CSEF

Centre for Studies in Economics and Finance

WORKING PAPER NO. 737

Exploring Choice Errors in Children

Daniele Caliari, Valentino Dardanoni, Carla Guerriero,

Paola Manzini and Marco Mariotti

November 2024

University of Naples Federico II **University of Salerno** **Bocconi University, Milan**

CSEF - Centre for Studies in Economics and Finance
DEPARTMENT OF ECONOMICS AND STATISTICS – UNIVERSITY OF NAPLES FEDERICO II
80126 NAPLES - ITALY
Tel. and fax +39 081 675372 – e-mail: csef@unina.it

ISSN: 2240-9696

WORKING PAPER NO. 737

Exploring Choice Errors in Children

Daniele Caliari[*], Valentino Dardanoni[†], Carla Guerriero[‡], Paola Manzini[§] and Marco Mariotti[**]

Abstract

Choice mistakes may be consequential. While we have plentiful evidence on adult behaviour, children's choices are much less studied, yet not only may they shed light on adult behaviour, but they are themselves important, as potentially leading to low educational attainment, unhealthy food choices, and risky behaviours. In this paper, we study experimentally how children's choice consistency and ability to avoid mistakes change with age. We study choice by primary school children in two (ubiquitous) domains: riskless and risky choice. We elicit complete choice functions over deterministic choices, while for lotteries we introduce a novel experimental design, documenting as a particular type of framing effect, consistent with correlation neglect, so far only studied in adults. With plentiful evidence of choice errors in adults, unsurprisingly choice errors and inconsistencies abound in children - strikingly though, in some cases already by age 10-11 children display error rates which are close to those observed in adults. Our results are well captured by a model of limited, stochastic consideration. Our experiment is rich enough to highlight the shape that potential interventions could take, aiming at increasing children's consideration capacity. Different socioeconomic backgrounds seem to matter, though, reassuringly, the gap does tend to close over time.

JEL Classification: D01, D90.

Keywords: correlation neglect, bounded rationality, violations of first order stochastic dominance.

Acknowledgments: We thank Lorenzo Neri 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.

[*] WZB - Berlin Social Science Centre. Email: daniele.caliari@wzb.eu.

[†] Università di Palermo. Email: valentino.dardanoni@unipa.it.

[‡] University of Naples Federico II and CSEF. Email: carla.guerriero@unina.it.

[§] School of Economics, University of Bristol. Email: p.manzini@bristol.ac.uk.

[**] School of Economics and Finance, Queen Mary University of London. Email: m.mariotti@qmul.ac.uk.

1 Introduction

Rationality is a cornerstone assumption in economics. While by and large economic policies are drawn with the assumption of "rational agents" in mind, increasing evidence to the contrary shows that people's response to incentives may be influenced by a wide range of cognitive biases. In this paper, we focus on rationality in decision-making intended as *choice consistency and error avoidance* in young children of different ages. Being able to assess the extent of choice mistakes is important in general, as e.g. depending on the type of mistake, individuals may end up with lower overall welfare as compared to their starting point, from ending up in the 'wrong' occupation to being money-pumped, for example. While we have plentiful evidence on adults, focusing on children is important, as childhood development may shed light on adult behaviour, uncovering whether (bounded) rationality persists from childhood or evolves with age/schooling, and at what stage it converges to that observed in adults. Moreover studying error avoidance in children is important on its own, as the role of children and adolescents as economic agents is being increasingly recognised, and so is understanding what mechanisms drive economically relevant behaviour.

In this paper we study how this key form of (ir)rationality varies with age and socioeconomic background, looking at primary school children's ability to avoid mistakes in simple settings in two (ubiquitous) domains: riskless and risky choice. To study error avoidance in riskless choice, we ask children to pick their favourite from a menu of options: by eliciting choices from a variety of subsets of the same universal set of items we are able to investigate children's consistency in choice (or lack of it). To study error avoidance in risky choice, we design a novel elicitation method to investigate children's ability to recognise first order stochastic dominance as well as a particular type of framing effect, which is consistent with correlation neglect (i.e. overvaluing the probability of the union of two events by discounting their common source/correlation). Our experimental subjects are primary school children enrolled in four state schools in an urban context that differ by the socioeconomic characteristics of their catchment areas. By taking a snapshot of the same task performed by children of various ages and across schools with heterogeneous catchment areas, we are able to see how children's choice errors vary, if at all, with age and socioeconomic background across risky and riskless domains.

Our contribution.

While we defer a full discussion of how our paper relates to the extant literature, in a nutshell, the novelty of our contribution can be summed up as follows:

1. unlike previous studies, for the case of riskless choice we go beyond binary comparisons by eliciting the *full choice function*; this enables us to study consistency in choice in the simplest of setups (i.e. we do not need to employ budget sets);

2. we document a theoretical choice 'anomaly', correlation neglect, as yet not studied in children, implementing a novel experimental setup;

3. we provide and estimate a structural model of children's choice behaviour;

4. while we do not claim to have a balanced sample reflecting the general income distribution of the population, our subject pool is heterogeneous, consisting of children in primary state schools from across catchment areas from varying socioeconomic backgrounds.

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 year olds) are already comparable to those seen in adults.[1] Error rates fall as age increases and as socioeconomic status (as proxied by the school's location) and parental educational attainment improve. We then show that, on the one hand, the gap in error rates between children from more and less privileged backgrounds reduces with age. On the other hand, however, when considering tests of so-called 'fluid intelligence', namely children's performances in Raven matrices, we find that the performances *diverge* with age: children start off performing equally badly, but then those in schools located in wealthier neighbourhoods improve their Raven scores more than children attending schools in more deprived areas. This evidence aligns with the existing literature on early educational intervention (**?**) which shows that the window of opportunity to intervene is relatively narrow;[2] and on the short-term effect of moving to wealthier neighbourhoods (**?**). 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 (**?**, **?**, **?**).

What could explain these differences? The decision-theoretic approach we take is that children may fail to appraise all alternatives available. We model children's choices using the 'stochastic consideration set' model (see **?**), according to which each item in a menu is not necessarily considered for sure, but with some probability captured by a consideration parameter. The consideration parameter can be taken as a proxy for (an index for) rationality, as in the presence of stable, unvarying preferences, a fully rational agent can be represented as having full consideration (consideration parameter approximating unity), while someone choosing uniformly can be represented as not considering anything (consideration parameter approximating zero). We follow **?** to estimate the consideration parameters in the corresponding structural model and find that in all domains this parameter mirrors the error rates: consideration improves with age; is higher in pupils from the most well-off schools; and the gap in consideration

[1] This is partly based on existing results in the literature, partly on a small incentivised test of correlation neglect and violations of First Order Stochastic Dominance in a subpopulation of adults - more details in section 5 in the paper.

[2] **?** find evidence that additional years of schooling (in a natural experiment where the compulsory school age in England was raised from 15 to 16) have no significant effect on the quality of decision-making (while having significant positive effects on qualifications obtained and future earnings). If error avoidance in children as young as 11, as in our experiment, approximates that of adults, then possibly this confirms that any intervention on children should be early. (see **?**).

parameter across schools reduces over 'simple' domains but persists in the case of 'difficult choices' (i.e. in tests of correlation neglect).

1.1 Related literature

Our paper fits the recent small but expanding literature on experiments with children. Recent comprehensive reviews are **?** and **?** - the papers closest to our study are **?** and (**?**. 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 year olds) were presented with 11 choices of either three or seven bundles consisting each of bags of crisps and fruit juice cartons. They investigate violations of the Generalised Axiom of Revealed Preferences (loosely speaking, it rules out preference reversals in chains of choices, or equivalently violations of transitivity), and consistent with our findings, uncover that rationality improves with age, with the response from older children being quite similar to that of adults (undergraduate students in this case). While the bundles offered to the children were taken from/consistent with budget sets, understandably the choice menu was not framed as a budget (i.e. no mention of prices and of income); on the other hand, given the intention to approximate budgets, there isn't a grand set including all bundles (which would have been a pretty complex set to choose from, as it would have to include 28 bundles), and more importantly, as noted in **?**, 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 **?** (in riskless and risky choices as well as social/group choices). However, these do not allow checking for the presence of menu effects, as we do (while we do not cover the social preference dimension). Hence we view our contribution as complementary to these important papers.[3]

As noted, in our paper we also investigate violations of first-order stochastic dominance (FOSD) and the correlation neglect phenomenon,[4] hence our study is also related to the experimental literature on FOSD violations and correlation neglect, though none of the extant papers explores these issues with children, Virtually all the literature documenting FOSD violations relies on covert implementation, with questions framed so that the FOSD relation isn't apparent - two notable exceptions are **?** and **?**.[5] Interestingly, both papers find violations of FOSD in university students to be low, and much lower than in our findings for children. The experimental literature on correlation neglect (see e.g. **?**, **?**, **?**), is also focused entirely on the adult population, which is unsurprising given the intrinsic complexity of the notion of correlation. The simplest

[3] Another paper studying risk preferences in children is **?**. This experiment too has children choosing from binary sets, with subjects being older children (average age 13.78) than our experimental cohort. Another recent study of risk preferences in children is **?**, which also focuses on binary choices in the lottery domains, and enriches the analysis by also investigating personality traits.

[4] 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 **?**.

[5] In **?** terminology, our presentation is such that the best strategy is 'obviously dominant'.

setup, which is closest to ours, is that used in **?**, who consider the case of school choice.[6] They simulate in the lab the case of applicants to university places, where the same test mark is used to process applications. Applicants have to select two universities, with different entry requirements. If the top two schools differ only slightly in entry requirements, risk neutral and risk averse applicants should select the top school and the safe choice, yet **?** 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, with only a handful picking the dominated strategy and roughly half selecting. This is very similar to the choice behaviour we observe in our older children, as we show below.

The paper is organized as follows. In section 2, we introduce our two experimental designs. In section 3, we present our stochastic consideration model, provide identification results, and describe the estimation technique. In section 4, we describe our dataset with information about the experimental procedure and the different sample sizes. In section 5, we present our results for both parts of our experiment. Finally, in section 7 and 8, we conclude with some discussion on potential further research. The Appendix contains omitted proofs, several further results, robustness checks, and a description of the experiment using screen-shots and instructions. Finally, in the Online Appendix, we include a different estimation approach and more robustness checks.

2 Experimental design

As noted, in our experiment we wished to explore three types of choice mistakes: failure of choice consistency in riskless choice, and failures of dominance and correlation neglect in risky choice. We describe our approach to testing each of these below.

2.1 Errors in riskless choice.

First of all, we look at the simplest possible setup for 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, and pick consistently 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. When choices satisfy WARP, as noted choice behaviour can be described as if a preference is being maximised. When WARP fails, then failures of rationality can be attributed to one of *only two* possible

[6]Other early experimental papers documenting correlation neglect in economics are **?**, on portfolio choice, allocating 'wealth' to financial assets with differing, state-dependent returns; and on belief formation in the sense of predicting from correlated signals in the case of **?**.

forms of non rational behaviour: cyclical choice or menu effects.[7] Cyclical choices are problematic as they can turn any decision maker into a money pump; menu effects are problematic as the decision maker can in principle be manipulated into choosing what the designer of the menu prefers.

More in detail, 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.[8] The questions were presented in random order,[9] with a distractor task[10] between each choice.

After both sets of 11 questions, we elicited preference intensities for each item on a Likert scale from 1 to 5.[11] Lastly, subjects answered eight Raven's Coloured Progressive Matrices (RCPM), presented in increasing difficulty order.[12] 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 the Appendix. We will refer to this deterministic choice task as the pens/pencils task.

Figure 1: Sets of pencils and pens.

2.2 Errors in risky choice - testing for failures of first order stochastic dominance and for 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 objects 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

[7] See **?** for this taxonomy of errors.

[8] This part of the experiment was conducted through tablet computers. The game was delivered on a website created specifically for the experiment and hosted on a local server. In the beginning, a teacher inserted the demographic data about the pupil in the tablet to ensure the correct identification of each pupil. After this preparatory phase, the pupils were instructed, and then upon receiving the tablet, they started the experiment. The code is available here: *https://github.com/DCaliari/RationalityInChildren.git*

[9] 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.

[10] A short nature video (about 10 seconds) was played after each choice question.

[11] This task is similar to the "ranking task" (**?**), where in addition we allow for indifferences. This way we are able to devise a preference intensity index, as we describe more in detail in section 5 and in appendix B.2.

[12] 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.

coin drop game that induces a uniform distribution over the outcomes. 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.[13] A lottery takes the form of a strip of eight squares, either white or yellow, placed below the pockets, see Figure 2.

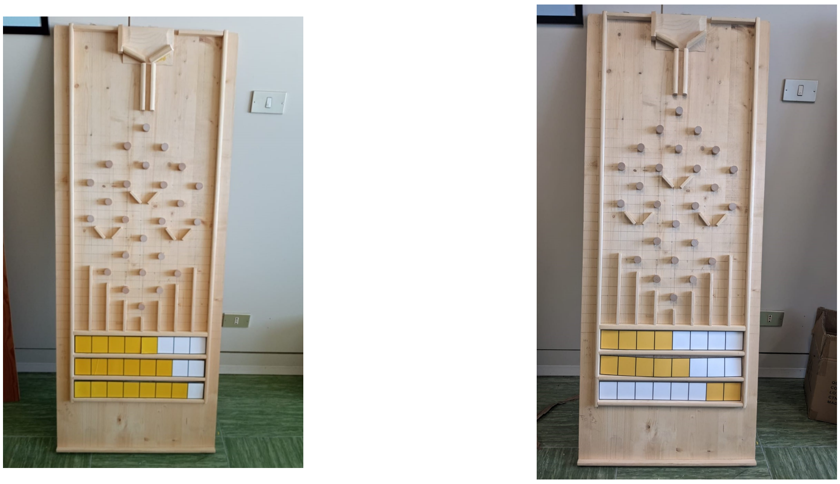

Figure 2: The coin drop game: the FOSD test lotteries are on the left, the correlation neglect test lotteries are on the right.

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 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 statewise dominance between the lotteries, to make it more easily recognisable).[14] 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 coin drop game'. Children performed their choice on a piece of paper reporting 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 of the yellow squares, and no prize otherwise.[15]

The right panel in Figure 2 displays the lotteries that we used to test for correlation neglect. Correlation neglect is the failure to react "correctly" to the correlation of

[13]The pin placement follows **?**.

[14]Most FOSD tests are non transparent, with a notable exception being **?**, who test FOSD failures in university students. They find violations of FOSD in university students to be low, and much lower than in our findings, which we report in section 5.3.

[15]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) children recorded their choices on answer sheets (see Figure 3; actual answer sheets did not have the "independent" and "correlated" labels).

