Developmental Meritocracy *

Pablo Brañas Garza
Universidad de Loyola

Isabelle Brocas
University of Southern California and CEPR

Juan D. Carrillo
University of Southern California and CEPR

October 2023

Abstract

A sample of 2,883 children and teenagers aged 5 to 18 in the United States and Spain participate in a third party dictator game where they can influence the rewards of the winner and the loser of a competition. We explore three methods for determining the competition's winner (talent, effort, and luck) and two methods to allocate rewards (redistribution from winner to loser and addition to winner and loser). Participants are reluctant to transfer rewards from the winner to the loser, resulting in more meritocratic allocations in the redistribution scenario across all age groups. In contrast, the addition scenario exhibits strong age-related effects, with ex-post fairness peaking during middle school and then declining. These trends are consistent across populations and across socioeconomic backgrounds. Merit-based allocations are more frequent when participants compete in a task that requires talent. Allocation decisions are influenced by experience, which reduces the gap between redistribution and addition. They are also affected by perceived performance, personality traits, and affective processes.

Keywords: field experiment, developmental decision making, third-party dictator, inequality.
JEL Classification: C93, D63.

*We thank the members of the Los Angeles Behavioral Economics Laboratory (LABEL) and the Loyola Behavioral Lab for their insights and to Chris Crabbe for his outstanding programming skills. Special thanks to Pablo Montero, Monica Vasco and Javier Gonzalez for excellent research assistance. We also thank the editor and referees for their constructive comments. We are grateful to the staff of the Lycée International de Los Angeles, Thomas Starr King and Colegios Salesianos –in particular Emmanuelle Acker, Nordine Bouriche, Andrés Corbacho, Adriana Díaz, Anneli Harvey, Mathieu Mondange, Manuel Redondo and Monica Soliss– for their help and support in running the experiment in their schools. The study was conducted with the University of Southern California IRB approval UP-12-00528. We gratefully acknowledge the financial support of the National Science Foundation grant SES-2315770, the Spanish Ministry of Economy and Competitiveness (PID2021-126892NB-100), the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement number 101095175 and the UK Government.

1 Introduction

Redistributive politics is a cornerstone of policy debates. The preference for redistribution within a society is influenced by a myriad of factors including culture (Alesina and Giuliano, 2011) and political inclinations (Thorisdottir et al., 2007; Pontusson and Rueda, 2008). Nonetheless, three factors stand out when assessing an individual's inherent tolerance for inequality: the personal consequences of inequality, the source of that inequality, and the redistribution method. The experimental literature on dictator and ultimatum games has extensively documented instances of altruism, egalitarianism, envy as well as a willingness to punish unequal treatments (Engel, 2011; Oosterbeek et al., 2004; Fehr and Schmidt, 2006). What is perhaps more subtle is how inequalities stemming from luck, effort, and talent are perceived and accepted differentially. Those who attribute inequality primarily to luck tend to have a lower tolerance for it (Fong, 2001; Konow, 2000). These tendencies have been observed both in empirical studies and laboratory experiments.

Although preferences and redistributive concerns have been the object of substantial experimental research,[1] several important questions have been only partially explored. First, to what extent does an individual's intrinsic tolerance for inequality rely on the *origin* of that inequality? More precisely, what level of inequality are people wiling to accept (i) when we strip choices of trade-offs between self-interest, efficiency and fairness concerns and (ii) when we modify the reasons for inequality? Second, how does our attitude toward inequity change depending on the specific redistributive mechanisms in place, such as subsidies and taxes, cultural factors or economic status? Third, how does the tolerance to inequality develop from childhood to adulthood? Is it an innate trait, acquired over time, shaped by environmental factors or personal characteristics?

Some of these questions have been studied in the recent literature. Of special interest for our purpose is the research by the FAIR group in Norway. Cappelen et al. (2022) and Almås et al. (2020) task adult participants in the role of impartial observers to choose allocations for others. Cappelen et al. (2022) show that impartial observers tend to opt for a more equitable distribution when disparities in outcomes are more likely attributed to luck rather than productivity. Almås et al. (2020) extended this investigation by considering a large representative sample of adult Americans and Norwegians. They showed that a significant proportion of Americans perceive both luck-based and productivity-based inequalities as fair, whereas a substantial fraction of Norwegians regard both as unfair. A developmental approach is provided by Almås et al. (2010) who studied dictator choices among children and adolescents. They found that, as children age, they become more accepting of inequalities that arise from differences in performance. Taken together, these studies suggest that attitudes toward inequity change with age and are influenced by both

[1] Prominent examples include Charness and Rabin (2002) on tests to determine the nature of social preferences, Cappelen et al. (2013) on fairness views about risk-taking and Andreoni et al. (2020) on ex-ante vs. ex-post fairness.

the source of inequity and environmental factors.

The current study aims to explore the developmental trajectory of individuals' inherent attitudes toward inequality. It takes an experimental approach to scrutinize three pivotal factors influencing these attitudes: (i) the origin of inequality, such as luck, effort, and talent; (ii) the mechanisms for reducing inequality, such as subsidies and taxes; and (iii) the personal experience with the sources of inequality. Furthermore, we investigate how these attitudes are influenced by cultural and economic factors as well as individual traits. For this purpose, we recruited a large population of children and adolescents (as in Almås et al. (2010)). These participants play the role of impartial observers responsible for determining the allocation of resources among other players engaged in a competitive task (as in Cappelen et al. (2022) and Almås et al. (2020)).

Our study incorporates three competition tasks, each aimed at capturing different sources for inequality: *luck*, *effort* and *talent*, respectively. In each of these tasks, the player who emerges as the victor receives a higher payoff than the losing player. We have adopted two methods to ascertain these payoffs: some impartial observers are tasked with deciding how to allocate additional payoffs between the winner and the loser (referred to as our *addition* treatment), while others are assigned the responsibility of redistributing payoffs from the winner to the loser (referred to as our *redistribution* treatment). Impartial observers also participate as competitors, where another impartial observer is responsible for determining their earnings. In this setup, each player is required to select an allocation (as an outside observer) both before and after experiencing the competitive task (as a competitor). This manipulation enables us to gauge the impact of personal experience on individuals' attitudes toward inequality.

The project is organized in two separate experiments. In the first one, we recruited 465 children and teenagers ranging from 5 to 16 years old from middle- to high-income families and 138 teenagers aged 10 to 14 from very low- to low-income families in the United States. Participants in this experiment compete in the *effort* condition and allocate resources under the *addition* treatment. For the second experiment, we recruited 2,280 children and teenagers aged 6 to 18 years old from low- to middle- income families in Spain. We organized the children in groups, with one group replicating the first experiment. Meanwhile, the other groups were subjected to the five other combinations of conditions and treatments. Additionally, in this experiment, we introduced two novel child-friendly non-cognitive tasks: a pictorial personality questionnaire and a test designed to assess affective theory of mind.

We obtain the following insights. First, participants are reluctant to transfer rewards from the winner to the loser, resulting in more meritocratic allocations in the redistribution scenario than in the addition scenario across all age groups. Second, only the redistribution method exhibits age-related effects. Specifically, ex-post fairness peaks during middle school and declines thereafter. These trends are consistent across populations and socioeconomic backgrounds. Third, merit-based allocations are more frequent when participants compete in a task that requires talent, whereas we found no distinction between redistribution concerns under luck v. effort. Fourth, our study underscores the impact of experience and perceived performance on allocation decisions, substantially narrowing the gap between redistribution and addition scenarios. In summary, we tend to embrace inequity, respect talent, and generally prefer not to impose taxes on others from an early age. These tendencies appear to be minimally influenced by environmental factors such as family income or country of origin but are shaped by an individual's personal experiences. Last, allocation decisions are shaped by personality traits and affective processes, often aligning with trends observed in adults concerning their political inclinations.

The paper is organized as follows. In section 2, we present the design of experiments 1 and 2. In section 3, we discuss the main results, including the effect on choice of age, origin of inequality, allocation mechanism, and experience. In section 4, we analyze the influence of personality and affective theory of mind on choices. In section 5, we briefly compare the behavior of middle schoolers from different cultural and socioeconomic backgrounds. In section 6, we present some concluding remarks.

2 Experimental design

We recruit children and adolescents to study the decision of a third-party dictator –or impartial observer– who allocates rewards between two competitors. We adopt the guidelines presented in Brocas and Carrillo (2020a) to address the methodological challenges that arise in studies with young participants. We conduct two experiments using the same paradigm but different treatments, variants, procedures and locations. The designs are discussed separately.

2.1 Experiment 1. United States.

Population. Experiment 1 involves two groups of children. The main group, LILA, consists of 465 students from Kindergarten to 10th grade at Lycée International de Los Angeles (LILA), a French-English bilingual private school in Los Angeles. The second group, KING, consists of 138 middle school children (6th to 8th grade) from Thomas Starr King Middle School, a public school in Los Angeles. There are two control adult groups: USC consists of 72 college students from the University of Southern California (A) and TEACHERS consists of 34 teachers at LILA (T). Students in our main population (LILA) are from caucasian families of upper-middle socioeconomic status. After graduating, they typically attend well-ranked colleges (including USC). Table 1 summarizes the 709 subjects in this sample.

Procedures. We ran sessions at LILA and KING in classrooms during school hours and at Los Angeles Behavioral Economics Laboratory (LABEL) at USC's Economics department for the control undergraduate population. In each school-age session, a combination

	LILA											KING			USC	TEACHERS
Grade	K	1	2	3	4	5	6	7	8	9	10	6	7	8	A	T
#	55	49	46	53	40	31	54	67	22	14	34	77	43	18	72	34

Table 1: Number of participants by grade and population in Experiment 1

of male and female participants from either the same or consecutive grade levels was present. Procedures were identical in all cases, except for payments as detailed below.

The experiment comprises two games programmed with 'oTree' (Chen et al., 2016) and implemented on touchscreen PC tablets. The focus of this study is the first game, the third-party dictator. After a short break, participants play a two-person dynamic game of complete information. The findings of the second game are discussed in Brocas and Carrillo (2023). The two games are distinct enough that we are not worried about potential cross-contamination. However, as a precautionary measure, we always conducted the two games in the same order, anonymously re-matching subjects between the two games. Additionally, we refrained from disclosing any results from the first game, including outcomes and payoffs, until the conclusion of the second game.

Third-party dictator game. Each participant acts as a *third-party dictator* who chooses how to allocate 4 tokens between two competitors of an *effort* task, a winner who has already secured 3 tokens and a loser who has secured 1 token. We require the dictator to allocate a minimum of 1 token to each individual, so that they effectively choose between three options (x, y) where x represents the number of tokens given to the winner and y the amount of tokens given to the loser. The "meritocratic" choice $(3, 1)$ replicates the incentive scheme of the experimenter. The "equitable" choice $(2, 2)$ provides equal payments to both competitors. The "compensatory" choice $(1, 3)$ reverses payments so as to equalize final payoffs. The object of the study is to determine whether age affects the tendency to reinforce, maintain, or eliminate inequalities.[2]

Timing. We first told participants that they were going to play a competition task and a dictator task.[3] For the competition task, participants were randomly and anonymously grouped in pairs and they competed in an effort-intensive task that required patience and concentration. Following Gill and Prowse (2012), we employed sliders. Participants faced a screen displaying 39 non-aligned sliders, each starting in a random position between 0 and 100 (but different from 50). In each pair, the individual who positioned more sliders at the median value (50) within a 90-second timeframe won.[4] Prior to starting in the game,

[2] We considered starting with a $(4, 2)$ allocation, with only two tokens to be divided. However, we concluded that requiring a positive transfer to each participant would make dictators more comfortable.

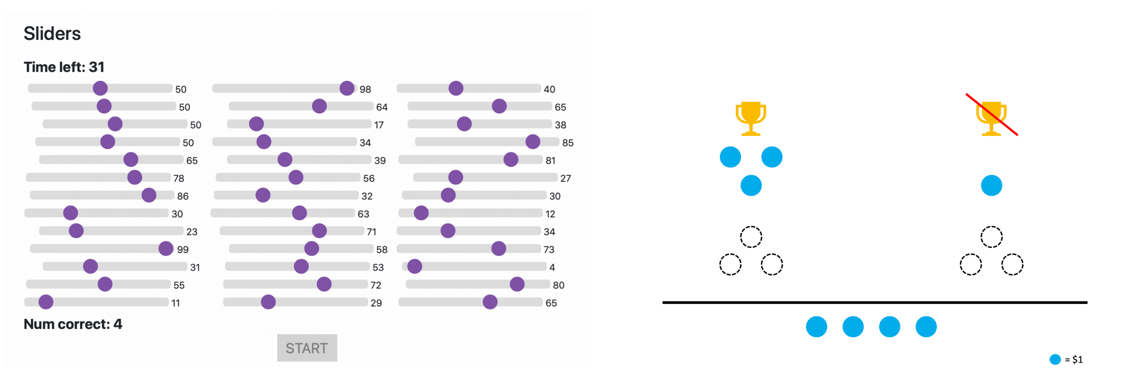
[3] To avoid priming, the tasks were called "the slider task" and "the observer task", respectively. At no point in the experiment participants heard the word 'dictator'.

[4] In case of a tie, the subject who took less time to reach that number of sliders won. We never told

we provided participants with an opportunity to familiarize themselves with the task by requesting them to adjust three sliders to their respective median positions at their own pace. Figure 1a provides a screenshot of the effort competition task showing four sliders already adjusted at 50 and a timer indicating that 31 seconds are left.

Second, we explained the rules of the dictator task, highlighting the following elements. We told the participants that all would play both the dictator and the competition tasks and their decisions would be private. We also told them that each task would be conducted with distinct sets of participants, ensuring that if a person acted as the dictator for two competitors, neither of those competitors would serve as the dictator for each other. We informed them that all participants would first play the dictator task, allocating tokens between a future winner and a future loser of the competition task, then play the competition task. Last, we told them that feedback would be provided only at the end of the experiment and they would only learn whether they won or lost the competition task and the number of tokens given by the anonymous dictator.

Third, participants played the dictator task. We placed particular emphasis on introducing the task through a straightforward and visually intuitive graphical interface. Figure 1b presents a screenshot of the dictator task with 3 initial tokens for the winner and 1 for the loser.



(a) Competition task (effort): participants had 90 seconds to position as many sliders as possible at the median location (50)

(b) Dictator task (addition): 4 tokens (bottom) had to be allocated between winner and loser (with at least one for each person)

Figure 1: Screenshots of Experiment 1

At the bottom of the screen, dictators could see their tokens. To implement an allocation, they simply tapped on the dashed token silhouettes of the winner (represented by a trophy) or the loser (represented by a crossed-out trophy) and a token moved from the bottom to the new location. Notice that the tokens were presented in a manner that facilitated easy

subjects the performance difference between winner and loser.

comparison for young participants between the winner's and loser's amounts. Subjects could move around tokens as many times as they wanted. The allocation became final only after the participant pressed the OK button, at which point the screen transitioned to a blank page with a "please wait" sign displayed.

Fourth, participants played the competition task. Upon reaching the end of the 90-second period, participants learned only the number of sliders correctly positioned at 50, an information that was visible to them throughout the task.

Fifth, after playing the competition task, participants were informed that they had the option to modify their decision in the dictator task. We presented a screen displaying their prior choice, with the option to either select OK to retain that allocation or CHANGE to make a new selection. We also emphasized that no further opportunities for modification would be available after this point. The full set of instructions is included in Appendix A.

Remarks. While we considered the possibility of having each participant play only one task (dictator or competition), we deemed it impractical since it would substantially reduce the number of observations for the task of primary interest. Based on our pilot study, we also discovered that informing children that they would play both tasks was helpful, particularly for younger participants. It allowed them to better take the perspective of both a winner and a loser, aiding their understanding of how it feels to be in each role.

Performance in the competition task may potentially influence the dictator's decision. Our timing allows us to investigate this possibility. By conducting the dictator task prior to the competition task, we can determine the ex-ante preferences of the impartial observer. Then, by permitting an ex-post reallocation, we can explore how one's experience with the task impacts their choices for others. These changes may also reveal behavioral shifts once participants have gained a better understanding of the task.

