 1st-2nd grade and 3rd-4th grade. Overall, each additional year of school has an average effect of 0.037 (p-value < 0.01), which is equivalent to roughly two additional brackets in parental educational attainment. The correlation with the performance in the Raven test becomes only mildly significant, and its magnitude reduces considerably. Nonetheless, the difference between a zero and a perfect score is equivalent to about three years of schooling. Finally, even if we cannot disentangle the different channels of the effect of schools, i.e., peer effects, teachers' quality, neighborhood quality, etc, we find that overall it is smaller than a year of schooling and not significant.

D.2 Risky choice

As noted, it is not possible to identify the consideration parameter in the correlated coin drop tasks; however, we can study a proxy for it. In the independent task, we are mostly interested in failures of first order stochastic dominance - hence it seems appropriate to study the ratio between the probability of choosing (correctly) the dominant lottery vis a vis the probability of making a different choice, as shown by the logistic regression in equation 21. In the correlated task, what is more interesting is the ratio between the probability of choosing the dominant lottery vis a vis the probability of choosing the 'correlation neglect' lottery (see results on Implication **2'** in the previous section) - hence we focus only on the regression in equation 22 and not on the choices of ℓ_6^C which regarded only 15% of the pupils.

We use the same set of independent variables as in the pencils/pens task, apart from dropping preference intensities (not relevant for this task due to its different nature) and introducing, in equation 22, a dummy that takes value one if the pupil has chosen the dominant lottery in the independent task. This variable is, as expected, positively related to choosing optimally in the correlated task.

$$\log\left(\frac{p(\ell_7^I|\mathbf{X}_{i,s},\mathbf{Z}_{i,s},\rho_s)}{p(\ell_6^I+\ell_5^I|\mathbf{X}_{i,s},\mathbf{Z}_{i,s},\rho_s)}\right)=\alpha+\beta_1\mathbf{X}_{i,s}+\beta_2\mathbf{Z}_{i,s}+\rho_s+\varepsilon_{i,s} \tag{21}$$

$$\log\left(\frac{p(\ell_7^C|\mathbf{X}_{i,s},\mathbf{Z}_{i,s},\rho_s)}{p(\ell_5^C|\mathbf{X}_{i,s},\mathbf{Z}_{i,s},\rho_s)}\right)=\alpha+\beta_1\mathbf{X}_{i,s}+\beta_2\mathbf{Z}_{i,s}+\rho_s+\varepsilon_{i,s} \tag{22}$$

The results are reported in Tables 13 and 14. In both regressions, once the school fixed effect is included (specification 8), only the coefficient related to the child's fifth grade is significant (p-value < 0.1 in Table 13 and p-value < 0.01 in Table 14). The magnitude of these coefficients is large, with the odds ratio increasing about fourfold between pupils of third and fifth grade.[41]

[41] In a separate online appendix, we replicate the results also including the two other schools; there, the regression coefficients are larger and more significant, pointing to the fact that the weaker results in this sub-sample might depend on its smaller size. We also run the same regression excluding pupils who failed the independent task and find similar correlation patterns.

Finally, the independent task, being a transparent test of FOSD, also serves as a comprehension test. The odds ratio doubles for pupils who successfully answer the independent task.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	1.117			-0.135	0.754			-0.418
	(0.720)			(0.971)	(0.693)			(1.027)
Fourth Grade		0.322		0.285		0.387		0.341
		(0.584)		(0.500)		(0.556)		(0.493)
Fifth Grade		1.015		1.044*		1.094*		1.111*
		(0.632)		(0.575)		(0.590)		(0.601)
Parental education			0.386*	0.445**			0.256	0.312
			(0.204)	(0.223)			(0.235)	(0.248)
Gender	0.00810	0.0941	0.0826	0.167	-0.0306	0.0662	0.0601	0.167
	(0.398)	(0.379)	(0.402)	(0.400)	(0.412)	(0.383)	(0.418)	(0.419)
School Fixed Effect					0.862*	0.989**	0.668	0.730
					(0.456)	(0.472)	(0.500)	(0.512)
Constant	0.608	0.898**	0.0917	-0.426	0.452	0.398	0.184	-0.219
	(0.539)	(0.434)	(0.656)	(0.978)	(0.581)	(0.540)	(0.662)	(0.945)
Observations	272	272	217	217	272	272	217	217
Log-likelihood	-262.46	-259.59	-206.27	-202.12	-259.82	-256.02	-205.12	-200.90
AIC	536.91	535.18	424.53	428.23	535.63	532.05	426.24	429.79
BIC	558.55	564.03	444.81	468.79	564.48	568.10	453.28	477.11

Table 13: Independent task, age, parental education, and cognitive abilities.

We estimate a binomial Logit in which the dependent variable is a dummy that takes value 1 if the pupils chose ℓ_7^I and 0 otherwise. The independent variables are the same as in Table 16 apart from preference intensity. The regression models are estimated in Stata using the function *logit*. *** <0.01, ** <0.05, *<0.1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	1.320**			1.064*	1.127**			0.885
	(0.572)			(0.549)	(0.567)			(0.574)
Fourth Grade		0.498		0.516		0.539		0.553
		(0.433)		(0.422)		(0.409)		(0.412)
Fifth Grade		1.302***		1.357***		1.358***		1.399***
		(0.437)		(0.442)		(0.413)		(0.424)
Parental education			0.148	0.188			0.0563	0.105
			(0.138)	(0.145)			(0.139)	(0.141)
Gender	-0.543*	-0.461	-0.655**	-0.571*	-0.574**	-0.514*	-0.691**	-0.597*
	(0.296)	(0.303)	(0.320)	(0.328)	(0.283)	(0.291)	(0.310)	(0.321)
Succ. Ind. Task	0.820***	0.834***	0.892***	0.874***	0.792***	0.787***	0.856***	0.844***
	(0.274)	(0.257)	(0.293)	(0.313)	(0.288)	(0.272)	(0.302)	(0.321)
School Fixed Effect					0.427	0.611*	0.493	0.462
					(0.394)	(0.344)	(0.435)	(0.352)
Constant	-1.188**	-0.960**	-0.818	-2.283***	-1.266**	-1.282***	-0.748	-2.145***
	(0.465)	(0.384)	(0.520)	(0.731)	(0.507)	(0.492)	(0.511)	(0.703)
Observations	272	272	217	217	272	272	217	217
Log-likelihood	-255.45	-249.50	-201.60	-191.67	-254.22	-247.03	-200.19	-190.21
AIC	526.91	519.00	419.20	411.35	528.43	518.05	420.39	412.43
BIC	555.76	555.06	446.24	458.67	564.49	561.32	454.19	466.50

Table 14: Correlation neglect, age, parental education, and cognitive abilities.

We estimate a multinomial Logit in which the dependent variable represents the pupil's choices in the Correlated task. The baseline is represented by the pupils who choose ℓ_5^C and we only report the coefficients related to equation 22. The independent variables are the same as in Table 16 apart from preference intensity and Success Ind. Task. This latter is a dummy that takes the value 1 if the pupil chooses the best lottery in the Independent task. The regression models are estimated in Stata using the function *mlogit*. *** <0.01, ** <0.05, *<0.1.

D.3 Relevant factors associated with the Raven scores

We now replicate our regression analysis on the Raven scores. We document a slightly different non-linear development by grade with the majority of the effect happening between 2nd and 3rd grades. The average effect of a year of schooling is 0.085 (p-value < 0.01). The effect of the parental educational attainment is significant but with a smaller size than in previous results, i.e. after controlling for school fixed effect, one-fourth of a year of schooling. On the other hand, the role of the school seems crucial for the development of fluid intelligence with an effect that is very significant and roughly equivalent to a year of schooling. This result mirrors the divergent path reported in Figure 11, and it is starkly different from our finding in Tables 16 to 14.

	(1)	(2)	(3)	(4)	(5)	(6)
Second Grade	0.0958***		0.0922**	0.0670**		0.0743**
	(0.0313)		(0.0344)	(0.0264)		(0.0335)
Third Grade	0.235***		0.240***	0.213***		0.225***
	(0.0312)		(0.0327)	(0.0333)		(0.0364)
Fourth Grade	0.271***		0.269***	0.254***		0.260***
	(0.0331)		(0.0366)	(0.0292)		(0.0357)
Fifth Grade	0.335***		0.345***	0.322***		0.335***
	(0.0330)		(0.0314)	(0.0278)		(0.0335)
Parental education		0.0383***	0.0352***		0.0191*	0.0198**
		(0.0130)	(0.00936)		(0.0110)	(0.00973)
Gender	0.00213	-0.0370	7.30e-05	-0.00150	-0.0406	-0.00449
	(0.0214)	(0.0273)	(0.0208)	(0.0198)	(0.0263)	(0.0197)
School Fixed Effect				0.0950***	0.0904**	0.0745***
				(0.0168)	(0.0432)	(0.0195)
Constant	0.363***	0.452***	0.252***	0.334***	0.470***	0.277***
	(0.0230)	(0.0544)	(0.0474)	(0.0291)	(0.0499)	(0.0493)
Observations	499	404	404	499	404	404
R-squared	0.243	0.037	0.277	0.276	0.061	0.293

Table 15: Raven's scores, age, and parental education.

The dependent variable is normalised - in [0,1] - Raven's scores. The covariates are coded such that the constant represents the consideration parameter of a pupil who is male, has both parents with the lowest educational attainment, has scored the lowest in the Raven test, is attending the 1st grade in **School L**, and is indifferent between all alternatives. The coefficient of parental education is the effect of improving the degree of at least one of the parents. Grade can be interpreted as a one-year age effect. The only control here is Gender, which is uncorrelated with Raven's scores. The regression models are estimated in Stata (Robust clustered standard errors at the session level). *** <0.01, ** <0.05, *<0.1.

D.4 Linear trends

Based on Figures 7-11, in the paper we have discussed whether pupils in School L and H have different development patterns. Here, we estimate a simple regression to test whether there are the trend reported in the figures are significantly different. To do so, we run a regression where the dependent variables are those described in previous sections, while the covariates are simply: grade, a dummy representing School L and H, and the interaction between these two covariates. The coefficient of the interaction term measure the difference in slope between the linear trends in the two schools. The results confirm our discussion in the paper. Namely, in the pencils and pens task there is a convergence pattern with children in School L catching up to those in School H. In the lotteries tasks, instead, we find no difference between schools. Finally, comparing Raven scores we find a divergent pattern with children in the schools starting at similar levels, but then with those in School H scoring developing significantly quicker than those in School L.