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.[16] 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 fact that correlation between the two events (being admitted to either university) depends on the same mark.

To test for correlation neglect in children, we used the lotteries depicted in the right panel in Figure 2 and in Figure 3, which we refer to as 'correlated'. Now children had to pick **two** out of three lotteries, with a win realising if **any** of the yellow squares **across the two** chosen lotteries corresponded to the pocket with the dropped coin. As can be seen in the picture, one strip/lottery had the four leftmost squares in yellow, and the other four in white; one had the five leftmost squares in yellow, the three rightmost ones in white; and one had the two rightmost squares in yellow, and the rest in white. Here picking the two lotteries with the four or five leftmost squares in yellow evidences 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.

Note that each pair of lotteries in the correlated set maps into the same probability of winning as a corresponding lottery in the independent set. For instance, the top two lotteries in the correlated set produce a win in 5 out of 8 cases, as does the top independent lottery. Similarly, the bottom two lotteries in the correlated set produce a win in 7 out of 8 cases, as does the bottom lottery in the independent set; and the extreme lotteries in the correlated set produce a win in 6 out of 8 cases, as does the middle lottery in the independent set. 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).

[16]We borrow this example from **?**.

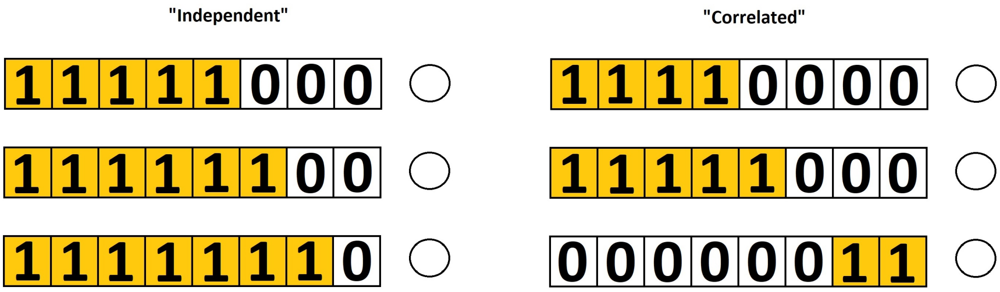

Figure 3: The answer sheets in the two tasks.

We run the independent task prior to the correlated task; note that the independent task is also a test of first order stochastic dominance (and also doubles as a comprehension test). Finally, all three lotteries were played out, and prizes were assigned accordingly.

3 Modelling errors in choice - stochastic consideration set

In this section, we describe the theoretical foundations of our experimental design and our subsequent analysis. Our experimental setup concerns mostly discrete choice problems, in which each subject has to select one element out of a menu of alternatives. Given a generic finite set A of choice alternatives, we denote by A_i a generic subset, and by $\mathcal{A}$ the collection of all possible subsets. A (deterministic) *choice function* C is a map that associates to each menu an element of that menu, with the interpretation that $C(A)$ is the alternative chosen in set A. Stochastic choice functions allow for variability in choice: a stochastic choice function is a probability distribution over choices: $p : A \times \mathcal{A} \mapsto [0, 1]$ such that $\sum_{a \in A} p(a, A) = 1$.

We model discrete choice using the Stochastic Consideration Set (SCS) model introduced in **?**, following **?**'s conditional version.[17] While in the standard choice model, a rational agent picks the alternative that ranks highest in her preference, in the SCS model an agent may fail to consider some alternatives, hence she picks the most preferred alternative *among those considered*. More precisely, in this model, the decision maker considers each feasible alternative a in (finite) menu A with probability $\gamma_a \in (0, 1]$, which is independent of the probability of considering other alternatives, and then selects the most preferred one according to the asymmetric, complete and transitive preference $\succ$. We will refer to $\succ$ as the agent's *type*, as in **?**, and let P denote the collection of all possible types. Then the probability that type $\succ$ chooses alternative a

[17] In its original formulation, the SCS model allows for non choice, while **?**'s version rules it out, hence it is more suited to model limited consideration in an experiment with forced choice as ours.

from menu A , which we denote by $p^{\succ}(a, A)$, is expressed by:

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

where at the numerator the first term is the probability that a is considered, while the second term is the probability that no alternative better than a is considered (for in that case a would never be chosen). 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). We now turn to showing how this general setup applies to the three experimental tasks. It is going to be convenient to start from the coin drop games, as each task involves a single choice, and then move to the pen/pencil choice task.

3.1 The coin drop game - independent lotteries task

This is the simplest setup in which to model choice behaviour: it is natural to assume homogeneity in preferences, i.e. that larger win probabilities are preferred to lower win probabilities. Then we can drop the preference superscript without risk of confusion. We denote by ℓ_z^I a lottery with z winning outcomes, and by L^I the set of independent lotteries (that is $L^I = \{\ell_5^I, \ell_6^I, \ell_7^I\}$).

For a standard rational agent, the three lotteries are ordered by first order stochastic dominance: the dominant lottery ℓ_7^I should be chosen with probability 1. Hence as a preliminary test of rationality, we have:

Implication 0. *If the model holds with $\gamma_n^I = \gamma^I = 1$ for all n, then $p(\ell_7^I, L^I) = 1$ and $p(\ell_6^I, L^I) = 0 = p(\ell_5^I, L^I)$.*

The analysis changes in the presence of limited consideration. Letting γ_n^I denote the consideration probability for the lottery with n winning outcomes in this treatment, we have:

$$p\left(\ell_5^I, L^I\right) = \frac{\gamma_5^I (1 - \gamma_6^I)(1 - \gamma_7^I)}{1 - (1 - \gamma_5^I)(1 - \gamma_6^I)(1 - \gamma_7^I)} \tag{2}$$

$$p\left(\ell_6^I, L^I\right) = \frac{\gamma_6^I (1 - \gamma_7^I)}{1 - (1 - \gamma_5^I)(1 - \gamma_6^I)(1 - \gamma_7^I)} \tag{3}$$

$$p\left(\ell_7^I, L^I\right) = \frac{\gamma_7^I}{1 - (1 - \gamma_5^I)(1 - \gamma_6^I)(1 - \gamma_7^I)} \tag{4}$$

Since we have two independent equations in three unknown, the model with alternative-dependent consideration parameters is under-identified. However we can achieve identification in the case of lottery independent consideration parameter: given the observed probabilities of choice of each lotteries, any of these equations suffice to identify the unobserved consideration parameter γ^I. Then the following condition is necessary for the model, hence a testable prediction:

Implication 1. *If the model holds with $\gamma_n^I = \gamma^I$ for all n, then:*[18]

$$\frac{p\left(\ell_7^I, L^I\right)}{p\left(\ell_6^I, L^I\right)} = \frac{p\left(\ell_6^I, L^I\right)}{p\left(\ell_5^I, L^I\right)} \tag{5}$$

Implication 2. *If the model holds and consideration is monotonically non decreasing in n, then*

$$p\left(\ell_7^I, L^I\right) > p\left(\ell_6^I, L^I\right) > p\left(\ell_5^I, L^I\right) \tag{6}$$

3.2 The coin drop game - correlated lotteries task

In the correlated lotteries task subjects pick *two* lotteries. Quite irrespective of the model, 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.[19] We let L^C denote the set of correlated lotteries, that i $L^C = \{\ell_2, \ell_4, \ell_5\}$, 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:

$$p\left(\ell_5^C, L^C\right) = \frac{\gamma_4^C \gamma_5^C \left(1 - \gamma_2^C\right)}{\gamma_2^C \gamma_4^C + \gamma_2^C \gamma_5^C + \gamma_4^C \gamma_5^C - 2\gamma_2^C \gamma_4^C \gamma_5^C} \tag{7}$$

$$p\left(\ell_6^C, L^C\right) = \frac{\gamma_2^C \gamma_4^C \left(1 - \gamma_5^C\right)}{\gamma_2^C \gamma_4^C + \gamma_2^C \gamma_5^C + \gamma_4^C \gamma_5^C - 2\gamma_2^C \gamma_4^C \gamma_5^C} \tag{8}$$

$$p\left(\ell_7^C, L^C\right) = \frac{\gamma_2^C \gamma_5^C}{\gamma_2^C \gamma_4^C + \gamma_2^C \gamma_5^C + \gamma_4^C \gamma_5^C - 2\gamma_2^C \gamma_4^C \gamma_5^C} \tag{9}$$

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.[20]

[18] This is verified by setting $\gamma_n^I = \gamma^I$ for all n in equations (2)-(4) where $\frac{p\left(\ell_7^I, L^I\right)}{p\left(\ell_6^I, L^I\right)} = \frac{1}{1-\gamma} = \frac{p\left(\ell_6^I, L^I\right)}{p\left(\ell_5^I, L^I\right)}$

[19] This differs from the setup in **?**, where the focus is on misunderstanding the integration of separate problems, the usual issue addressed in choice bracketing.

[20] Note that the probability of not choosing anything is the probability of considering at most one lottery, which is 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.

From the above it is clear that any given choice probability for the payoff maximising pair $\ell_7^C = \{\ell_2, \ell_5\}$ is compatible with very different choice probabilities of the suboptimal pairs; in particular for low values of γ_2^C and high values of γ_5, probability "transfers" to the "correlation neglect" pair $\ell_5^C = \{\ell_4, \ell_5\}$. So the magnitude of γ_2^C is a proxy for correlation neglect - however within the setup we can't estimate alternative dependent consideration parameters - so it may be convenient to provide a measure of correlation neglect that does not depend on the unobservables.

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

$$\iota = \frac{p\left(\ell_7^C, L^C\right) - p\left(\ell_5^C, L^C\right)}{p\left(\ell_7^C, L^C\right) + p\left(\ell_5^C, L^C\right)} \in [-1, 1] \tag{10}$$

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 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:*

$$\frac{p\left(\ell_5^C, L^C\right)}{p\left(\ell_6^C, L^C\right)} = 1 \textit{ and } \frac{p\left(\ell_7^C, L^C\right)}{p\left(\ell_i^C, L^C\right)} > 1, i = 5, 6 \tag{11}$$

Implication 2'. *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,*[21] *that is:*

$$\frac{p\left(\ell_5^C, L^C\right)}{p\left(\ell_6^C, L^C\right)} > 1 \textit{ and } \frac{p\left(\ell_7^C, L^C\right)}{p\left(\ell_6^C, L^C\right)} > 1 \tag{12}$$

To test the implication of our model, we estimate it as follows. Let γ be a triple of consideration parameters and Y_n the number of subjects choosing lottery ℓ_n then the maximum-likelihood problem becomes:

$$\max_{\gamma} \log \mathcal{L}\left(\gamma; \mathbf{Y}\right) = \sum_{\ell_n \in A} Y_n \log(p(\ell_n, A)) \tag{13}$$

We now turn to the formal analysis of the riskless choice task.

3.3 Stochastic consideration and choice from multiple sets

The pens/pencils task involved for each subject the elicitation of choices from a collection of sets: for each domain (either pens or pencils) we asked each subject to perform

[21] This holds in the 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 $p\left(\ell_7^C, L^C\right) > p\left(\ell_6^C, L^C\right)$.

11 choices, selecting the preferred item from each of 11 possible sets. For each experimental subject we have a specific realisation of such choices. Let C denote a generic realisation (i.e. a choice function), and let $\mathcal{C}$ denote the collection of all possible choice functions. A Mixture Choice Function[22] 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

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

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

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

More in general the mixture choice function can be expressed as

$$\boldsymbol{\mu} = \mathbf{Z}\boldsymbol{\pi}, \tag{14}$$

where $\boldsymbol{\mu} = \left[\mu(C_1), \mu(C_2), ..., \mu\left(C_{|\mathcal{C}|}\right)\right]'$, 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 $\boldsymbol{\pi} = \left[\pi(\succ_1), \mu(\succ_2), ..., \mu\left(\succ_{|P|}\right)\right]'$ is the vector form of the (unknown) type distribution.[23] The specific form that the type conditional probabilities take depend on the model postulated.

3.3.1 Identification of the preference and cognitive parameters

Our dataset consists of a finite sample $(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_I)$ of observed choices made by I experimental subjects; sample units are heterogeneous pupils each choosing from the same set of menu choices, that is, for each pupil we observe two choice functions, which

[22] See **?**.

[23] Equation (14) can be written more extensively as

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

we analyse separately. With four alternatives we have 24 possible preference orderings. We wish to identify the type distribution $\pi(\succ)$ and the type conditional choice probabilities $p^{\succ}$. First of all, we establish that the model is generically identified:

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 A. It relies on arguments in **?** (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 in the case of our model, which then ensures that the rest of the proof by AMR also holds for our case.

Then since the theoretical model is identified from the collection of binary and ternary sets (see **?**), 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.

Next, we move to the estimation of our model.

3.3.2 Estimation: Finite Mixture Model

To begin with, we estimate our model (type distribution $\pi(\succ)$ and consideration parameter γ) from the deterministic choice dataset by maximum likelihood, following an iterated two-step procedure as in **?**: at each step we proceed to maximise the likelihood with respect to one unknown conditional on the estimated values of the other (with an arbitrary imputation to kick off the process), iterating until convergence. The outcome of the maximisation problem is, therefore, an optima consideration parameter γ^* and the estimated population distribution $\hat{\pi}(\omega)$.

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 $\mathbf{y}_i = \left(y_{iA_1}, y_{iA_2}, \ldots y_{iA_{|\mathcal{A}|}}\right)$.[24] The individual likelihood that observation $\mathbf{y}_i$ could have been generated by type $\succ$ is given by:

$$\mathcal{L}_i(\succ; \boldsymbol{y}_i) = p(\boldsymbol{y}_i) = \prod_{A \in \mathcal{A}} p(y_{iA}) \tag{15}$$

[24] Each subject i makes a single choice from each of the eleven choice menus in each of the pen or pencils domains (six binary choices, four 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 ω, 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 ω-th column of $\mathbf{Z}$ (as defined in footnote 17).

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

$$\mathcal{L}_i\left(\succ;\boldsymbol{y}_i\right)=\sum_{\succ\in P}\pi\left(\succ\right)\left[\prod_{A\in\mathcal{A}}p^{\succ}\left(y_{iA}\right)\right]. \tag{16}$$

When the model is identified (that it, 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)):

$$\max\ \log\mathcal{L}\left(\pi,\gamma;\left(\succ;\boldsymbol{y}\right)\right)=\sum_{i=1}^{|I|}\log\left(\sum_{\succ\in P}\pi\left(\succ\right)\prod_{A\in\mathcal{A}}p^{\succ}\left(y_i\right)\right). \tag{17}$$

4 Data

Our sample consists of 676 primary school children from four state primary schools in different districts of a large city in Italy, Naples. It comprises classes across all years for which we obtained consent to conduct the experiments from the school president, the school board, and the teachers.[25] The schools differ in catchment areas, providing a good degree of socio-economic variability.[26]. Our analysis focuses on the two schools among the four that are 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. The remaining two schools will be used for robustness checks and the analysis will be replicated comprehensively in Appendix C. 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.