Payments and duration. Following Brocas and Carrillo (2020a), we used different mediums of payment for different ages, in order to equalize as best as possible the *value of rewards* across individuals, and not the rewards themselves. Students from 6th grade and above and adults earned $1 per token paid at the end of the experiment in cash (USC) or with an amazon e-giftcard (LILA, KING and TEACHERS, since cash transfers are not allowed in the schools). For students in K to 5th grade, we set up a shop with 20 to 25 pre-screened, age-appropriate toys and stationery (bracelets, erasers, figurines, cars, trading cards, apps, calculators, earbuds, gel pens, etc.). Before the experiment, we took children to the shop and showed the toys they were playing for. We described the token prices of each toy and, for the youngest subjects, we explicitly conveyed that acquiring more tokens would translate into obtaining more toys. At the end of the experiment, we informed participants of their token earnings and accompanied them to the shop to exchange their tokens for toys.[5]

[5]The procedure emphasizes the importance of accumulating tokens while making the experience enjoy-

The game studied in this paper lasted about 15 minutes. The entire experiment, including the other game and payment, never exceeded one school period (50 minutes). Participants earned between 2 and 6 tokens in this game, therefore \$2 to \$6 for 6th graders and above. Average earnings in the entire experiment (and not including show-up fees) were \$12.0 (LILA), \$12.8 (KING), \$12.4 (USC) and \$13.7 (TEACHERS). To compensate for the opportunity cost of time and following LABEL procedures, we paid an extra \$5 show-up fee to the adult control groups (USC and TEACHERS). The pricing of tokens was structured in a way that ensured every elementary school participant could acquire a minimum of three toys. Nonetheless, there was a noticeable variation in the number and variety of toys chosen. On average, we spent approximately \$5 per child, which is generous compared to typical experiments involving participants in these age groups.

Questionnaire. At the end of the dictator game, we asked the participants to report how difficult they felt the slider task was ('Hard', 'Medium', 'Easy' or 'I don't know') and how well they thought they performed compared to their competitor ('Better', 'Equal', 'Worse' or 'I don't know'). These questions were not incentivized. We chose this timing intentionally to mitigate any potential demand effect that might prompt participants to alter their allocation in response to the task questions. At the end of the experiment we collected information regarding "gender", "grade", "date of birth", "number of siblings" and "favorite subject at school".

2.2 Experiment 2. Spain.

Experiment 1 raised questions that we addressed in a follow-up study. Experiment 2 introduced four modifications to the previous setup: variants, population, payments, and additional non-cognitive tasks. The rationale behind these changes is discussed below.

Variants. We adopted a 3×2 between-subject design. We explored three conditions. They describe different procedures for determining the winner of the competition, each of which encapsulates one of the three fundamental elements affecting performance in conventional economic models: the *effort* task described in Experiment 1, a *talent* task, and a *luck* task.[6] In the *talent* task, we presented (depending on the participants' grade) 8 to 12 age- and culture-appropriate multiple-choice general knowledge questions, covering various subjects such as mathematics, biology, history, geography, sports, and entertainment.

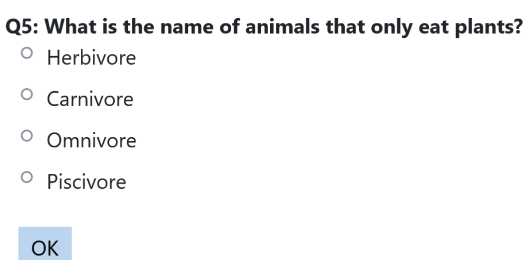
able. In our experience, children at this age perfectly understand the concept of money, but they find toys more attractive and immediately gratifying than cash.

[6]Canonical models of performance in labor economics, industrial organization, political economics, and organizations are typically summarized by $y = f(e, \theta) + \eta$. In this equation, y = performance or output, e = effort, costly choice or investment (sometimes referred to as a 'hidden action' variable), θ = talent, ability or productivity (sometimes referred to as a 'hidden information' variable), and η is a stochastic component (see e.g., Laffont and Tirole (1986); Holmström (1999)). While Experiment 1 investigates differences in y due to e, Experiment 2 distinguishes between differences in y that come from e, θ or η.

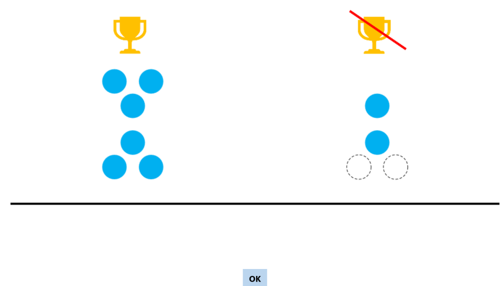
Participants had 12 to 18 seconds to select one of four possible answers. The participant in the pair who answered the most correct answers won the competition. Before playing the game, participants responded to a practice question. Figure 2a displays a screenshot of the talent competition task (translated from the original Spanish version). In the *luck* task, participants were simply informed that the computer would randomly select, with equal probability, one winner and one loser from each pair. The objective was to investigate whether the perception of meritocracy varies depending on its source or origin.

We also explored two treatments, which corresponded to different methods for allocating tokens. First, we retained the same token *addition* method as in Experiment 1. Second, we considered a *redistribution* method, characterized by an initial sharing of (6,2) and the possibility to redistribute 0, 1, or 2 tokens from the winner to the loser. This second formulation is closer to the existing literature (Cappelen et al., 2022; Almås et al., 2020).[7] Although these two designs are formally equivalent, we hypothesized that the framing of the allocation method could lead to behavioral differences. Indeed, *addition* might be perceived as a system of subsidies to both participants whereas *redistribution* might be best associated to a redistributive taxation from the wealthy to the needy. Figure 2b presents a screenshot of the dictator task under the *redistribution* method.

Overall, Experiment 2 comprises six variants. They encompass all combinations of the three conditions for the selection of the winner (*effort*, *talent*, and *luck*) and the two treatments for the allocation of tokens (*addition* and *redistribution*).



(a) Competition task (talent): participants answer age-appropriate multiple-choice general knowledge questions



(b) Dictator task (redistribution): 0, 1, or 2 tokens can be transferred from winner to loser

Figure 2: Additional screenshots for Experiment 2

Population. We enrolled a large sample of 2,280 school-age students of low- to middle-income families from the Spanish equivalent of grades 1 to 12 in the US education system, from four schools located in four cities in the southwestern region of Spain. It is referred

[7]Typical studies start with maximal inequality $(n, 0)$, so that the third-party can only reduce it. Sometimes inequalities are allowed to be reversed, although such an option is virtually never chosen in practice.

to as the SPAIN sample. All schools belong to the same private network of catholic schools sponsored by the Spanish government (the Salesianos).[8] Given that the oldest school-age participants are already considered young adults at 18 years old, we decided not to include an undergraduate control group comprising individuals only slightly older (aged 18 to 22 years old) but with a different background. The experiment was integrated into the school's academic curriculum, and our Institutional Review Board (IRB) stipulated an opt-out procedure. Importantly, there were no opt-outs, effectively eliminating any potential selection bias. The only students who did not participate were those who were absent on the day of the experiment (5.7%). Table 2 summarizes the distribution of the 2,280 participants by grade, treatment and condition.

	SPAIN												
Grade	1	2	3	4	5	6	7	8	9	10	11	12	Total
addition / effort *	28	30	32	31	20	24	54	27	23	50	23	0	342
addition / talent	22	39	25	23	52	47	30	29	51	25	31	24	398
addition / luck	36	22	22	24	24	50	26	80	54	46	29	45	458
redistribution / effort	19	28	23	25	45	22	54	52	31	28	23	34	384
redistribution / talent	32	8	22	21	49	43	27	26	31	25	29	20	333
redistribution / luck	15	22	33	33	25	26	62	29	28	43	23	26	365
Total	152	149	157	157	215	212	253	243	218	217	158	149	2280

* identical treatment as Experiment 1

Table 2: Number of participants by grade, treatment, and condition in Experiment 2

Payments. While we have consistently emphasized the importance of age-appropriate payments, even if they involve different mediums, for Experiment 2, we identified a payment method that is uniform across all age groups and highly appreciated by all participants. Schools in SPAIN have integrated cafeterias where students can purchase drinks, food and snacks at discounted prices. The cafeteria is very popular and widely accessed during recess and lunch time. We created our own currency that could be spent on any item. Currency had no expiration date, although most participants spent their endowments within a week. For simplicity, we kept the conversion rate of 1 token = 1€. Considering the differences in exchange rate and cost of living, incentives in Experiment 2 are slightly higher than in Experiment 1.

Emotional intelligence and personality. We hypothesize that choices in the dictator game are affected by non-cognitive abilities, empathy in particular, which is itself related to emotional intelligence and personality. To test this possibility, we adapted, validated and conducted novel, child-friendly versions of two well-known non-cognitive tasks: an

[8] The Salesianos educational network has a global presence in 132 countries. The network has 59 schools in Spain alone serving almost 40,000 students (see `https://www.salesianos.es/escuelas`).

affective Theory-of-mind task (a-ToM) and a Big 5 personality questionnaire (Big5). In a-ToM, individuals are presented with images of a person's eyes and are tasked with identifying the most fitting adjective among four options that describe the emotion being conveyed. In Big5, the personality of the participant (extraversion, agreeableness, openness, conscientiousness, and neuroticism) is elicited through a Likert scale questionnaire. Our versions of a-ToM and Big5 are described in detail in sections 4.1 and 4.2.

Experiment 2 follows this specific timeline: the third party dictator task is followed by a new game of incomplete information that we analyze in a different paper (Alfonso et al., 2023).[9] We then implement the affective theory-of-mind task, and the Big 5 personality questionnaire.

3 Results

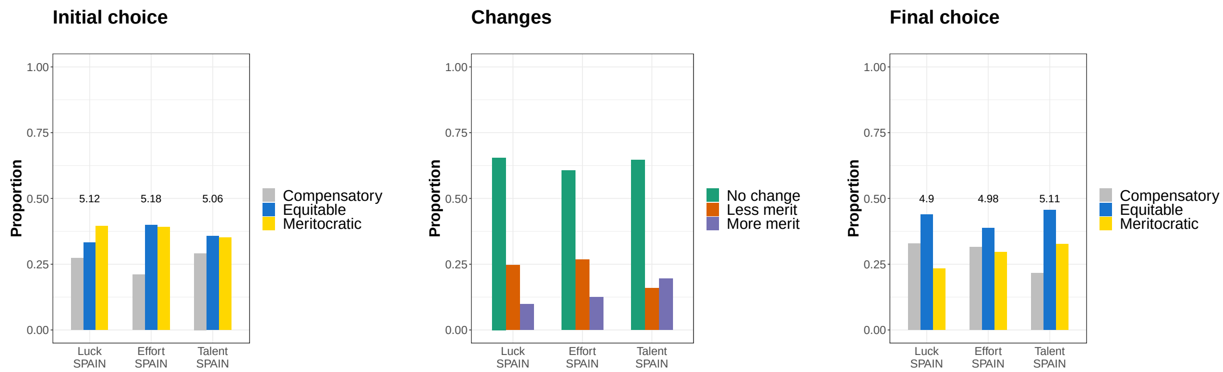
Due to the size and heterogeneity of our samples, there are numerous approaches available for organizing and examining our data. We decided to focus first and primarily on our two main populations of children: LILA (K to 10th grade, 465 participants, 1 variant, Experiment 1) and SPAIN (1st to 12th grade, 2,280 participants, 6 variants, Experiment 2). We anticipate both commonalities and variations in behavior, influenced by disparities in cultural and socioeconomic circumstances. We provide a comparison with the middle schoolers at KING in section 5, and a brief account of the behavior of our adult populations (USC and TEACHERS) in Appendix B5. Note that a small number of responses were inaccurately recorded, and a small number of students provided ages that fell outside the typical age ranges for their respective grades (always less than 1%). To maintain the integrity of our analyses, we excluded these observations from the affected data sets.

3.1 Aggregate choices

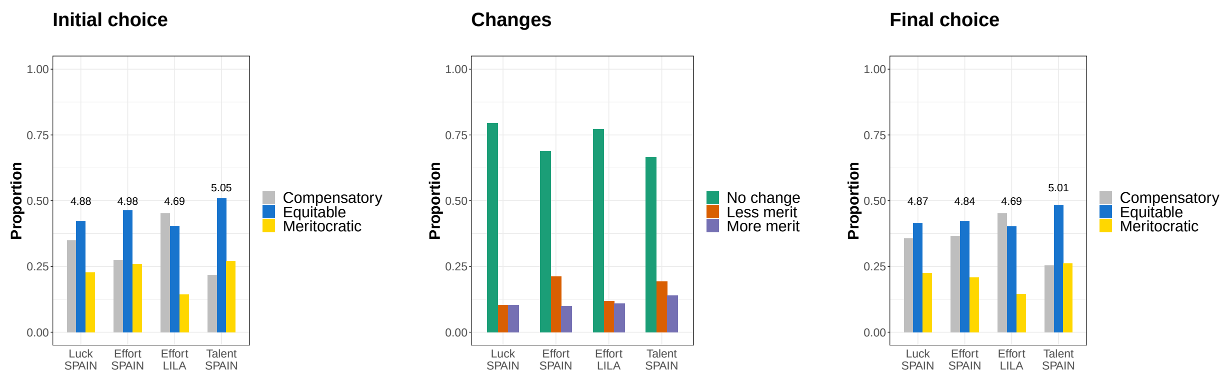
In our first data examination, Figure 3 displays the aggregate proportions of meritocratic (6,2), equitable (5,3) and compensatory (4,4) choices for each population (LILA and SPAIN), each winner selection procedure (*effort*, *luck* and *talent*) and each token allocation method (*addition* and *redistribution*). In each case, we include the initial decisions (left), proportion of changes (center) and final decisions (right).

As shown in Figure 3a, initial aggregate behavior under the *redistribution* method is very similar across winner selection procedures, with a slightly lower fraction of compensatory choices in *effort* compared to *talent* (t-test of average payment to winner, $p = 0.04$). More than a third of participants (35% to 39%) change their allocation after experiencing the task (*effort* and *talent*) or perhaps after giving it some more thought (*luck*). Changes

[9] The game is different from the second game studied in Experiment 1 but the order of presentation of games and the procedures (e.g., no results revealed until the end of the experiment) are identical.



(a) Redistribution method



(b) Addition method

Figure 3: Aggregate initial decisions (left), changes (center) and final decisions (right) with the redistribution (top) and addition (bottom) methods, for the three winner selection (luck, effort, talent) procedures and the two populations (LILA and SPAIN). The average payoff of the winner is reported for each case.

imply significantly higher transfers to the loser except in *talent* (t-test of average difference between final and initial payments to winner, p $<$ 0.001 in *luck* and *effort*).

The picture looks very different in Figure 3b, which displays results for the *addition* treatment and includes the *effort* condition in LILA. The initial allocations in SPAIN are very similar for *effort* and *luck*, while the allocation to the winner is higher in *talent* (t-tests between *luck* and *talent*, p $<$ 0.001) and lower in LILA *effort* (t-tests between LILA *effort* and all conditions in SPAIN, p $<$ 0.001). Overall, compensation to the loser is significantly more common in *addition* than in *redistribution* for *luck* and *effort* (t-tests of differences between average payments to winner in *addition* and *redistribution*, p $<$ 0.001). This aligns with the intuitive idea that allocating resources to the losing party becomes more straightforward when they do not have to be "subtracted" from the winner. The percentage of participants altering their allocation in *addition* is lower compared to *redistribution*, but it remains substantial (21% to 33%). The most notable shift in behavior emerges for the SPAIN population in *effort* (t-tests of difference between final and initial payments, $p < 0.001$), resulting in the intriguing outcome that the final allocations under *effort* become fairly alike in both populations.

In the SPAIN population, the payment awarded to the winner is notably higher when using the *redistribution* method for both *effort* and *luck* (t-tests of difference between treatments, $p < 0.001$). However, this significant difference does not hold for *talent*. This observation reflects both the reluctance to deduct from the winner in the *redistribution* method and the reluctance to compensate the loser in *talent*, even when employing the *addition* method. These disparities tend to disappear after participants gain experience, with the exception of *effort* (t-tests of difference between treatments, $p = 0.015$).

