Why do children pass in the centipede game? Cognitive limitations v. risk calculations *

Isabelle Brocas
University of Southern California and CEPR

Juan D. Carrillo
University of Southern California and CEPR

September 2023

Abstract

Children and adolescents from 8 to 16 years old play the centipede game in the laboratory, where non-equilibrium behavior (passing) can occur for two reasons: an inability to backward induct (cognitive limitation) or a decision to best respond to the empirical risk and take a measured chance (behavioral sophistication). We find that logical abilities develop gradually. While young participants are (as expected) least likely to perform backward induction, those who do, tend to over-estimate the ability of their peers to behave similarly. With age, participants gradually learn to think strategically and to best respond to their beliefs about others. Overall, the centipede game is an ideal test case for studying the development of abilities, as it disentangles the causes for passing in young children and in teenagers. Interestingly, shrewdness does not transform into earnings, and we document for the first time a game of strategy where payoffs monotonically *decrease* with age. Finally, experience heavily disciplines behavior in all age groups.

Keywords: developmental decision-making, centipede game, backward induction, risk-taking.

JEL Classification: C72, C90.

*We thank the members of the Los Angeles Behavioral Economics Laboratory (LABEL) for their insights and to Chris Crabbe for his outstanding programming skills. We are grateful to the staff of the Lycée International de Los Angeles (LILA) –in particular Emmanuelle Acker, Nordine Bouriche, Mathieu Mondange and Anneli Harvey– for their help and support running the experiment in their school. 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.

1 Introduction

Behavior in strategic games results from the combination of cognitive ability, inferences about others and preferences. Game theory provides normative predictions of behavior under the assumption of selfish preferences and perfect foresight. However, we often observe departures from these predictions. The centipede game is an excellent example of a game where theoretical predictions yield low payoffs and are counterintuitive to most individuals. In a typical centipede game, two players face two escalating piles of money and take a finite number of turns deciding whether to stop (in which case the player takes the largest pile, leaves the smallest one to the partner and ends the game) or to pass (in which case the partner faces a similar decision with larger piles of money). Backward induction prescribes stopping at each stage, which prevents players from taking advantage of the jointly increasing rewards.

Since its introduction by Rosenthal (1981), the centipede game has fascinated theorists, who describe it as a paradigmatic paradox of backward induction (see e.g., Aumann (1992); Reny (1992); Ben-Porath (1997)). Authors have discussed several intuitive reasons why participants may decide to pass initially: limited cognition (inability to perform backward induction), non-selfish motivations (preferences for fairness, cooperation and efficiency) and inferences about mutual uncertainty (about the players' cognitive ability and preferences). It is worth noting that once the first player has passed in their first opportunity, the second player cannot rely on backward induction to predict the rival's subsequent moves. This provides an argument for the second player to pass in their first opportunity which, in turn, offers a motivation for the first player to pass in the first place. More generally, under some assumptions, the theoretical literature has shown that passing may be consistent with certain definitions of rationality (Reny, 1992).

The experimental literature, initiated with McKelvey and Palfrey (1992), overwhelmingly sides with the intuitive expectation and against backward induction: few participants stop immediately... but they do not continue all the way to the end, either. As in most robust paradoxes, deviations are attenuated but not eliminated with experience, and they depend on specific elements, such as game length, payoff manipulations and number of players (Krockow et al., 2016). The persistent departures could indicate that people are not able to apply backward induction logic. This hypothesis has been investigated in two studies. Palacios-Huerta and Volij (2009) argue that experts (chess Grandmasters) are more likely to play the backward induction equilibrium than students. Levitt et al. (2011) did not share that result. More importantly, they found no relationship between ability to backward induct (measured in a challenging dominance solvable task) and the decision to stop immediately. Their result suggests that cognitive limitations are not the main driver of early passing. This is in contrast to other strategic settings where equilibrium behavior largely depends on cognitive ability (Gill and Prowse, 2016; Fe et al., 2022).

Our study adopts an ontogenic approach. It leverages documented age-related changes in both cognition and Theory-of-Mind (ToM) –the ability to read the rival's intentions and form beliefs– to better disentangle the contribution of each of these two factors to choices in the centipede game. We recruit children and teenagers and study the change in behavior with age in a version of the centipede game. Our hypotheses rely on known developmental changes in cognition and ToM abilities throughout development. We run 5 rounds of a centipede game with linearly increasing payoffs for the player who stops and constant payoffs for the one who does not. Rounds 1, 3, 4 and 5 have ten stages whereas round 2 has only four stages. The short four-stage game serves as a diagnostic tool for the ability to backward induct. Therefore, contrasting behavior in rounds 1 and 2 allows us to disentangle between passing due to limited cognition and passing for other strategic motives at different ages. Studying the evolution of behavior across long games (rounds 1, 3, 4 and 5) allows us to determine how age and the gradual development of abilities affects learning over the course of the experiment.