[25]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 personally administered by Carla Guerriero, along with a team of trained interviewers selected for their extensive experience in conducting scientific research with children. These interviewers were all educated to degree level and were aged between 30 and 46 years.

[26]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. The other three schools are located in different parts of Pozzuoli, a town within the wider metropolitan area around Naples. One, which we will refer to as School H, is in a privileged area of Pozzuoli.

	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.

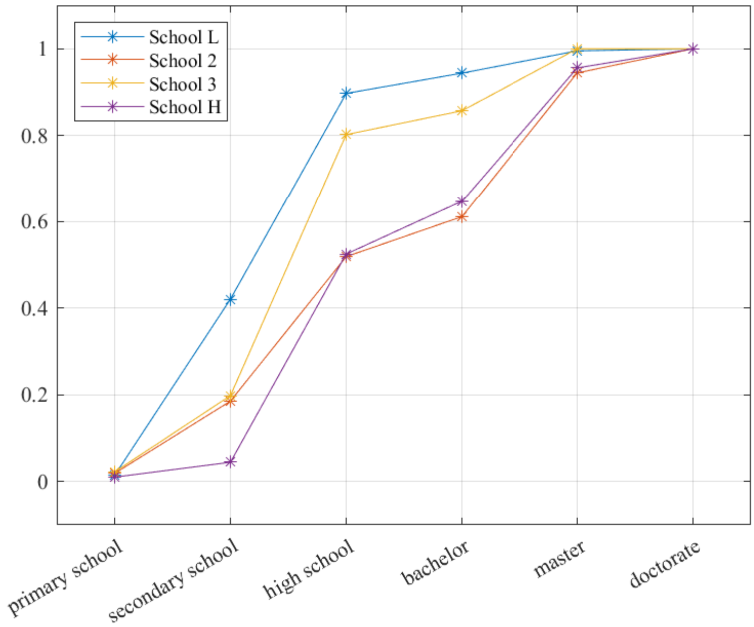

Figure 4: Cumulative distributions of parental education

The pen/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.[27]

[27]Our complete dataset comprises 732 primary school children. Due to COVID-19, the data collection for the pen/pencils 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).

5 Results

5.1 Overview of the results

Recall that, in our take, error avoidance translates into choice consistency in the pen/pencils task, and in choosing the dominant lottery in the coin drop tasks. We find that consistency increases with age in both tasks, as can be seen in Figure 5, where the left-hand panel plots the percentage of children whose choices in the pen/pencil task are compatible with the maximisation of a preference, and the right-hand panel plots the proportion of children who picked the dominant lottery in the two coin drop tasks.

In the pen/pencil task difference in proportions between the youngest and oldest children is almost fourfold, from 14% to 47%, and is statistically significant (p-value < 0.01). In the independent coin drop task the difference is surprisingly small, from 59% to 65% (p-value > 0.1) while in the correlated coin drop task the difference is again large, from 32% to 56%, and significant (p-value < 0.01).

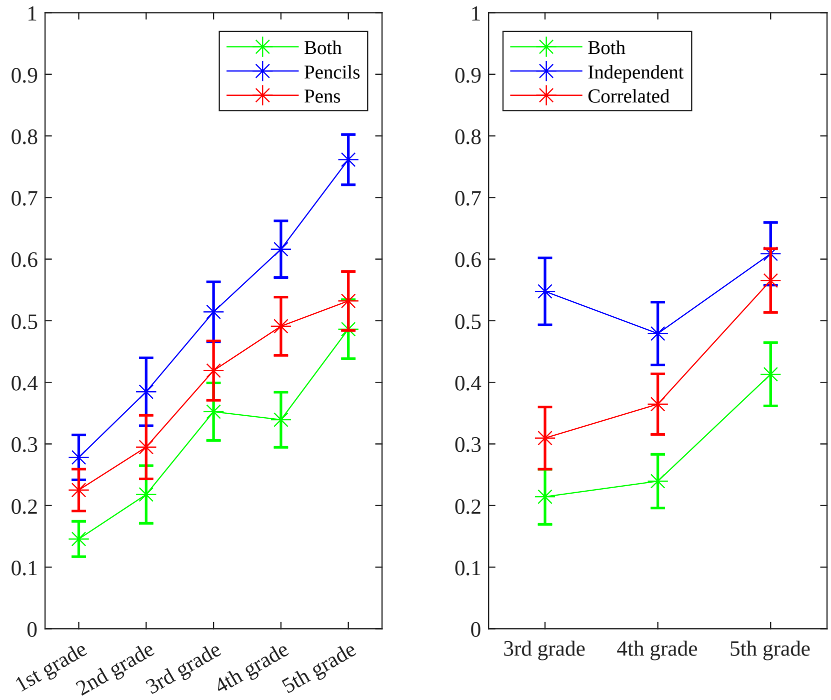

Figure 5: Percentage of rational pupils in the two experiments

5.2 Errors in the Pen/Pencils task

We report our results distinguishing by domain, i.e. pencils and pens. In light of Proposition 1, we estimated a unique consideration parameter in each domain (either pens or pencils), distinguishing by grade and school. Our results are collected in Table 2, and plotted in Figures 6 and 7. Figure 6 reports the distribution of the preference types at the school level showing an overall similarity of preferences across schools. 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 either the eraser on top of each pencil (Duck, Ladybug, Frog, and Shark) or the pen colour (Yellow, Red, Blue, and Green).

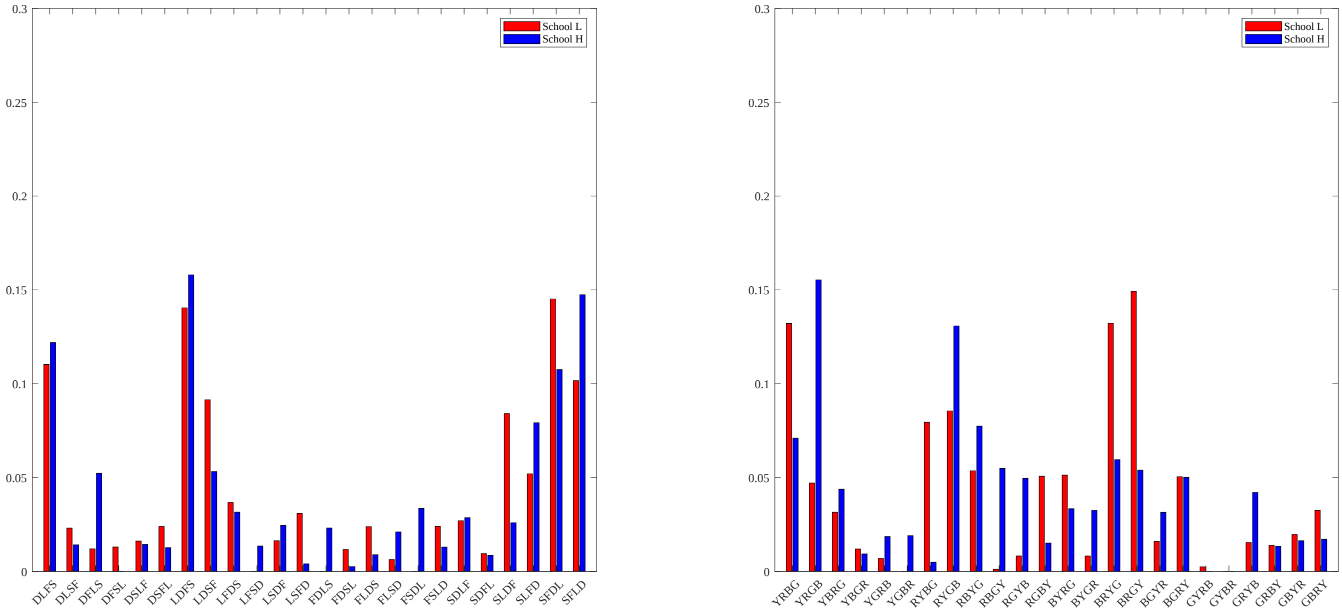

Figure 6: Distribution of the preference types by school.

In Table 2 we report the estimates of the consideration parameter, both at the level of each subpopulation (by grade) and overall in each school.

	Pencils	Pens	Pencils	Pens	
1st Grade	0.67	0.63	0.83	0.77	1st Grade
2nd Grade	0.78	0.72	0.81	0.77	2nd Grade
3rd Grade	0.77	0.72	0.90	0.85	3rd Grade
4th Grade	0.90	0.88	0.91	0.87	4th Grade
5th Grade	0.92	0.86	0.95	0.90	5th Grade
School	0.81	0.76	0.88	0.84	School
	School L		School H		

Table 2: Estimated values of γ, pen/pencils tasks

These are plotted in Figure 7, where from left to right the estimates regard the choices between (i) pencils, and (ii) pens.

As can be seen, the estimated consideration parameter is larger the older the children and this applies to both schools. Since in our model, a less preferred alternative can only be chosen if the better one(s) is not considered, the consideration parameter acts as a proxy for the ability to avoid errors; our estimates show that this is much more pronounced in School H than in School L, with this difference being statistically significant for both pencils and pens (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, as 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 in both schools), by the time children reach fifth grade the differences between schools that are signification in the 1st grade (p-value < 0.01) are no more statistically significant (p-value > 0.1).[28]

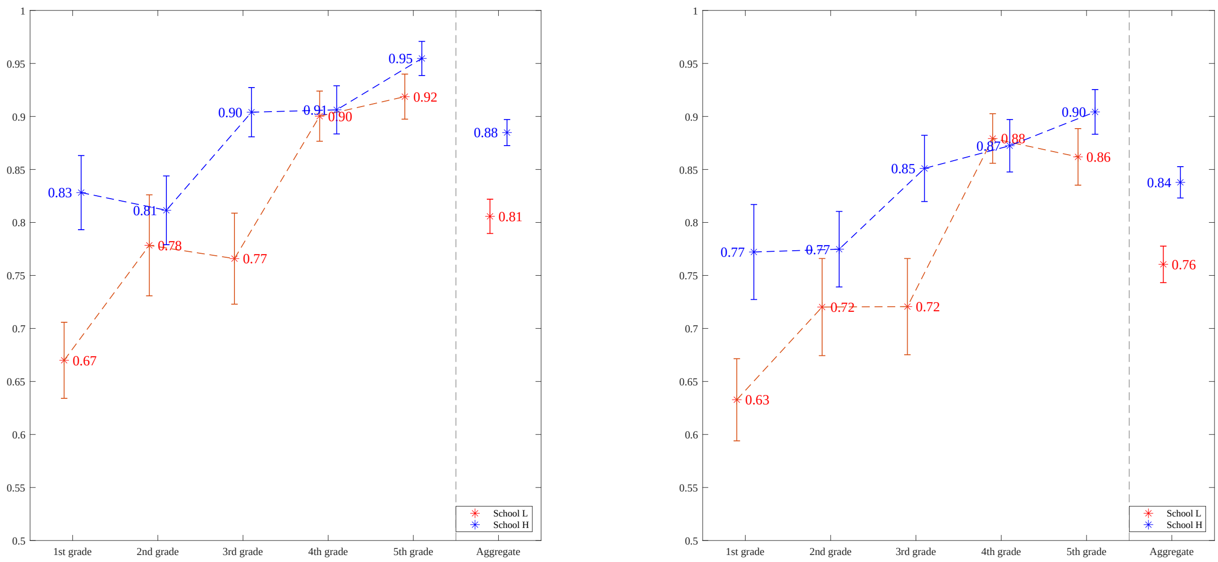

Figure 7: Consideration parameter γ by school and grade

5.3 Errors in the coin drop games

The results of our coin-drop tasks are reported in Figure 8, which displays the joint distribution of choices in the Correlated and Independent tasks, across grades and schools (recall that these tasks were only carried out with children in third to fifth grade).

[28]Finally we note that the ability to avoid error is similar across genders (while preferences are distinct), full details in appendix B.1.

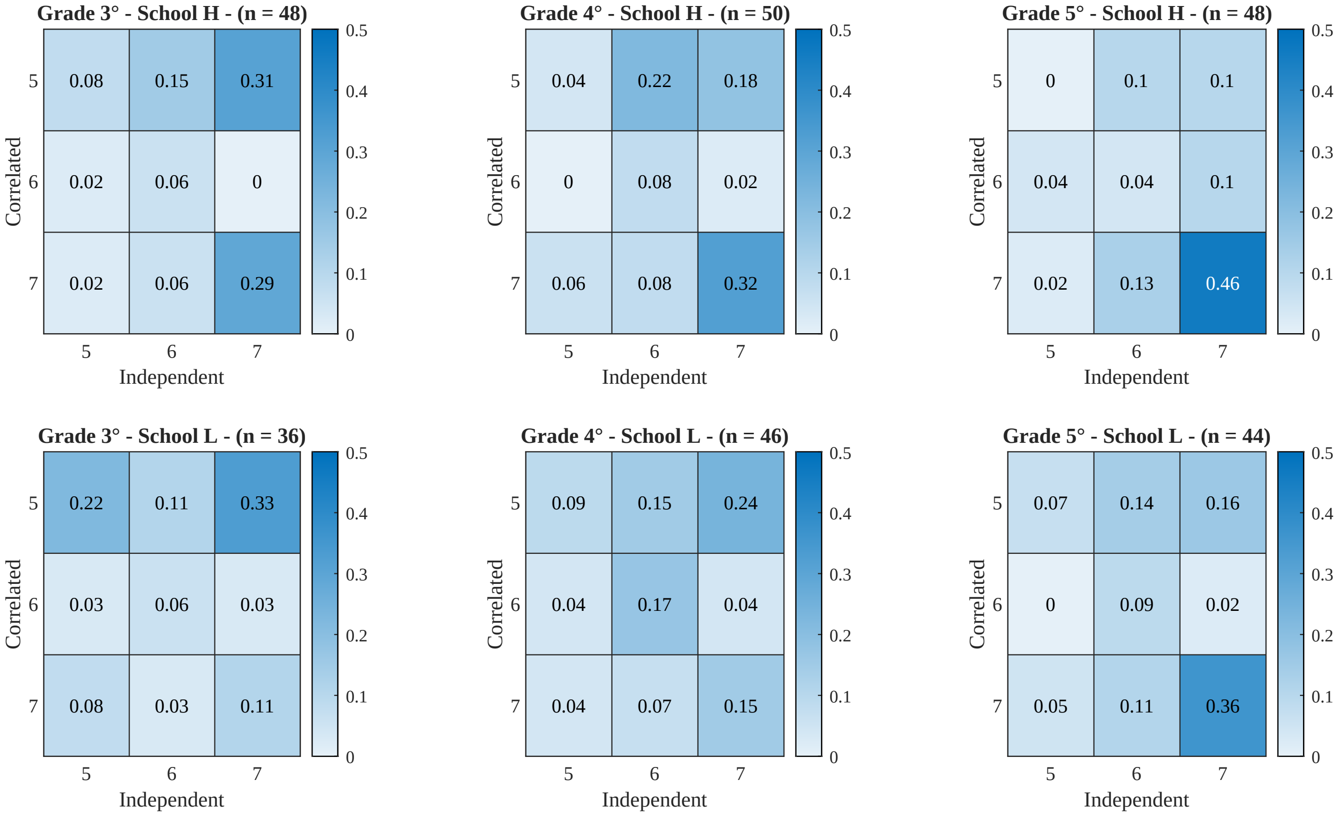

Figure 8: Joint distribution of the "risky choices" by school and grade

While there are stark differences between grades and schools, there are is also considerable commonality in patterns, which we now explore.