	Pencils	Pens	Raven	Independent Task	Correlated Task
Grade	0.0543***	0.0514***	0.0716***	0.647	0.801**
	(0.00595)	(0.00869)	(0.00597)	(0.498)	(0.381)
School	0.128***	0.110**	0.0179	1.963	1.137
	(0.0421)	(0.0531)	(0.0377)	(2.695)	(1.799)
Grade*School	-0.0209*	-0.0172	0.0254**	-0.262	-0.112
	(0.0106)	(0.0133)	(0.0101)	(0.644)	(0.426)
Constant	0.659***	0.632***	0.289***	-1.641	-3.661**
	(0.0199)	(0.0303)	(0.0246)	(2.083)	(1.667)
Observations	499	499	499	272	272
R-sqaured	0.139	0.114	0.275		

Table 16: Linear trends: pencils task, pens task, Raven scores, independent and correlated lotteries.

The first three columns report linear regressions with consideration parameters and Raven scores as dependent variables. The last two columns, instead, report multinomial logit regression focusing only on the coefficients of rational choices in the independent and correlated task as dependent variables. The covariates are grades normalized from 0 to 4, a dummy that takes value 0 for School L and 1 for School H, and an interaction term. The regression models are estimated in Stata (Robust clustered standard errors at the experimental session level). *** <0.01, ** <0.05, *<0.1.

D.5 Preference intensity

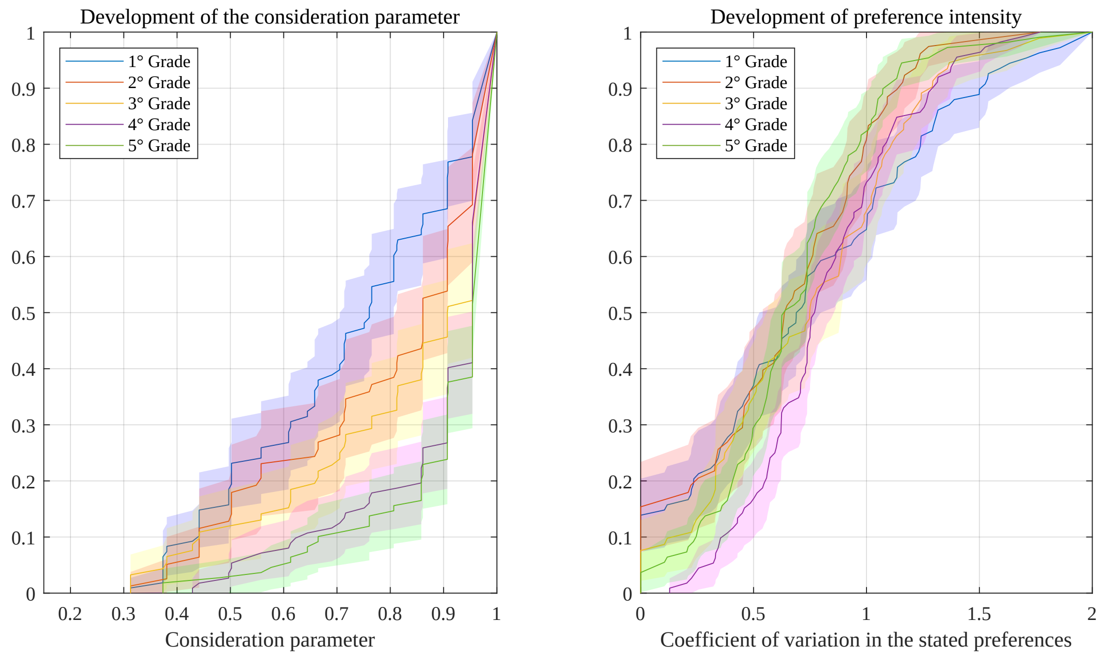

Figure 25: Evolution of the Consideration Parameter and Preference Intensity.
Notes: The left panel shows the cumulative distributions of the consideration parameter averaged between pencils and pens from 1st to 5th grade. The right figure shows the cumulative distributions of preference intensity measured as the coefficient of variation of the stated preferences summed between pencils and pens from 1st to 5th grade. The empirical cumulative distributions are estimated using the Matlab function 'ecdf' with confidence bounds at the default level of 5%.

In the left panel of Figure 25, we plot the cumulative distribution of the estimated individual consideration parameter by grade. The figures show a clear order of stochastic dominance. In the right panel of Figure 25, we plot the cumulative distributions of our measure of preference intensity. Every pupil assigned to each alternative a value $v_i(a) \in \{1, 2, 3, 4, 5\}$ with the possibility to assign the same value to two different alternatives. Our measure of preference intensity is the coefficient of variation of the stated values, namely $cv_i = \frac{\sigma(v)}{\bar{v}}$ where $\sigma(v)$ is the standard deviation and $\bar{v}$ is the mean of v. We then report the sum between the coefficient of variation measured for pencils and pens. We show that there is virtually no difference by age. Both the mean and median cv are unchanged from 1st (mean = 0.75, median = 0.70) to 5th grade (mean = 0.69, median = 0.64). However, the distributions are different with higher variance among younger pupils. The standard deviation drops from 0.54 in the 1st grade to 0.36 in the 5th grade, and the difference is statistically significant, F-test of equal variance, $p < 0.01$. Note that, for instance, older pupils rarely like all goods equally (3.7% in the 5th grade), while this behaviour is widespread among younger pupils (13.9% in the 1st grade).

D.6 Risk Preferences

After answering the choices over the lotteries, but before being paid, the pupils answered a bomb-like task to measure their risk preferences (Andreoni et al., 2020). This task was performed using eight cards (see Figure 26), seven representing sheep and one representing a wolf. The instructions are simple: when cards are turned, if the wolf is one of the cards turned the pupil wins nothing, otherwise the pupil wins many goods equal to the number of cards turned. The pupils reported their choices on a paper. They could choose to turn 0 to 8 cards, i.e. we did not exclude the dominated options 0 and 8. However, only 5.5% of the pupils chose one of these options. In this task, 4 represents risk neutrality, 1-3 represents risk aversion, and 5-7 represents risk-seeking behaviour.[42]

Figure 26: Cards for the Bomb task to measure risk preferences.

Figure 27 shows that females are significantly more risk-averse than males (Wilcoxon rank-sum test, p-value < 0.01). The distributions are also well-behaved with reasonable inter-quantile ranges, and overall, pupils are mildly risk-averse. These results are in line with previous evidence from the literature (Andreoni et al. (2020), Piovesan & Willadsen (2021)).[43] Apart from gender, we do not find any significant correlation between risk preferences and other variables such as parental education, age, cognitive abilities, or school.

[42]We also collected data on the risk preferences of the pupils in Fall 2022. Since in the first wave, the research assistants paid the lottery game before the elicitation of the risk preferences, we opted to re-elicit the risk preferences in Spring 2023.

[43]The correlation between the risk preferences collected in Fall 2022 and Spring 2023, even under the protocol failure in Fall 2022, is +0.14 (p-value < 0.01). The distributions are very similar both in aggregate and by gender (Chi-squared tests and t-tests all have p-values > 0.1), and also in Fall 2022, we find differences between genders (Wilcoxon rank-sum test, p-value < 0.05).

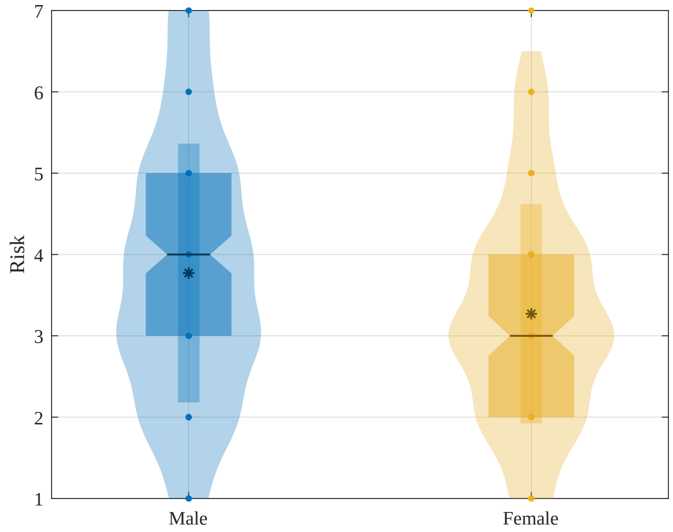

Figure 27: Risk by gender.

D.7 Coin-drop task with undominated lotteries

In Fall 2022, we collected data on a more complex version of the coin-drop task (see Figure 28). The lotteries are now defined on a different outcome space, i.e. $\{0, 1, 2, 3\}$. In the Independent coin drop task, each pupil chooses one lottery out of the following: $\ell_1^I = (3, 3/8; 2, 1/8; 0, 4/8)$, $\ell_2^I = (3, 3/8; 1, 4/8; 0, 1/8)$, and $\ell_3^I = (2, 4/8; 1, 3/8; 0, 1/8)$. Note that these lotteries are not ranked by FOSD. In the Correlated coin drop task, each pupil chooses two of the following: $\ell_3 = (3, 3/8)$, $\ell_2 = (2, 4/8)$, and $\ell_1 = (1, 7/8)$. As in section 4.2.2, we denote $\ell_1^C = \{\ell_3, \ell_2\}$, $\ell_2^C = \{\ell_3, \ell_1\}$, and $\ell_3^C = \{\ell_2, \ell_1\}$. Similarly to the risky choices described in section 2, combining every two correlated lotteries yields exactly the independent ones.

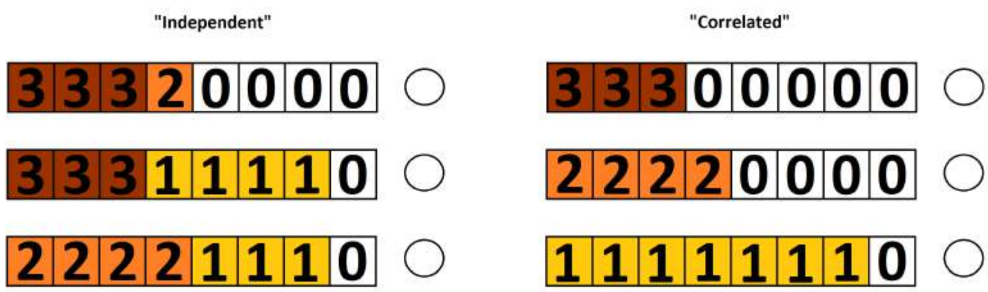

Figure 28: The answer sheet in the two tasks.

We randomized the pupils into the two tasks (between-subject design). Figure 29 shows clear signs of correlation neglect. In the Independent task, the majority of pupils choose ℓ_2^I, while in the Correlated task, the three lotteries have similar proportions of choices. The two distributions are significantly different (chi-square test, p-value < 0.01).