There has been an increased recent interest in studying decision making by children and adolescents in experimental economics (see Sutter et al. (2019) and List et al. (2021) for excellent surveys). The study of games of strategy has revealed some interesting developmental trajectories that track the development of preferences (see e.g., Murnighan and Saxon (1998); Harbaugh and Krause (2000)), reasoning and ToM (see e.g., Sher et al. (2014); Czermak et al. (2016)) and largely correlates with measures of cognitive ability (Fe et al., 2022). Also, while children master the most basic false belief ToM tasks by age 5, the more general ToM ability continues to develop throughout adolescence (Royzman et al., 2003). Finally, young children (5 to 8 years old) already backward induct in very simple two-step settings (Brocas and Carrillo, 2020b) and they show gradual improvements in their ability to foresee future events and input them in current calculations although, in some cases, such ability reaches a plateau (Brocas and Carrillo, 2021).

The first and major finding is that the amount of passing the first time the long and short games are played decreases monotonically with age, from elementary school to young adulthood. Differences are due to age-related differences in *both* the ability to perform backward induction and the ability to read the rival's intentions. For elementary school children (ages 8 to 11), finding and playing the equilibrium are intimately related. Children either find and play the equilibrium in both the long and short games (the minority), or they do not play it in either game (the majority). In other words, backward induction ability drives behavior. However, this correlation disappears starting in middle school. While the fraction of participants who stops immediately in the short game increases significantly with age, this choice does not predict their behavior in the long game. It suggests that, just like the experts in Levitt et al. (2011), passing for children at age 11 and above is often a deliberate decision for strategic considerations. As they grow, participants become more able to evaluate the empirical risk associated with passing and fine tune their stopping time, a consequence of the increased development of ToM.

The second related finding is the monotonic decrease in payoffs with age (from elementary school to adulthood) both in the long and in the short game. To our knowledge, this is a first in a game of strategy. The astute reader may find it unsurprising, given that higher payoffs of younger participants are a consequence of their collectively larger deviations. One should notice, however, that such argument could also apply to other games where joint deviations increase the payoffs of players (e.g., prisoner's dilemma), and yet it has never been documented previously. More generally, high ability is often associated to a better empirical reading of the situation, which results in higher payoffs (Proto et al., 2019, 2022). Under such definition, one would expect more passing and larger gains as individuals get older, at least in the long game.

Our third finding relates to the change in behavior during the experiment. Participants of all ages stop earlier as they play more rounds. This tendency is a natural result of the asymmetry of incentives: 'losing' one round pushes subjects to preempt their rival in the next whereas 'winning' is unlikely to trigger deferral. However, it does not result in complete unraveling; by the end of the game the majority of participants still stop between the second and fourth stage. Interestingly, the change is more pronounced for the younger participants. It implies that choices and payoffs are very similar across ages by the end of the experiment. Nevertheless, initial choices are still informative about subsequent behavior: participants who play Nash in the first two rounds (stop immediately) continue to stop earlier than their peers in the remaining rounds.

The paper is organized as follows. In section 2, we detail our population and discuss our design choices. In section 3, we report the choices and payoffs in the first round of the long and short versions of the game, with age as the main treatment factor. In section 4, we investigate the evolution in behavior during the five rounds of the experiment. Concluding remarks and alleys for future research are collected in section 5.

2 The game

2.1 Design and procedures

The paper studies the behavior of children and adolescents in the well-known centipede game, which was first introduced by Rosenthal (1981) and first studied in the laboratory by McKelvey and Palfrey (1992). Since working with young participants presents important methodological challenges, we have developed some methodological guidelines in Brocas and Carrillo (2020a) which we closely follow in this paper.[1] In particular, we develop a graphical, story-based version of the game.

Population. The experiment was conducted with 315 school-age students from 3rd to 10th grade at the Lycée International de Los Angeles (LILA), a private school in Los Angeles.[2] We also included a control adult population (A) consisting of 72 college students from the University of Southern California (USC).