5.3.1 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% (supporting our focusing on older children). Disaggregating by school and grade we see quite clearly that the ability to avoid errors improves with age, and is larger in School H than in School L. 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, **?** is a notable exception. 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 one with probability $\frac{5}{9}$ and the other with probability $\frac{6}{9}$, about 9% of their subjects chose the dominated lottery. In our case 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. We replicated the questions in an online survey with a population of adult (average self-declared age: 58 years old) female participants from a craft private Italian Facebook group (presenting the questions as in Figure 3). Surprisingly to us, in this non-student sample, the error rate is even higher than in our school children: only 20 out of 54 respondents answered correctly (37%). Our data show a difference in the ability to avoid errors between ages and schools, which is statistically greater in School H than in School L. In all cases however our Implication **2**, i.e. $p\left(\ell_7^I, L^I\right) > p\left(\ell_6^I, L^I\right) > p\left(\ell_5^I, L^I\right)$, is confirmed.

As we noted, 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 **1**: 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**
Relative 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

In view of Table 3, the estimates of the uniquely estimated consideration parameter are reported in Figure 9. T-tests comparing our estimates report significant differences in the aggregate between schools (p-value < 0.05), and in the 3rd grade (p-value < 0.05), while as in the riskless choice, differences in the 5th grade are not statistically significant.

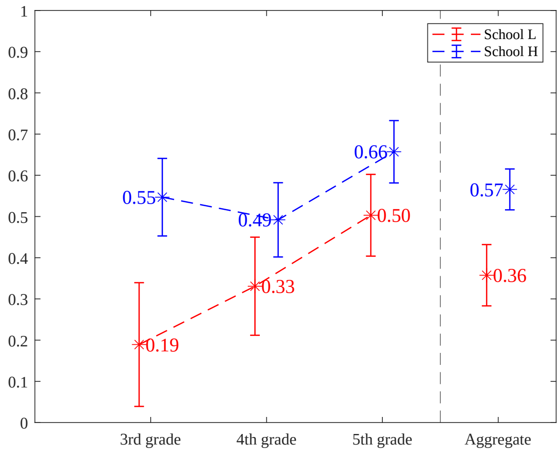

Figure 9: Estimated consideration parameter in the Independent Coin drop task

5.3.2 The Correlated Coin Drop task

As clear from Figure 8, there is a substantial amount of correlation neglect: the choice patterns in this task are very different than in the Independent choice task, 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 **?**. [29] 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 **1′**, which required the choice probabilities for lotteries $\ell_5^C = \{\ell_4, \ell_5\}$ and $\ell_6^C = \{\ell_4, \ell_2\}$ should 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 oldest group of children, in 5th grade, in School H.[30]

[29] **?** 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 **?**).

[30] 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				Aggregate
LRT (p-value)	**0.0000**	**0.0423**	**0.0138**	**0.0000**
Observed probabilities				
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				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 **2'** 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 some support for the assumption that the consideration parameter is non decreasing in the number of states in which a prize is won.

In summary then 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) signaling a clear development in rationality by age. On the other hand, when controlling for grades, differences across schools are never statistically significant.

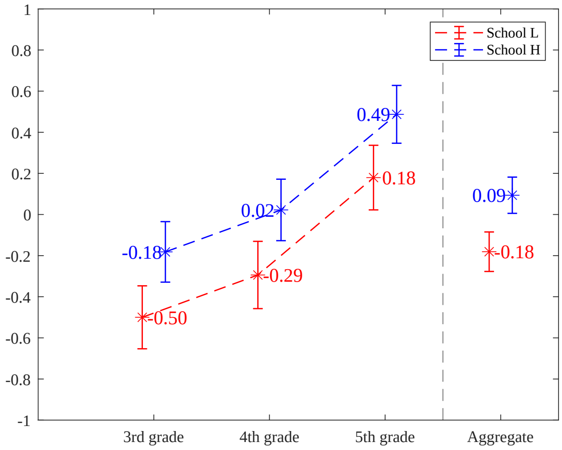

Figure 10: Index of correlation neglect ι

6 The determinants of the consideration parameter

In this section we study how the covariates we have available might influence the consideration parameter γ in the various tasks, focusing on individual-level estimates.[31]

6.1 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, and we average across the pen and pencil tasks.[32] 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.[33] 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 a Likert scale as mentioned in section 2.1.[34] We also include

[31] Individual estimates are recovered by maximizing the likelihood in equation 16 which is achievable since we have multiple observations for each pupil. While Proposition 1 establishes identification in the pen/pencils task, these individual estimates may be noisy due to the small sample size. In the Online Appendix, we replicate our results by obtaining individual estimates via the application of Bayes' theorem to the estimates resulting from the maximum likelihood in equation 17 applied to our sub-samples (by schools and grades) and presented in Figure 6, 7, and Table 2. The results are everywhere robust and, if anything, correlations are greater.

[32] We aggregate the two tasks for reasons of parsimony and because the results when separating the two are very similar. We report these results in the Online Appendix.

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

[34] See the appendix B.2 for descriptive statistics on preference intensity and its calculation.

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 pupil's preferences. Both gender and preference intensity[35] allow us to control for any residual effect of the pupil's preferences that is not captured by the model. The equation we estimate is:

$$\hat{\gamma}_{i,s} = \alpha + \beta_1 \mathbf{X}_{i,s} + \beta_2 \mathbf{Z}_{i,s} + \rho_s + \varepsilon_{i,s} \tag{18}$$

Results are presented in Table 5 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.

[35]Preference intensity is a particularly important control variable. Preferences enter our model ordinally and, therefore, do not account for different degrees of preference intensity. However, it is reasonable to think that children may develop their preferences with age and that as preferences become clearer the errors will decrease.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	0.190***			0.0753	0.175***			0.0659
	(0.0399)			(0.0450)	(0.0406)			(0.0471)
2nd Grade		0.0678**		0.0669*		0.0524*		0.0601*
		(0.0318)		(0.0345)		(0.0311)		(0.0356)
3rd Grade		0.0928***		0.0669*		0.0817***		0.0630*
		(0.0335)		(0.0380)		(0.0277)		(0.0353)
4th Grade		0.150***		0.128***		0.142***		0.126***
		(0.0241)		(0.0289)		(0.0237)		(0.0284)
5th Grade		0.182***		0.154***		0.175***		0.152***
		(0.0223)		(0.0260)		(0.0199)		(0.0260)
Parental education			0.0230**	0.0196**			0.0138	0.0134*
			(0.00951)	(0.00732)			(0.00902)	(0.00753)
Preference Intensity	0.107***	0.108***	0.112***	0.112***	0.101***	0.101***	0.106***	0.107***
	(0.0242)	(0.0216)	(0.0292)	(0.0230)	(0.0254)	(0.0228)	(0.0294)	(0.0236)
Gender	-0.0243	-0.0131	-0.0252	-0.00702	-0.0251*	-0.0139	-0.0268	-0.00838
	(0.0148)	(0.0144)	(0.0165)	(0.0138)	(0.0145)	(0.0138)	(0.0160)	(0.0134)
School Fixed Effect					0.0368	0.0487***	0.0439	0.0315*
					(0.0230)	(0.0163)	(0.0263)	(0.0187)
Constant	0.664***	0.670***	0.691***	0.574***	0.659***	0.654***	0.704***	0.587***
	(0.0342)	(0.0226)	(0.0466)	(0.0432)	(0.0340)	(0.0210)	(0.0452)	(0.0455)
Observations	499	499	404	404	499	499	404	404
R-squared	0.128	0.180	0.090	0.206	0.137	0.196	0.100	0.211

Table 5: Consideration, age, parental education, and cognitive abilities.

The dependent variable is $\hat{\gamma}_i$ being the average between the estimated consideration parameters using the questions about pencils and pens. 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, which is 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.[36] 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. Specification (4) shows that each additional year of schooling has an average effect of 0.0376 (p-value < 0.01) which is equivalent to two additional brackets in parental educational attainment; while the effect of the performance in the Raven test becomes non-significant and its magnitude reduces considerably. The analysis of individual grades shows a more or less linear development with improvements that are included between 0.0678 (1st to 2nd grade) and 0.025 (2nd to 3rd grade). Finally, specification (8) shows that parental educational attainment has a smaller (roughly a third of the effect of a year of schooling) and less significant effect once accounting for the school-fixed effect. In this respect, even if we cannot disentangle the effect of schools, i.e. peer effects, teachers' quality, neighbourhood quality, etc...; we find that overall it is comparable to approximately one school grade.

6.2 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 interested mostly 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 19. 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 20 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 pen/pencils task, apart from dropping preference intensities (not available for this task due to its different nature) and introducing, in equation 20, 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.

$$\log\left(\frac{p(\ell_7^I|\mathbf{X}_{i,s},\mathbf{Z}_{i,s},\rho_s)}{p(\ell_6^I+\ell_5^I|\mathbf{X}_{i,s},\mathbf{Z}_{i,s},\rho_s)}\right)=\alpha+\beta_1\mathbf{X}_{i,s}+\beta_2\mathbf{Z}_{i,s}+\rho_s+\varepsilon_{i,s} \tag{19}$$

$$\log\left(\frac{p(\ell_7^C|\mathbf{X}_{i,s},\mathbf{Z}_{i,s},\rho_s)}{p(\ell_5^C|\mathbf{X}_{i,s},\mathbf{Z}_{i,s},\rho_s)}\right)=\alpha+\beta_1\mathbf{X}_{i,s}+\beta_2\mathbf{Z}_{i,s}+\rho_s+\varepsilon_{i,s} \tag{20}$$

[36] In the Appendix, we further analyse the relationship between gender and our estimates showing that despite boys and girls have substantially different preferences, their error rates develop similarly.

Results are reported in Tables 6 and 7. 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 6 and p-value < 0.01 in Table 7). The magnitude of these coefficients is large with the odds ratio increasing about fourfold between pupils of third and fifth grade. As mentioned in section 2, 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.[37] Finally, in Table 7, we document the only significant correlation between gender and one of our measures. The odds ratio is about 50% lower for girls, an effect that can be speculatively connected to differences in risk preferences influencing the probability of considering ℓ_C^2.[38]

	(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 6: 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 5 apart from preference intensity. The regression models are estimated in Stata using the function *logit*. *** <0.01, ** <0.05, *<0.1.

[37] See appendix C for the analysis when 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. In the Online Appendix, we also run the same regression excluding pupils who failed the independent task and find similar correlation patterns.

[38] In the appendix B.2, we document a significant difference in risk preferences between boys and girls similar to previous studies, both using the standard bomb-like task **?** and the coin-drop task).

	(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 7: 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 20. The independent variables are the same as in Table 5 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.

7 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 the case for performance in the Raven tests. Figure 11 depicts the Raven score by grade, normalised to the [0,1] interval. 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 (unpaired t-test, p-value < 0.01). We also document a more non-linear development by grade with the majority of the effect happening between 2nd and 3rd grade.

Finally, the effect of the parental educational attainment, similar to previous results, 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 larger than a year of schooling. This result mirrors the divergent path reported in Figure 11.

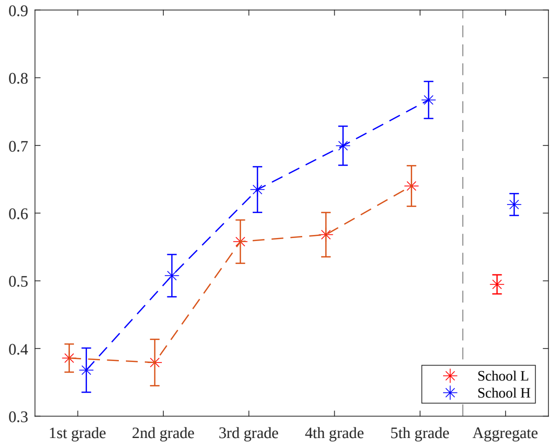

Figure 11: Normalized Raven's scores by grades and schools.

	(1)	(2)	(3)	(4)	(5)	(6)
2nd Grade	0.0958***		0.0922**	0.0670**		0.0743**
	(0.0313)		(0.0344)	(0.0264)		(0.0335)
3rd Grade	0.235***		0.240***	0.213***		0.225***
	(0.0312)		(0.0327)	(0.0333)		(0.0364)
4th Grade	0.271***		0.269***	0.254***		0.260***
	(0.0331)		(0.0366)	(0.0292)		(0.0357)
5th Grade	0.335***		0.345***	0.322***		0.335***
	(0.0330)		(0.0314)	(0.0278)		(0.0335)
Parental education		0.0383***	0.0373***		0.0191*	0.0211**
		(0.0130)	(0.00981)		(0.0110)	(0.0101)
Gender	-0.000966	-0.0370	-0.00161	-0.00267	-0.0406	-0.00506
	(0.0219)	(0.0273)	(0.0207)	(0.0199)	(0.0263)	(0.0198)
School Fixed Effect				0.0963***	0.0904**	0.0764***
				(0.0174)	(0.0432)	(0.0208)
Constant	0.383***	0.452***	0.262***	0.340***	0.470***	0.280***
	(0.0229)	(0.0544)	(0.0429)	(0.0239)	(0.0499)	(0.0426)
Observations	499	404	404	499	404	404
R-squared	0.233	0.037	0.268	0.269	0.061	0.285

Table 8: Raven's scores, age, and parental education.

The dependent variable is normalized - 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 the improvement of the degree of at least one of the parents. Grade can be interpreted as 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. The dependent variable is normalized - in [0,1] - Raven's scores. The independent variables are normalized such that the constant represents the consideration parameter of a male pupil whose parents have no education, who is in the 1st grade, and who is in **School L**. The coefficient of parental education is the effect of the improvement of the degree of at least one of the parents. Grade can be interpreted as 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.