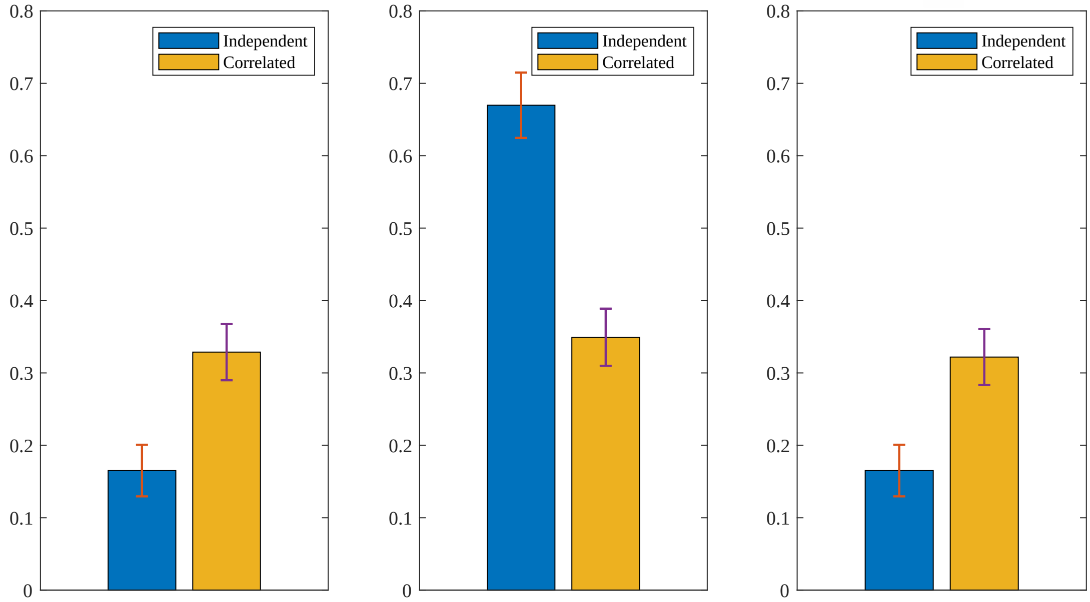

Figure 29: Distribution of choices among the three lotteries in the independent and correlated tasks.
Notes: The left panel shows the proportion of pupils choosing ℓ_1^I, ℓ_1^C, the middle panel those choosing ℓ_2^I, ℓ_2^C, and the right panel those choosing ℓ_3^I, ℓ_3^C.

We now investigate whether we can rationalize the choices of the pupils using the risk preferences measured by the Bomb card game six months later.[44] This exercise is a validation of the coin-drop game as an experimental design for choices between lotteries. In the first panel of Figure 30, we show that the pupils who choose ℓ_1^I are significantly more risk-loving than the remaining pupils (Wilcoxon rank-sum test, p-values < 0.05, w.r.t to pupils who choose ℓ_2^I and ℓ_3^I resp.) indicating that the coin-drop game is indeed an effective tool and the pupils had understood the task. In the second panel, we focus on the correlated lotteries task and find that the correlation with risk preferences is attenuated. However, the pupils who choose ℓ_1^C are still more risk-loving than the remaining pupils (Wilcoxon rank-sum test, p-value > 0.1 and p-value < 0.05, w.r.t to pupils who choose ℓ_2^C and ℓ_3^C resp.).

Finally, we confirm the differences in risk preferences between genders of Figure 27 in the Independent task with 33.5% of males choosing ℓ_1^I while only 26.4% of females doing so. This difference is not due to lower comprehension among female participants as we find no significant differences in the correct answers to a comprehension question both in Fall 2022 (63% vs 64% for males and females resp.) and Spring 2023 (60% vs 56%). Even focusing only on pupils who answered the comprehension question correctly, 30% of males chose ℓ_1^I while only 22% of

[44]The results are robust if we use the risk preferences measured in the first wave, however, we think it is cleaner to use the risk preferences measured in the second wave.

females did so.

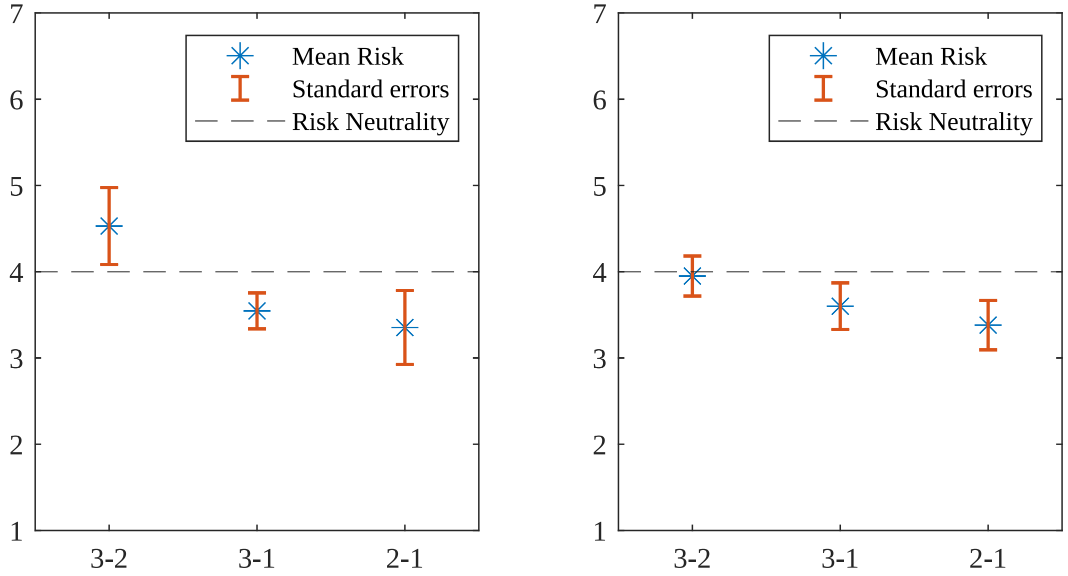

Figure 30: Number of cards chosen broken down by the choices in the coin-drop task.

E Replication using the entire set of schools

We replicate Figures 6, 7, 22, 9, 10, 11, 23, 24 and Tables 16, 13, 14, 12 including all schools. We aggregate the remaining two schools as **Schools M** to achieve a critical mass of children.

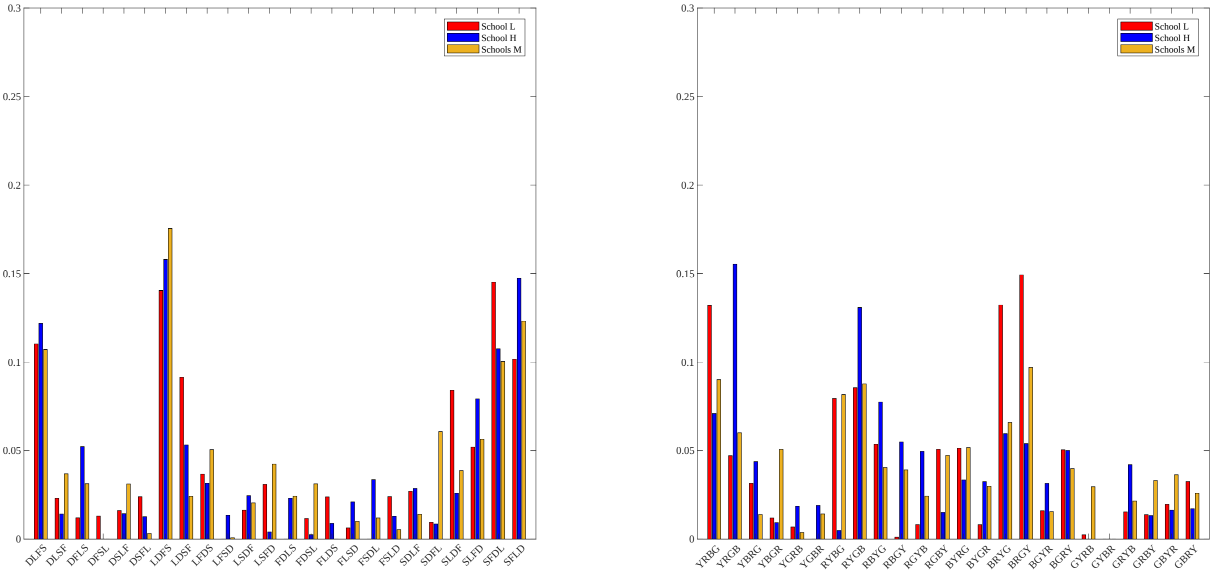

Figure 31: Distribution of preference types by schools (pencils and pens).

Notes: The estimates regarding the preference types between pencils and pens are from left to right. Chi-square tests show no differences across schools (pencils, comparisons School L-H, L-M, and H,M yield p-values > 0.1; pens, comparisons School L-H yields p-value = 0.085 while both comparisons L-M and H-M yield p-values > 0.1).

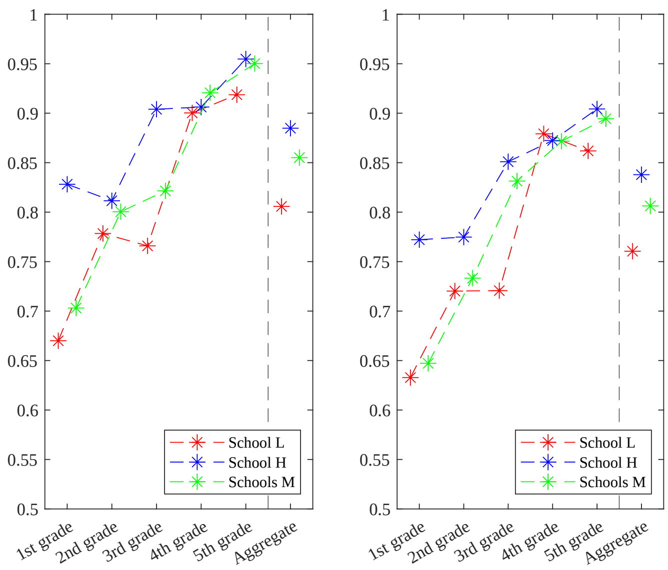

Figure 32: Consideration parameter by schools (pencils and pens).

Notes: Estimates of γ in sub-populations characterized by grades and schools. The estimates regarding the choices between pencils and pens are from left to right. The γ^* are estimated using MATLAB function *fmincon*.