With some exceptions (see e.g., Cobo-Reyes et al. (2020)), experiments with children and adolescents typically do not feature an adult population and, instead, rely on prior research for a comparison. We believe it is helpful to include an adult control group that follows *identical* procedures to establish a behavioral benchmark (Brocas and Carrillo, 2020a). This is especially important when procedures are slightly modified, as it is the case here. Ideally, the control population should also be as similar as possible to the treatment population of children. We argue that a private university like USC is a reasonable match for LILA.[3] Table 1 summarizes the participants by grade and age in our sample.

	LILA								USC
Grade	3	4	5	6	7	8	9	10	A
Age	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	18-23
# indiv.	53	40	31	54	67	22	14	34	72

Table 1: Summary of participants by grade and population

Procedures. We ran 27 and 6 sessions at LILA and USC with 8 to 14 participants each.

[1] In a nutshell, the principles are: (i) adapt the length and procedures to a population with limited attention span; (ii) offer age-appropriate incentives (possibly different at different ages); (iii) present the task in a way that subjects are not required to possess strong analytical skills to participate (e.g., graphical interfaces and simple instructions); (iv) understand, describe and compare the children population, and (v) include a benchmark adult comparison group whenever possible.

[2] High schoolers from 11th and 12th grade did not participate in the study because they were taking or preparing for national exams during this period.

[3] After high school, a large fraction of students from LILA go to well-ranked colleges in North America, including USC and universities in the UC system.

Sessions at LILA were run in classrooms during school hours with individual partitions to preserve anonymity. Sessions at USC were run at Los Angeles Behavioral Economics Laboratory (LABEL) in the Department of Economics at USC. For each school-age session, we tried to have male and female participants from the same grade, but for logistic reasons we sometimes had to mix participants from two consecutive grades. Procedures were identical in all cases, except for payments as explained below.

The experiment consisted of two games programmed in 'oTree' (Chen et al., 2016) and implemented on touchscreen PC tablets through a wireless closed network. We started with a third-party dictator game. After a short break, we moved to the centipede game. The findings of the dictator game are discussed in a different article (Brocas and Carrillo, 2022). The two games are sufficiently different that we are not concerned about cross contamination. Nevertheless, to avoid any potential issues, we always performed the two games in the same order (dictator followed by centipede) with random and anonymous re-matching of subjects between the two games. Most importantly, we did not announce any result regarding the first game until the second game was finished.

Centipede Game. In developmental game theoretic studies, it is key to provide a simple, graphical interface and a story which is sound, accessible and appealing to children and adolescents. This is all the more important when the age span is large, as in our study (8 to 16 years old). With this goal in mind, we developed our version of the Centipede Game, called "Going Down the Street". Figure 1 presents a screenshot of the game from the perspective of the second player at the second decision node. In our narrative, the blue and orange players walk together down a street and must decide in which house they enter. When arriving at a house, the player whose color matches that of the house decides for both whether to enter or continue to the next house. If they reach the last house, there is no possibility of continuing. Whenever they enter a house, players collect their color coded payoff and the game ends. (Appendix A provides the full set of instructions).

Rounds and stages. Participants were matched in pairs, assigned a role as player 1 or player 2 (blue or orange in our game) and played in Round 1 (R1) the ten-stage Centipede game described in Figure 1. Then they were randomly rematched, kept the same role and played in Round 2 (R2) a four-stage version of the same game, with only the first four houses. After that, they played Rounds 3, 4 and 5 (R3, R4, R5) of the original ten-stage version with alternating roles and random re-matching between rounds.[4]

Contrasting behavior in R1 and R2 allows us to disentangle between different motives

[4]This means that a player had one role in R1, R2 and R4 and the other role in R3 and R5.