8 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. To do this we go beyond binary choices in two domains, eliciting the full choice function over deterministic choices, and studying choice behaviour over lotteries. The latter requires the introduction of a new design to document for children a phenomenon, correlation neglect, so far only studied in adults. 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. Our results are well captured by a model of stochastic consideration for these children: their mistakes are compatible with limited consideration, that is failure to consider all the available options. - however, our experiment is rich enough to highlight the shape that potential interventions could take, operating both on increasing children's consideration capacity. Different socioeconomic backgrounds seem to also impact the ability to avoid errors significantly, though the gap between children does tend to close over time - whether this is due to schooling or growing is not possible to ascertain, from our data, but warrants further investigation.

A Proof of Proposition 1

Following AMR (?), we construct three sub-collections $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$ comprising, respectively, the grand set, the collection of all binary sets, and the collection of all ternary sets. These enable us to identify all 24 types.[39] 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 this, from each such matrix, we extract a 24×24 minor that has 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 and 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.

More formally, let $\mathcal{A}_1 = \{X\}$, $\mathcal{A}_2 = \{A \subset X : |A| = 2\}$, and $\mathcal{A}_3 = \{A \subset X : |A| = 2\}$, and correspondingly construct the restriction of each type conditional mixture choice function to the collection $\mathcal{A}_i$, with $i = 1, 2, 3$. For collection $\mathcal{A}_1$ first, the restriction collapses to the four possible choice functions corresponding to $C(X)$, hence at most the Kruskal[40] 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

$$\begin{bmatrix} a & ab & ab^2 & ab^3 \\ ab & ab^2 & ab^3 & a \\ ab^2 & ab^3 & a & ab \\ ab^3 & a & ab & ab^2 \end{bmatrix} \tag{21}$$

where

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

That is, b is the ratio in type conditional choice probabilities between two consecutively ranked alternatives. 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. ?), it follows that minor (21) has a full rank, hence full Kruskal rank.

[41] Next consider $\mathcal{A}_2$, which collects all binary sets, and construct the 24×24 minor selecting, for each type, the unique row corresponding to the choice functions selecting

[39] Let $\kappa_i = \prod_{A \in \mathcal{A}_i} |A|$. Corollary 3 in AMR states that r types can be identified, up to label swapping, as long as $\sum_{i=1}^{3} \min\{\kappa_i, r\} \geq 2r + 2$.

[40] The Kruskal a matrix is the largest number k such that every set of k rows is linearly independent.

[41] 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 (21) is

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

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, 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, hence 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 choice function that, for each preference type, selects 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.

B Additional results

B.1 Preference distribution and consideration parameters by gender

Here we report the estimates disaggregated by grade and gender in Figure B.1, together with the distribution of preferences, where again we estimate the consideration parameter separately by class and gender. The top-left and bottom-left panels display the estimates regarding the choices between pencils and pens. A similar procedure as for Figure 7 has been followed to estimate the standard errors. The top-right and bottom-right panels display the estimates of the preference distribution for pencils and pens.

We show that in none of the grades we find a significant difference in the consideration parameter between females and males with both displaying a similar development path. However, on the right panels, we show the distribution of the preferences from the sub-populations of females and males aggregated by grades. Preferences on the x-axis are shortened such that, for instance, $D \succ L \succ F \succ S$ reads $DLFS$. Although

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.

females and males have similar consideration parameters, they display completely different preferences with the former showing more interest in the *Ladybug*, *Duck* pencils and the *Yellow, Red* pens while the latter showing more interest in the *Shark* pencil and the *Blue* pen.

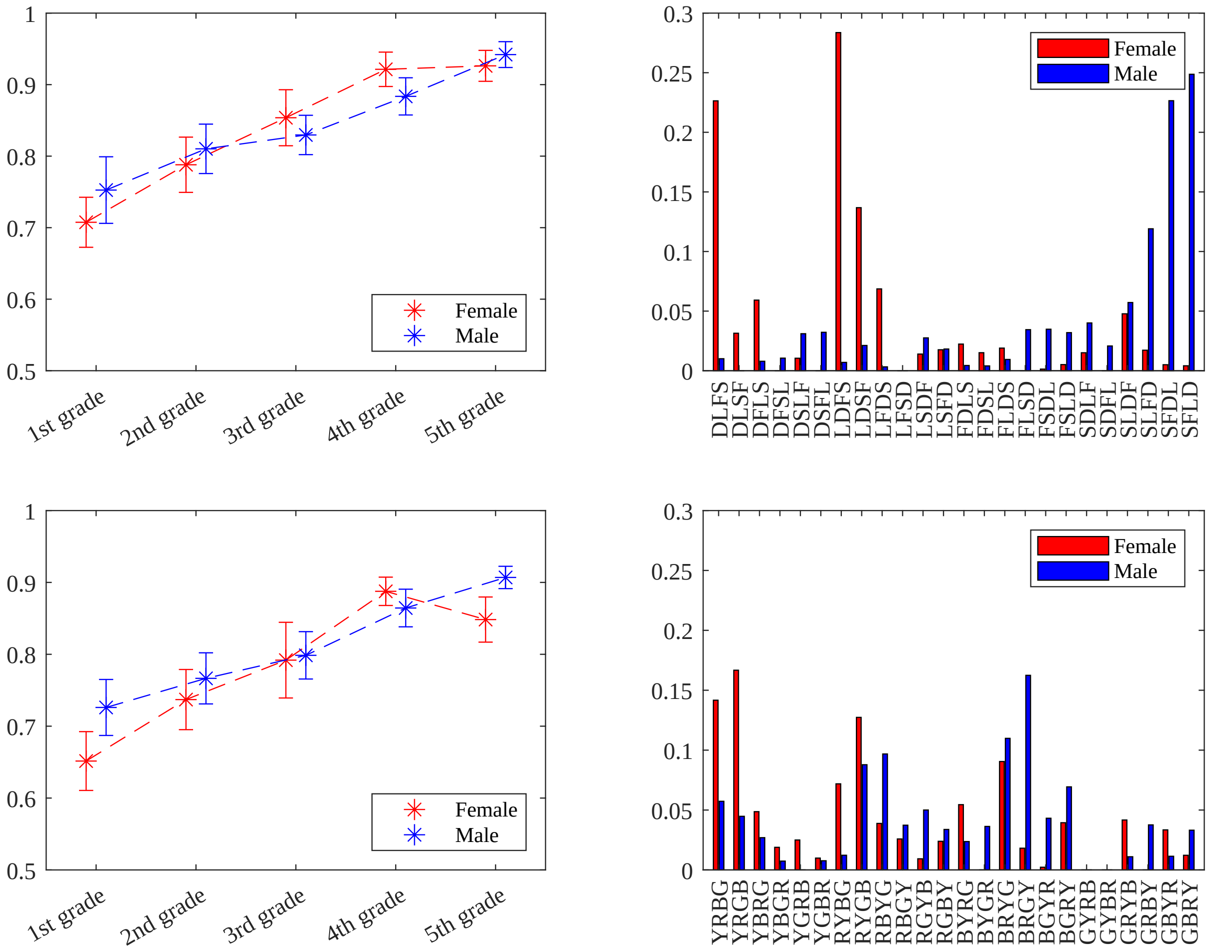

Figure 12: Estimated consideration parameter and preference distribution by grade and gender. Pencils at the topSchool H at the top

B.2 Preference intensity

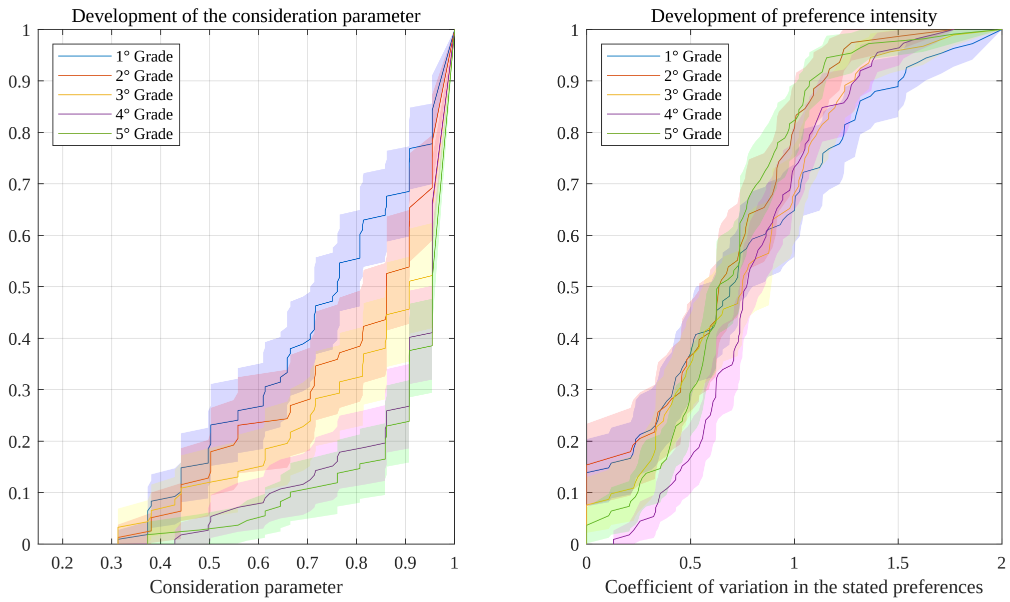

Figure 13: Evolution of the Consideration Parameter and Preference Intensity.

Notes: The left panel shows the cumulative distributions of $\hat{\gamma}$ 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 13, we plot the cumulative distribution of the estimated individual consideration parameter $\hat{\gamma}$ by age. The distributions of $\hat{\gamma}$ show a clear order of stochastic dominance. In the right panel of Figure 13, 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, $p < 0.001$. Note that, for instance, only rarely do older pupils like all goods equally (3.7% in the 5th grade), while this behaviour is widespread among younger pupils (13.9% in the 1st grade).

B.3 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 (**?**). This task was performed using eight cards (see Figure 14), 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 14: Cards for the Bomb task to measure risk preferences.

Figure 15 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 (**?**, **?**).[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).

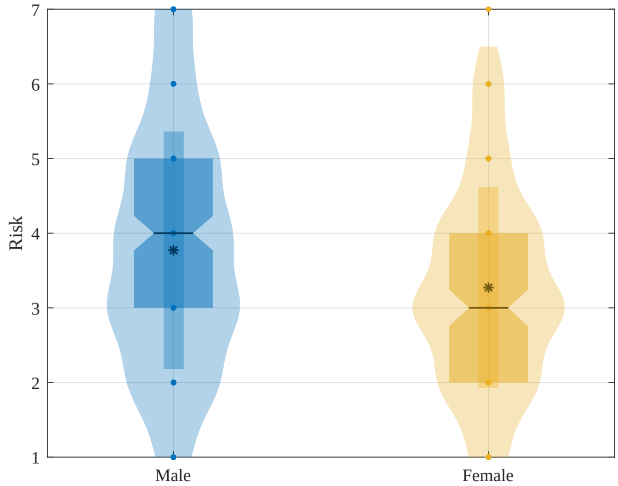

Figure 15: Risk by gender.

B.4 Coin-drop task with undominated lotteries

In Fall 2022, we collected data on a more complex version of the coin-drop task (see Figure 16). 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 3.2, 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.

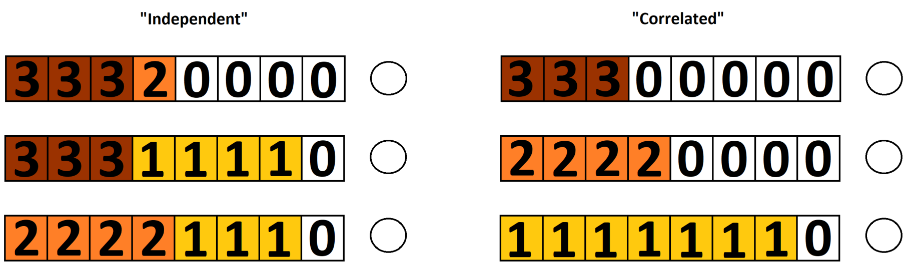

Figure 16: The answer sheet in the two tasks.

We randomized the pupils into the two tasks (between-subject design). Figure 17 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).

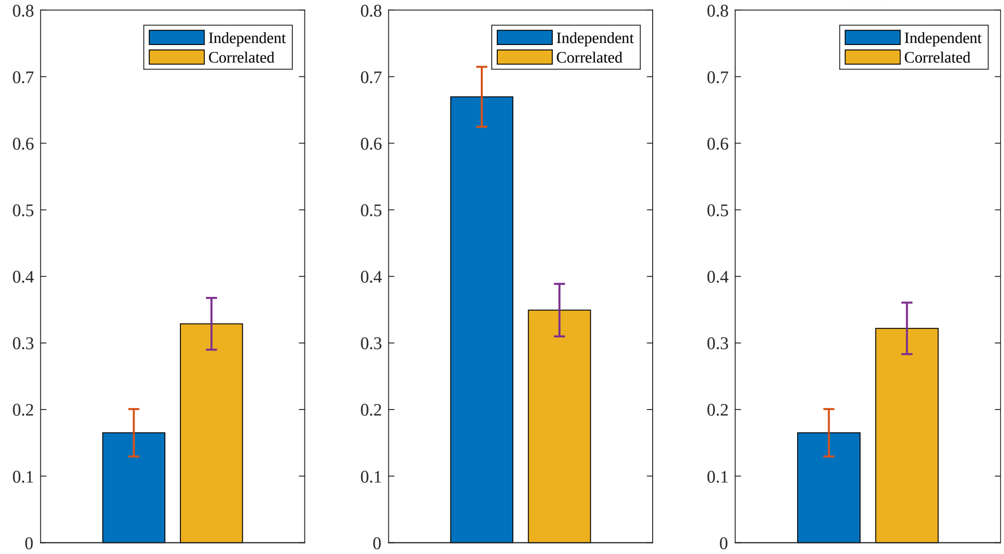

Figure 17: Distribution of choices among the three lotteries in the independent and correlated tasks.

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 18, we show that the pupils who choose ℓ_1^I are significantly more risk-loving than the remaining pupils (Wilcoxon rank-sum test, p-value < 0.03 and p-value < 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 15 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 females did so.

[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.