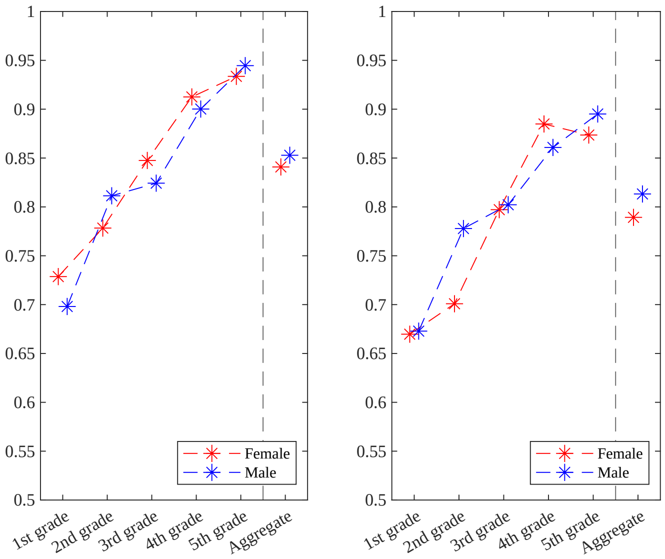

Figure 33: Consideration parameter by gender using all schools

Notes: Estimates of γ in sub-populations characterized by grades and genders. The estimates regarding the choices between pencils and pens are from left to right. The γ^* are estimated using MATLAB function *fmincon*.

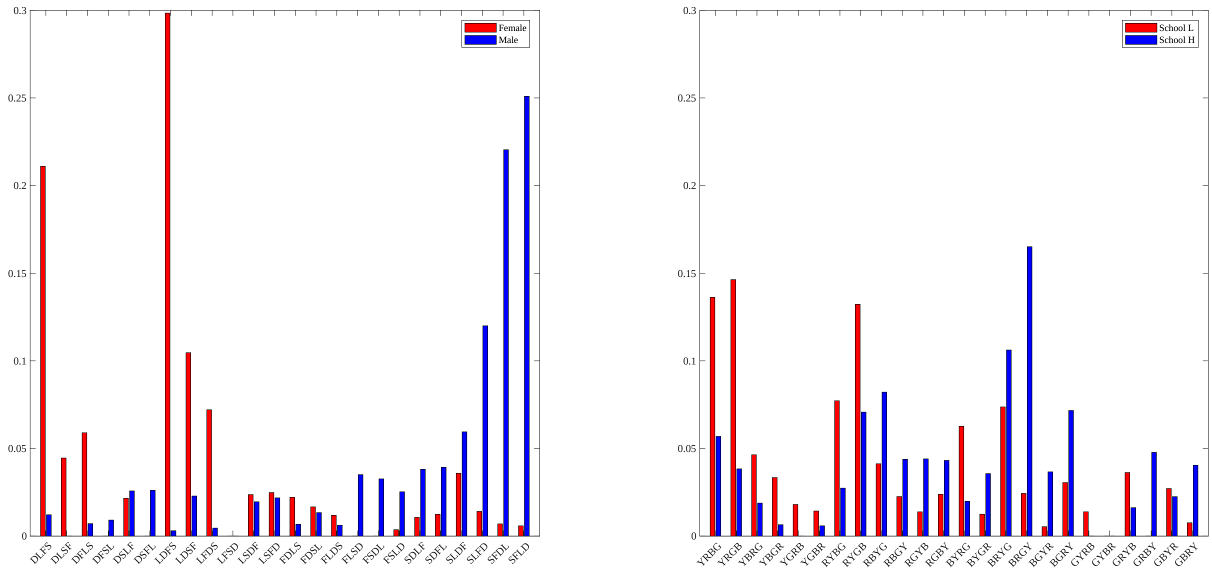

Figure 34: Distribution of preference types by gender (pencils and pens)

Notes: The estimates regarding the preference types between pencils and pens are from left to right. Chi-square tests show significant differences (pencils, p-value < 0.01; pens, p-value = 0.03).

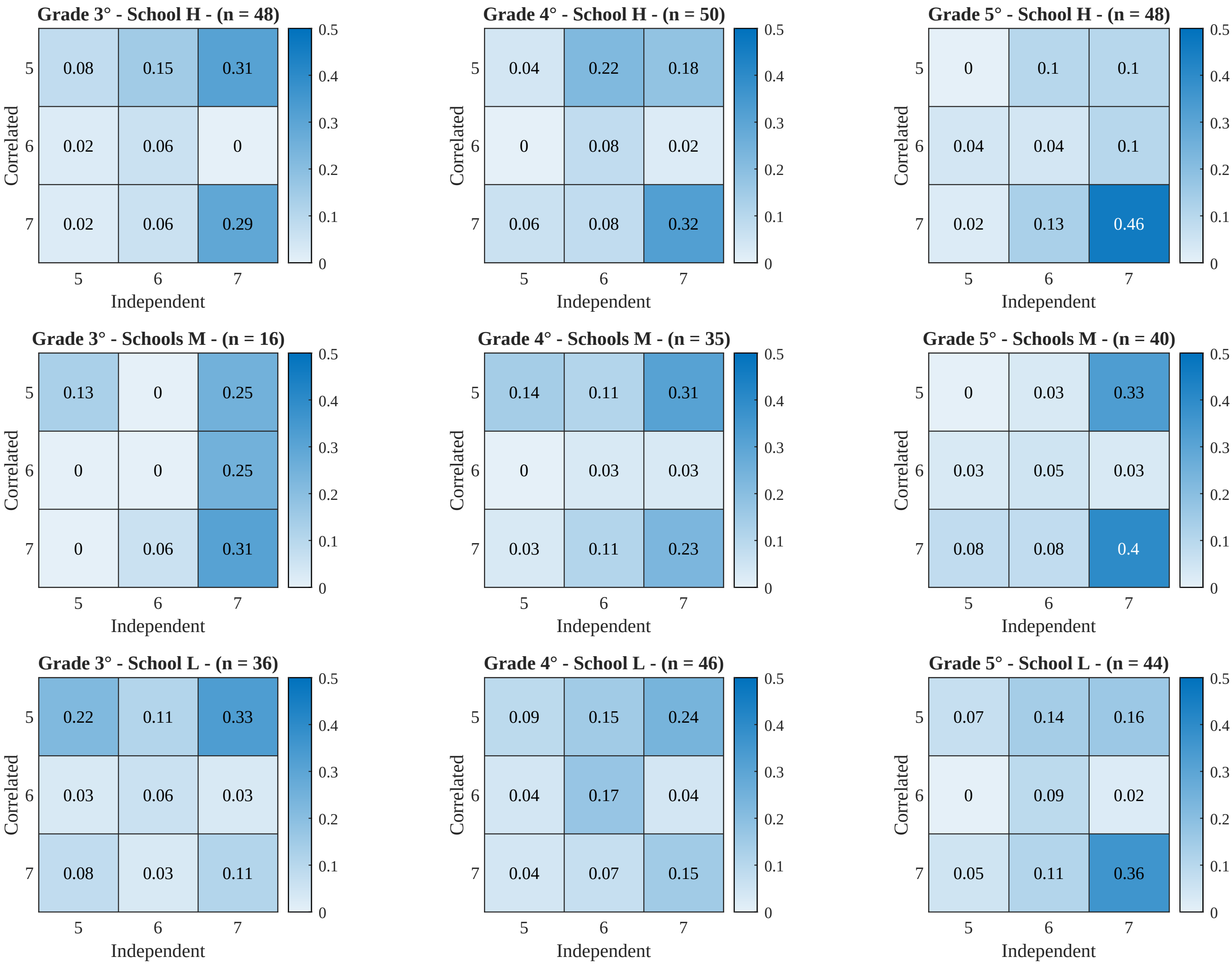

Figure 35: Joint distribution of the 'risky choices' by school and grade

Notes: This figure reports the joint distribution of choices in the Correlated and Independent Tasks. In the Independent task, the pupil should choose 7 regardless of correlation neglect. In the Correlated task, the pupil chooses 5 with correlation neglect and 7 without. The top row regards **School H**, grades 3rd to 5th from left to right. The middle row regards **Schools M**, grades 3rd to 5th from left to right. The bottom row regards **School L**, grades 3rd to 5th from left to right.

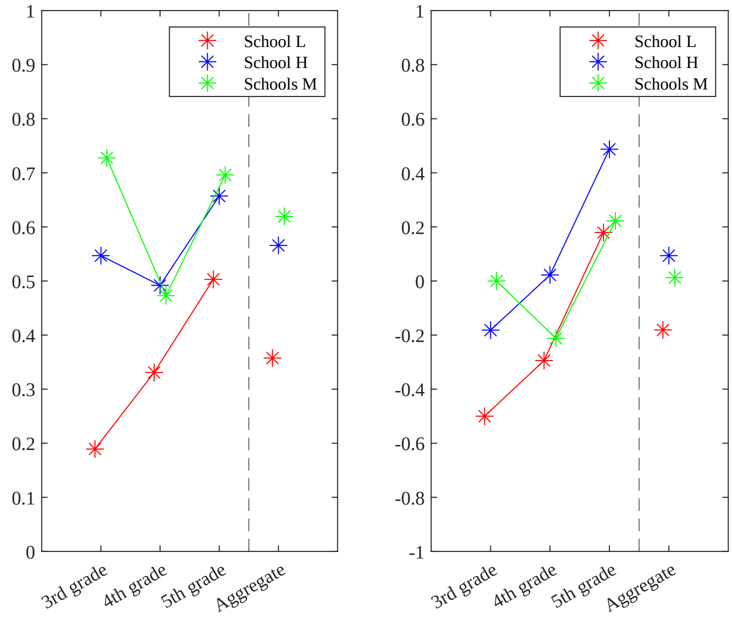

Figure 36: Development of the consideration parameter in choice with risk.

Notes: On the left, we report the estimates of the unique consideration parameter by grades and in the aggregate for **School L**, **Schools M**, and **School H**. On the right, we report the index of correlation neglect.