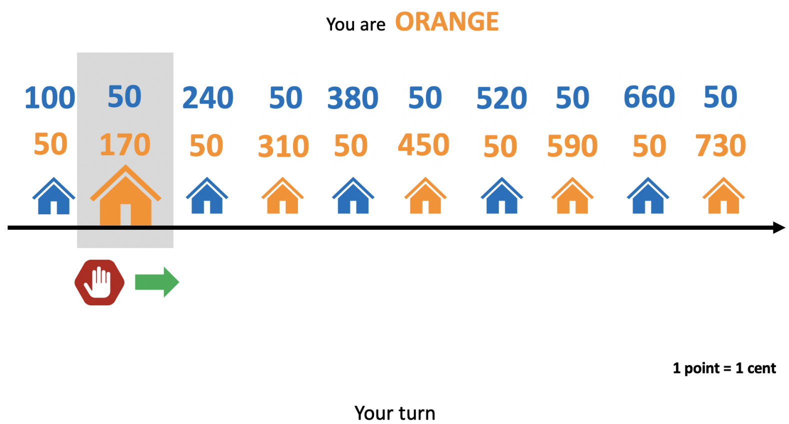

Figure 1: Screenshot of "Going down the street" game

for early passing. Indeed, while R2 serves as a diagnostic test for the ability to perform backward induction, R1 helps reveal other drivers of behavior.[5] In particular, participants familiar with backward induction but who consider other strategic factors are less likely to pass in the first stage of R2 than in the first stage of R1. Since players receive feedback and have the opportunity to adjust their strategies, including R3, R4 and R5 helps us study whether and how quickly participants in the different age groups make these adjustments.

Remark. Our setting uses different parameters compared to the original experiment (McKelvey and Palfrey, 1992): ten (instead six) stages for the long game, a constant (instead of increasing) payoff for the participant who does not stop, and a linearly (instead of exponentially) increasing payoff for the participant who stops.[6] Ten stages enriches the number of passing options. A constant payoff for the "loser" removes the possibility that both participants strictly win by passing several times which, compared to the traditional setting, reduces the role of social preferences as a driver of behavior. Indeed, while a strict efficiency maximizer would choose to pass, it would result in severe losses for oneself. A linearly increasing payoff for the "winner" ensures that the empirical variance in payments, to which children are particularly sensitive, is high but not massive. We conjecture that the first feature increases the incentives to pass in the early stages for strategic considerations

[5]Given our population, we found that a short centipede was the best way to test for backward induction, as it is most comparable, average in difficulty, and does not require new instructions.

[6]Researchers have studied experimentally many variants of the game, including length (McKelvey and Palfrey, 1992), payoff structure (Fey et al., 1996), incentives (Parco et al., 2002), number of players (Rapoport et al., 2003) and game presentation (Nagel and Tang, 1998). Qualitative properties of the empirical behavior are usually robust to such modifications (see Krockow et al. (2016) for a survey).

while the second and third feature decrease those incentives. Of course, the key feature of course is that all participants –including our control group– play the exact same game.

Payments and duration. Following the guidelines discussed in Brocas and Carrillo (2020a), we used different mediums of payment for different ages. This comes at added effort but we view it as a key choice. Indeed, for an incentive system to be optimal it must equalize the *value of rewards* across individuals, not the rewards themselves. Money is usually the most adequate medium of payment precisely because it is valued most similarly by participants. However, this is not the case when age is a factor. Young children prefer desirable objects for their immediate enjoyment rather than the equivalent amount of money, which they understand and appreciate, but it is likely to be administered by the parents.[7] School-age students from 6th grade and above and control adults earned $0.01 per point paid immediately at the end of the experiment in cash (at USC) or with an amazon giftcard (at LILA, given that cash transfers are not allowed in the school). For elementary school students (grades 3 to 5), we set up a shop with 20 to 25 pre-screened, age-appropriate toys and stationery that children find appealing (bracelets, erasers, figurines, die-cast cars, trading cards, apps, calculators, earbuds, scented pens, etc.).[8] Before the experiment, we took the children to the shop, showed the toys they were playing for and explained their point prices. At the end of the experiment, subjects learned their point earnings and were accompanied to the shop to exchange points for toys.[9]

The game studied in this paper lasted around 30 minutes. The entire experiment never exceeded one school period (50 minutes). Average monetary earnings in the Centipede Game were $8.25 (LILA grade 6 and above) and $7.75 (USC). Participants also earned $2 to $6 in the other game, and there was a $5 show-up fee paid only to the adult population to correct for differences in the opportunity cost of time. For children in grades 3 to 5, point prices were calibrated in a way that all children obtained at least three toys, although there was large variance in number and type. We spent on average $5 per child in toys, which is considerably higher than most experiments with elementary school children.

[7] In other words, just like in an experiment with chimpanzees and humans it would not make sense to give money to the former or sips of fruit juice to the latter, rewards must be adapted to the age of our participants.

[8] While the age cut-off between toys and money is arbitrary, it coincides with school practices: only after 6th grade the school offers amazon giftcards as prizes for performance in intra-school activities (math puzzles, art shows, literature competitions, etc.).

[9] The procedure emphasizes the importance of accumulating points while making the experience enjoyable. Children at this age are familiar with this method of accumulating points that are subsequently exchanged for rewards since it is commonly employed in arcade rooms and fairs.