In summary, the greatest level of meritocracy is observed in the initial allocations under *redistribution*, which is precisely the case commonly examined in the literature. On average, we notice greater allocations to the loser after participants experience the task (final choice), with an even more pronounced effect in *addition*, where there is not a predefined default choice.

3.2 Age-related changes in behavior

The previous picture is incomplete as it overlooks the primary variable of our study, namely the shifts in choices from childhood to adulthood. Although the initial choice is arguably the purest decision, we opted to center our analysis on the final choice because we believe that disparities between *luck*, *effort*, and *talent* hold greater significance after participants have engaged in the competition. At the same time, the ultimate allocations may be influenced by the participant's own performance, even if they act as impartial observers (a possibility we explore in section 3.3). Therefore, and also for the sake of completeness, we report an analysis of initial choices in Appendices B1 and B2.

We start with the *redistribution* method, which is the procedure most commonly employed in the literature (Cappelen et al., 2022; Almås et al., 2020) and then investigate our original *addition* method.

3.2.1 Final choices in the redistribution treatment

We first examine the scenario in which the experimenter allocates all resources (6 to the winner and 2 to the loser), and the observer has the option to redistribute 0, 1 or 2 tokens from the winner to the loser. This particular treatment is exclusively carried out in the SPAIN population. We report OLS regressions, where the dependent variable is the final allocation to the winner. For the independent variables, we include the Age in months and its quadratic term Age^2, to account for potential non-linearities. We use *effort* –the condition which is common in Experiments 1 and 2– as the default condition and we add dummies for *luck* and *talent*. We also include dummy variables for gender (Male = 1), whether the participant has at least one sibling (Siblings = 1), and school favorite topic to capture self-reported analytical inclination (STEM = 1).[10] To ensure comparability with the *addition* treatment (in section 3.2.2), we chose not to account for emotional intelligence and personality factors, since this information was not collected in the LILA dataset.This issue is studied in section 4. Figure 4 presents the results of the regressions (left) and depicts the quadratic best fit of the age variable (right).

	SPAIN	
	(1)	(2)
Age	0.003	0.002
	(0.003)	(0.005)
Age^2	−0.00001	−0.00001
	(0.00002)	(0.00002)
Luck	−0.075	−0.080
	(0.052)	(0.054)
Talent	0.133**	0.129**
	(0.057)	(0.057)
Male		0.110**
		(0.047)
Siblings		0.014
		(0.066)
STEM		−0.019
		(0.048)
Constant	−1.176***	−1.182***
	(0.335)	(0.338)
Obs.	1,078	1,078
Adj. R^2	0.009	0.011

*p<0.1; **p<0.05; ***p<0.01

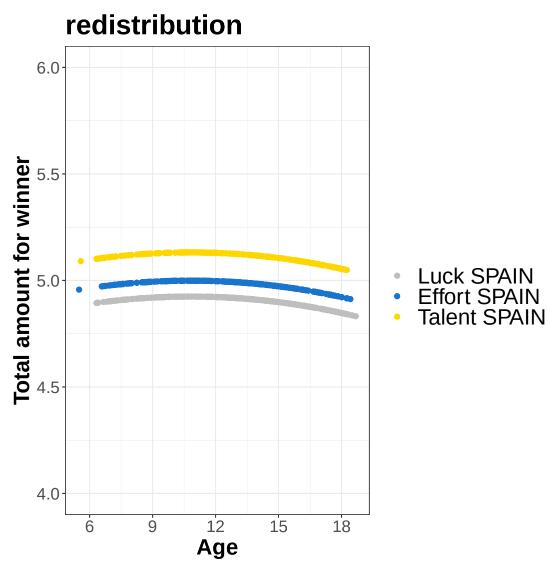


Figure 4: OLS regressions of final allocation to winner (left) and representation of quadratic best fit (right) as a function of age for *luck*, *talent* and *effort* (SPAIN) in the *redistribution* treatment.

Decisions in the *redistribution* treatment remain consistent regardless of age, with reallocation typically hovering around one token, that is, 50% of the maximum possible amount. This results in an average final distribution of (5,3). We anticipated that the allocation to the winner would be notably greater in *talent* than in *effort*, and greater in *effort* than in *luck*. This expectation aligns with the intuition that meritocracy is most embraced when it comes to natural ability and least embraced when it pertains to good

[10]STEM refers to a preference for Mathematics or Science/Technology. Consistent with the curriculum of the school, the three other categories offered are Languages, History/Geography and Arts/Music, globally referred to as 'Arts & Humanities'.

fortune. We found some evidence supporting the first effect but not the second. It is conceivable that after engaging in the slider task, participants perceived it more as an inconsequential task with outcomes based on chance rather than a challenging, effort-intensive activity deserving of merit-based compensation.

As in the literature on inequality acceptance in adults (Almås et al., 2020), we observe a notable gender effect, with males exhibiting a higher degree of meritocracy compared to females. This result stands in contrast to the findings of Almås et al. (2010) who reported no gender effect on inequality acceptance in children and adolescents. Additionally, our expectation that children with siblings would allocate more to the losing party, due to their familiarity with adult-induced redistribution (typically in favor of the younger child), was not supported by our results. Furthermore, we observed no impact of analytical orientation on redistribution. When we conduct separate regressions for the three winner selection methods, the results remain consistent: age continues to show no significant impact on redistribution, and males tend to be more meritocratic than females, except in the *luck* condition. (see Appendix B3).

3.2.2 Final choices in the addition treatment

We now shift our focus to the scenario with no predefined default allocation, where some resources are under the observer's discretion and must be deliberatively divided between the winner and loser. This corresponds to the original *effort* condition conducted at LILA in Experiment 1. We report OLS regressions separately for LILA (columns (1) and (2)) and SPAIN (columns (3) and (4)), using the same variables as in Figure 4. Notice that age brackets are different across populations (K to 10 in LILA and 1 to 12 in SPAIN). Figure 5 summarizes the findings.

The regression analysis reveals intriguing and consistent age-related dynamics in the *addition* treatment. The allocation to the winner follows a U-shaped pattern in both populations, with a substantial and highly significant convexity coefficient. In simpler terms, when participants are required to actively determine the allocation, ex-post fairness, regardless of merit, experiences a significant rise during elementary school, reaching its peak in the early (LILA) and late (SPAIN) middle-school years, and subsequently declines.

At the same time, there are similarities with the previous treatment. Most notably, meritocracy in SPAIN is most pronounced in *talent*, while there are no significant differences between *effort* and *luck*. In both populations, males exhibit higher levels of meritocracy than females, and the presence of siblings does not yield a significant effect. However, we do observe more redistribution by participants with an analytical inclination in LILA.

It is interesting to notice that the evolution with age in *effort* has a similar U-shape in both populations, despite the large SES and cultural differences. And yet, LILA participants appear to redistribute more than their peers in SPAIN (we provide a more in-depth comparison of these two groups in section 3.2.3). More generally, the consistency of the

	LILA		SPAIN	
	(1)	(2)	(3)	(4)
Age	−0.027***	−0.027***	−0.022***	−0.022***
	(0.006)	(0.006)	(0.004)	(0.004)
Age^2	0.0001***	0.0001***	0.0001***	0.0001***
	(0.00003)	(0.00003)	(0.00001)	(0.00001)
Male		0.116*		0.204***
		(0.066)		(0.043)
Siblings		−0.054		0.063
		(0.074)		(0.058)
STEM		−0.152**		−0.015
		(0.067)		(0.043)
Luck			0.031	0.040
			(0.053)	(0.053)
Talent			0.164**	0.161**
			(0.054)	(0.054)
Constant	6.398***	6.451***	6.513***	6.393***
	(0.385)	(0.387)	(0.297)	(0.283)
Obs.	465	465	1,177	1,177
Adj. R^2	0.040	0.049	0.038	0.055

*p<0.1; **p<0.05; ***p<0.01

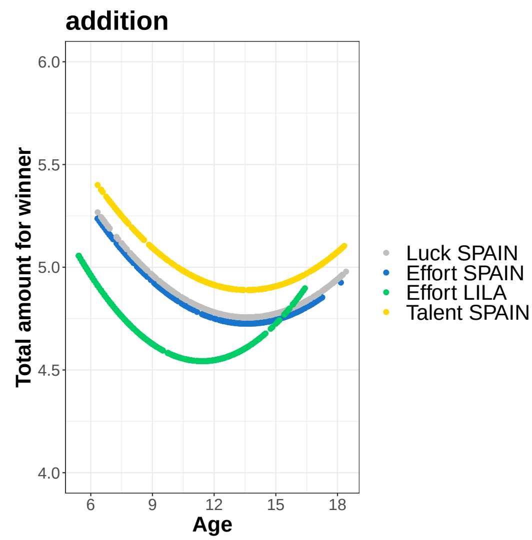


Figure 5: OLS regressions of final allocation to winner (left) and representation of quadratic best fit (right) as a function of age, winner selection method and population in the *addition* treatment.

U-shaped age pattern in the allocation to the winner represents a robust and reliable finding. The significance of this pattern persists even when separate regressions are conducted for the three winner selection methods. (see Appendix B4 for the details).[11] It seems that the age effect disappears only when the process of redistribution involves taking money away from the winner.

Finally, when comparing Figures 4 and 5, it becomes evident that there is a greater inclination toward allocating resources to the loser in *addition* compared to *redistribution*, especially among participants in grades 4 and above (see also section 3.4). These observations align with the conclusions drawn in section 3.1 and they are consistent with the ideas presented in List (2007), where participants are more generous under the "give" scenario than the "take". Nonetheless, we expected a more pronounced anchoring difference between the willingness to give to the loser and the willingness to deduct from the winner.

3.2.3 Population comparison: LILA v. Spain

Although a comprehensive study of the shared and distinct social preferences between our two populations falls outside the purview of this project, we believe it holds merit to

[11] Such exercise also reveals a larger meritocratic tendency for the youngest participants in *luck*.

compare their perspectives on meritocracy within the (common) *effort* condition in the *addition* treatment. We therefore focus on this variant and replicate the same analysis as outlined in Figure 5, adding a population dummy variable (LILA =1). The regression outcomes and the best quadratic fit are reported in Figure 6.

	(1)	(2)
Age	-0.023^{***}	-0.024^{***}
	(0.005)	(0.005)
Age^2	0.0001^{***}	0.0001^{***}
	(0.00002)	(0.00002)
LILA	-0.161^{***}	-0.157^{***}
	(0.054)	(0.054)
Male		0.202^{***}
		(0.051)
Siblings		-0.020
		(0.063)
STEM		-0.078
		(0.052)
Constant	6.350^{***}	6.338^{***}
	(0.304)	(0.307)
Obs.	794	794
Adj. R^2	0.037	0.053

*p<0.1; **p<0.05; ***p<0.01

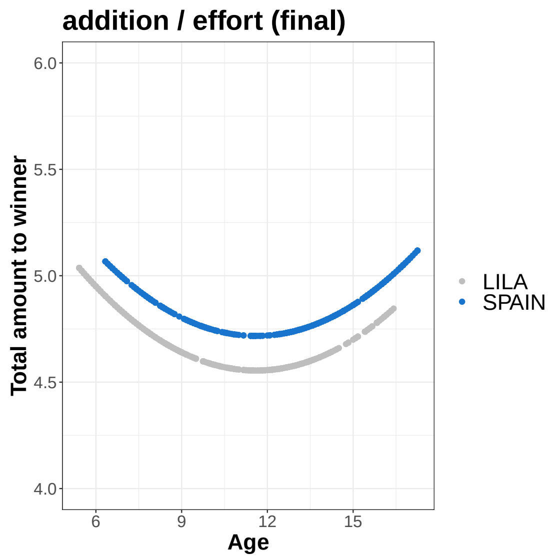


Figure 6: OLS regressions of final allocation to winner (left) and representation of quadratic best fit (right) for the *addition* / *effort* variant in LILA and SPAIN.

There is a higher willingness to redistribute in LILA compared to SPAIN, accompanied by the familiar U-shaped age trajectory. The difference is small but statistically significant. This similarity across populations, however, may mask some subtle differences. Specifically, participants in grades 4 and above in SPAIN are more equitable and less compensatory compared to their counterparts in LILA. Conversely, participants in grades 3 and below in SPAIN are more meritocratic and less equitable compared to those in LILA (chi-square tests, $p = 0.007$ for grades 4 and above, $p = 0.005$ for grades 3 and below). This disparity contributes to the higher level of observed redistribution in LILA than in SPAIN. The precise reasons for these variations are challenging to pinpoint, as our two populations differ significantly in terms of educational, cultural, and socioeconomic conditions.

Finally, it is also interesting to briefly compare the initial choices of these two populations. Figure 7 presents the same information as Figure 6 for the initial decision.

While the quadratic fit in LILA is practically identical for initial and final choices, there is a significant disparity in meritocracy in SPAIN before and after experiencing the task. The amount given to the winner in SPAIN is significantly reduced after experience, resulting in partial convergence between the two groups. This result implies that while cultural factors may initially influence our preferences regarding redistribution and ideals,

	(1)	(2)
Age	-0.025^{***}	-0.026^{***}
	(0.005)	(0.005)
Age^2	0.0001^{***}	0.0001^{***}
	(0.00002)	(0.00002)
LILA	-0.308^{***}	-0.302^{***}
	(0.053)	(0.052)
Male		0.170^{***}
		(0.051)
Siblings		-0.003
		(0.062)
STEM		-0.056
		(0.051)
Constant	6.649^{***}	6.623^{***}
	(0.300)	(0.304)
Obs.	805	805
Adj. R^2	0.071	0.082

*p<0.1; **p<0.05; ***p<0.01

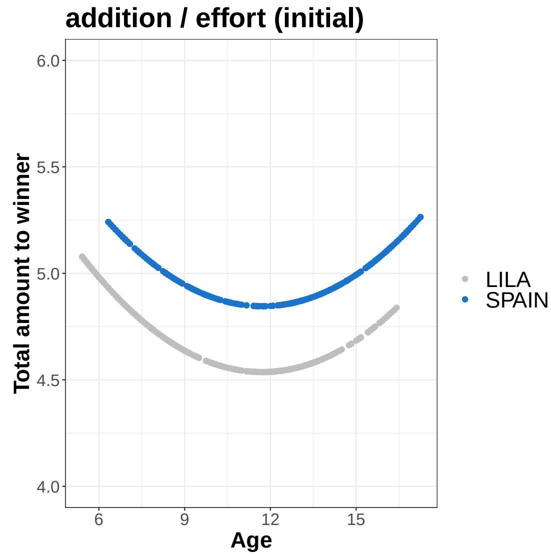


Figure 7: OLS regressions of initial allocation to winner (left) and representation of quadratic best fit (right) for the *addition / effort* variant in LILA and SPAIN).

experience appears to guide individuals towards finding common ground and adjusting their views. Although our study is not specifically designed to investigate the underlying causes of these adjustments or adaptations, we find this result intriguing. Indeed, it implies the presence of an intrinsic or innate inclination to respond to experiences, which suggests that institutions or cultures that do not align with these innate tendencies may have limited capacity to shape or influence our attitudes.

3.3 Revision of allocation: initial v. final choice

An analysis of the initial decisions can be found in Appendices B1 and B2. In this section, we explore potential explanations for the shifts in choices between initial and final decisions. During the experiment, we included two sets of non-incentivized questions where participants could express their *perceptions* regarding (i) the task's level of difficulty and (ii) their performance relative to others. These questions were not included in the *luck* condition, as the winner was randomly determined. We anticipate observing self-serving biases among participants. Those who perceive their performance as favorable and find the task easy are likely to lean towards more meritocratic allocations (and vice versa). It is important to acknowledge the exogenous limits to this tendency, as individuals who initially opt for compensatory and meritocratic allocations cannot shift towards more and less redistribution, respectively.