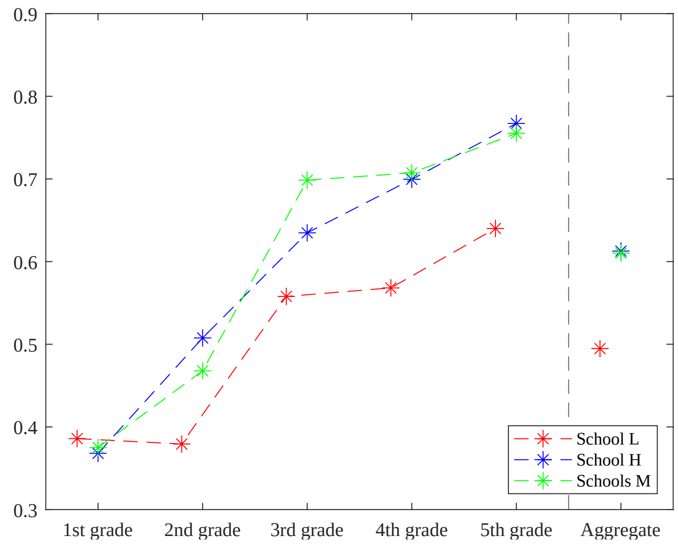

Figure 37: Normalised Raven's scores by grades and schools.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	0.211***			0.0839*	0.199***			0.0765*
	(0.0370)			(0.0419)	(0.0373)			(0.0429)
Second Grade		0.0803**		0.0959***		0.0692**		0.0921***
		(0.0306)		(0.0315)		(0.0308)		(0.0323)
Third Grade		0.103***		0.0838**		0.0969***		0.0831***
		(0.0288)		(0.0323)		(0.0251)		(0.0310)
Fourth Grade		0.158***		0.132***		0.152***		0.132***
		(0.0215)		(0.0271)		(0.0213)		(0.0273)
Fifth Grade		0.200***		0.172***		0.194***		0.172***
		(0.0190)		(0.0252)		(0.0188)		(0.0258)
Parental education			0.0264***	0.0231***			0.0200**	0.0201***
			(0.00835)	(0.00704)			(0.00864)	(0.00718)
Preference Intensity	0.186***	0.184***	0.183***	0.183***	0.182***	0.177***	0.176***	0.179***
	(0.0326)	(0.0300)	(0.0410)	(0.0327)	(0.0335)	(0.0306)	(0.0411)	(0.0331)
Gender	-0.00881	0.00218	-0.00343	0.0115	-0.00962	0.000711	-0.00461	0.0107
	(0.0136)	(0.0132)	(0.0156)	(0.0138)	(0.0135)	(0.0130)	(0.0155)	(0.0137)
School Fixed Effect					0.0308	0.0411***	0.0406	0.0207
					(0.0210)	(0.0139)	(0.0270)	(0.0170)
Constant	0.668***	0.671***	0.702***	0.558***	0.658***	0.654***	0.700***	0.561***
	(0.0337)	(0.0227)	(0.0377)	(0.0357)	(0.0342)	(0.0223)	(0.0390)	(0.0371)
Observations	676	676	550	550	676	676	550	550
R-squared	0.121	0.177	0.075	0.203	0.127	0.187	0.083	0.205

Table 17: Consideration parameter, age, parental education, and cognitive abilities (pencils task).

Notes: The dependent variable is the estimated consideration parameter in the pencils task. The covariates are coded such that the constant represents the consideration parameter of a pupil who is male, has both parents with the lowest educational attainment, has scored the lowest in the Raven test, is attending the 1st grade in **School L**, and is indifferent between all alternatives. The coefficient of parental education is the effect of the improvement of the degree of at least one of the parents. The coefficient of Raven's scores is the difference between the highest and lowest-scoring pupils. Grade can be interpreted as a one-year age effect. Preference intensity takes values between 0 and 2. Gender is 0 for males and 1 for females. School is 0 for **School L** and 1 for either **Schools M** or **School H**. The regression models are estimated in Stata (Robust clustered standard errors at the session level). *** <0.01, ** <0.05, *<0.1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	0.192***			0.0523	0.181***			0.0426
	(0.0372)			(0.0446)	(0.0386)			(0.0454)
Second Grade		0.0625*		0.0795**		0.0530		0.0747**
		(0.0344)		(0.0312)		(0.0338)		(0.0314)
Third Grade		0.105***		0.0877**		0.0998***		0.0869**
		(0.0323)		(0.0355)		(0.0295)		(0.0338)
Fourth Grade		0.170***		0.152***		0.164***		0.152***
		(0.0254)		(0.0287)		(0.0260)		(0.0284)
Fifth Grade		0.187***		0.173***		0.181***		0.172***
		(0.0241)		(0.0277)		(0.0234)		(0.0274)
Parental education			0.0213**	0.0180**			0.0147	0.0140*
			(0.00940)	(0.00749)			(0.00929)	(0.00813)
Preference Intensity	0.173***	0.176***	0.176***	0.199***	0.168***	0.169***	0.171***	0.195***
	(0.0419)	(0.0372)	(0.0469)	(0.0392)	(0.0426)	(0.0376)	(0.0471)	(0.0398)
Gender	-0.0209	-0.0111	-0.0196	-0.00650	-0.0216	-0.0124	-0.0209	-0.00767
	(0.0138)	(0.0136)	(0.0168)	(0.0147)	(0.0139)	(0.0136)	(0.0167)	(0.0147)
School Fixed Effect					0.0259	0.0353**	0.0418	0.0267
					(0.0242)	(0.0165)	(0.0277)	(0.0188)
Constant	0.661***	0.655***	0.702***	0.567***	0.653***	0.641***	0.700***	0.572***
	(0.0318)	(0.0264)	(0.0448)	(0.0409)	(0.0322)	(0.0262)	(0.0454)	(0.0419)
Observations	676	676	550	550	676	676	550	550
R-squared	0.083	0.142	0.052	0.159	0.086	0.149	0.060	0.162

Table 18: Consideration parameter, age, parental education, and cognitive abilities (pens task). *Notes*: The dependent variable is the estimated consideration parameter in the pens task. The covariates are coded such that the constant represents the consideration parameter of a pupil who is male, has both parents with the lowest educational attainment, has scored the lowest in the Raven test, is attending the 1st grade in **School L**, and is indifferent between all alternatives. The coefficient of parental education is the effect of the improvement of the degree of at least one of the parents. The coefficient of Raven's scores is the difference between the highest and lowest-scoring pupils. Grade can be interpreted as a one-year age-effect. Preference intensity takes values between 0 and 2. Gender is 0 for males and 1 for females. School is 0 for **School L** and 1 for either **Schools M** or **School H**. The regression models are estimated in Stata (Robust clustered standard errors at the session level). *** <0.01, ** <0.05, *<0.1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	1.388**			0.201	0.990			-0.0850
	(0.655)			(0.836)	(0.631)			(0.885)
Fourth Grade		0.168		-0.0106		0.160		0.0141
		(0.517)		(0.460)		(0.491)		(0.469)
Fifth Grade		0.883*		0.852*		0.867*		0.860*
		(0.526)		(0.479)		(0.488)		(0.483)
Parental education			0.253	0.292			0.146	0.196
			(0.167)	(0.179)			(0.189)	(0.194)
Gender	-0.182	-0.0930	-0.259	-0.187	-0.251	-0.171	-0.312	-0.226
	(0.302)	(0.287)	(0.306)	(0.307)	(0.322)	(0.299)	(0.328)	(0.325)
School Fixed Effect					0.776*	0.873**	0.698	0.651
					(0.410)	(0.413)	(0.447)	(0.438)
Constant	0.583	1.126***	0.730	0.178	0.404	0.634	0.656	0.270
	(0.499)	(0.415)	(0.601)	(0.866)	(0.539)	(0.526)	(0.584)	(0.845)
Observations	363	363	293	293	363	363	293	293
Log-likelihood	-340.52	-337.78	-271.64	-266.38	-337.35	-333.35	-268.99	-264.29
AIC	693.03	691.55	555.28	556.76	690.70	686.70	553.98	556.57
BIC	716.40	722.71	577.36	600.93	721.86	725.64	583.42	608.10

Table 19: Independent task, age, parental education, and cognitive abilities.

We estimate a binomial Logit in which the dependent variable is a dummy that takes value 1 if the pupils chose ℓ_7^I and 0 otherwise. The independent variables are the same as in Table 17 apart from preference intensity. The regression models are estimated in Stata using the function *logit*. *** <0.01, ** <0.05, *<0.1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	1.560***			1.422***	1.394***			1.284**
	(0.526)			(0.501)	(0.520)			(0.522)
Fourth Grade		0.344		0.445		0.344		0.457
		(0.347)		(0.334)		(0.324)		(0.326)
Fifth Grade		1.139***		1.026***		1.141***		1.033***
		(0.355)		(0.360)		(0.337)		(0.356)
Parental education			0.0955	0.118			0.0288	0.0753
			(0.106)	(0.112)			(0.105)	(0.109)
Gender	-0.493**	-0.407*	-0.485*	-0.439*	-0.524**	-0.462**	-0.528**	-0.464*
	(0.239)	(0.238)	(0.263)	(0.264)	(0.223)	(0.223)	(0.252)	(0.254)
Succ. Ind. Task	0.601**	0.610***	0.652***	0.605**	0.568**	0.554**	0.598**	0.572**
	(0.236)	(0.232)	(0.245)	(0.264)	(0.249)	(0.240)	(0.261)	(0.273)
School Fixed Effect					0.329	0.479	0.478	0.314
					(0.358)	(0.310)	(0.369)	(0.322)
Constant	-1.246***	-0.761**	-0.539	-2.108***	-1.317***	-1.022**	-0.583	-2.055***
	(0.423)	(0.330)	(0.404)	(0.563)	(0.468)	(0.430)	(0.421)	(0.546)
Observations	363	363	293	293	363	363	293	293
Log-likelihood	-342.99	-339.16	-276.75	-267.09	-342.05	-337.14	-275.27	-266.20
AIC	701.97	698.33	569.49	562.18	704.11	698.27	570.53	564.40
BIC	733.13	737.27	598.94	613.70	743.05	745.01	607.34	623.28

Table 20: Correlation neglect, age, parental education, and cognitive abilities.

Notes: We estimate a multinomial Logit in which the dependent variable represents the pupil's choices in the Correlated task. The baseline is represented by the pupils who choose ℓ_5^C and we only report the coefficients related to equation 22. The independent variables are the same as in Table 17 apart from preference intensity and Success Ind. Task. This latter is a dummy that takes the value 1 if the pupil chooses the best lottery in the Independent task. The regression models are estimated in Stata using the function *mlogit*. *** <0.01, ** <0.05, *<0.1.

F Robustness check, relaxing the uniqueness of γ.

In this section, we explore a different estimation approach. We discretise the type space Ω by letting γ take two, three, and four possible values. Hence with 24 possible orderings and three values of the consideration parameter, the type space has cardinality $|\Omega| = \{48, 72, 96\}$. Again, we estimate the type distribution $\pi(\omega)$ and the type conditional choice probabilities p^{ω}, with $\omega = (\succ, \gamma)$ being a pair of preferences and consideration parameters. The estimation approach is equivalent to the one described in section **??** but now the outcome of the maximisation is a vector of consideration parameters $\hat{\gamma}$ and the estimated population distribution $\hat{\pi}(\omega)$.