Clustering. To increase the statistical power, some of our analysis groups the school-age participants into three naturally clustered age-groups: grades 3-4-5 (**C1**, ages 8-11, 124 participants), grades 6-7 (**C2**, ages 11-13, 121 participants) and grades 8-9-10 (**C3**, ages 13-16, 70 participants).[10] The control adult population consists of USC undergraduates (**C4**, ages 18-23, 72 participants) and it is included only for relevant comparisons. Regressions use either the age in months of the participants or dummies for age-group.[11] Unless otherwise noted, when comparing aggregate choices we perform two-sided t-tests of mean differences. Standard errors are clustered at the individual level whenever appropriate and we use a p-value of 0.05 as the benchmark threshold for statistical significance.

2.2 Hypotheses

We formulate three hypotheses based on current knowledge regarding behavior by adults on the centipede game and behavior by children and adolescents in other backward induction games.

Hypothesis 1 *Stopping (and consequently earnings) changes monotonically with age:*
(a) *In R1, older participants stop later and earn more than younger participants.*
(b) *In R2, older participants stop earlier and earn less than younger participants.*

Hypothesis 2 *Within each age group, there is no correlation between early stopping in R1 and R2.*

Hypothesis 3 *Participants of all ages stop earlier as they play more rounds of the game.*

As discussed in section 1, there are two possible reasons for passing: inability to backward induct (cognitive limitation) or decision to best respond to the empirical risk and take a measured chance (behavioral sophistication). The second motive is not present in a very short (4-stage) centipede game, R2, where initial passing is empirically detrimental. Thus, in R2, passing is driven exclusively by cognitive limitations. Since the cognitive ability to backward induct increases with age (Brocas and Carrillo, 2021), we expect less passing as individuals get older (**H1b**). By contrast, in a long (10-stage) centipede game,

[10] While it could be argued that 8th graders should be grouped with the other middle schoolers, we chose otherwise mainly to reach a similar sample size in all school-age groups (LILA has recently expanded the size of elementary and middle school, which explains the higher number of participants in those grades). Results are similar (though statistical significance is affected) if we consider other grouping methods.

[11] For regressions with age, we do not use the adult participants to avoid having the coefficient estimates driven by the age of the control group.

R1, strategic passing in initial stages is, on expectation, optimal. Again, we expect that older participants will be more sophisticated and read better the situation than their younger, more naïve counterparts. In this case, it will result in more passing (**H1a**).

Overall, we expect average earnings to increase with age in R1 due to a better empirical read of the game by older participants. We also expect average earnings to decrease with age in R2, with the group of younger players financially benefitting from their lower cognitive ability.

Our next hypothesis addresses more directly whether the strategy of early passing in the long game is due to behavioral sophistication or cognitive limitation. Since the first effect is not present in the short game, the correlation between choices in R1 and R2 can address it. There are arguments for both hypotheses in the literature on adults (Palacios-Huerta and Volij, 2009; Levitt et al., 2011). When it comes to individuals who are developing their cognitive abilities, we expect that passing in the long game will often reflect empirically motivated strategic sophistication and therefore will not be followed by passing in the short game. Thus, as summarized in **H2**, we do not expect a correlation in any of our age-groups between early passing in R1 and early passing in R2.

Finally, **H3** predicts a behavioral change towards the equilibrium as the game is played repeatedly, independently of the participant's age. In this game, half the participants 'lose' in a given round, which provides incentives to stop earlier the next time. On the flip side, the participants who 'win' a given round need to form a counterfactual belief of what could have happened had they waited another turn. This asymmetric structure of learning, specific to this game, should result in earlier stopping over the course of the experiment for individuals of all ages, since they can all perform this inductive logic argument.

3 Initial choices in long and short games

3.1 Choice and earnings as a function of age

We first study stopping behavior by age-group in R1 and R2. We compute the proportion of pairs who stopped at any given stage, with the understanding that only one of the individuals in the pair (player 1 or player 2) had a choice in each stage. For visual ease, we group the R1 stopping decision in four categories: stages 1-2, 3-4, 5-6, 7-10. The results are presented in Figure 2.