To study the effect of task perception on the decision to change allocations, we consider each initial choice separately (meritocratic, equitable, compensatory) as well as each treatment and population (*redistribution* in SPAIN, *addition* in SPAIN, *addition* in LILA). For initial meritocratic (Merit.) and compensatory (Comp.) choices, we perform Poisson regressions, where we interpret the absolute value of the change in the allocation to the winner between the initial and final choice as counts $(0, 1, 2)$.[12] For initial equitable choices (Equit.), we perform a multinomial regression of the likelihood to decrease (-1) or increase $(+1)$ the final allocation of the winner, that is, move to compensatory and meritocratic choices, respectively. Thus, a positive coefficient in Merit. and Equit. (-1) implies an increase in the allocation to the loser whereas a positive coefficient in Comp. and Equit. $(+1)$ implies an increase in the allocation to the winner. We include the familiar dependent variables Age, Age^2, Male, Siblings, STEM. We do not include data from *luck* since no information is elicited between the initial and final choice in that condition. For the SPAIN population, we pool data from both the *effort* and *talent* conditions and include a dummy *talent* to increase power.

Regarding our variables of interest, we use "equal / I don't know" as the benchmark and introduce dummies for "better" and "worse" than the rival in the case of participants' self-reported perceived performance. Similarly, we set "same / I don't know" as the benchmark and incorporate dummy variables for "hard" and "easy" for participants' self-reported perceived difficulty. Table 3 presents the regressions.

As these regressions reveal, perceived own performance serves as a strong predictor of behavioral changes across all three conditions. Specifically, the belief of having performed "better" than others is linked to a reduced inclination toward redistribution from the winner to the loser. This manifests as fewer shifts away from meritocratic choices, fewer transitions from equitable to compensatory allocations, more transitions from equitable to meritocratic allocations, and more shifts away from compensatory choices. Conversely, the belief of having performed "worse" than others is associated with a greater propensity for reallocation to the loser. This results in fewer transitions from equitable to meritocratic allocations and fewer shifts away from compensatory choices. This effect occurs in all conditions, although it is more significant in some treatments (*addition* in SPAIN) than in others. The influence of perceived difficulty is somewhat similar but less pronounced. For instance, a 'hard' task is associated with fewer shifts from equitable to meritocratic choices. However, these effects are not as frequent or consistent as those observed in the case of reported performance. In summary, perceived performance, and to a lesser extent, perceived difficulty provide evidence in support of our self-serving bias hypothesis. This self-serving or egocentric bias is in line with some current findings in the literature. For example, Neuber (2021) found that individuals engaged in distributing rewards to others often adhere to norms that favor their own interests, even in situations where their choices exclusively impact others.

[12] Changes are necessarily non-positive for initial meritocratic and non-negative for initial compensatory.

	redistribution (SPAIN)				*addition* (SPAIN)				*addition* (LILA)			
	Merit.	Equit. (−1)	Equit. (+1)	Comp.	Merit.	Equit. (−1)	Equit. (+1)	Comp.	Merit.	Equit. (−1)	Equit. (+1)	Comp.
Age	−0.013	−0.854*	−0.839*	0.017	0.014	−0.280	−0.876	−0.025	0.088**	0.283	3.017*	−0.043*
	(0.016)	(0.499)	(0.448)	(0.022)	(0.022)	(0.466)	(0.547)	(0.025)	(0.044)	(0.508)	(1.564)	(0.026)
Age^2	0.0004	−0.030	0.030*	−0.0001	−0.0001	0.009	0.035	0.0001	−0.0003*	−0.011	−0.138*	0.0002*
	(0.0001)	(0.019)	(0.018)	(0.0001)	(0.0001)	(0.018)	(0.022)	(0.0001)	(0.0002)	(0.024)	(0.073)	(0.0001)
Male	−0.314*	0.610	0.464	0.139	−0.428**	−0.015	0.541	0.425	−0.159	−0.234	1.919**	−0.319
	(0.179)	(0.376)	(0.400)	(0.224)	(0.218)	(0.351)	(0.438)	(0.288)	(0.455)	(0.429)	(0.895)	(0.314)
Siblings	0.068	1.167	−0.187	−0.141	0.027	−0.384	−0.301	0.568	0.355	0.611	−0.478	−0.482
	(0.245)	(0.793)	(0.545)	(0.279)	(0.274)	(0.541)	(0.672)	(0.439)	(0.522)	(0.514)	(0.761)	(0.313)
STEM	0.183	−0.628*	−0.662	0.057	0.134	0.007	0.667	−0.337	0.885*	0.342	−0.046	−0.309
	(0.182)	(0.381)	(0.413)	(0.227)	(0.211)	(0.357)	(0.450)	(0.287)	(0.461)	(0.418)	(0.731)	(0.305)
Talent	−0.139	−1.242***	−0.455	0.465**	−0.526**	0.085	0.111	0.156				
	(0.185)	(0.417)	(0.404)	(0.232)	(0.222)	(0.360)	(0.455)	(0.290)				
Better	−0.446*	−0.921*	0.757*	0.543**	−0.856***	−0.208	0.759	−0.231	−0.595	−1.052*	−0.838	0.609*
	(0.260)	(0.552)	(0.401)	(0.243)	(0.299)	(0.420)	(0.510)	(0.342)	(0.452)	(0.625)	(0.792)	(0.327)
Worse	0.325	0.424	−0.639	−1.123*	0.331	1.528***	−13.966***	−0.766	0.316	0.801	−11.444***	−0.593
	(0.222)	(0.449)	(0.811)	(0.617)	(0.339)	(0.493)	(0.723)	(0.620)	(0.832)	(0.516)	(0.0002)	(0.466)
Hard	0.381	−1.380	−0.341	0.586	−0.110	0.904	−1.001	−0.374	0.194	0.410	−13.178***	−0.330
	(0.258)	(0.859)	(0.848)	(0.437)	(0.332)	(0.701)	(0.998)	(0.638)	(1.124)	(0.632)	(0.00003)	(0.556)
Easy	0.024	−0.839*	0.227	0.187	−0.222	0.034	0.406	0.474	−0.113	1.141**	−0.331	0.258
	(0.200)	(0.456)	(0.400)	(0.236)	(0.235)	(0.382)	(0.472)	(0.295)	(0.465)	(0.485)	(0.787)	(0.319)
Constant	0.235	4.222	4.169	−2.312	−0.672	0.681	2.638	−0.428	−6.595**	−3.991	−18.380**	1.518
	(1.191)	(3.302)	(2.920)	(1.490)	(1.485)	(3.085)	(3.550)	(1.931)	(2.690)	(2.628)	(8.263)	(1.605)
Obs.	266	272	272	166	191	356	356	177	67	188	188	210
AIC	526	443	443	321	343	576	576	260	119	261	261	266

*p<0.1; **p<0.05; ***p<0.01

Table 3: Poisson and multinomial regressions of change in allocation to winner as a function of initial choice (meritocratic, equitable, compensatory) by treatment and population.

3.4 Token allocation: redistribution v. addition

We have shown in section 3.1 that while the initial payment awarded to the winner is, on aggregate, larger in the redistribution treatment, experience results in similar final payments. Here, we reexamine this conclusion to account for age effects. Specifically, we report similar OLS regressions as before, for which we pool the data of treatments *addition* and *redistribution*, and we include a dummy treatment variable.

Controlling for age effects and other sources of individual heterogeneity, the winner of the competition obtains a small but highly significant additional payoff in the *redistribution* treatment compared to the *addition* treatment, after experiencing the task in all three conditions. Notice that the age trajectory when both treatments are pooled together has a slight convex shape, which is the result of combining a strong convex shape in *addition* with a flat shape in *redistribution*.

This result shows that redistributive methods exercise a significant effect on the willingness to reallocate wealth, independently of its origin. While the allocations chosen by the dictators tell us something about the utility they derive over others' final distribu-

	(1)	(2)
Age	-0.011^{***}	-0.011^{***}
	(0.003)	(0.003)
Age^2	0.00003^{***}	0.00003^{***}
	(0.00001)	(0.00001)
Luck	-0.023	-0.024
	(0.038)	(0.038)
Talent	0.142^{***}	0.138^{***}
	(0.039)	(0.039)
Redistribution	0.092^{***}	0.094^{***}
	(0.031)	(0.031)
Male		0.161^{***}
		(0.032)
Siblings		0.042
		(0.044)
STEM		-0.019
		(0.032)
Constant	5.713^{***}	5.657^{***}
	(0.223)	(0.225)
Observations	2,255	2,255
Adjusted R^2	0.020	0.030
Note:		*p<0.1; **p<0.05; ***p<0.01

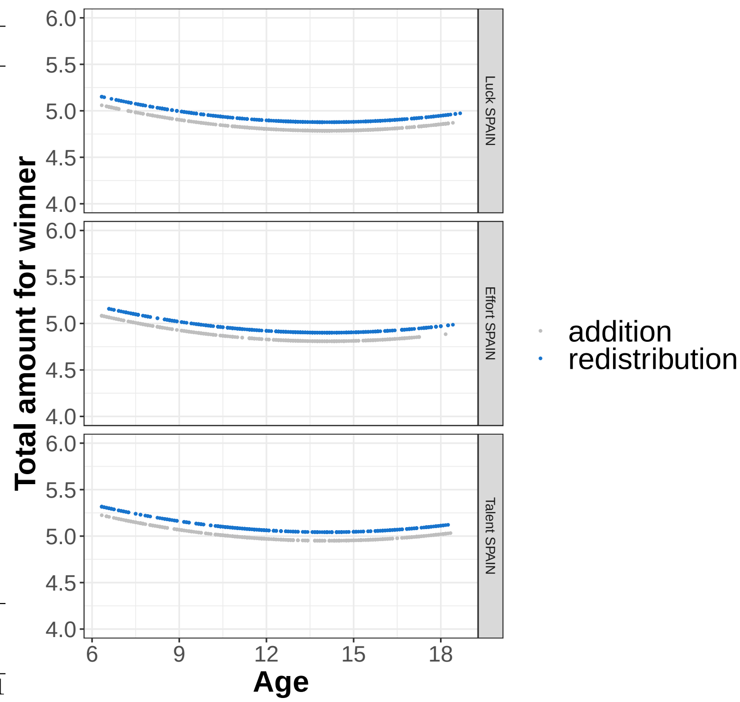


Figure 8: OLS regressions of final allocation to winner (left) and representation of quadratic best fit (right) as a function of age, treatment and condition.

tions, differences across redistributive methods show that their utility is also affected by the process that shapes these distributions. Our participants are generally reluctant to take resources away. The idea that people do not like to take responsibility for hurting others even if it leads to a better social outcome is reminiscent of the literature in ethics in psychology as, for example, in the well-known trolley problem (Foot, 1967).

4 Non-cognitive abilities

Recent research has found that cognitive abilities shape our (time and risk) preferences (see e.g., Dohmen et al. (2010, 2018)), although the exact mechanism is still not fully understood (Amador-Hidalgo et al., 2021). In Experiment 2, we took a different route and hypothesized a link between non-cognitive abilities –personality and capacity to empathize with others– and (social) preferences. To investigate this connection, we introduced two novel assessments: an incentivized affective Theory-of-mind task (a-ToM) and a Big 5 personality questionnaire (Big5). These assessments were adapted and validated specifically for children and adolescents, making them novel and unique tools for researchers interested in studying developmental decision-making processes. In this section, we provide a comprehensive overview of our a-ToM task in section 4.1 and our Big5 personality questionnaire in section 4.2. Additionally, we analyze the performance evolution with respect to age in these two tasks. Finally, we delve into the connections between emotional intelligence, personality traits, and meritocratic decision-making in section 4.3.

4.1 Affective theory of mind (a-ToM)

The "Read the Mind in the Eyes Task" (RMET), developed by Baron-Cohen et al. (1997), is a psychological assessment designed to measure an individual's ability to interpret complex emotional and mental states. During the task, participants view images of a person's eyes and are required to select the most appropriate adjective from a set of four options to describe the emotion conveyed. The test is often used to assess an individual's theory of mind, which is the ability to attribute mental and emotional states to others. The RMET has been widely used in research on social cognition. In recent studies, there has been a growing emphasis on the distinction between Affective Theory of Mind (a-ToM), which involves recognizing and empathizing with emotions, and Cognitive Theory of Mind (c-ToM), which centers on understanding the cognitive processes driving human behavior. Within this framework, the RMET is a preferred tool for assessing a-ToM.

Originally, the RMET was designed for use with adults. Accordingly, the emotions depicted in the test may involve complex nuances that younger children might not fully grasp (e.g. "suspicious", "thoughtful") or be expressed in a language excessively complicated for children (e.g. "flustered", "skeptical"). In studies involving children, adaptations often feature subsets of the original images. However, given our goal of examining the developmental trajectory of a-ToM, it is essential to employ the same task across all age groups. In response to this need, we decided to create and validate a novel a-ToM task, which was tested with both adults and children.

Our adaptation features 16 images capturing the expressive eyes of a 12 years-old child who conveys 16 fundamental emotions, all described in simple language. We call this task the DeRMET (Developmental Read the Mind in the Eyes Test). For each image, we provide four possible answers, one of which most accurately represents the depicted emotion.[13] Participants were incentivized, earning 0.10€ for every correct response. Figure 9 presents one of the images with the four possible answers as seen by our participants (65% of the population provided the correct answer, namely "sad"). Detailed instructions, in-

[13]We validated our task by enlisting the participation of 43 teenagers and adults, who were asked to identify emotions in 20 original images. Additionally, they were requested to provide feedback on the perceived difficulty of identifying these emotions. Following this validation process, we removed four emotions that participants encountered difficulty in identifying. The task and validation materials were registered with the Open Science Framework (OSF) to ensure transparency and accessibility.

cluding the complete set of images, possible responses, correct answers, and percentage of participants who chose correctly each of them are available in Appendix A2.

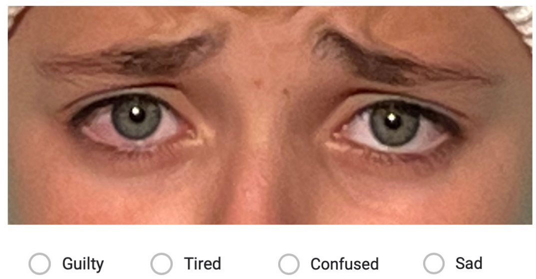


Figure 9: An example of an image and possible answers in the DeRMET

Using this modified task in our population offers three advantages when compared to the original task (Baron-Cohen et al., 1997). By sourcing all images from a single individual, we made the task simpler and more consistent. By employing a child as a model to depict emotions, we made the images more relatable to children and teenagers. Finally, by focusing on simpler emotions and using a straightforward language, we made it suitable for any age, including some very young participants.[14]

Figure 10 summarizes the distribution of the number of correct answers in the a-ToM task by grade in SPAIN, with the grade average displayed at the top.

Like many tasks that involve a combination of sensory perception and the ability to interpret social cues, performance in a-ToM is not constant across age groups. The improvement is large and significant between grades 1 and 5 (8.1 to 11.3, for an average increase of 0.80 per grade, Pearson Correlation Coefficient between score and age PCC = 0.59, $p < 0.001$). The improvement is small but significant between grades 5 and 12 (11.3 to 12.8, for an average increase of 0.21 per grade, PCC = 0.16, $p < 0.001$). It is consistent with an RMET study examining age-related changes of a-ToM in adolescents and adults (Moor et al., 2012). Most importantly, very few participants obtain a perfect score and the aggregate behavior is nowhere close to random guessing for any grade (if individuals guessed randomly, 81% would obtain a score of 5 or less). This suggests that the task is appropriately challenging and suitable for participants across the entire age range of the study. Furthermore, the percentage of correct answers for each image varies between 45.9% and 97.1%. This implies that the answer we designated as "correct" consistently aligns with the choice made by the majority of participants.[15] Finally, we find that females are

[14] Other authors have developed variations of the RMET task, or emotional intelligence tasks based on facial expressions (Mattek et al., 2021; Pahnke et al., 2020). To our knowledge, none are designed to study the developmental trajectory of a-ToM.