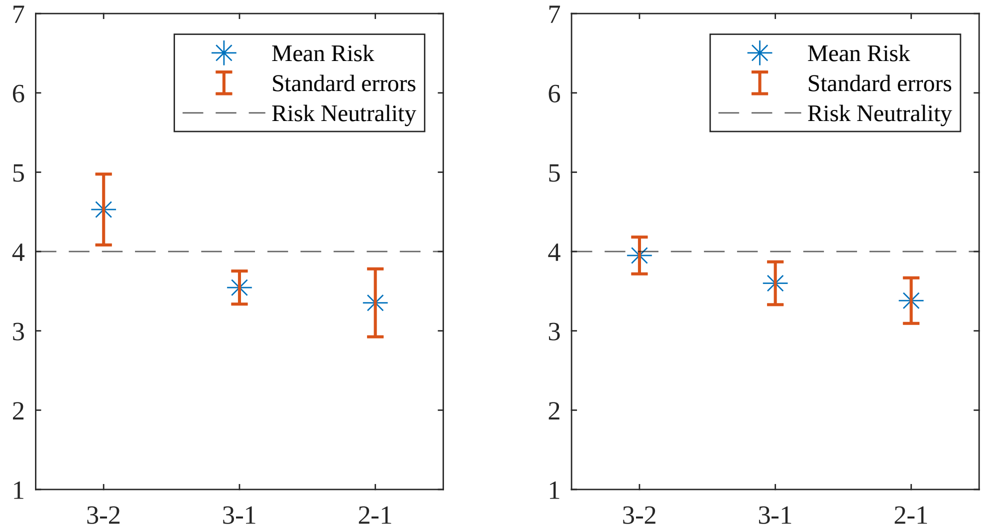

Figure 18: Number of cards chosen broken down by the choices in the coin-drop task.

C Replication using the entire set of schools

We replicate Figures 6, 7, 8, 9, 10, 11, and Tables 5, 7 including all schools. To achieve a critical mass of children we aggregate the remaining two schools as **Schools M**.

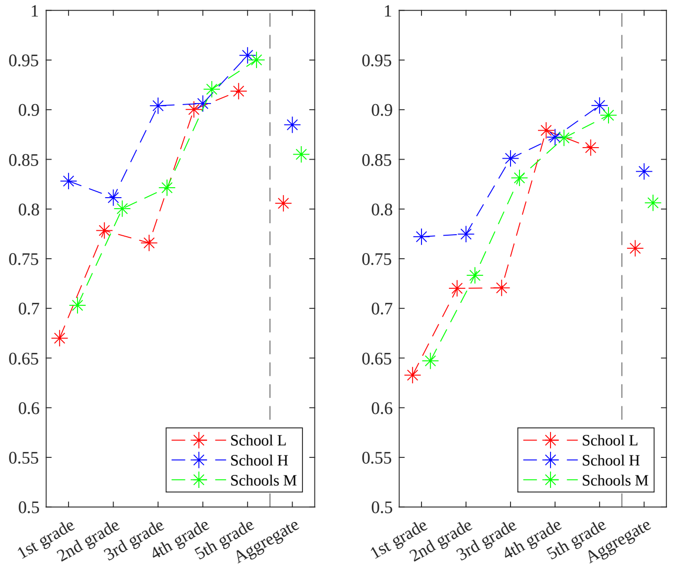

Figure 19: Consideration parameter γ by school and grade

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*.

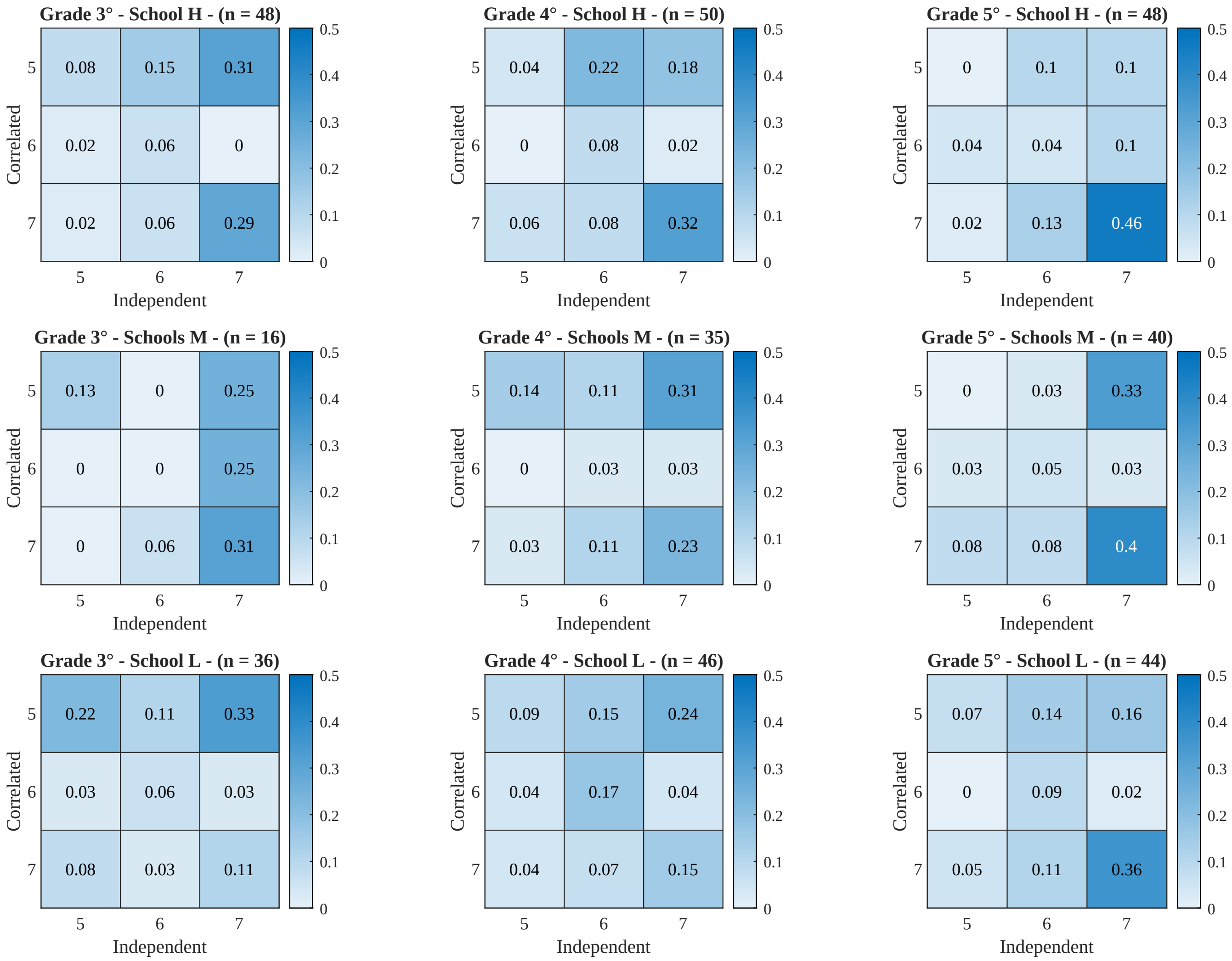

Figure 20: Joint distribution of the "risky choices" by school and grade

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.

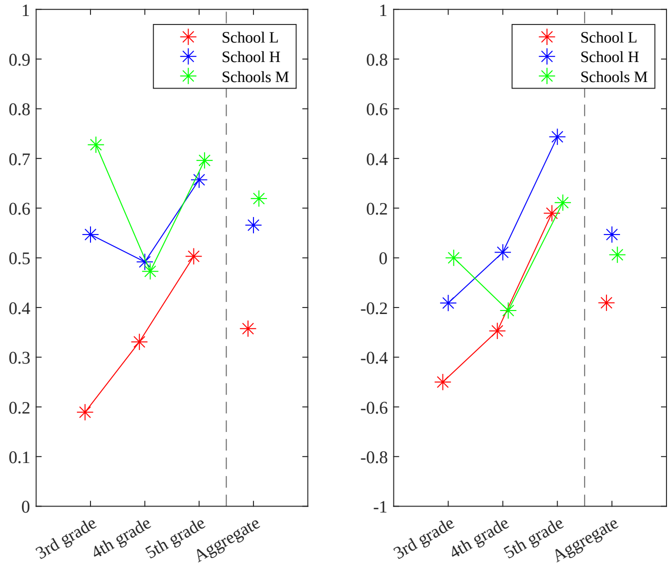

Figure 21: Development of the consideration parameter in choice with risk.

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.

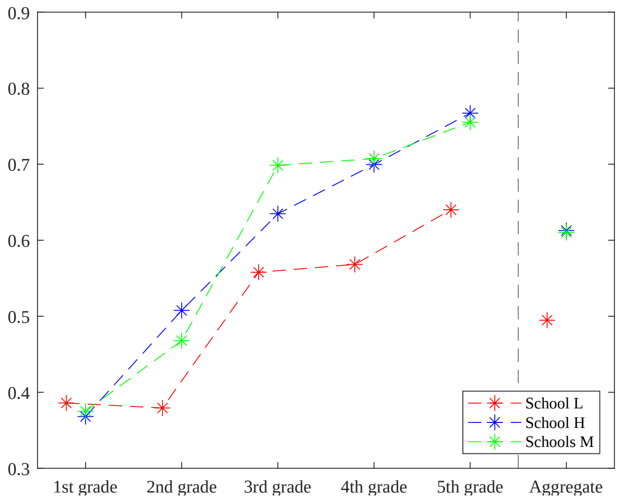

Figure 22: Normalized Raven's scores by grades and schools.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	0.209***			0.0690*	0.196***			0.0581
	(0.0351)			(0.0385)	(0.0362)			(0.0395)
Grade		0.0497***		0.0418***		0.0488***		0.0422***
		(0.00461)		(0.00492)		(0.00447)		(0.00477)
Parental education			0.0248***	0.0211***			0.0178**	0.0166**
			(0.00834)	(0.00632)			(0.00820)	(0.00647)
Constant	0.677***	0.684***	0.719***	0.591***	0.667***	0.661***	0.716***	0.593***
	(0.0304)	(0.0196)	(0.0375)	(0.0357)	(0.0306)	(0.0178)	(0.0385)	(0.0366)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
School FE	×	×	×	×	✓	✓	✓	✓
Observations	676	676	550	550	676	676	550	550
R-squared	0.113	0.184	0.062	0.207	0.119	0.195	0.073	0.211

Table 9: Consideration parameter, age, parental education, and cognitive abilities.

Notes: The dependent variable is the estimated consideration parameter γ. 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 pupil. Grade can be interpreted as one-year age-effect. Preference intensity takes values between 0 and 2. Gender is 0 for male and 1 for female. School is 0 for **School L** and 1 for **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	1.560***			1.434***	1.394***			1.293**
	(0.526)			(0.505)	(0.520)			(0.527)
Grade		0.580***		0.512***		0.581***		0.515***
		(0.178)		(0.178)		(0.169)		(0.176)
Parental education			0.10	0.12			0.03	0.07
			(0.106)	(0.112)			(0.105)	(0.109)
Constant	-1.246***	-0.864***	-0.539	-2.141***	-1.317***	-1.125***	-0.583	-2.083***
	(0.423)	(0.314)	(0.404)	(0.567)	(0.468)	(0.410)	(0.421)	(0.551)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
School FE	×	×	×	×	✓	✓	✓	✓
Observations	363	363	293	293	363	363	293	293
Log-likelihood	-361.25	-339.58	-276.75	-267.16	-342.05	-337.55	-275.27	-266.26
AIC	701.97	695.17	569.49	558.32	704.11	695.11	570.53	560.51
BIC	733.13	726.32	598.94	602.49	743.05	734.05	607.34	612.04

Table 10: 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 20. The independent variables are the same as in Table 5 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 Screenshots and Instructions

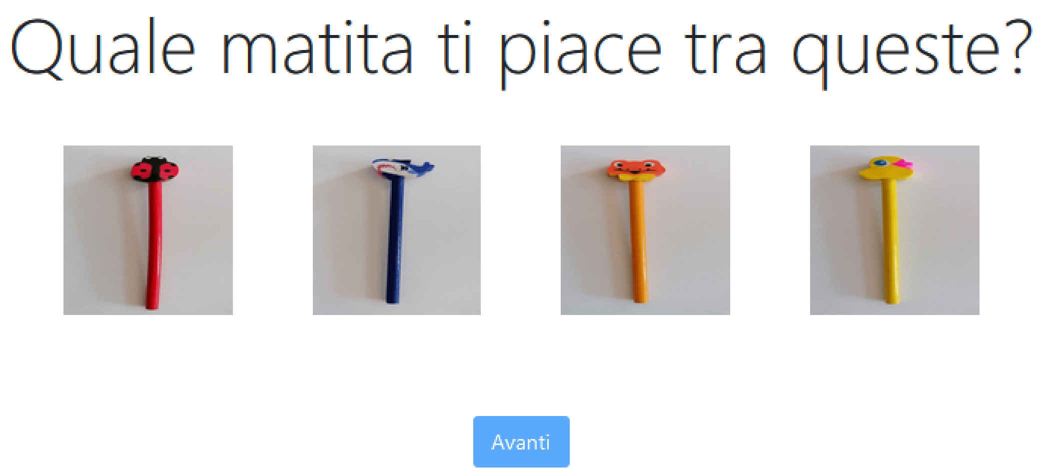

Figure 23: Screenshot for the problem in which all the pencils were available.

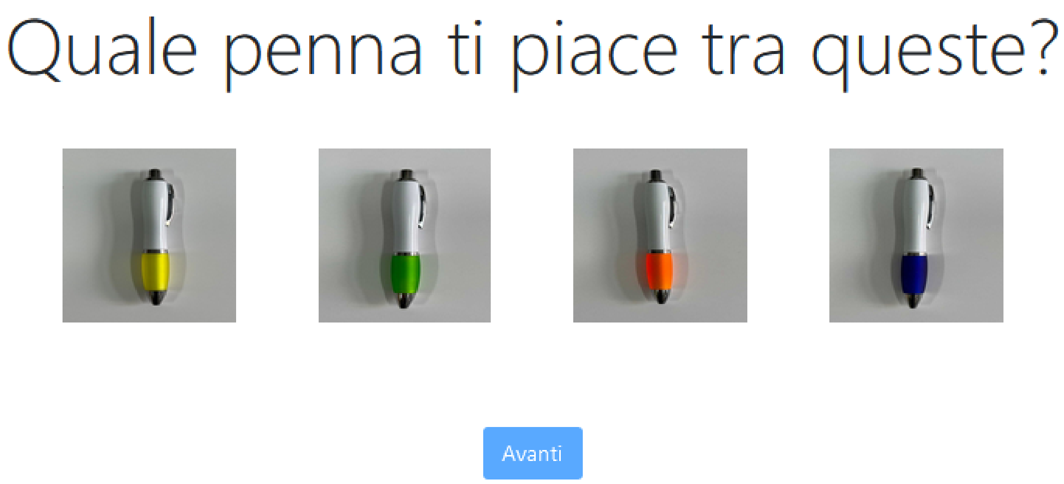

Figure 24: Screenshot for the problem in which all the pens were available.