In Figure 38, we replicate Figure 7 for all model specifications. We find that the patterns are remarkably similar with only a change in scaling with a slightly higher estimate in the case of unique γ. In Tables 21 and 22, we report the estimated mean value $E(\hat{\gamma})$, the log-likelihood, and the Akaike, Bayesian Information Criteria (AIC and BIC) for our 12 sub-populations: five grades in **School L** and **School H**, and the two aggregate estimations for each school.

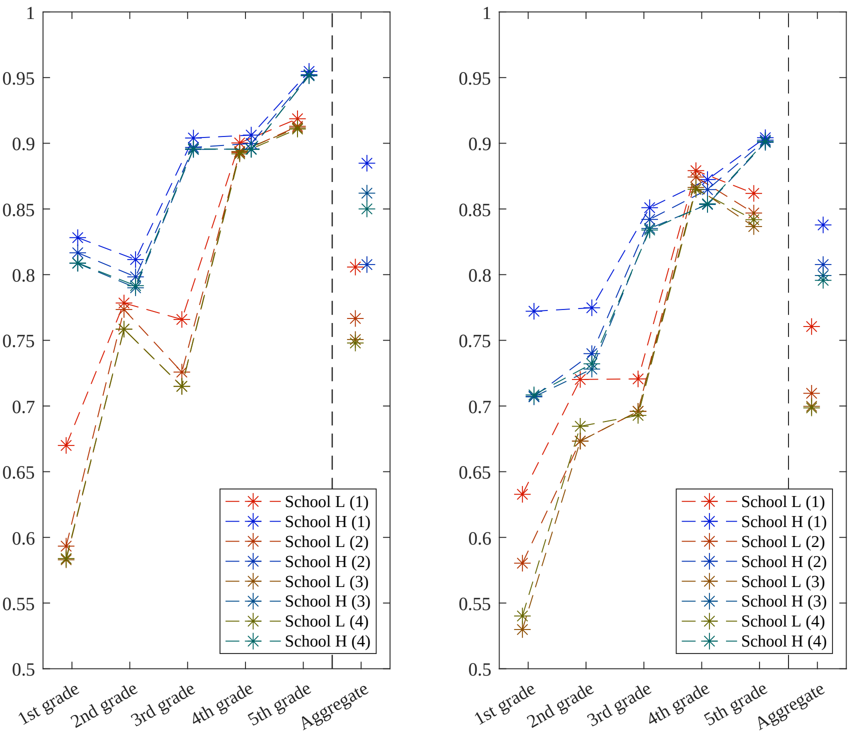

Figure 38: Consideration parameter γ by school and grade for all model specifications

Notes: Estimates of $E(\hat{\gamma})$ in sub-populations characterised by grades and schools for all model-specifications. From left to right, the estimates regard the choices between (i) pencils and (ii) pens. The γ^* are estimated using MATLAB function *fmincon*. In the legend, '(j)' identifies the model specification with j values of γ.

	One value of γ				Two values of γ				
	$E(\gamma)$	$\log \mathcal{L}$	AIC	BIC	$E(\gamma)$	$\log \mathcal{L}$	AIC	BIC	
School L - 1st grade	0.67	-616.91	1283.81	1340.73	0.59	-577.81	1255.62	1369.46	**School L - 1st grad**
School L - 2nd grade	0.78	-221.51	493.02	527.21	0.77	-200.06	500.12	568.48	**School L - 2nd grac**
School L - 3rd grade	0.77	-325.36	700.73	743.57	0.73	-293.41	686.83	772.51	**School L - 3rd grad**
School L - 4th grade	0.90	-318.70	687.40	737.58	0.89	-292.76	685.53	785.89	**School L - 4th grad**
School L - 5th grade	0.92	-294.14	638.27	689.78	0.91	-258.08	616.16	719.18	**School L - 5th grad**
School L - aggregate	0.81	-1882.03	3814.06	3902.60	0.77	-1711.11	3522.21	3699.28	**School L - aggregat**
School H - 1st grade	0.83	-244.54	539.07	578.66	0.82	-218.45	536.89	616.07	**School H - 1st grad**
School H - 2nd grade	0.81	-350.17	750.33	797.63	0.80	-324.61	749.23	843.82	**School H - 2nd grac**
School H - 3rd grade	0.90	-281.58	613.15	661.45	0.90	-257.07	614.14	710.73	**School H - 3rd grac**
School H - 4th grade	0.91	-326.01	702.01	753.09	0.90	-296.43	692.85	795.01	**School H - 4th grac**
School H - 5th grade	0.95	-226.48	502.96	551.26	0.95	-207.08	514.16	610.75	**School H - 5th grac**
School H - aggregate	0.88	-1497.64	3045.28	3132.71	0.81	-1392.36	2884.72	3059.58	**School H - aggrega**

Table 21: Model selection for different specifications of the model in the pencils questions.

	One value of γ				Two values of γ				
	$E(\gamma)$	$\log\mathcal{L}$	AIC	BIC	$E(\gamma)$	$\log\mathcal{L}$	AIC	BIC	
School L - 1st grade	0.63	-644.52	1339.04	1395.96	0.58	-616.86	1333.71	1447.55	**School L - 1st grad**
School L - 2nd grade	0.72	-236.46	522.92	557.10	0.67	-223.84	547.69	616.05	**School L - 2nd grac**
School L - 3rd grade	0.72	-334.13	718.26	761.10	0.70	-310.61	721.23	806.91	**School L - 3rd grad**
School L - 4th grade	0.88	-344.25	738.49	788.68	0.87	-321.09	742.17	842.54	**School L - 4th grad**
School L - 5th grade	0.86	-361.62	773.25	824.76	0.85	-334.96	769.93	872.95	**School L - 5th grad**
School L - aggregate	0.76	-2022.58	4095.16	4183.69	0.71	-1905.44	3910.88	4087.94	**School L - aggregat**
School H - 1st grade	0.77	-280.79	611.58	651.16	0.71	-255.87	611.73	690.91	**School H - 1st grad**
School H - 2nd grade	0.77	-376.86	803.72	851.01	0.74	-347.94	795.88	890.47	**School H - 2nd grac**
School H - 3rd grade	0.85	-335.88	721.77	770.06	0.84	-298.36	696.72	793.32	**School H - 3rd grac**
School H - 4th grade	0.87	-356.49	762.97	814.05	0.86	-327.45	754.91	857.06	**School H - 4th grac**
School H - 5th grade	0.90	-293.21	636.41	684.71	0.90	-273.19	646.39	742.98	**School H - 5th grac**
School H - aggregate	0.84	-1719.38	3488.75	3576.18	0.81	-1593.35	3286.70	3461.56	**School H - aggrega**

Table 22: Model selection for different specifications of the model in the pens questions.

G Robustness check, approximate Bayesian approach

In section **??**, we have estimated individual consideration parameters individually. This approach may have drawbacks related to the limited sample employed and the excessive variability within the individual maximum likelihood estimator. Here, we propose a robustness check based on an empirical Bayesian approach known as 'approximate' Bayesian (Kass & Steffey, 1989).

We first estimate the population in 10 sub-populations defined by schools and grades using three values of the consideration parameter to allow variation at the individual level. Tables 21 and 22 show that three values, denote them $\{L, M, H\}$ seem to provide a fairly substantial increase in the log-likelihood while a fourth value does not add much to it. Using the resulting estimate $\hat{\pi}(\omega)$ as the prior distribution we then use Bayes' Theorem to obtain the posterior distribution for each pupil $\hat{\pi}_i(\omega)$ as shown in equation 23.

$$\hat{\pi}_i(\omega) = \frac{p_i^{\omega}(C_i)\hat{\pi}(\omega)}{\sum\limits_{\omega} p_i^{\omega}(C_i)\hat{\pi}(\omega)} \tag{23}$$

Then as point-estimate of the individual consideration parameter, $\hat{\gamma}_i$, we take its expectation as shown in equation 24.

$$\hat{\gamma}_i = \sum\nolimits_{j\in\{L,M,H\}} \left(\sum_{\succ} \hat{\gamma}_j \hat{\pi}_i\left(\omega\right) \right) \tag{24}$$

We replicate Tables 16 and 12 with very similar results.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	0.214***			0.0188	0.182***			0.00364
	(0.0575)			(0.0574)	(0.0542)			(0.0582)
Second Grade		0.172***		0.177***		0.148***		0.164***
		(0.0414)		(0.0455)		(0.0422)		(0.0470)
Third Grade		0.192***		0.182***		0.175***		0.175***
		(0.0348)		(0.0397)		(0.0229)		(0.0331)
Fourth Grade		0.309***		0.307***		0.297***		0.305***
		(0.0253)		(0.0316)		(0.0239)		(0.0287)
Fifth Grade		0.225***		0.242***		0.214***		0.239***
		(0.0391)		(0.0374)		(0.0279)		(0.0339)
Parental education			0.0212	0.0179**			0.00538	0.00707
			(0.0127)	(0.00797)			(0.0114)	(0.00931)
Preference Intensity	0.168***	0.152***	0.183***	0.153***	0.150***	0.131***	0.161***	0.136***
	(0.0500)	(0.0401)	(0.0580)	(0.0472)	(0.0511)	(0.0398)	(0.0571)	(0.0457)
Gender	0.0102	0.0313*	0.0111	0.0409**	0.00850	0.0286*	0.00846	0.0376*
	(0.0178)	(0.0171)	(0.0209)	(0.0198)	(0.0176)	(0.0167)	(0.0203)	(0.0191)
School Fixed Effect					0.0762**	0.0744***	0.0761*	0.0557**
					(0.0350)	(0.0209)	(0.0416)	(0.0249)
Constant	0.500***	0.434***	0.546***	0.359***	0.488***	0.419***	0.568***	0.387***
	(0.0506)	(0.0299)	(0.0662)	(0.0485)	(0.0514)	(0.0247)	(0.0607)	(0.0470)
Observations	499	499	404	404	499	499	404	404
R-squared	0.081	0.216	0.043	0.237	0.104	0.238	0.061	0.246

Table 23: Consideration, age, parental education, and cognitive abilities (pencils task - approx. Bayes).