[15] If this were not the case, it would bring into question what constitutes a correct answer: the one we anticipate or the one that individuals assert. In our experiment, the least clearcut case is image 15, where

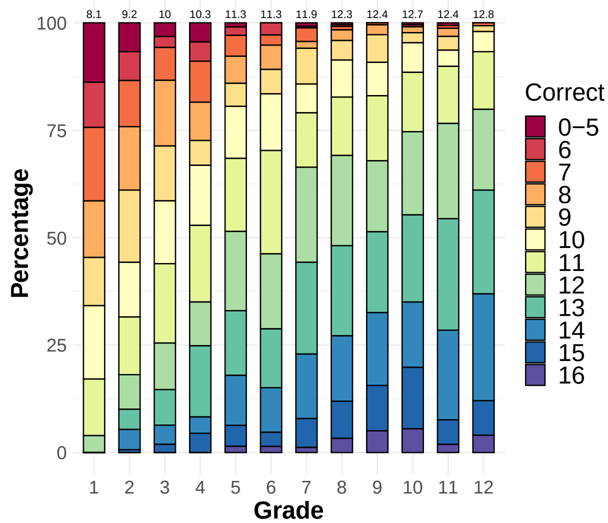


Figure 10: Distribution of correct answers in DeRMET by grade (SPAIN)

better than males at judging emotions (t-test of comparison of mean scores, $p < 0.001$). This aligns with a study of the RMET conducted over 57 countries (Greenberg et al., 2023) and a meta-analysis of the task (Kirkland et al., 2013). In section 4.3, we utilize the a-ToM performance as an input to explain the decisions made in the dictator game.

4.2 Personality traits (Big5)

In psychology, personality has long been acknowledged as a significant influencer of human behavior. A common objective is to categorize individuals across the five major personality traits known as the Big Five: extraversion, agreeableness, openness, conscientiousness, and neuroticism. The Big Five model resulted from the contributions of many independent researchers (Fiske, 1949; Norman, 1967; McCrae and Costa, 1987; Goldberg, 1993). Many studies have employed a wide array of self-report verbal questionnaires, varying greatly in length and complexity, to examine individual traits and how they relate to behavior and life outcomes. Verbal questionnaires can pose significant challenges for young participants, particularly when subtle emotional states are involved. In the case of studies involving children, assessments often rely on variants completed by parents or caregivers. However, for the developmental trajectory of personality, it is crucial to employ a single assessment tool to ensure meaningful comparisons across various age groups.

To address these challenges and ensure consistency in testing individuals aged 6 to 18, we developed a novel, child-friendly version of the Big Five questionnaire, which we call the

46% of participants answer correctly and 32% provide the same wrong answer among the remaining three.

"Pictorial Big 5". In our test, each trait is represented by a set of images at opposite ends of the trait spectrum. Participants are presented with a 5-point Likert scale and asked to select the option that best describes how they perceive themselves.[16] Figure 11 presents the images employed for evaluating the "neuroticism" trait, and Appendix A3 provides the complete personality test, along with instructions and a comprehension example.

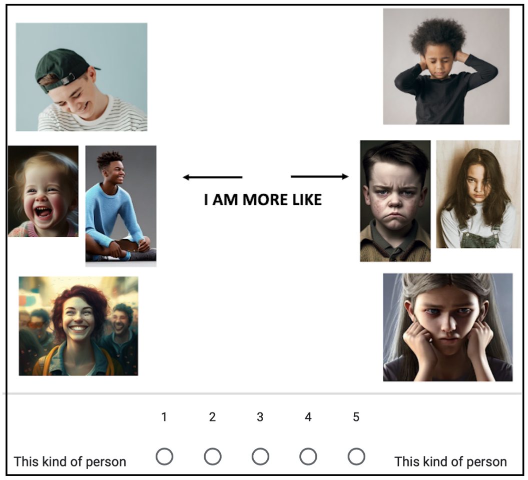


Figure 11: Pictorial Big Five images used to evaluate the individual's level of "neuroticism", with five possible responses ordered from least (1) to most (5).

The Pictorial Big Five questionnaire offers two advantages. Firstly, it does not rely on verbal comprehension, making it suitable for very young children and individuals from diverse linguistic backgrounds, without the need for intricate translations. Secondly, each question presents a natural choice between two explicit polar ends, simplifying the decision-making process for participants.[17] The primary limitation is that we offer just one question for each of the five personality traits, allowing no space for cross-validation. Naturally, this questionnaire is not incentivized. We made it clear that there are no correct or incorrect answers and that responses are confidential.

Figure 12 presents the average score by grade for each personality trait.

[16]We validated this task by enlisting 40 teenagers and adults, who completed our Pictorial Big 5 as well as a 25-question variant of the Five-Factor Model created for adolescents (Rogers and Glendon, 2018). The pictorial design captured well the five personality traits. Indeed, we obtained significant correlations between the responses in the verbal and pictorial versions for Extroversion, Conscientiousness and, to a lesser extent, Neuroticism. For Agreeableness and Openness, the pictorial design predicted the score through the two questions that were most accurately represented by our images. The task and validation materials were registered with the Open Science Framework (OSF) to ensure transparency and accessibility.

[17]This contrasts with verbal questions that typically present only one extreme (e.g., "I Am interested in people") and ask the individual how much they agree with the statement. In our experience, children understand the question better if they can picture the two polar cases.

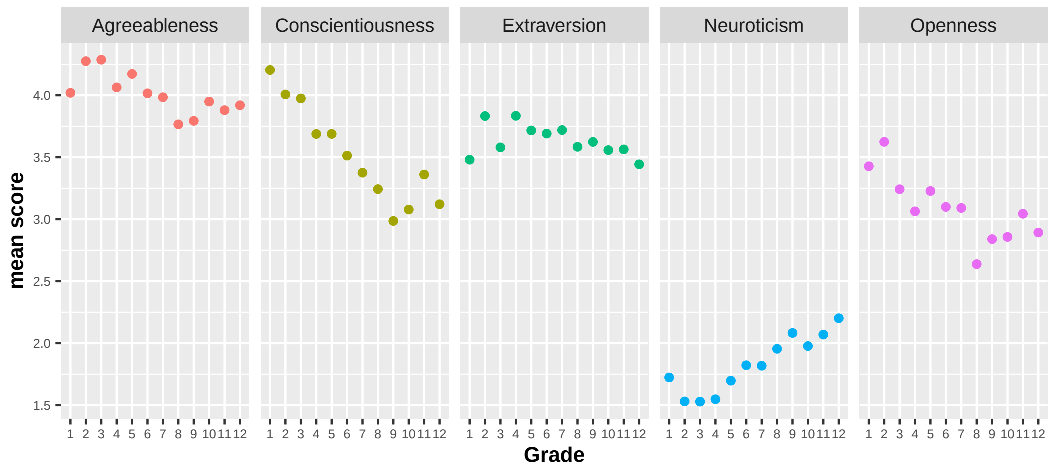


Figure 12: Average score in the Pictorial Big 5 by trait and grade (SPAIN)

We find significant trends in the reports of all five traits with, most notably, strongly decreasing levels of consciousness (PCC between score and age $= -0.27$, $p < 0.001$) and openness (PCC $= -0.16$, $p < 0.001$) and strongly increasing levels of neuroticism (PCC $=$ 0.20, $p < 0.001$) as the individuals age.[18] While the developmental changes are intriguing, our primary focus in this paper lies in investigating the connection between personality and third-party redistribution, an issue we delve into in the following section.

4.3 Non-cognitive abilities and decision-making

In the present analysis, we investigate the potential interplay between ability to interpret emotions, individual personality traits, and meritocratic decision-making in SPAIN. To achieve this, we employ multinomial regressions, with the equitable allocation as the default choice. The dependent variables we examine include "Merit." which represents the likelihood of opting for the meritocratic allocation over the equitable one, and "Comp." which signifies the likelihood of selecting the compensatory allocation instead of the equitable one. To enhance statistical power, we pool data from all variants, incorporating dummy variables for both the token allocation treatment *redistribution* (while the default is *addition*) and the winner selection condition (*luck*, *talent*) (while the default is *effort*). Our established control variables (Age, Age^2, Male, Siblings, STEM) are also included in the analysis. For each individual, we assess the number of correct answers in the DeRMET test (a-ToM) and their responses to the five personality questions. We carry out these

[18]Developmental shifts in the expression and maturation of traits have been documented in longitudinal data (Tackett, 2006). Our result is partly consistent with the finding that mean levels of agreeableness, conscientiousness, and openness decline from late childhood to late adolescence (Soto and Tackett, 2015).

regressions for both the initial choices (columns 1 and 2) and the final choices (columns 3 and 4). The results are presented in Table 4.

	Initial choice		Final choice	
	Merit.	Comp.	Merit.	Comp.
Age	−0.321**	−0.308**	−0.846***	−0.415***
	(0.129)	(0.134)	(0.136)	(0.132)
Age^2	0.011**	0.012**	0.031***	0.016***
	(0.005)	(0.005)	(0.005)	(0.005)
Male	0.472***	0.062	0.680***	0.019
	(0.109)	(0.112)	(0.118)	(0.107)
Siblings	−0.018	−0.137	0.009	−0.138
	(0.147)	(0.148)	(0.155)	(0.142)
STEM	0.0001	−0.021	−0.192*	−0.122
	(0.107)	(0.111)	(0.114)	(0.106)
a-ToM	−0.059**	0.003	−0.022	0.008
	(0.024)	(0.025)	(0.025)	(0.024)
Extraversion	−0.002	0.059	−0.063	0.030
	(0.049)	(0.051)	(0.051)	(0.049)
Agreeableness	−0.136**	−0.082	−0.150***	−0.064
	(0.055)	(0.059)	(0.058)	(0.056)
Conscientiousness	0.071*	0.061	0.158***	0.065
	(0.043)	(0.045)	(0.046)	(0.042)
Neuroticism	0.034	−0.008	0.055	−0.040
	(0.059)	(0.062)	(0.062)	(0.060)
Openness	0.028	0.109**	−0.014	0.022
	(0.042)	(0.044)	(0.045)	(0.042)
redistribution	0.683***	0.107	0.277**	−0.096
	(0.104)	(0.108)	(0.110)	(0.102)
luck	0.110	0.460***	−0.140	−0.027
	(0.125)	(0.130)	(0.136)	(0.120)
talent	−0.068	0.053	−0.044	−0.532***
	(0.128)	(0.136)	(0.135)	(0.130)
Constant	2.078**	0.864	4.943***	2.472***
	(0.859)	(0.901)	(0.888)	(0.876)
Obs.	2255	2255	2255	2255
AIC	4,730	4,730	4,664	4,664

*p<0.1; **p<0.05; ***p<0.01

Table 4: Multinomial regressions of meritocratic and compensatory initial and final choices in SPAIN to explain the effect of emotional intelligence and personality

Participants with higher a-ToM scores exhibit a notable decrease in their likelihood to initially opt for the meritocratic allocation over the equitable one. This implies that a better grasp of others' emotions correlates with increased empathy towards those who are less advantaged, even when the inequality arises from their own actions. However, this effect diminishes when it comes to the final choice. As individuals engage with the task and gain experience, their decisions are influenced by their own performance (as discussed in section 3.3), which seems to counterbalance the emotional considerations.

Personality traits also play a significant role in shaping decisions. Specifically, conscientiousness is linked to a greater likelihood of making meritocratic choices, while agreeableness is associated with a preference for less meritocratic choices, both in the initial and final decisions. Additionally, openness appears to be connected to an inclination for compensatory choices, although this association is observed mainly during the initial stage. These findings are consistent with existing literature that aims to establish connections between personality traits and political preferences. Specifically, they align with previous research that has linked conscientiousness with a preference for right-wing political parties and openness with a preference for left-wing political parties (Vecchione et al., 2011; Furnham and Fenton-O'Creevy, 2018; Ekstrom and Federico, 2019).[19]

Finally, it is noteworthy that the previously highlighted trend of males showing a stronger preference for meritocratic allocations remains significant even after accounting for emotional intelligence and personality factors.

5 Behavior in middle school across populations

Lastly, we make a comparative analysis of the behavior of middle schoolers (grades 6, 7, and 8, ages 11 to 14) across different populations. In addition to the schools previously examined, LILA and SPAIN, we include data from 138 middle schoolers from Thomas Starr King Middle School (KING), a public school located in Los Angeles. These students took part in Experiment 1, that is, in the *addition / effort* variant.

These three schools exhibit vastly different educational environments. They differ in terms of the curriculum offered, with LILA employing a bilingual approach while SPAIN and KING follow a monolingual model. Class sizes also vary significantly, with around 20 students per class in LILA, approximately 30 in SPAIN, and 35 or more in KING. Furthermore, the size of the schools differs substantially, with approximately 200 middle schoolers in LILA and SPAIN, compared to a much larger population of 2000 in KING. In terms of peer groups, many LILA and SPAIN students stay together from kindergarten to 12th grade, while KING comprises middle schoolers who come from various elementary schools located

[19] We performed an analysis of the probability of change between the first and second decision similar to Table 3 but with the independent variables of Table 4. Our findings indicated that individuals who initially made meritocratic choices and scored higher on Conscientiousness were less inclined to modify their allocations. Participants who initially made equitable choices and had higher Conscientiousness scores were more likely to increase the allocation of the winner. Conversely, individuals who initially made equitable choices and had higher a-ToM scores were more inclined to reduce the allocation of the winner. This shows that non-cognitive traits affect how experience shapes behavior.

throughout Los Angeles.

Cultural disparities are also quite significant, with LILA having a predominantly white American and French population living in the United States, KING being primarily composed of Latinos with some white Americans and Asians living in the United States, and SPAIN consisting of Spaniards residing in Spain.

Perhaps most crucial for our research objectives, the socioeconomic backgrounds of these three populations are remarkably diverse. LILA families typically belong to the upper-middle socioeconomic status bracket, SPAIN families fall within the lower and middle socioeconomic status ranges, and KING families predominantly have low or very low socioeconomic status, with 75% living at or below the U.S. poverty level.

In our prior research, we demonstrated significant differences between LILA and KING regarding the rate of equilibrium behavior in dominance-solvable games (Brocas and Carrillo, 2021). We suggested that one major factor contributing to these differences was the contrasting learning environments in these schools. This time, we anticipated substantial variations in choices across schools, primarily due to the marked differences in socioeconomic backgrounds.

Figure 13 illustrates the initial and final allocation choices made by the 138 middle school students at KING for the *addition / effort* variant. For comparison, we also incorporate the choices of the 143 middle schoolers from LILA and 105 middle schoolers from SPAIN. Given the limited number of observations in certain grades (e.g., 22 participants in grade 8 at LILA, 18 in grade 8 at KING, and 24 in grade 6 at SPAIN), we decided to aggregate all middle school grades together for our analysis.

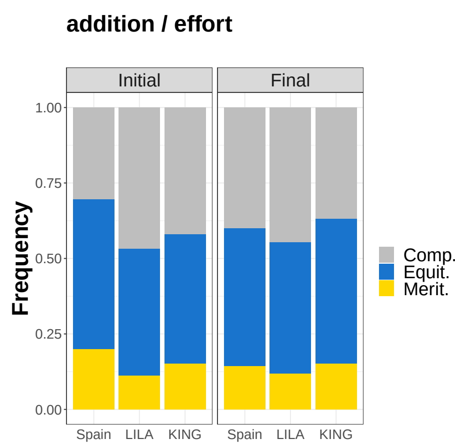


Figure 13: Initial and final allocation choices of middle schoolers (grades 6 to 8, ages 11 to 14) in SPAIN, LILA and KING for the *addition / effort* variant.

Behavior exhibits a striking degree of similarity across all three schools. In each case, a range of 30% to 47% of the population opts for the compensatory allocation, while 11% to 20% choose the meritocratic allocation. Importantly, there are no statistically significant differences observed across schools, both for initial and final choices.

The uniformity in choices across schools is quite surprising, considering the substantial disparities in the educational, social, cultural, and economic environments of the three groups of pre-adolescents in our sample. It suggests the presence of a common factor influencing their perceptions of inequality, which intriguingly tends to peak around this age. It appears that the values children hold regarding inequality and redistribution remain similar regardless of their educational background, family income, and even their personal life prospects.