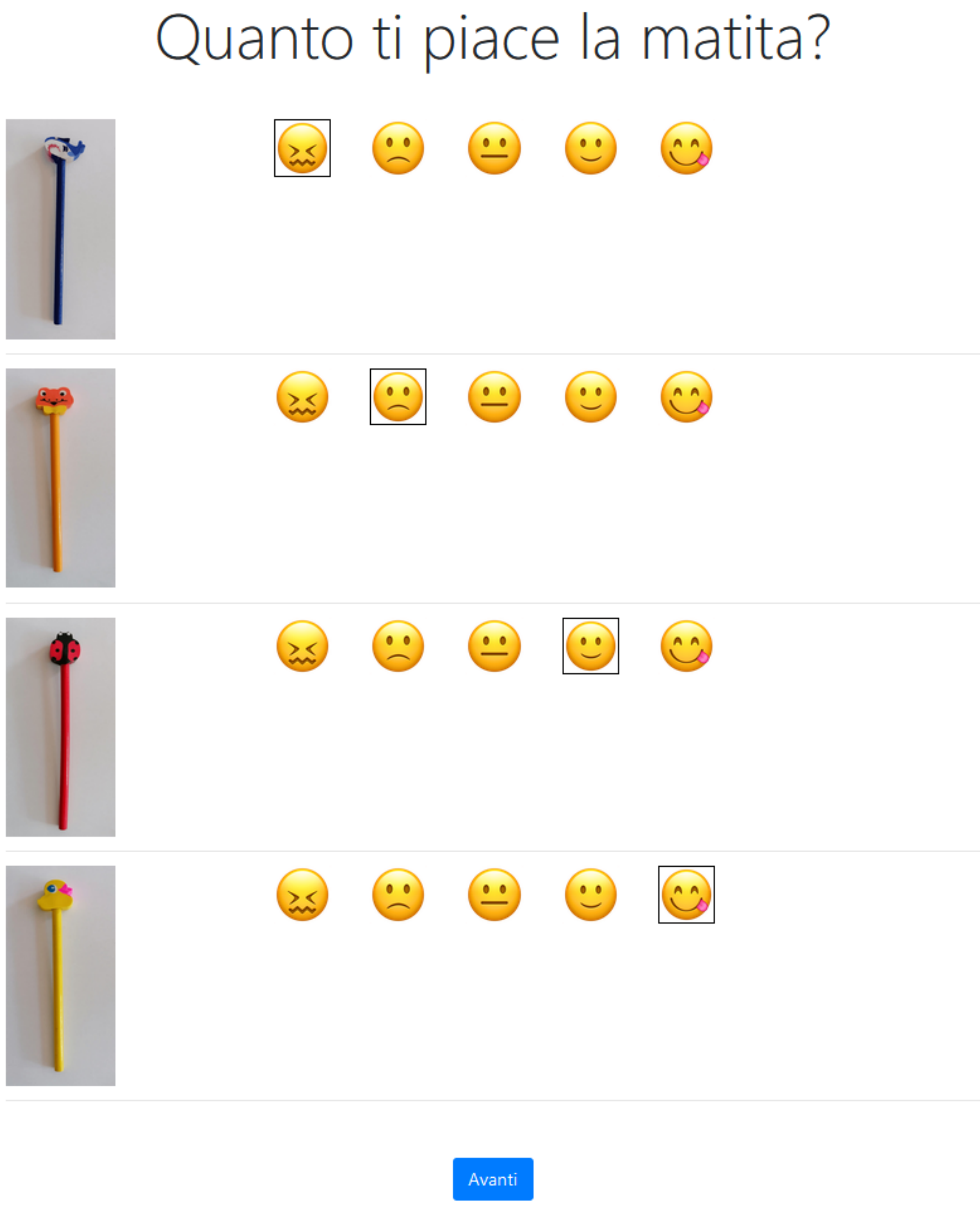

Figure 25: Screenshot for the Likert-scale task for pencils.

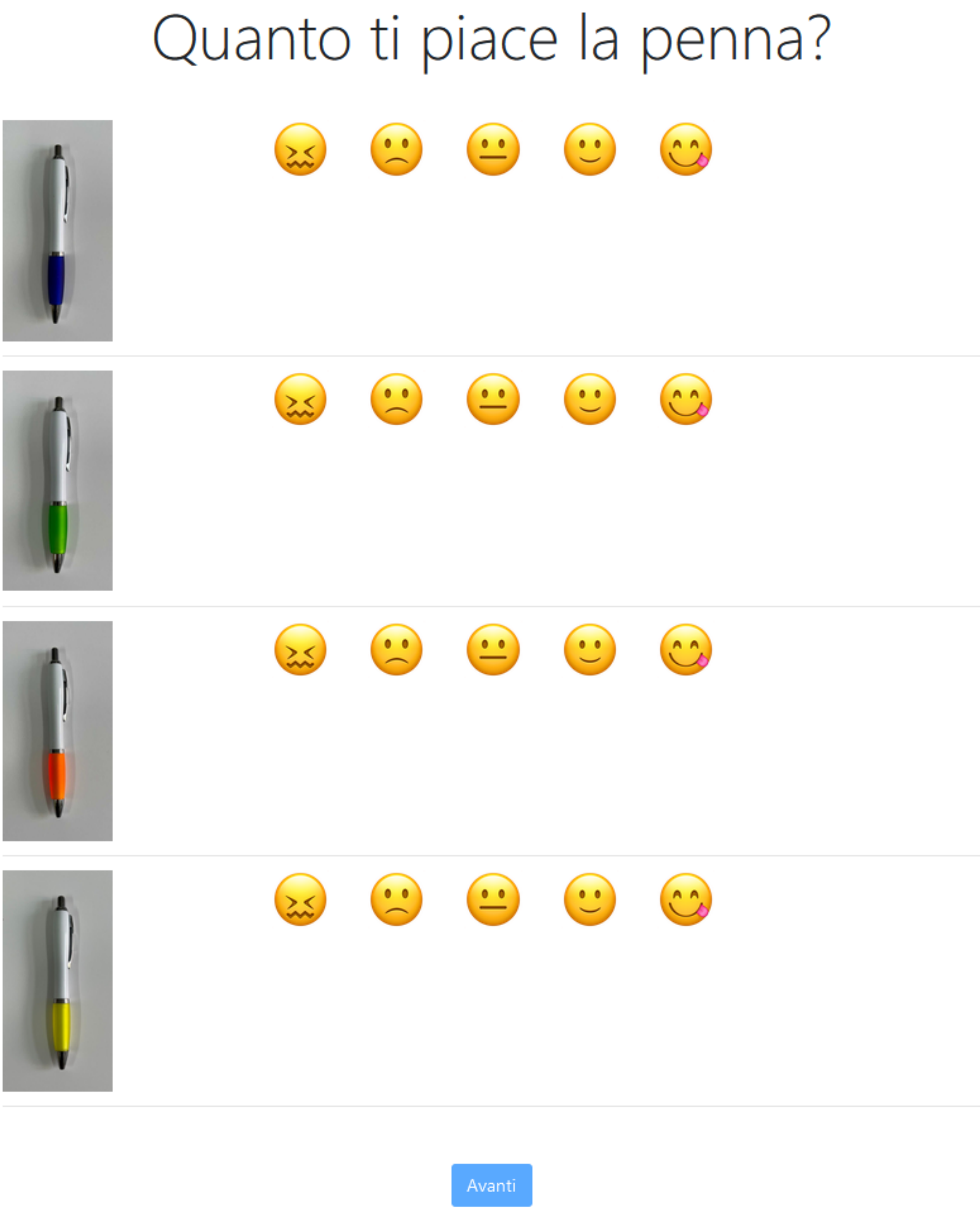

Figure 26: Screenshot for the Likert-scale task for pens.

Figure 27: Screenshot for one of the Raven question.

Italiano

Per noi, ricercatori della Facoltà di Economia, è importante capire le preferenze di voi bambini, per poi analizzare gli spostamenti dei soldi legati alle vostre scelte. Quindi oggi ci darete una mano partecipando a questo esperimento e avrete delle scelte da compiere.

Durante l'esperimento vedrete un breve filmato della durata di pochi secondi, in cui appariranno animali di ogni sorta, come oche e galline. Vedrete che il video dura pochi secondi. Una volta terminato, comparirà una scelta da fare.

Vedrete due, tre o quattro penne di colori diversi, e il vostro compito sarà scegliere la penna che più vi piace, puntando il dito su di essa. Anche se vi verrà presentata più volte la stessa domanda, dovrete sempre indicare la penna del colore che preferite. Alla fine del gioco, avrete maggiori probabilità di ricevere la penna che avete scelto più spesso. Questi sono i colori delle penne disponibili.

Dopo le penne, avrete le matite, con un piccolo pupazzetto sulla parte superiore (per esempio una coccinella, una rana, ecc.). Anche in questo caso, dovrete puntare il dito sulla matita che vi piace di più, seguendo lo stesso criterio utilizzato per le penne.

Inglese

For us, researchers at the Faculty of Economics, it is important to understand the preferences of you children, to then analyze the movements of money related to your choices. So today you will give us a hand by participating in this experiment and you will have choices to make.

During the experiment you will see a short film lasting a few seconds, in which all sorts of animals will appear, such as geese and chickens. You will see that the video lasts a few seconds. Once it is finished, a choice to make will appear.

You will see two, three or four pens of different colors, and your task will be to choose the pen that you like the most, by pointing your finger at it. Even if you are asked the same question several times, you must always indicate the pen of the color you prefer. At the end of the game, you will have a greater chance of receiving the pen that you have chosen most often. These are the colors of the available pens.

After the pens, you will have the pencils, with a little doll on the top (for example a ladybug, a frog, etc.). Again, you will have to point your finger at the pencil that you like the most, following the same criteria used for the pens.

Figure 28: Instruction (read aloud) for the pencils/pens task.

Italiano:

Questo gioco consiste nel lanciare una biglia in quello che vedete qui, chiamato Pachinko, molto simile a quello che voi conoscete come flipper. La biglia cade in un imbuto e rimbalza su dei pioli, con la possibilità di finire in uno degli otto slot sottostanti.

Dovete scegliere tra le tre righe davanti a voi, la riga con cui volete partecipare, sapendo che se la biglia finisce in uno slot corrispondente a un riquadro giallo, vincerete una matita. Se finisce in un riquadro bianco, non vincerete nulla. La probabilità che la biglia vada in uno degli otto slot è uguale (somministratore indica gli slot).

A questo punto, segnate la vostra scelta mettendo una “X” nel cerchio corrispondente alla riga con cui volete partecipare. L'amministratore vi mostrerà i tre cerchi.

Adesso faremo un secondo gioco simile al primo, ma questa volta parteciperete con due righe invece di una. Avrete sempre la possibilità di vincere una matita. Potete scegliere combinazioni come la prima e la seconda riga, la prima e la terza, o la seconda e la terza. Ricordate: se la biglia finisce in un riquadro giallo di una delle righe selezionate, vincerete una matita; se finisce in corrispondenza di un riquadro bianco, non vincerete nulla; il quadrato giallo può essere quello della prima o della seconda riga che avete selezionato, non ha importanza.

Inglese:

This game involves launching a marble into what you see here, called Pachinko, which is very similar to what you know as a pinball machine. The marble falls into a funnel and bounces off pegs, with the possibility of landing in one of the eight slots below.

You need to choose from the three rows in front of you, the row you want to participate with, knowing that if the marble lands in a slot corresponding to a yellow square, you will win a pencil. If it lands in a white square, you will not win anything. The probability of the marble landing in one of the eight slots is equal (the administrator indicates the slots).

At this point, mark your choice by putting an "X" in the circle corresponding to the row you wish to participate with. The administrator will show you the three circles.

Now we will play a second game similar to the first, but this time you will participate with two rows instead of one. You will still have the chance to win a pencil. You can choose combinations like the first and second rows, the first and third, or the second and third. Remember: if the marble lands in a yellow square from either of the selected rows, you will win a pencil; if it lands in a white square, you won't win anything. The yellow square can be from either the first or the second row you selected; it doesn't matter.

Figure 29: Instruction (read aloud) for the coin-drop game.

E Online Appendix

E.1 Robustness check [1]: relaxing the uniqueness of γ.

In this section, we explore a different estimation approach. We discretize 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}. The estimation approach is equivalent to the one described in section 3 but now the outcome of the maximization is a vector of consideration parameters $\hat{\gamma}$ and the estimated population distribution $\hat{\pi}(\omega)$.

In Figure 30, 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 11 and 12, 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.