The developmental trajectory is *qualitatively* similar in R1 compared to R2, featuring a decrease in the tendency to stop at the latest stages. In R1, participants become more inclined to stop at the intermediate stages (3-4 and 5-6) with age rather than at the

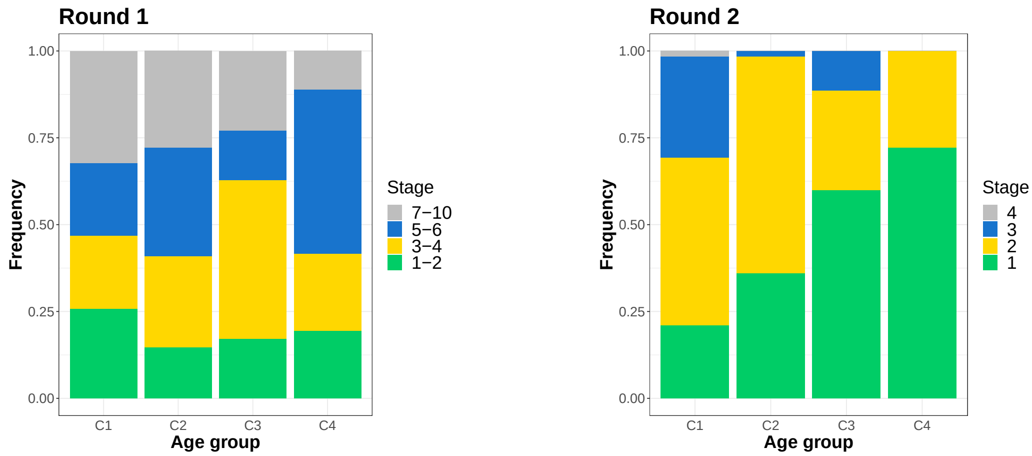

Figure 2: Stopping decision in Round 1 (left) and Round 2 (right)

later stages (7-10). The stopping distributions in C1 and C4 as well as C3 and C4 are significantly different (chi-squared tests, p = 0.021 and p = 0.014 respectively). Stopping in the participant's first opportunity (1-2) is roughly constant and rather infrequent in all age-groups. In R2, participants also stop significantly earlier as they grow older, with the average stopping stage moving from 2.11 in **C1** to 1.28 in **C4**, (**C1** different from all age groups, p $<$ 0.0001, **C2** different from **C4**, p = 0.013). This also reflects an increased understanding of the strategic forces at play: as they age, participants learn to anticipate how the game will unravel if they don't stop immediately and to apply backward induction reasoning. Choices in R1 and R2 are, however, quantitatively very different. The increase in early stopping with age is much more pronounced in R2 than in R1. Also, the likelihood of immediate stopping is, as any behavioral theory would predict, much higher in R2 than in R1, and this difference increases with age. Indeed, the percentages of immediate stopping among player 1 participants in R1 and R2 are, 8% and 21% in **C1**, 5% and 36% in **C2**, 6% and 60% in **C3**, and 5% and 72% in **C4**. The behavior of adults in R2 confirms that social preferences or aversion to opportunity gains is an unlikely explanation for early passing given the parameters adopted in our version of the game.

To refine this analysis, we focus on the school age population and we conduct OLS regressions of the number of stages before stopping (columns 1 and 2) as a function of *Age* (in months), gender (*Male* = 1), the interaction between age and gender, and the participant's role (*Player2* = 1). To construct our independent variable, we group together two consecutive stages (1-2, 3-4, etc.), to correct for the fact that players 1 and 2 can only stop at the odd and even stages, respectively. We also run OLS regressions of the payoff of the participant who stops as a function of the same variables (columns 3 and 4), this time without grouping two stages together.[12] We do not include our control undergraduate students in these regressions to not bias the age coefficients of the regression in their direction. The results are presented in Table 2.

	Stages before stop		Payoff of winner	
	R1	R2	R1	R2
Age	-0.012*	-0.005**	-1.651*	-0.642**
	(0.006)	(0.001)	(0.748)	(0.200)
Male	-2.036*	-0.185	-285.1*	-25.9
	(1.015)	(0.278)	(142.1)	(38.9)
Age×Male	0.009	0.001	1.290	0.071
	(0.007)	(0.002)	(0.968)	(0.266)
Player2	-0.469*	-0.292***	4.320	29.2***
	(0.180)	(0.050)	(25.2)	(6.98)
const.	5.015***	2.015***	662.1***	242.0***
	(0.789)	(0.214)	(110.5)	(29.9)
Adj. R^2	0.138	0.250	0.092	0.206
# obs.	158	158	158	158

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2: OLS of stopping stage and payoffs in R1 and R2 in school-age population

The regressions strengthen the previous results. Indeed, the stopping stage significantly decreases with age in R1 (p = 0.029), and the effect is much stronger in R2 (p = 0.002). Interestingly, males stop earlier than females in R1 but not in R2. This suggests that there is no gender differences in the ability to respond purely strategically (in R2) while gender-related motives are at play in R1. Table 2 also reveals that player 2 chooses less frequently to pass than player 1, although this should be interpreted with caution, since it is conditional on being given the same opportunities (which is endogenous to the model).