6 Conclusion

We have presented the findings of a study involving children and teenagers aged 5 to 18 in the United States and Spain. Participants could affect the allocation of rewards between a winner and a loser. We examined three different conditions for determining the competition's winner: *effort*, *talent*, and *luck*. We also investigated two methods for allocating rewards: *redistribution* from winner to loser or *addition* to winner and/or loser. In either case, the final outcomes could only be (6,2), (5,3) or (4,4).

We made several key observations. First, participants in all age groups displayed a reluctance to actively reduce tokens from the winner, leading to more meritocratic allocations in *redistribution*. In contrast, the *addition* treatment revealed slightly less meritocratic choices, accompanied by very noticeable age-related effects, characterized by a rise in ex-post fairness that reached its peak during middle school, followed by a subsequent decrease. Importantly, these patterns were strikingly similar in the United States and Spain. We also observed significant similarities among different socioeconomic levels within the United States. Regardless of the token allocation method, we found that allocations were significantly more meritocratic when the winner was determined through the *talent* task. These allocation decisions were sensitive to participants' past experiences and their perception of performance. They were also notably influenced by individual characteristics and non cognitive traits. A few comments are in order.

The initial token distribution influences the decisions. This behavior is reminiscent of the well-known status quo bias (Samuelson and Zeckhauser, 1988; Kahneman et al., 1991), where individuals tend to prefer maintaining the current state of affairs. However, it could also be attributed to the specific nature of the redistribution choice, which involves taking from someone perceived as deserving. The consistent observation of this behavior across scenarios involving luck, effort, and talent suggests an inherent aversion to take tokens away. The stronger effect in the talent scenario indicates a heightened discomfort in taking from individuals who deserve it. The persistence of the effects across all ages suggests the potential for these tendencies to develop early in life or even to be innate.

The age-related trajectory in the *addition* treatment is more directly comparable to existing developmental studies on other-regarding concerns. The initial increase in allocations to the loser is consistent with existing studies in elementary school: young children tend to be selfish (even spiteful) and they gradually become fair (Fehr et al., 2008; Brocas et al., 2019). The decrease in allocations to the loser when transitioning from middle school to high school mirrors the observed departures from fairness and towards more unequal allocations to account for other aspects such as efficiency, opportunity cost of giving or image concerns (Almås et al., 2010; Fehr et al., 2013; Brocas and Carrillo, 2020b). More generally, prior research has emphasized age-related differences in people's willingness to increase or decrease the payoff of others while keeping their own constant. By contrast, our approach allows us to study the factors that influence merit-based splits by outside observers, and how these factors evolve with age.

Our results regarding the winner selection method indicate that participants made little differentiation between *luck* and *effort*, but they categorically distinguished *talent*. It is possible that, even after gaining experience, they did not view *effort* as requiring significant skills, perceiving it as akin to a game of chance. However, they unequivocally regarded *talent* as deserving of rewards in the event of a win. Through their choices, participants conveyed a willingness to accept some degree of inequality when it is based on a skill that is challenging to acquire.

The findings also highlight the significance of experience. While individuals initially act based on their ideals when determining what is appropriate for winners and losers, 20% to 40% of them reassess their choices when given the opportunity to experience what the players were tasked to do. Crucially, their post-experience decisions tend to be influenced by their own perception of performance. This observation aligns with the principles of experiential learning as discussed in the literature (Kolb et al., 2014), which underscores that true learning occurs when individuals engage in a cycle of concrete experience. Experience can exert two distinct types of effects. On the one hand, it provides factual and impartial insights regarding roles and outcomes. On the other hand, the feedback garnered from experience can potentially trigger more subjective decisions. In our scenario, while participants gain knowledge about the task, theoretically enhancing the informativeness of their decisions, their own perceived performance acts as a reference point shaping these decisions. We conjecture that experience should play a role in other decision-making paradigms where participants are given an opportunity to act in a different role than their own. Due to its dual-sided nature, experience can either enhance collective behavior or create polarization among individuals.

Our findings indicate that environmental factors have a limited overall impact. While cultural factors may have initially influenced preferences and ideals related to redistribution in Spain compared to the US, these differences were dampened in light of experience. Variations in socioeconomic status among teenage participants in the US did not influence allocation decisions either.[20] Research has previously shown that people tend to tolerate higher levels of income inequality when such inequality is more prominent, and their tolerance adjusts to the actual level of inequality. This adaptive mechanism leads to the establishment of social norms (Schröder, 2017), which are evident in numerous empirical studies indicating that some degree of inequality is deemed acceptable by society (Schokkaert and Devooght, 2003; Konow, 2000). Recent findings also suggest that predispositions toward redistribution are consistent across countries, with differences in redistribution patterns likely reflecting variations in cultural stereotypes (Aarøe and Petersen, 2014).

The variability in attitudes toward inequality aligns with findings from previous studies. Existing literature often highlights that disagreement tends to revolve around the source of inequality (Cappelen et al., 2010), which resonates with our observations of differences in the *talent* condition, where a larger proportion of participants favor meritocratic allocations. These perspectives are sometimes linked to political affiliations and individual characteristics (Alesina and Giuliano, 2011; Luttmer and Singhal, 2011; Almås et al., 2020). In our experiment, we identify that this individual heterogeneity is associated with non-cognitive skills. Participants with a higher capacity for empathy tend to exhibit lower levels of merit-based choices and are less reliant on their own experiences when making redistributive decisions. Similarly, individuals who score higher on traits like agreeableness and openness tend to lean towards greater redistribution.

It is noteworthy that males tend to exhibit a higher degree of merit-based choices compared to females. This observation could be linked to the documented tendency of adult women to be more generous than adult men in dictator games (Eckel and Grossman, 1998; Andreoni and Vesterlund, 2001; Brañas-Garza et al., 2018). However, it is important to highlight that these findings have not been consistently replicated in children (Fehr et al., 2013; Brocas et al., 2019; Brocas and Carrillo, 2020b). Additionally, some research has suggested that females are more inclined toward left-wing political ideologies than males (Edlund and Pande, 2002; Furnham and Fenton-O'Creevy, 2018). Although our study was not specifically designed to investigate gender differences, and we lack the appropriate tools to draw definitive conclusions, the robustness of the correlation we observed suggests that males and females may approach concerns related to the well-being of others differently. Interestingly, we also identified a significant gender-related difference in affective Theory of Mind, indicating that females tend to more accurately perceive and interpret emotions.

Inequity is a multifaceted concept, influenced by factors such as origin and available mechanisms, which shape our attitudes toward it. Our study represents a pioneering effort to jointly explore the roles played by these factors, examining how their contributions evolve as we grow, and how they differ across various cultural and economic contexts. From an early age, we embrace inequity, respect talent, and are reluctant to impose taxes on

[20] This contrasts with Bauer et al. (2014), who found that children of parents with low education are less altruistic and more spiteful.

others. Our results also suggest that these tendencies are minimally influenced by wealth and culture-specific conditions but are occasionally shaped by our personal experiences. Designing successful redistributive policies should take these observations into account.

Our study also adds to the research examining the foundational skills and traits that underlie decision-making. Existing literature on other-regarding concerns reports large variability in behavior but fails to pinpoint the reasons for these differences. While cross-population comparisons in other studies suggest that environmental factors are likely to influence our attitudes, relatively little attention has been devoted to examining individual traits. In our research, we have demonstrated that both affective processing and personality factors consistently influence how we express our attitudes towards inequity. We believe that a thorough and methodical assessment of cognitive and non-cognitive traits is essential for comprehending heterogeneity in decision-making and, ultimately, for shedding light on the causal mechanisms behind behavior and outcomes.

The redistributive policy implications of this work are significant. In the redistribution treatment, children take less from the winners resulting in more unequal final allocations. This immediately leads us to consider that, even among children, redistribution between the more and less fortunate is more complicated in a "taxation" situation because the median voter is unlikely to approve it. However, redistribution using "new" funds seems to be much more acceptable.

A second result with clear policy implications is that meritocracy is much stronger in situations where talent rather than luck is involved. This is concerning because talent (at least in our setup) represents the cultural level of the individual, which is either inherent or brought from home, whereas effort reflects what the child struggles with but eventually overcomes. With such views, inequalities are likely to be perpetuated.

References

Lene Aarøe and Michael Bang Petersen. Crowding out culture: Scandinavians and americans agree on social welfare in the face of deservingness cues. *The Journal of Politics*, 76(3):684–697, 2014.

Alberto Alesina and Paola Giuliano. Preferences for redistribution. In *Handbook of social economics*, volume 1, pages 93–131. Elsevier, 2011.

Antonio Alfonso, Pablo Brañas Garza, Isabelle Brocas, Juan D Carrillo, Juan B Gonzalez, and Maria Jose Vazquez. The development of rationality in games with hidden information. Work in progress, 2023.

Ingvild Almås, Alexander W Cappelen, Erik Ø Sørensen, and Bertil Tungodden. Fairness and the development of inequality acceptance. *Science*, 328(5982):1176–1178, 2010.

Ingvild Almås, Alexander W Cappelen, and Bertil Tungodden. Cutthroat capitalism versus cuddly socialism: are americans more meritocratic and efficiency-seeking than scandinavians? *Journal of Political Economy*, 128(5):1753–1788, 2020.

Luis Amador-Hidalgo, Pablo Brañas Garza, Antonio M Espín, Teresa García-Muñoz, and Ana Hernández-Román. Cognitive abilities and risk-taking: Errors, not preferences. *European Economic Review*, 134:103694, 2021.

James Andreoni and Lise Vesterlund. Which is the fair sex? gender differences in altruism. *The Quarterly Journal of Economics*, 116(1):293–312, 2001.

James Andreoni, Deniz Aydin, Blake Barton, B Douglas Bernheim, and Jeffrey Naecker. When fair isn't fair: Understanding choice reversals involving social preferences. *Journal of Political Economy*, 128(5):1673–1711, 2020.

Simon Baron-Cohen, Therese Jolliffe, Catherine Mortimore, and Mary Robertson. Another advanced test of theory of mind: Evidence from very high functioning adults with autism or asperger syndrome. *Journal of Child psychology and Psychiatry*, 38(7):813–822, 1997.

Michal Bauer, Julie Chytilová, and Barbara Pertold-Gebicka. Parental background and other-regarding preferences in children. *Experimental Economics*, 17:24–46, 2014.

Pablo Brañas-Garza, Valerio Capraro, and Ericka Rascon-Ramirez. Gender differences in altruism on mechanical turk: Expectations and actual behaviour. *Economics Letters*, 170:19–23, 2018.

Isabelle Brocas and Juan D Carrillo. Studying decision-making in children: challenges and opportunities. *Journal of Economic Behavior & Organization*, 2020a.

Isabelle Brocas and Juan D Carrillo. The development of social strategic ignorance and other regarding behavior from childhood to adulthood. *Journal of Behavioral and Experimental Economics*, 85:101524, 2020b.

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

Isabelle Brocas and Juan D Carrillo. Why do children pass in the centipede game? cognitive limitations v. risk calculations. USC Working Paper, 2023.

Isabelle Brocas, Juan D Carrillo, T Dalton Combs, and Niree Kodaverdian. The development of consistent decision-making across economic domains. *Games and Economic Behavior*, 116:217–240, 2019.

Alexander W Cappelen, Erik Ø Sørensen, and Bertil Tungodden. Responsibility for what? fairness and individual responsibility. *European Economic Review*, 54(3):429–441, 2010.

Alexander W Cappelen, James Konow, Erik Ø Sørensen, and Bertil Tungodden. Just luck: An experimental study of risk-taking and fairness. *American Economic Review*, 103(4):1398–1413, 2013.

Alexander W Cappelen, Johanna Mollerstrom, Bjørn-Atle Reme, and Bertil Tungodden. A meritocratic origin of egalitarian behaviour. *The Economic Journal*, 132(646):2101–2117, 2022.

Gary Charness and Matthew Rabin. Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869, 2002.

Daniel L Chen, Martin Schonger, and Chris Wickens. otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97, 2016.

Ramon Cobo-Reyes, Jose J Dominguez, Fernando García-Quero, Brit Grosskopf, Juan A Lacomba, Francisco Lagos, Tracy Xiao Liu, and Graeme Pearce. The development of social preferences. *Journal of Economic Behavior & Organization*, 2019.

Thomas Dohmen, Armin Falk, David Huffman, and Uwe Sunde. Are risk aversion and impatience related to cognitive ability? *American Economic Review*, 100(3):1238–1260, 2010.

Thomas Dohmen, Armin Falk, David Huffman, and Uwe Sunde. On the relationship between cognitive ability and risk preference. *Journal of Economic Perspectives*, 32(2): 115–134, 2018.

Catherine C Eckel and Philip J Grossman. Are women less selfish than men?: Evidence from dictator experiments. *The economic journal*, 108(448):726–735, 1998.

Lena Edlund and Rohini Pande. Why have women become left-wing? the political gender gap and the decline in marriage. *The Quarterly Journal of Economics*, 117(3):917–961, 2002.

Pierce D Ekstrom and Christopher M Federico. Personality and political preferences over time: Evidence from a multiwave longitudinal study. *Journal of personality*, 87(2): 398–412, 2019.

Christoph Engel. Dictator games: A meta study. *Experimental Economics*, 14(4):583–610, 2011.

Ernst Fehr and Klaus M Schmidt. The economics of fairness, reciprocity and altruism–experimental evidence and new theories. *Handbook of the economics of giving, altruism and reciprocity*, 1:615–691, 2006.

Ernst Fehr, Helen Bernhard, and Bettina Rockenbach. Egalatarianism in young children. *Nature*, 454(28):1079–1084, 2008.

Ernst Fehr, Daniela Glätzle-Rützler, and Matthias Sutter. The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence. *European Economic Review*, 64:369–383, 2013.

Donald W Fiske. Consistency of the factorial structures of personality ratings from different sources. *The Journal of Abnormal and Social Psychology*, 44(3):329, 1949.

Christina Fong. Social preferences, self-interest, and the demand for redistribution. *Journal of Public economics*, 82(2):225–246, 2001.

Philippa Foot. The problem of abortion and the doctrine of the double effect. *Oxford Review*, (5), 1967.

Adrian Furnham and Mark Fenton-O'Creevy. Personality and political orientation. *Personality and Individual Differences*, 129:88–91, 2018.

David Gill and Victoria Prowse. A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, 102(1):469–503, 2012.

Lewis R Goldberg. The structure of phenotypic personality traits. *American psychologist*, 48(1):26, 1993.

David M Greenberg, Varun Warrier, Ahmad Abu-Akel, Carrie Allison, Krzysztof Z Gajos, Katharina Reinecke, P Jason Rentfrow, Marcin A Radecki, and Simon Baron-Cohen. Sex and age differences in "theory of mind" across 57 countries using the english version of the "reading the mind in the eyes" test. *Proceedings of the National Academy of Sciences*, 120(1):e2022385119, 2023.

Bengt Holmström. Managerial incentive problems: A dynamic perspective. *The review of Economic studies*, 66(1):169–182, 1999.

Daniel Kahneman, Jack L Knetsch, and Richard H Thaler. Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives*, 5(1):193–206, 1991.

Rena A Kirkland, Eric Peterson, Crystal A Baker, Stephanie Miller, and Steven Pulos. Meta-analysis reveals adult female superiority in" reading the mind in the eyes test". *North American Journal of Psychology*, 15(1), 2013.

David A Kolb, Richard E Boyatzis, and Charalampos Mainemelis. Experiential learning theory: Previous research and new directions. In *Perspectives on thinking, learning, and cognitive styles*, pages 227–247. Routledge, 2014.

James Konow. Fair shares: Accountability and cognitive dissonance in allocation decisions. *American economic review*, 90(4):1072–1091, 2000.

Jean-Jacques Laffont and Jean Tirole. Using cost observation to regulate firms. *Journal of political Economy*, 94(3, Part 1):614–641, 1986.

John A List. On the interpretation of giving in dictator games. *Journal of Political economy*, 115(3):482–493, 2007.