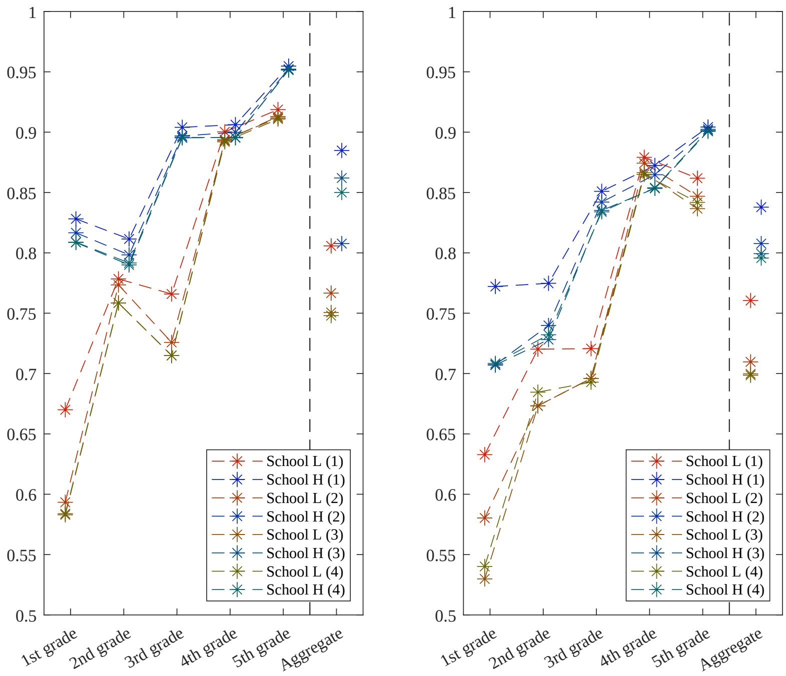

Figure 30: Consideration parameter γ by school and grade for all model specifications

Notes: Estimates of $E(\hat{\gamma})$ in sub-populations characterized 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 - 2nd grade	0.78	-221.51	493.02	527.21	0.77	-200.06	500.12	568.48
School L - 3rd grade	0.77	-325.36	700.73	743.57	0.73	-293.41	686.83	772.51
School L - 4th grade	0.90	-318.70	687.40	737.58	0.89	-292.76	685.53	785.89
School L - 5th grade	0.92	-294.14	638.27	689.78	0.91	-258.08	616.16	719.18
School L - aggregate	0.81	-1882.03	3814.06	3902.60	0.77	-1711.11	3522.21	3699.28
School H - 1st grade	0.83	-244.54	539.07	578.66	0.82	-218.45	536.89	616.07
School H - 2nd grade	0.81	-350.17	750.33	797.63	0.80	-324.61	749.23	843.82
School H - 3rd grade	0.90	-281.58	613.15	661.45	0.90	-257.07	614.14	710.73
School H - 4th grade	0.91	-326.01	702.01	753.09	0.90	-296.43	692.85	795.01
School H - 5th grade	0.95	-226.48	502.96	551.26	0.95	-207.08	514.16	610.75
School H - aggregate	0.88	-1497.64	3045.28	3132.71	0.81	-1392.36	2884.72	3059.58

	Three values of γ				Four values of γ			
	$E(\gamma)$	$\log \mathcal{L}$	AIC	BIC	$E(\gamma)$	$\log \mathcal{L}$	AIC	BIC
School L - 1st grade	0.58	-577.42	1304.85	1475.60	0.58	-577.42	1354.84	1582.51
School L - 2nd grade	0.76	-195.13	540.25	642.80	0.76	-195.13	590.25	726.98
School L - 3rd grade	0.71	-289.77	729.53	858.05	0.72	-289.77	779.53	950.89
School L - 4th grade	0.89	-287.61	725.21	875.76	0.89	-286.19	772.37	973.10
School L - 5th grade	0.91	-253.95	657.89	812.42	0.91	-253.54	707.09	913.13
School L - aggregate	0.75	-1690.64	3531.29	3796.88	0.75	-1690.42	3580.83	3934.96
School H - 1st grade	0.81	-214.63	579.27	698.03	0.81	-214.01	628.02	786.37
School H - 2nd grade	0.79	-321.77	793.55	935.43	0.79	-319.87	839.73	1028.92
School H - 3rd grade	0.90	-252.18	654.36	799.24	0.90	-252.16	704.33	897.51
School H - 4th grade	0.90	-290.59	731.17	884.40	0.90	-290.55	781.10	985.40
School H - 5th grade	0.95	-201.70	553.40	698.28	0.95	-201.68	603.36	796.54
School H - aggregate	0.86	-1366.42	2882.83	3145.12	0.85	-1363.32	2926.64	3276.36

Table 11: 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 - 2nd grade	0.72	-236.46	522.92	557.10	0.67	-223.84	547.69	616.05
School L - 3rd grade	0.72	-334.13	718.26	761.10	0.70	-310.61	721.23	806.91
School L - 4th grade	0.88	-344.25	738.49	788.68	0.87	-321.09	742.17	842.54
School L - 5th grade	0.86	-361.62	773.25	824.76	0.85	-334.96	769.93	872.95
School L - aggregate	0.76	-2022.58	4095.16	4183.69	0.71	-1905.44	3910.88	4087.94
School H - 1st grade	0.77	-280.79	611.58	651.16	0.71	-255.87	611.73	690.91
School H - 2nd grade	0.77	-376.86	803.72	851.01	0.74	-347.94	795.88	890.47
School H - 3rd grade	0.85	-335.88	721.77	770.06	0.84	-298.36	696.72	793.32
School H - 4th grade	0.87	-356.49	762.97	814.05	0.86	-327.45	754.91	857.06
School H - 5th grade	0.90	-293.21	636.41	684.71	0.90	-273.19	646.39	742.98
School H - aggregate	0.84	-1719.38	3488.75	3576.18	0.81	-1593.35	3286.70	3461.56

	Three values of γ				Four values of γ			
	$E(\gamma)$	$\log \mathcal{L}$	AIC	BIC	$E(\gamma)$	$\log \mathcal{L}$	AIC	BIC
School L - 1st grade	0.53	-611.05	1372.09	1542.84	0.54	-610.03	1420.05	1647.72
School L - 2nd grade	0.67	-223.84	597.69	700.24	0.68	-218.45	636.89	773.62
School L - 3rd grade	0.70	-310.61	771.23	899.75	0.69	-308.82	817.64	988.99
School L - 4th grade	0.87	-315.91	781.83	932.38	0.86	-314.07	828.13	1028.87
School L - 5th grade	0.84	-327.45	804.90	959.43	0.84	-326.42	852.84	1058.89
School L - aggregate	0.70	-1894.80	3939.61	4205.20	0.70	-1892.51	3985.03	4339.15
School H - 1st grade	0.71	-249.26	648.51	767.28	0.71	-249.23	698.46	856.82
School H - 2nd grade	0.73	-343.31	836.61	978.50	0.73	-342.11	884.23	1073.41
School H - 3rd grade	0.83	-294.84	739.68	884.56	0.83	-294.50	789.00	982.19
School H - 4th grade	0.85	-322.79	795.59	948.81	0.85	-322.67	845.34	1049.65
School H - 5th grade	0.90	-271.45	692.90	837.79	0.90	-269.05	738.10	931.28
School H - aggregate	0.80	-1581.06	3312.13	3574.42	0.80	-1578.31	3356.63	3706.35

Table 12: Model selection for different specifications of the model in the pens questions.

E.2 Robustness check [2]: individual estimates via Bayes' Theorem

In section 5, 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 the application of Bayes' theorem to our population estimates.

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 11 and 12 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 22.

$$\hat{\pi}_i(\omega) = \frac{p_i^{\omega}(C_i)\hat{\pi}(\omega)}{\sum_{\omega} p_i^{\omega}(C_i)\hat{\pi}(\omega)} \tag{22}$$

Then as point-estimate of the individual consideration parameter, $\hat{\gamma}_i$, we take its expectation as shown in equation 23.

$$\hat{\gamma}_i = \sum\nolimits_{j\in\{L,M,H\}} \left(\sum_{\succ} \hat{\gamma}_j \hat{\pi}_i(\omega) \right) \tag{23}$$

We replicate Table 5 in Section 5 with very similar results.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	0.236***			0.0464	0.208***			0.0289
	(0.0521)			(0.0528)	(0.0487)			(0.0543)
Grade		0.0655***		0.0648***		0.0639***		0.0653***
		(0.00936)		(0.00937)		(0.00751)		(0.00857)
Parental education			0.0229*	0.0203**			0.00787	0.00878
			(0.0128)	(0.00922)			(0.0113)	(0.00946)
Constant	0.511***	0.497***	0.566***	0.402***	0.503***	0.472***	0.587***	0.425***
	(0.0461)	(0.0332)	(0.0684)	(0.0614)	(0.0465)	(0.0319)	(0.0636)	(0.0603)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
School FE	×	×	×	×	✓	✓	✓	✓
Observations	499	499	404	404	499	499	404	404
R-squared	0.125	0.224	0.074	0.252	0.144	0.252	0.092	0.264

Table 13: Consideration, age, parental education, and cognitive abilities.

Notes: The dependent variable is $\hat{\gamma}_i$ estimated as in equation 23 being the average between the estimated consideration parameters using the questions about pencils and pens. 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 **??**. The regression models are estimated in Stata (Robust clustered standard errors at the session level). *** <0.01, ** <0.05, *<0.1.

E.3 Robustness check [3]: 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	1.125***			0.520**	1.044***			0.420
	(0.196)			(0.257)	(0.200)			(0.260)
Grade		0.232***		0.183***		0.233***		0.194***
		(0.0339)		(0.0445)		(0.0336)		(0.0445)
Parental education			0.0946**	0.0869*			0.0310	0.0294
			(0.0465)	(0.0476)			(0.0544)	(0.0548)
Constant	0.953***	1.027***	1.215***	0.591***	0.926***	0.925***	1.320***	0.714***
	(0.152)	(0.134)	(0.202)	(0.220)	(0.152)	(0.136)	(0.207)	(0.227)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
School FE	✗	✗	✗	✗	✓	✓	✓	✓
Observations	499	499	404	404	499	499	404	404

Table 14: Consideration, age, parental education, and cognitive abilities (Beta-regression).

Notes: The dependent variable is $\hat{\gamma}_i$ estimated as in equation 23 being the average between the estimated consideration parameters using the questions about pencils and pens. 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 **??**. The regression models are estimated in Stata using the command *betareg* with a logistic transformation. *** <0.01, ** <0.05, *<0.1.

E.4 Robustness check [4]: considering data from May 2022.

As discussed in section 4, we started collecting our data on the pens/pencils 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. 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	0.236***			0.112***	0.212***			0.104**
	(0.0325)			(0.0382)	(0.0329)			(0.0407)
Grade		0.0511***		0.0403***		0.0482***		0.0394***
		(0.00569)		(0.00733)		(0.00535)		(0.00729)
Parental education			0.0323***	0.0245***			0.0184**	0.0170**
			(0.00975)	(0.00765)			(0.00858)	(0.00715)
Constant	0.611***	0.637***	0.626***	0.507***	0.606***	0.624***	0.648***	0.524***
	(0.0291)	(0.0218)	(0.0481)	(0.0331)	(0.0282)	(0.0207)	(0.0455)	(0.0349)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
School FE	×	×	×	×	✓	✓	✓	✓
Observations	528	528	427	427	528	528	427	427
R-squared	0.167	0.223	0.122	0.271	0.186	0.243	0.142	0.277

Table 15: Consideration, age, parental education, and cognitive abilities (including May 2022).

Notes: The dependent variable is $\hat{\gamma}_i$ estimated as in equation 23 being the average between the estimated consideration parameters using the questions about pencils and pens. 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 **??**. The regression models are estimated in Stata (Robust clustered standard errors at the session level). *** <0.01, ** <0.05, *<0.1.

E.5 Robustness check [5]: splitting pencils and pens.

In our main specification (section 5), we consider the mean consideration parameter estimated using choices among pencils and pens. Here, we run the same regressions separating the two domains.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	0.207***			0.0965**	0.190***			0.0873*
	(0.0408)			(0.0470)	(0.0404)			(0.0488)
Grade		0.0452***		0.0353***		0.0442***		0.0356***
		(0.00571)		(0.00691)		(0.00530)		(0.00695)
Parental education			0.0265**	0.0226**			0.0171	0.0162*
			(0.0102)	(0.00853)			(0.0106)	(0.00927)
Constant	0.675***	0.691***	0.699***	0.582***	0.669***	0.673***	0.713***	0.595***
	(0.0366)	(0.0244)	(0.0451)	(0.0475)	(0.0366)	(0.0234)	(0.0442)	(0.0500)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
School FE	×	×	×	×	✓	✓	✓	✓
Observations	499	499	404	404	499	499	404	404
R-squared	0.120	0.156	0.078	0.180	0.129	0.173	0.087	0.184

Table 16: Consideration, age, parental education, and cognitive abilities (pencils).

Notes: The dependent variable is $\hat{\gamma}_i$ estimated as in equation 23 being the average between the estimated consideration parameters using the questions about pencils and pens. The independent variables are normalized such that the constant represents the attention parameter of a male pupil whose parents have no education, who has scored the lowest in the Raven's test, who is in the 1st grade, who is indifferent between all alternatives, and who is in **School L**. The controls are gender and preference intensity *only for pencils*. 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 **??**. 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	0.176***			0.0574	0.160***			0.0467
	(0.0434)			(0.0517)	(0.0450)			(0.0539)
Grade		0.0451***		0.0398***		0.0441***		0.0401***
		(0.00658)		(0.00731)		(0.00617)		(0.00706)
Parental education			0.0201*	0.0173**			0.0105	0.0105
			(0.0103)	(0.00813)			(0.00978)	(0.00866)
Constant	0.670***	0.667***	0.699***	0.582***	0.664***	0.650***	0.711***	0.595***
	(0.0344)	(0.0239)	(0.0507)	(0.0458)	(0.0343)	(0.0227)	(0.0492)	(0.0478)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
School FE	×	×	×	×	✓	✓	✓	✓
Observations	499	499	404	404	499	499	404	404
R-squared	0.094	0.143	0.070	0.161	0.101	0.154	0.079	0.165

Table 17: Consideration, age, parental education, and cognitive abilities (pens).

Notes: The dependent variable is $\hat{\gamma}_i$ estimated as in equation 23 being the average between the estimated consideration parameters using the questions about pencils and pens. 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 *only for pens*. 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 **??**. The regression models are estimated in Stata (Robust clustered standard errors at the session level). *** <0.01, ** <0.05, *<0.1.

E.6 Robustness check [6]: 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_{25} (correlated task) are chosen and zero otherwise.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Raven's scores	0.248*			0.0852	0.206*			0.0505
	(0.122)			(0.175)	(0.119)			(0.175)
Grade		0.0310		0.0159		0.0336		0.0190
		(0.0617)		(0.0556)		(0.0596)		(0.0573)
Parental education			0.0609*	0.0603*			0.0420	0.0432
			(0.0302)	(0.0314)			(0.0357)	(0.0355)
Constant	0.399***	0.525***	0.361***	0.289**	0.379***	0.463***	0.371***	0.316**
	(0.0821)	(0.0735)	(0.117)	(0.133)	(0.0891)	(0.0970)	(0.111)	(0.124)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
School FE	×	×	×	×	✓	✓	✓	✓
Observations	272	272	217	217	272	272	217	217
R-squared	0.015	0.004	0.021	0.024	0.023	0.017	0.029	0.031

Table 18: 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 5 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**			0.250**	0.253**			0.229**
	(0.109)			(0.101)	(0.109)			(0.106)
Grade		0.118***		0.116**		0.121***		0.118**
		(0.0420)		(0.0432)		(0.0395)		(0.0425)
Parental education			0.0396	0.0436			0.0254	0.0332
			(0.0283)	(0.0283)			(0.0292)	(0.0282)
Constant	0.138	0.197**	0.190*	-0.106	0.121	0.137	0.200**	-0.0883
	(0.0848)	(0.0710)	(0.0948)	(0.126)	(0.0907)	(0.0825)	(0.0917)	(0.121)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
School FE	×	×	×	×	✓	✓	✓	✓
Observations	272	272	217	217	272	272	217	217
R-squared	0.098	0.115	0.108	0.165	0.106	0.131	0.112	0.168

Table 19: 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 5 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.

E.7 Robustness check [7]: 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 20: 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 20. The independent variables are the same as in Table 5 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 21: 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 20. The independent variables are the same as in Table 5 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.