In terms of payoff, this behavioral trajectory results in a decrease in payoffs from age 8 to age 16: younger children end up stopping later and getting higher payoffs compared to their older peers. Notice that the payoff (Π) of the participant who stops is a linear transformation of the stopping stage (t): $\Pi = 30 + 70t$. It is therefore natural that regressions on payoffs and on stopping stage yield very similar findings. Significance levels

[12] In our formulation, the payoff of the participants who do not stop before their rival is constant and equal to 50 independently of the stopping stage, so we can ignore them for our purposes.

are not identical only because we group stopping stages in pairs but we consider payoffs separately. In turn, this methodology allows us to unveil an interesting effect of role: player 2 stops, on average, at an earlier *opportunity* (columns 1 and 2) but at a later overall *stage* (columns 3 and 4), hence obtaining a higher payoff conditional on being the player who stops.

Summing up, our results provide support for **H1b** but not **H1a**. Consistent with their more developed cognitive abilities, older participants stop earlier in the short game as an unsatisfying yet individually optimal strategy. Surprisingly, they also stop earlier in the long game, thereby foregoing the risky but ultimately paying strategy of multiple passing.

3.2 Factors of behavior

Since participants keep the same role in R1 and R2, we can use the choice of player 1 in the first round of R1 and R2 (stop S or continue C) to infer drivers of behavior. Arguably, individuals who play Nash (SS) strictly apply backward induction. Those who stop immediately in R2 but continue in R1 (CS) seem to know both how to apply backward induction logic *and* also read the intentions of others and best respond. Those who wait in both games (CC) are likely motivated by seeking high rewards ignoring the risk involved (which is quite significant for the case of R2). Finally, individuals who stop in R1 but not in R2 (SC) probably miss all strategic aspects of the game. Figure 3 reports the proportion of player 1 types in each age group.

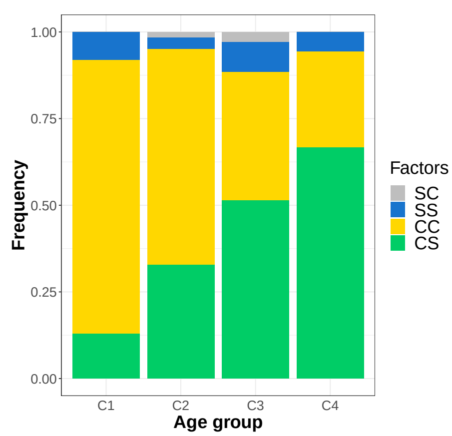

Figure 3: Joint behavior in first round of R1 and R2

The proportion of players who understand backward induction and possess Theory of mind (*CS*) increases with age while the proportion of players motivated by risky rewards (*CC*) decreases. By contrast, *SS* and *SC* are infrequent in all age groups. The distribution of factors in **C1** is significantly different from all other age groups ($p < 0.03$). The distribution in **C2** is significantly different from **C4** ($p = 0.007$) and there is no difference between **C3** and **C4**. Also, and consistent with the previous analysis, the proportion of players who pass in the first stage of R1 among those who stop immediately in R2 (formally, the proportion of *CS* types over *CS* and *SS* types) is significantly smaller in **C1** (61.5%) compared to **C2**, **C3** and **C4** altogether (89.9%, $p = 0.026$).

3.3 Choice across rounds

To further investigate the choices of individuals across rounds, we run a Probit regression of the first player's choice in the first stage of R2 (Stop(R2) = 1) as a function of the number of stages in which that player chose to pass in R1 (*PassR1*). We focus on the school-age population and we use **C2** as the benchmark age group. We include dummies for the other age-groups (*C1* and *C3*) as well as interaction terms, and a dummy for gender. It is important to acknowledge the imperfect nature of this exercise. Indeed, our data is censored given that an individual can only make a choice in stage t if the partner passed in stage $t-1$. Results are reported in Table 3 and sheds some interesting light on the relationship between cognition and behavior.