Notes: The dependent variable is the consideration parameter estimated as in equation 24. The covariates are coded such that the constant represents the consideration parameter of a pupil who is male, has both parents with the lowest educational attainment, has scored the lowest in the Raven test, is attending the 1st grade in **School L**, and is indifferent between all alternatives. Parental education is coded from 1 to 6 following the categories in Figure 4. The coefficient of Raven's scores is the difference between the highest and lowest-scoring pupils. Since the pupils answered 8 matrices, the coefficient can be divided by eight to obtain the effect of each correct answer. Grade can be interpreted as a one-year age effect. Specifications that contain parental education have a smaller sample size due to frictions in our survey as reported in Table 1. The regression models are estimated in Stata (Robust clustered standard errors at the session level). *** <0.01, ** <0.05, *<0.1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	0.263***			0.0741	0.238***			0.0588
	(0.0512)			(0.0569)	(0.0493)			(0.0583)
Second Grade		0.125***		0.125**		0.103**		0.114**
		(0.0453)		(0.0481)		(0.0419)		(0.0476)
Third Grade		0.182***		0.169***		0.166***		0.162***
		(0.0453)		(0.0475)		(0.0367)		(0.0428)
Fourth Grade		0.266***		0.248***		0.253***		0.245***
		(0.0359)		(0.0412)		(0.0344)		(0.0393)
Fifth Grade		0.284***		0.263***		0.273***		0.261***
		(0.0352)		(0.0401)		(0.0296)		(0.0383)
Parental education			0.0258*	0.0206**			0.0101	0.0106
			(0.0138)	(0.01000)			(0.0122)	(0.01000)
Preference Intensity	0.249***	0.256***	0.256***	0.273***	0.234***	0.237***	0.242***	0.263***
	(0.0605)	(0.0491)	(0.0676)	(0.0498)	(0.0619)	(0.0500)	(0.0679)	(0.0505)
Gender	-0.0423**	-0.0204	-0.0484**	-0.0165	-0.0438**	-0.0232	-0.0512**	-0.0197
	(0.0191)	(0.0177)	(0.0236)	(0.0197)	(0.0190)	(0.0173)	(0.0232)	(0.0194)
School Fixed Effect					0.0598*	0.0690***	0.0746*	0.0514**
					(0.0349)	(0.0219)	(0.0404)	(0.0248)
Constant	0.553***	0.512***	0.616***	0.406***	0.545***	0.498***	0.635***	0.430***
	(0.0447)	(0.0366)	(0.0705)	(0.0501)	(0.0450)	(0.0304)	(0.0665)	(0.0512)
Observations	499	499	404	404	499	499	404	404
R-squared	0.131	0.237	0.082	0.262	0.144	0.254	0.099	0.269

Table 24: Consideration, age, parental education, and cognitive abilities (pens task - approx. Bayes).

Notes: The dependent variable is the consideration parameter estimated as in equation 24. The covariates are coded such that the constant represents the consideration parameter of a pupil who is male, has both parents with the lowest educational attainment, has scored the lowest in the Raven test, is attending the 1st grade in **School L**, and is indifferent between all alternatives. Parental education is coded from 1 to 6 following the categories in Figure 4. The coefficient of Raven's scores is the difference between the highest and lowest-scoring pupils. Since the pupils answered 8 matrices, the coefficient can be divided by eight to obtain the effect of each correct answer. Grade can be interpreted as a one-year age effect. Specifications that contain parental education have a smaller sample size due to frictions in our survey as reported in Table 1. The regression models are estimated in Stata (Robust clustered standard errors at the session level). *** <0.01, ** <0.05, *<0.1.

H Robustness check, beta-regression on the consideration parameter.

Given that the domain of $\hat{\gamma}_i$ is the open interval (0,1), we replicate our results based on OLS regressions using a beta-regression with a logistic transformation and find very similar correlations.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	1.025***			0.561**	0.968***			0.522**
	(0.197)			(0.260)	(0.201)			(0.263)
Second Grade		0.238		0.242		0.179		0.216
		(0.162)		(0.177)		(0.164)		(0.179)
Third Grade		0.421***		0.290		0.370**		0.278
		(0.154)		(0.184)		(0.155)		(0.184)
Fourth Grade		0.614***		0.461***		0.591***		0.464***
		(0.147)		(0.177)		(0.147)		(0.176)
Fifth Grade		0.822***		0.597***		0.804***		0.599***
		(0.148)		(0.192)		(0.148)		(0.192)
Parental education			0.0934**	0.0819*			0.0549	0.0552
			(0.0468)	(0.0480)			(0.0542)	(0.0549)
Preference Intensity	0.516**	0.598***	0.618***	0.556**	0.487**	0.537***	0.576**	0.523**
	(0.205)	(0.205)	(0.233)	(0.231)	(0.205)	(0.205)	(0.235)	(0.233)
Gender	-0.0890	-0.0174	-0.0356	0.00562	-0.0934	-0.0297	-0.0454	-0.00233
	(0.0966)	(0.0973)	(0.109)	(0.110)	(0.0966)	(0.0973)	(0.109)	(0.110)
School Fixed Effect					0.144	0.243**	0.177	0.130
					(0.0990)	(0.0985)	(0.126)	(0.129)
Constant	1.122***	1.195***	1.308***	0.738***	1.098***	1.140***	1.367***	0.805***
	(0.150)	(0.149)	(0.201)	(0.231)	(0.150)	(0.149)	(0.205)	(0.241)
Observations	499	499	404	404	499	499	404	404

Table 25: Consideration, age, parental education, and cognitive abilities (pencils task - Beta-regression).

Notes: The dependent variable is the estimated consideration parameter in the pencils task. The covariates are coded such that the constant represents the consideration parameter of a pupil who is male, has both parents with the lowest educational attainment, has scored the lowest in the Raven test, is attending the 1st grade in **School L**, and is indifferent between all alternatives. The controls are gender and preference intensity. Parental education is coded from 1 to 6 following the categories in Figure 4. The coefficient of Raven's scores is the difference between the highest and lowest-scoring pupils. Since the pupils answered 8 matrices, the coefficient can be divided by eight to obtain the effect of each correct answer. Grade can be interpreted as a one-year age effect. Specifications that contain parental education have a smaller sample size due to frictions in our survey as reported in Table 1. The regression models are estimated in Stata using the command *betareg* with a logistic transformation. *** <0.01, ** <0.05, *<0.1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	0.755***			0.200	0.674***			0.0972
	(0.192)			(0.251)	(0.197)			(0.256)
Second Grade		0.249		0.322*		0.170		0.267
		(0.163)		(0.178)		(0.165)		(0.180)
Third Grade		0.469***		0.410**		0.402**		0.384**
		(0.156)		(0.186)		(0.157)		(0.185)
Fourth Grade		0.677***		0.653***		0.652***		0.667***
		(0.148)		(0.178)		(0.148)		(0.177)
Fifth Grade		0.792***		0.788***		0.765***		0.803***
		(0.151)		(0.196)		(0.151)		(0.195)
Parental education			0.0804*	0.0839*			0.0213	0.0298
			(0.0472)	(0.0484)			(0.0549)	(0.0554)
Preference Intensity	0.762***	0.826***	0.787***	0.843***	0.700***	0.730***	0.735***	0.787***
	(0.220)	(0.219)	(0.249)	(0.243)	(0.222)	(0.219)	(0.249)	(0.243)
Gender	-0.105	-0.0464	-0.117	-0.0164	-0.109	-0.0538	-0.129	-0.0323
	(0.0979)	(0.0987)	(0.110)	(0.111)	(0.0978)	(0.0985)	(0.110)	(0.111)
School Fixed Effect					0.193*	0.273***	0.270**	0.267**
					(0.101)	(0.100)	(0.128)	(0.132)
Constant	0.976***	0.908***	1.124***	0.511**	0.950***	0.851***	1.205***	0.648***
	(0.150)	(0.145)	(0.193)	(0.229)	(0.150)	(0.145)	(0.197)	(0.239)
Observations	499	499	404	404	499	499	404	404

Table 26: Consideration, age, parental education, and cognitive abilities (pens task - Beta-regression).

Notes: The dependent variable is the estimated consideration parameter in the pens task. The covariates are coded such that the constant represents the consideration parameter of a pupil who is male, has both parents with the lowest educational attainment, has scored the lowest in the Raven test, is attending the 1st grade in **School L**, and is indifferent between all alternatives. The controls are gender and preference intensity. Parental education is coded from 1 to 6 following the categories in Figure 4. The coefficient of Raven's scores is the difference between the highest and lowest-scoring pupils. Since the pupils answered 8 matrices, the coefficient can be divided by eight to obtain the effect of each correct answer. Grade can be interpreted as a one-year age effect. Specifications that contain parental education have a smaller sample size due to frictions in our survey as reported in Table 1. The regression models are estimated in Stata using the command *betareg* with a logistic transformation. *** <0.01, ** <0.05, *<0.1.

I Robustness check, considering data from May 2022.

As discussed in section 3, we started collecting our data on the pencils/pens task in May 2022 but the collection had to be stopped due to COVID regulations. Consequently, when we returned to the schools 27 pupils repeated the task. Our main specification considers only the data collected in the Autumn of 2022. If anything, we find even stronger correlations, including data from Spring 2022 and excluding the repeated observations from these 27 pupils.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	0.251***			0.127***	0.224***			0.120***
	(0.0329)			(0.0407)	(0.0317)			(0.0426)
Second Grade		0.0923**		0.0669*		0.0574		0.0496
		(0.0349)		(0.0350)		(0.0347)		(0.0348)
Third Grade		0.124***		0.0846**		0.109***		0.0777**
		(0.0313)		(0.0356)		(0.0258)		(0.0334)
Fourth Grade		0.165***		0.120***		0.145***		0.113***
		(0.0273)		(0.0337)		(0.0275)		(0.0341)
Fifth Grade		0.215***		0.162***		0.200***		0.156***
		(0.0242)		(0.0314)		(0.0231)		(0.0323)
Parental education			0.0345***	0.0256***			0.0194*	0.0180**
			(0.0104)	(0.00886)			(0.0101)	(0.00874)
Preference Intensity	0.199***	0.202***	0.221***	0.206***	0.183***	0.185***	0.199***	0.195***
	(0.0380)	(0.0375)	(0.0478)	(0.0386)	(0.0400)	(0.0395)	(0.0483)	(0.0405)
Gender	-0.00882	0.00110	-0.00981	0.00860	-0.0103	-0.00106	-0.0121	0.00664
	(0.0155)	(0.0159)	(0.0177)	(0.0149)	(0.0149)	(0.0151)	(0.0171)	(0.0146)
School Fixed Effect					0.0615**	0.0636***	0.0718**	0.0406*
					(0.0238)	(0.0181)	(0.0288)	(0.0232)
Constant	0.625***	0.646***	0.643***	0.518***	0.618***	0.639***	0.666***	0.539***
	(0.0295)	(0.0248)	(0.0438)	(0.0331)	(0.0282)	(0.0220)	(0.0420)	(0.0368)
Observations	528	528	427	427	528	528	427	427
R-squared	0.150	0.199	0.100	0.235	0.169	0.218	0.121	0.241

Table 27: Consideration, age, parental education, and cognitive abilities (pencils task - including May 2022).