Erzo FP Luttmer and Monica Singhal. Culture, context, and the taste for redistribution. *American Economic Journal: Economic Policy*, 3(1):157–79, 2011.

Sonny Mattek, Samantha Chavez, Julia L Berkowitz, M Ida Gobbini, and Robert S Chavez. 100 peers: An open set of controlled emotional facial expressions for use in psychology and neuroscience experiments. Unpublished, 2021.

Robert R McCrae and Paul T Costa. Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology*, 52(1): 81, 1987.

Bregtje Gunther Moor, Zdeňa A Op de Macks, Berna Güroğlu, Serge ARB Rombouts, Maurits W Van der Molen, and Eveline A Crone. Neurodevelopmental changes of reading the mind in the eyes. *Social cognitive and affective neuroscience*, 7(1):44, 2012.

Thomas Erik Neuber. Egocentric norm adoption. Technical report, ECONtribute Discussion Paper, 2021.

WT Norman. 2800 personality trait descriptors: Normative operating characteristics for a university population. department of psychology, university of michigan. *Ann Arbor*, 1967.

Hessel Oosterbeek, Randolph Sloof, and Gijs Van De Kuilen. Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental economics*, 7(2):171–188, 2004.

Rike Pahnke, Anett Mau-Moeller, Alfons O Hamm, and Alexander Lischke. Reading the mind in the eyes of children test (rme-ct): Development and validation of a complex emotion recognition test. *Frontiers in Psychiatry*, 11:376, 2020.

Jonas Pontusson and David Rueda. Inequality as a source of political polarization: A comparative analysis of twelve oecd countries. *Democracy, inequality, and representation*, pages 312–353, 2008.

Mary E Rogers and A Ian Glendon. Development and initial validation of the five-factor model adolescent personality questionnaire (ffm–apq). *Journal of Personality Assessment*, 100(3):292–304, 2018.

William Samuelson and Richard Zeckhauser. Status quo bias in decision making. *Journal of risk and uncertainty*, 1:7–59, 1988.

Erik Schokkaert and Kurt Devooght. Responsibility-sensitive fair compensation in different cultures. *Social Choice and Welfare*, 21(2):207–242, 2003.

Martin Schröder. Is income inequality related to tolerance for inequality? *Social Justice Research*, 30(1):23–47, 2017.

Christopher J Soto and Jennifer L Tackett. Personality traits in childhood and adolescence: Structure, development, and outcomes. *Current Directions in Psychological Science*, 24 (5):358–362, 2015.

Jennifer L Tackett. Evaluating models of the personality–psychopathology relationship in children and adolescents. *Clinical Psychology Review*, 26(5):584–599, 2006.

Hulda Thorisdottir, John T Jost, Ido Liviatan, and Patrick E Shrout. Psychological needs and values underlying left-right political orientation: Cross-national evidence from eastern and western europe. *Public Opinion Quarterly*, 71(2):175–203, 2007.

Michele Vecchione, Harald Schoen, José Luis González Castro, Jan Cieciuch, Vassilis Pavlopoulos, and Gian Vittorio Caprara. Personality correlates of party preference: The big five in five big european countries. *Personality and Individual Differences*, 51 (6):737–742, 2011.

Appendix A. Details of design and procedures

A1. Sample of instructions (variant *addition* / *effort*)

Today, we are going to play a few games with you. In all the games, you will earn points that will be converted into money and placed in your virtual wallet. At the end of the experiment we will send to your school email address an Amazon e-giftcard with the money you made. You will get many points today, so hopefully you will be able to get a nice gift card.

TROPHY GAME

There are two tasks in this game: the “Slider Task” and the “Observer Task.” The “Slider Task” is a competition. You will be matched in pairs with another student in the room, but you will not know who this person is (see Figure 14 for the corresponding slides).

[SLIDE 1]

You will see sliders with the cursor in a certain position between 0 and 100, in this example it is at 32. Your job will be to move the cursor exactly to the position 50. Whoever puts more cursors in 50 wins. Let’s try one on your screen. This is not for real. It’s just for you to know how it works.

[Start SLIDER TRAINING]

Put the cursor at 50. Once you have done it, press NEXT. When the time comes to play the Slider Task, you will see many sliders on your screen. It will look like this.

[SLIDE 2]

You will have 90 seconds to put as many cursors as possible at 50. You will have to work fast to beat the other person.

[SLIDE 3]

At the end of the 90 seconds, if you put more cursors at 50 in your screen than the other student puts in his, you will win the trophy game and receive 3 tokens. if you put fewer cursors than the other student, you will lose and get only 1 token. Each token is worth $1, so you will earn either $1 or $3. Is this task clear?

The other task you are going to play is the “Observer Task”. In this task, we give you 4 extra tokens that you have to allocate between a winner and a loser of the slider task.

[SLIDE 4]

This is you [point to the screen]. We give you 4 free tokens.

[SLIDE 5]

All you have to do is split the 4 tokens between the winner and the loser. There is only one rule. You have to give at least one token to each. So, do you want to give 3 tokens to the winner who already has 3 tokens and give 1 to the loser who already has 1 token? Or do you want to give 2 and 2? Or do you want to give 1 to the winner and 3 to the loser? Remember each token is worth $1. Your screen will look like this.

[SLIDE 6]

You allocate tokens by clicking on the dashed tokens. If you change your mind, you can move tokens from one player to another. Once you are happy, press OK. Remember that you will play both the Slider Task and the Observer Task. However, it is very important to understand that you will play with different persons. Let's look at this example.

[SLIDE 7]

You play the Observer Task and allocate tokens to two students who play the Slider Task against each other. You also play the Slider Task but with different people. So, you will earn 1 or 3 tokens with the Slider Task and also receive 1, 2 or 3 tokens from an Observer. Since each token is worth $1, you will get between $2 and $6 dollars just with this game.

OK. The computer will now select the groups.

We will start by playing the "Observer Task" and ask you to tell us how you would like to allocate your tokens between two students in this room who will play the "Slider Task" in just a moment. How many to the winner and how many to the loser.

[Start THIRD PARTY DICTATOR]

Please choose the allocation and press OK.

[Wait for participants to finish]

We will now play the "Slider Task". In a moment, you will see all the sliders. When you are ready press START and try to put as many cursors as possible at 50. At the end, you will see how many cursors you managed to put at 50.

[Start SLIDER TASK]

You will now see a reminder telling you what you chose in the Observer Task. Think about it. If you are happy with the choice you made, click OK. If you have changed your mind, click CHANGE and re-do the allocation. Then press OK.

[Start THIRD PARTY REALLOCATION]

On your screen, you will now see a questionnaire. Please answer the questions. We will tell you at the end of the session, whether you won the Slider Task and how many tokens you got in total.

[Start QUESTIONNAIRE 1 about difficulty and performance]

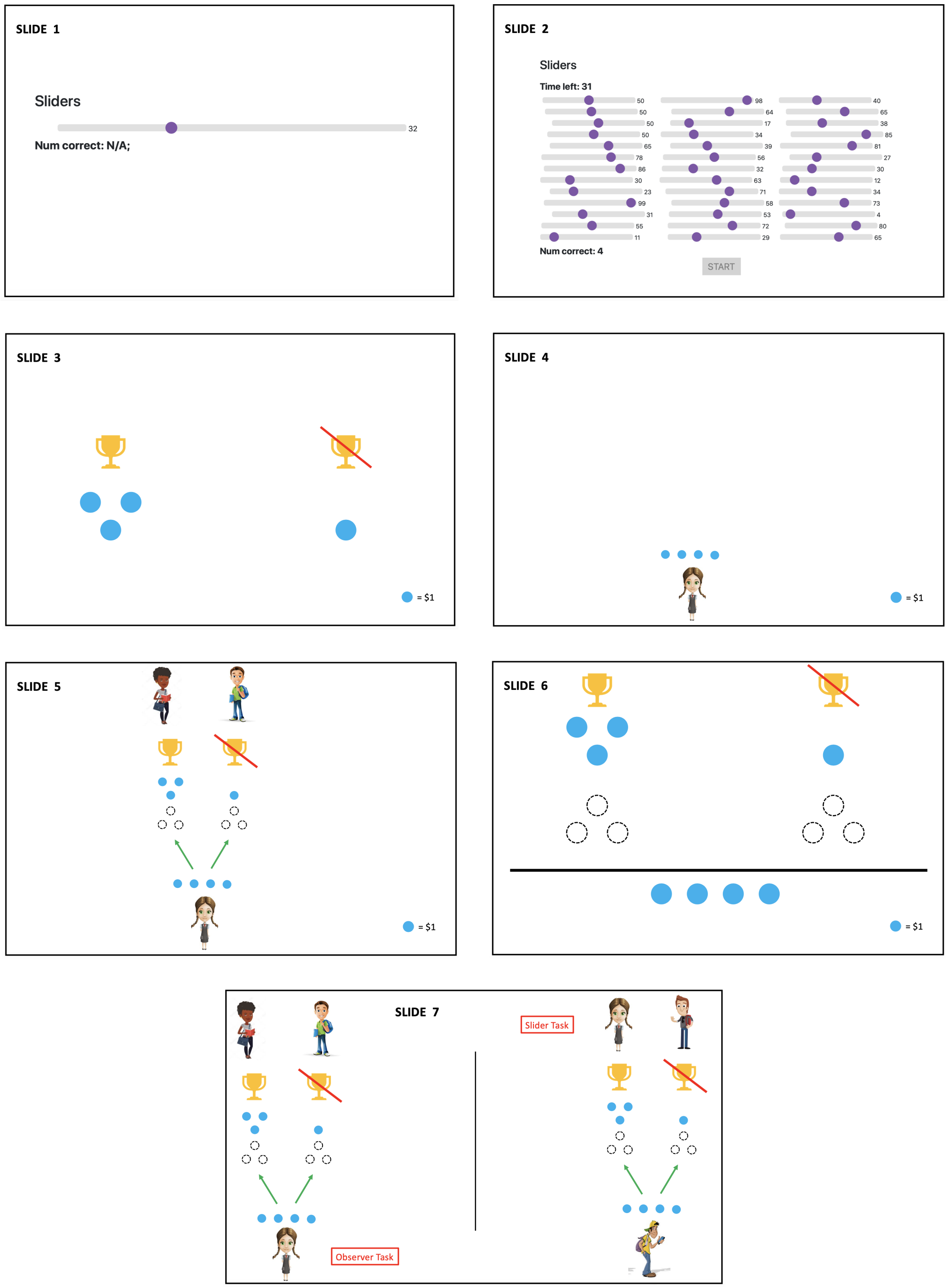


Figure 14: Slides projected on screen for instructions

A2. Developmental Read the Mind in the Eyes Test (DeRMET)

Instructions.

Below, you will see a series of images with a boy's eyes expressing a certain emotion. Look closely and choose, from the four options, the one that best represents the person's emotion. When you are sure of your answer, press NEXT, and the next image will appear. Once you press NEXT, you will not be able to change your answer. For each correct answer, you will receive 10 points, which is equivalent to 10 cents. The more correct answers, the more money you will earn.

Note to the instructor.

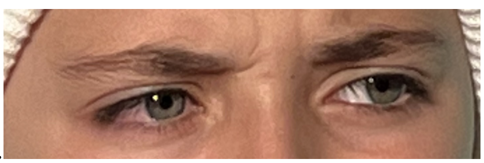
Each image is presented in a different page with four emotions to choose from (see the list below). Participant must select exactly one. After a choice is made, the next image appears on the screen and the participant cannot go back. There is no feedback between images.

List of images.

1. 

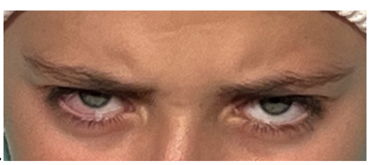
2. 

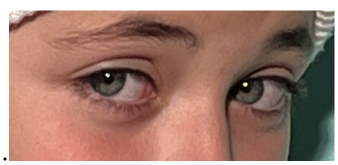
3. 

4. 

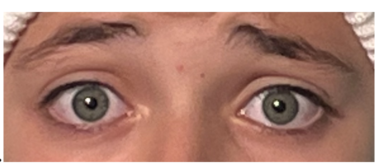
5. 

6. 

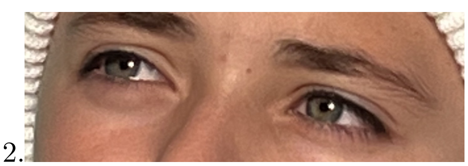
7. 

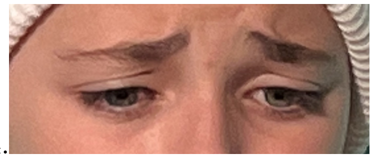
8. 

9. 

10. 

11. 

12. 

13.  14.  15. 

16. 

List of emotions for each image.

Image #	Emotions				% correct
1	1. Bored	2. Confused*	3. Hopeful	4. Guilty	55.2
2	1. Sneaky	2. Angry	3. Guilty	4. Tired*	79.4
3	1. Guilty	2. Tired	3. Confused	4. Sad*	64.7
4	1. Admiring	2. Disapproving*	3. Shocked	4. Bored	45.9
5	1. Surprised	2. Admiring	3. Happy*	4. Sneaky	74.3
6	1. Confused	2. Ashamed*	3. Scared	4. Tired	69.3
7	1. Angry*	2. Scared	3. Sad	4. Worried	97.1
8	1. Excited	2. Confused	3. Thinking	4. Flirting*	67.8
9	1. Excited*	2. Happy	3. Hopeful	4. Flirting	75.3
10	1. Disgusted	2. Sad	3. Angry	4. Scared*	88.2
11	1. Scared	2. Bored	3. Disgusted*	4. Angry	69.8
12	1. Surprised	2. Happy	3. Thinking*	4. Disapproving	86.3
13	1. Hopeful	2. Sneaky*	3. Thinking	4. Excited	86.8
14	1. Guilty	2. Bored	3. Disgusted	4. Worried*	59.7
15	1. Happy	2. Surprised	3. Hopeful*	4. Sneaky	45.9
16	1. Surprised*	2. Happy	3. Flirting	4. Admiring	66.7

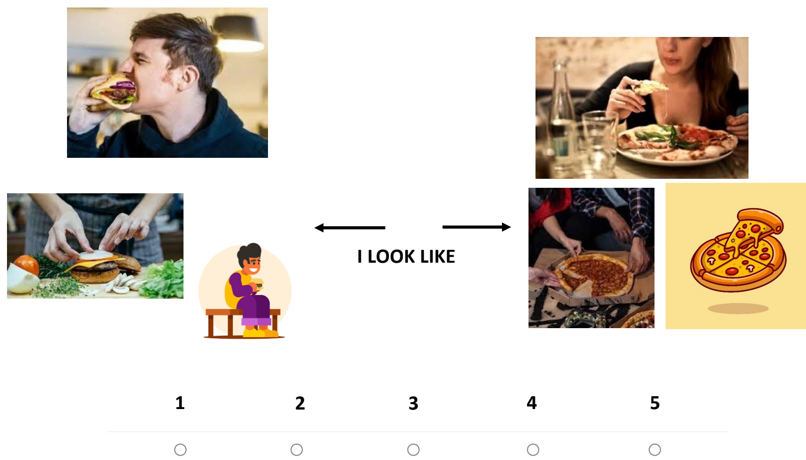
[* = correct answer. The last column reports for each image the aggregate empirical probability of finding the correct answer]

A3. Pictorial Big 5 personality questionnaire

Instructions.

Next, you will see a series of pictures. The group of pictures on the left represents one type of behavior, while the group of pictures on the right represents the opposite type of behavior. Look at the pictures in each group and choose the option that best represents you.