Notes: The dependent variable is the estimated consideration parameter in the pencils task. The covariates are coded such that the constant represents the consideration parameter of a pupil who is male, has both parents with the lowest educational attainment, has scored the lowest in the Raven test, is attending the 1st grade in **School L**, and is indifferent between all alternatives. Parental education is coded from 1 to 6 following the categories in Figure 4. The coefficient of Raven's scores is the difference between the highest and lowest-scoring pupils. Since the pupils answered 8 matrices, the coefficient can be divided by eight to obtain the effect of each correct answer. Grade can be interpreted as a one-year age effect. Specifications that contain parental education have a smaller sample size due to frictions in our survey as reported in Table 1. The regression models are estimated in Stata (Robust clustered standard errors at the session level). *** <0.01, ** <0.05, *<0.1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	0.224***			0.0995**	0.201***			0.0924*
	(0.0373)			(0.0483)	(0.0392)			(0.0506)
Second Grade		0.0772*		0.0651		0.0477		0.0504
		(0.0420)		(0.0395)		(0.0388)		(0.0385)
Third Grade		0.117***		0.0850**		0.104***		0.0789**
		(0.0351)		(0.0407)		(0.0317)		(0.0387)
Fourth Grade		0.175***		0.143***		0.157***		0.136***
		(0.0302)		(0.0360)		(0.0317)		(0.0359)
Fifth Grade		0.195***		0.162***		0.182***		0.157***
		(0.0272)		(0.0349)		(0.0259)		(0.0347)
Parental education			0.0311***	0.0226**			0.0172*	0.0161*
			(0.0106)	(0.00869)			(0.00955)	(0.00852)
Preference Intensity	0.227***	0.232***	0.244***	0.250***	0.212***	0.216***	0.229***	0.242***
	(0.0450)	(0.0412)	(0.0526)	(0.0441)	(0.0471)	(0.0424)	(0.0540)	(0.0455)
Gender	-0.0269	-0.0177	-0.0267	-0.00987	-0.0283*	-0.0196	-0.0288	-0.0115
	(0.0169)	(0.0165)	(0.0202)	(0.0178)	(0.0166)	(0.0161)	(0.0196)	(0.0174)
School Fixed Effect					0.0537*	0.0535***	0.0656**	0.0342
					(0.0269)	(0.0199)	(0.0290)	(0.0243)
Constant	0.617***	0.628***	0.631***	0.509***	0.610***	0.622***	0.649***	0.526***
	(0.0313)	(0.0285)	(0.0538)	(0.0428)	(0.0307)	(0.0274)	(0.0513)	(0.0447)
Observations	528	528	427	427	528	528	427	427
R-squared	0.128	0.178	0.096	0.212	0.141	0.190	0.112	0.216

Table 28: Consideration, age, parental education, and cognitive abilities (pens task - including May 2022).

Notes: The dependent variable is the estimated consideration parameter in the pens task. The covariates are coded such that the constant represents the consideration parameter of a pupil who is male, has both parents with the lowest educational attainment, has scored the lowest in the Raven test, is attending the 1st grade in **School L**, and is indifferent between all alternatives. Parental education is coded from 1 to 6 following the categories in Figure 4. The coefficient of Raven's scores is the difference between the highest and lowest-scoring pupils. Since the pupils answered 8 matrices, the coefficient can be divided by eight to obtain the effect of each correct answer. Grade can be interpreted as a one-year age effect. Specifications that contain parental education have a smaller sample size due to frictions in our survey as reported in Table 1. The regression models are estimated in Stata (Robust clustered standard errors at the session level). *** <0.01, ** <0.05, *<0.1.

J Robustness check, linear probability model for the coin-drop game.

Here, we replicate our results based on a logit specification with a linear probability model. To do so, the dependent variables are dummies that take value 1 if ℓ_7^I (independent task) and ℓ_7^C (correlated task) are chosen and zero otherwise.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	0.248*			0.0714	0.206*			0.0387
	(0.122)			(0.175)	(0.119)			(0.175)
Fourth Grade		-0.0674		-0.0853		-0.0613		-0.0791
		(0.105)		(0.107)		(0.106)		(0.112)
Fifth Grade		0.0595		0.0329		0.0647		0.0387
		(0.121)		(0.107)		(0.117)		(0.111)
Parental education			0.0609*	0.0586*			0.0420	0.0423
			(0.0302)	(0.0330)			(0.0357)	(0.0352)
Gender	-0.0386	-0.0224	-0.0270	-0.0149	-0.0427	-0.0286	-0.0311	-0.0185
	(0.0521)	(0.0463)	(0.0632)	(0.0580)	(0.0529)	(0.0479)	(0.0655)	(0.0602)
School Fixed Effect					0.0929	0.113	0.0976	0.0890
					(0.0845)	(0.0848)	(0.0975)	(0.0939)
Constant	0.399***	0.557***	0.361***	0.337**	0.379***	0.495***	0.371***	0.361**
	(0.0821)	(0.0868)	(0.117)	(0.153)	(0.0891)	(0.112)	(0.111)	(0.142)
Observations	272	272	217	217	272	272	217	217
R-squared	0.015	0.012	0.021	0.034	0.023	0.025	0.029	0.040

Table 29: Independent task, age, parental education, and cognitive abilities.

The dependent variable is a dummy that takes the value 1 if the pupils chose ℓ_7^I and 0 otherwise. The independent variables are the same as in Table 16 apart from preference intensity. The regression models are estimated in Stata (Robust clustered standard errors at the session level). *** <0.01, ** <0.05, *<0.1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	0.294**			0.246**	0.253**			0.225**
	(0.109)			(0.101)	(0.109)			(0.106)
Fourth Grade		0.0758		0.0811		0.0814		0.0847
		(0.0841)		(0.0799)		(0.0758)		(0.0785)
Fifth Grade		0.235***		0.232**		0.242***		0.236***
		(0.0834)		(0.0852)		(0.0785)		(0.0841)
Parental education			0.0396	0.0432			0.0254	0.0330
			(0.0283)	(0.0283)			(0.0292)	(0.0280)
Gender	-0.104*	-0.0824	-0.122**	-0.103*	-0.108*	-0.0895	-0.126**	-0.105*
	(0.0591)	(0.0582)	(0.0578)	(0.0560)	(0.0571)	(0.0553)	(0.0565)	(0.0549)
Succ. Ind. Task	0.240***	0.243***	0.266***	0.251***	0.232***	0.229***	0.260***	0.247***
	(0.0591)	(0.0549)	(0.0679)	(0.0674)	(0.0629)	(0.0578)	(0.0700)	(0.0688)
School Fixed Effect					0.0920	0.124*	0.0755	0.0569
					(0.0767)	(0.0644)	(0.0901)	(0.0706)
Constant	0.138	0.213***	0.190*	-0.0885	0.121	0.153	0.200**	-0.0720
	(0.0848)	(0.0762)	(0.0948)	(0.128)	(0.0907)	(0.0904)	(0.0917)	(0.122)
Observations	272	272	217	217	272	272	217	217
R-squared	0.098	0.117	0.108	0.166	0.106	0.132	0.112	0.169

Table 30: Correlation neglect, age, parental education, and cognitive abilities.

The dependent variable is a dummy that takes the value 1 if the pupils chose ℓ_7^C and 0 otherwise. The independent variables are the same as in Table 16 apart from preference intensity and Success Ind. Task. This latter is a dummy that takes the value 1 if the pupil chooses the best lottery in the Independent task. The regression models are estimated in Stata (Robust clustered standard errors at the session level). *** <0.01, ** <0.05, *<0.1.

K Robustness check, correlation neglect excluding pupils who failed the independent task.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	1.511**			1.538**	1.191*			1.109
	(0.594)			(0.738)	(0.635)			(0.811)
Fourth Grade		0.516		0.403		0.601		0.494
		(0.510)		(0.502)		(0.417)		(0.406)
Fifth Grade		1.487***		1.480***		1.594***		1.597***
		(0.519)		(0.573)		(0.443)		(0.487)
Parental education			0.144	0.192			-0.0144	0.0501
			(0.163)	(0.182)			(0.153)	(0.174)
Gender	-0.699*	-0.559	-0.816**	-0.564	-0.811**	-0.699*	-0.977***	-0.705*
	(0.389)	(0.384)	(0.364)	(0.393)	(0.377)	(0.394)	(0.368)	(0.403)
School Fixed Effect					0.704	0.940**	1.030**	0.999**
					(0.478)	(0.401)	(0.503)	(0.435)
Constant	-0.431	-0.135	0.157	-1.723*	-0.574	-0.682	0.157	-1.551*
	(0.450)	(0.413)	(0.743)	(0.980)	(0.530)	(0.448)	(0.694)	(0.896)
Observations	148	148	120	120	148	148	120	120

Table 31: Correlation neglect, age, parental education, and cognitive abilities.

We estimate a multinomial Logit in which the dependent variable represents the pupil's choices in the Correlated task. The baseline is represented by the pupils who choose ℓ_5^C and we only report the coefficients related to equation 22. The independent variables are the same as in Table 16 but we exclude pupils who failed the independent task. The regression models are estimated in Stata using the function *mlogit*. *** <0.01, ** <0.05, *<0.1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	1.532***			1.539***	1.321**			1.304**
	(0.559)			(0.592)	(0.606)			(0.660)
Fourth Grade		0.262		0.291		0.269		0.316
		(0.423)		(0.415)		(0.373)		(0.379)
Fifth Grade		1.031**		0.883**		1.032***		0.887**
		(0.415)		(0.433)		(0.387)		(0.416)
Parental education			0.101	0.128			0.0253	0.0764
			(0.123)	(0.133)			(0.120)	(0.126)
Gender	-0.535*	-0.476*	-0.590**	-0.479*	-0.593**	-0.555**	-0.677**	-0.544*
	(0.295)	(0.288)	(0.293)	(0.288)	(0.277)	(0.276)	(0.286)	(0.278)
School Fixed Effect					0.412	0.559	0.696	0.502
					(0.461)	(0.390)	(0.481)	(0.461)
Constant	-0.609	-0.0565	0.135	-1.501**	-0.726	-0.417	-0.0679	-1.503**
	(0.442)	(0.344)	(0.547)	(0.710)	(0.522)	(0.457)	(0.599)	(0.675)
Observations	211	211	174	174	211	211	174	174

Table 32: Correlation neglect, age, parental education, and cognitive abilities in all schools.

We estimate a multinomial Logit in which the dependent variable represents the pupil's choices in the Correlated task. The baseline is represented by the pupils who choose ℓ_5^C and we only report the coefficients related to equation 22. The independent variables are the same as in Table 16 but we exclude pupils who failed the independent task. The regression models are estimated in Stata using the function *mlogit*. *** <0.01, ** <0.05, *<0.1.