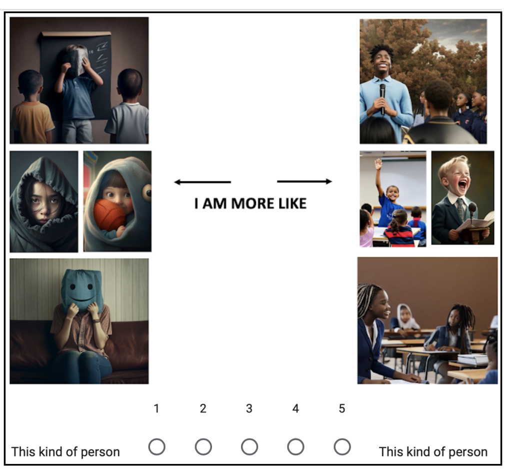
For example, imagine you saw this screen



We would be asking if you definitely prefer hamburgers (1), mostly hamburgers (2), both equally (3), mostly pizza (4) or definitely pizza (5). The questions are about your character or behavior. When you are sure, press NEXT, and another series of pictures will appear. There are no right or wrong answers, and please remember that your responses are confidential.

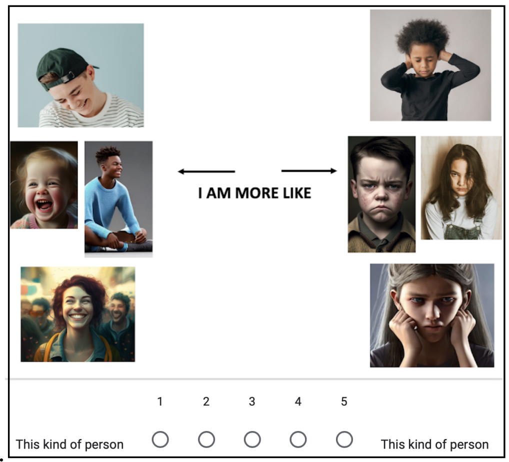
Questionnaire.

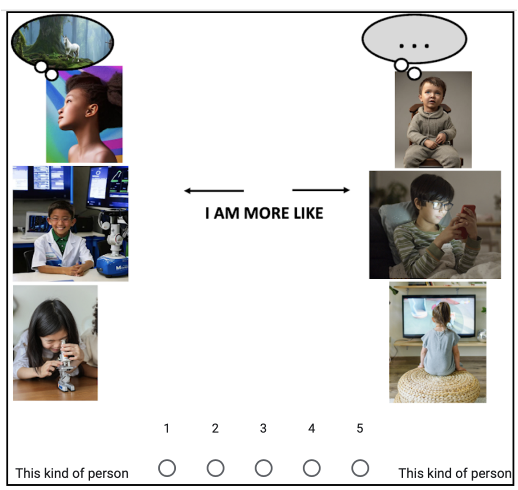
The groups of pictures presented to our participants are provided below. They correspond to the following traits: extraversion (1), agreeableness (2), conscientiousness (3), neuroticism (4) and openness (5).

1. 

2. 

3. 
I AM MORE LIKE
1 2 3 4 5
This kind of person
This kind of person

4. 
I AM MORE LIKE
1 2 3 4 5
This kind of person
This kind of person

5. 
I AM MORE LIKE
1 2 3 4 5
This kind of person
This kind of person

Appendix B. Additional Analyses

B1. Initial Choices in the *redistribution* treatment

In Figure 15, we present the same information as in Figure 4 except that we consider initial instead of final choices. It is important to remember that participants are not initially aware of the possibility to change the allocation after experiencing the task.

	SPAIN	
	(1)	(2)
Age	0.011**	0.010**
	(0.005)	(0.005)
Age^2	−0.00004**	−0.00003**
	(0.00002)	(0.00002)
Luck	−0.055	−0.061
	(0.058)	(0.058)
Talent	−0.113*	−0.117**
	(0.059)	(0.059)
Male		0.109**
		(0.049)
Siblings		0.041
		(0.069)
STEM		−0.008
		(0.050)
Constant	4.345***	4.325***
	(0.345)	(0.348)
Obs.	1,082	1,082
Adj. R^2	0.006	0.008

*p<0.1; **p<0.05; ***p<0.01

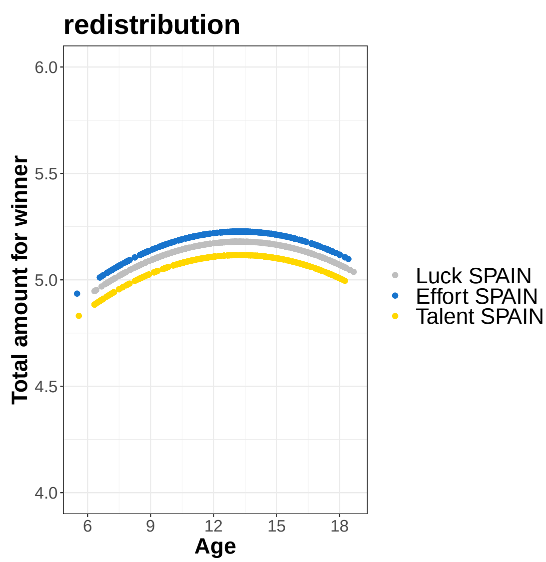


Figure 15: OLS regressions of **initial** allocation to winner (left) and representation of quadratic best fit (right) as a function of age for *luck*, *talent* and *effort* (SPAIN) in the *redistribution* treatment.

Initial choices look differently from final choices in *redistribution*. Participants in ages 10 to 14 initially start with lower levels of redistribution, which results in a slightly concave redistribution function. These participants decide to transfer more to the loser after experiencing the task, which implies an overall increase in redistribution and a flat function with no effect of age. The change is most noticeable in *effort* and *luck*, so that at the end of the game winners are slightly better-off if they achieved their success in the *talent* competition.

B2. Initial Choices in the *addition* treatment

Figure 16 presents the same information as Figure 5 except that we consider initial instead of final choices.

	LILA		SPAIN	
	(1)	(2)	(3)	(4)
Age	−0.029***	−0.030***	−0.014***	−0.014***
	(0.006)	(0.006)	(0.004)	(0.004)
Age^2	0.0001***	0.0001***	0.00004***	0.00004***
	(0.00003)	(0.00003)	(0.00001)	(0.00001)
Male		0.093		0.161***
		(0.066)		(0.043)
Siblings		0.055		0.025
		(0.074)		(0.058)
STEM		−0.079		−0.0004
		(0.066)		(0.043)
Luck			−0.082	−0.073
			(0.052)	(0.052)
Talent			0.079	0.078
			(0.053)	(0.053)
Constant	6.531***	6.503***	6.076***	5.974***
	(0.383)	(0.386)	(0.286)	(0.289)
Obs.	465	465	1,196	1,196
Adj. R^2	0.047	0.048	0.026	0.036

*p<0.1; **p<0.05; ***p<0.01

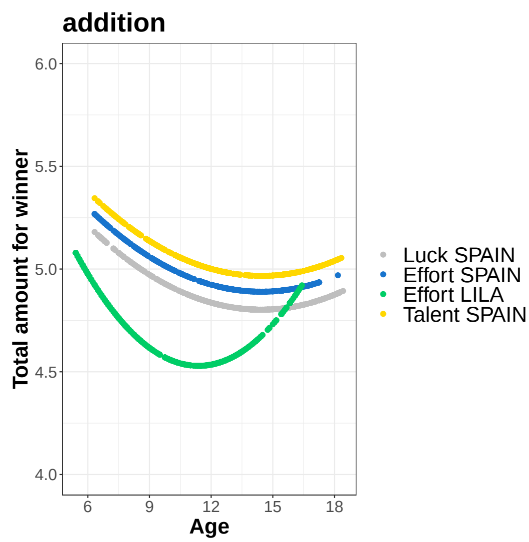


Figure 16: OLS regressions of **initial** allocation to winner (left) and representation of quadratic best fit (right) as a function of age, winner selection method and population in the *addition* treatment.

In contrast with the *redistribution* treatment, initial and final choices are very similar in the *addition* treatment. This is not surprising since, as described in Figure 3b, few participants change their allocation and those who do are equally likely to move in either direction. Figure 16 confirms that such pattern applies to all ages.

The differences between the *redistribution* and *addition* treatments emphasized in Appendices B1 and B2 are interesting. We conjecture that in *redistribution* there is a strong default (or anchor) that implies few transfers from winner to loser. This tendency to inaction needs to be overridden and such active change is more likely to occur after experiencing the task (final choice) and in cases where inequality between winner and loser might be perceived as "less fair" (luck and effort). By contrast, in *addition* the participant must make an active choice from the outset. This makes redistribution more common and changes between initial and final choices less frequent (but still not uncommon).

B3. Final choices in *redistribution* with separate regressions

Figure 17 presents the same analysis of final choices in the *redistribution* treatment as Figure 4, except that regressions are performed separately for the three winner selections methods.

	SPAIN *effort*		SPAIN *talent*		SPAIN *luck*	
	(1)	(2)	(3)	(4)	(5)	(6)
Age	0.008	0.007	−0.001	−0.002	0.002	0.002
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Age^2	−0.00002	−0.00002	0.00000	0.00000	−0.00001	−0.00001
	(0.00003)	(0.00003)	(0.00003)	(0.00003)	(0.00003)	(0.00003)
Male		0.191^{**}		0.150^{*}		−0.021
		(0.085)		(0.081)		(0.081)
Siblings		−0.016		−0.004		0.076
		(0.118)		(0.114)		(0.112)
STEM		−0.006		−0.110		0.032
		(0.088)		(0.082)		(0.081)
Constant	4.329^{***}	4.353^{***}	5.293^{***}	5.307^{***}	4.887^{***}	4.822^{***}
	(0.605)	(0.609)	(0.536)	(0.542)	(0.583)	(0.591)
Obs.	383	383	333	333	362	362
Adj. R^2	-0.0002	0.007	-0.002	0.003	-0.0003	-0.007

*p<0.1; **p<0.05; ***p<0.01

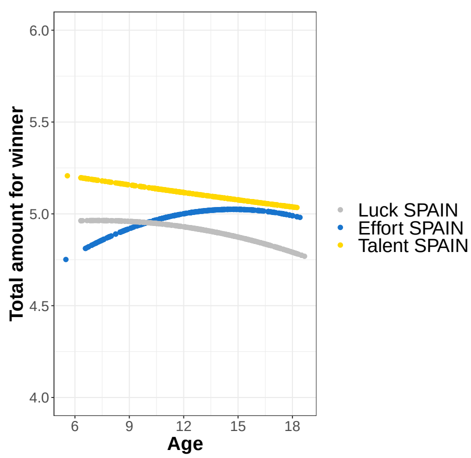


Figure 17: OLS regressions of final allocation and quadratic best fit in the *redistribution* treatment, performed separately on each winner selection method (luck, effort and talent).

The age trajectory (or rather lack of) is similar when regressions are performed separately. Perhaps the most noticeable difference is the surprising result that our youngest participants (age 9 and below) tend to redistribute less to the loser in luck than in effort. Interestingly, the documented tendency of males to choose more meritocratic allocations remains in place for effort and talent while it disappears when winners are chosen randomly.

B4. Final choices in *addition* with separate regressions

Figure 18 presents the same analysis of final choices in the *addition* treatment as Figure 5, except that regressions in SPAIN are performed separately for the three winner selections methods.

	LILA *effort*		SPAIN *effort*		SPAIN *talent*		SPAIN *luck*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age	-0.027^{***}	-0.027^{***}	-0.019^{**}	-0.019^{**}	-0.010	-0.012^{*}	-0.037^{***}	-0.037^{***}
	(0.006)	(0.006)	(0.009)	(0.009)	(0.007)	(0.007)	(0.006)	(0.006)
Age^2	0.0001^{***}	0.0001^{***}	0.0001^{**}	0.0001^{**}	0.00003	0.00004	0.0001^{***}	0.0001^{***}
	(0.00003)	(0.00003)	(0.00003)	(0.00003)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
Male		0.116^{*}		0.329^{***}		0.178^{**}		0.149^{**}
		(0.066)		(0.081)		(0.075)		(0.069)
Siblings		-0.054		0.082		0.058		0.030
		(0.074)		(0.113)		(0.098)		(0.094)
STEM		-0.152^{**}		0.043		0.028		-0.056
		(0.067)		(0.081)		(0.074)		(0.070)
Constant	6.398^{***}	6.451^{***}	6.104^{***}	5.840^{***}	5.800^{***}	5.770^{***}	7.855^{***}	7.760^{***}
	(0.385)	(0.387)	(0.595)	(0.597)	(0.520)	(0.521)	(0.447)	(0.453)
Obs.	465	465	329	329	395	395	453	453
Adj. R^2	0.040	0.049	0.009	0.053	0.002	0.011	0.116	0.120

*p<0.1; **p<0.05; ***p<0.01

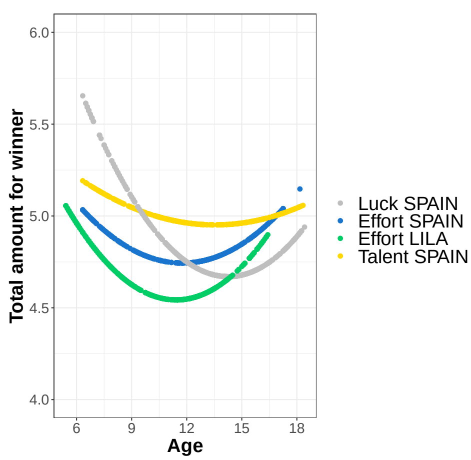


Figure 18: OLS regressions of final allocation and quadratic best fit in the *addition* treatment, performed separately on each population and winner selection method (luck, effort and talent).

Again, results are very similar when regressions are performed separately, which reinforces one of our main findings, namely the robustness of the convex age-trajectory of redistribution in *addition*. As in *redistribution* (see Figure 17), we again notice that younger participants allocate less to the loser in luck than in effort, with a surprisingly large effect in grades 1 and 2. Males are significantly more meritocratic in both populations and under all winner selection methods (this time, including luck).

B5. Behavior in our adult populations: USC and teachers

With some exceptions (e.g., Cobo-Reyes et al. (2019)), studies with children do not recruit an adult population. In Experiment 1, we included two adult control groups that follow identical procedures: 72 USC undergraduates (USC) and 34 teachers at LILA (TEACHERS). They constitute two interesting benchmarks. After graduating, the majority of students at LILA attend well-ranked colleges (including USC), so LILA and USC match 'reasonably' well. As for the other group, TEACHERS, they are older (often in their 30s or 40s) working professionals instead of students, but they live in the same environment as LILA.

Figure 19 presents the initial and final choices of the two adult populations in the *addition / effort* variant. For comparison, we also include the behavior of the 48 high schoolers in LILA (grades 9 and 10) and the 96 high schoolers in SPAIN (grades 9, 10 and 11).

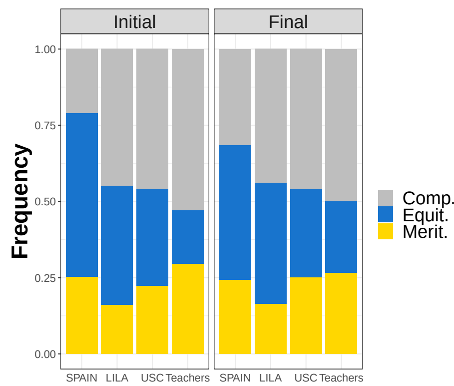


Figure 19: Initial and final allocation choices of high schoolers (SPAIN and LILA) and adults (USC and TEACHERS) for the *addition / effort* variant.

Behavior is similar between LILA high schoolers, USC and TEACHERS both for the initial and the final allocations (payoff for the winner between 4.71 and 4.79, with no statistically significant differences across populations or across decisions). There are also no differences in final allocations between SPAIN and any other population. There are two significant differences. One is between initial and final allocation in SPAIN (t-test, p = 0.048), which is similar to the one found for middle schoolers (see section 5). The other is between the initial allocations in LILA and SPAIN (t-test, p = 0.003), a result that has also been emphasized in section 3.2.3 when we considered the full age trajectory. The absence of statistically significant differences between high schoolers and adults (particularly between SPAIN and USC or TEACHERS in initial choice) can be partly due to a low statistical power due to a reduced number of observations, and therefore it should not be overemphasized. At the same time, similarities are not excessively surprising. Indeed, these students are approaching adulthood, so it is reasonable that their sharing preferences as impartial observers are close to that of their older peers when they face the exact same decisions.