We observe that repeated passing in R1 is a predictor of passing in the first stage of R2 only in our youngest age group (**C1**). Whenever our elementary school participants realize that the unraveling logic of the game dictates immediate stopping, they apply the argument *indiscriminately in both rounds*. In other words, if they manage to find the equilibrium strategy, the play it, without asking whether their partner is capable of the same logic or willing to apply it always. While foregoing these inferences is not payoff-detrimental in R2, it is in R1.[13]

By contrast, our middle- and high-schoolers are more discerning of the situation. Passing in the long game does not predict their behavior in the short version and is therefore not an indication of limited cognition. The behavior of participants who understand the incentives to stop immediately in R2 is still consistent with more sophisticated inferences about their partner's expected behavior in R1.[14]

[13] Ideally, and as counterbalance, we would like to run another treatment where participants play first the four-stage version and then the ten-stage version.

[14] As robustness checks, we show that the average stopping for participants in **C1** is significantly lower among individuals who stop immediately in R2 than among those who do not, whereas no differences exist

	Stop(R2)
PassR1	-0.256
	(.164)
C1	0.756
	(.643)
PassR1×*C1*	-0.998*
	(.410)
C3	0.490
	(.626)
PassR1×*C3*	-0.035
	(.266)
Male	0.396
	(.247)
const.	0.105
	(.419)
AIC	170.2
# obs.	158

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3: Individual Probit of stopping in R2 as a function of choice in R1

Overall, recall that **H2** stated that individuals who pass in the long game do so for strategic payoff considerations and therefore such behavior will not be predictive of their choice in the short game. This hypothesis is verified for our older participants (middle- and high-schoolers) but not for our youngest ones. In this latter group, finding and playing the equilibrium are two intimately related concepts: young children who understand backward induction will typically not have the theory of mind ability to realize that others will not be able or willing to apply the equilibrium logic to its ultimate consequences.

3.4 Summary

Our analysis reveals a clear developmental trajectory. In both rounds, older participants stop earlier than their younger peers, but these age-related differences are modulated by the length of the game, with a steeper gradient in the short than in the long game.

The short R2 round is instructive as a diagnostic tool for cognitive ability. The contrast between behavior in R1 and R2 reveals an important interplay between cognitive ability and Theory-of-Mind (ToM), the ability to read the rival's intentions and form beliefs which

in **C2** and **C3**. Also, adding **C4** to the regression in Table 3 and using a different age group as benchmark does not change the main result that stopping early in R1 predicts stopping in the first stage of R2 for **C1** only for **C1** (data omitted for brevity but available upon request).

develops throughout childhood and adolescence (Royzman et al., 2003). Because rivals take risks empirically in R1 but not in R2, it is optimal to depart from strict backward induction logic and factor that belief in the formulation of an empirical best response. In that respect, we obtain three key findings.

First, many participants in **C1** are lured by the high rewards in late stages and do not stop in either game. Still, a significant fraction (21%) behave consistently with backward induction logic in R2. Interestingly, these individuals fail ToM and do not anticipate that the majority of their peers will experiment in R1. As a result, they *over-apply* their skill. Also, the fact that the least cognitively able participants stop very late suggests that their reward seeking attitude is due to a lack of a proper cost-benefit trade-off of the risks involved rather than an intrinsic preferences for risks. This is consistent with the known tendency of children until age 11 to focus on salient features (Miller, 2002).

Second, while some participants in **C2** and **C3** are still unsuccessful at backward inducting in R2, the behavior of those who succeed does not drive their choice in R1. They have acquired enough ToM abilities to assess empirical risk with reasonable accuracy. This behavior is analogous to Levitt et al. (2011) who show that immediate stopping by chess Grandmasters in the centipede game is not related to their ability to apply backward induction reasoning in other diagnostic tasks. Last, participants in **C2** and **C3** are still prone to stop, on average, later than adults. However, while late stopping is best explained by the combination of a lack of cognition and salience effects in **C1**, it likely reflects significant markers of differential behavior during teenage time, such as impulsivity and immature risk avoidance processes (Li, 2017).

Finally, the behavioral trajectory described above implies that earnings monotonically decrease with age. While the result is expected in R2 (where the game configuration forces players to stop immediately or else incur in even higher losses), it is surprising in R1 since it precludes collective solutions that are payoff improving.

4 Behavior in long games

In this section, we analyze differences in behavior in the long games R1, R3, R4 and R5. We want to both identify systematic differences (or similarities) between age groups and reveal how experience affects behavior.

4.1 Unraveling within age groups

Figure 4 describes the evolution of the stopping strategy in each age group separately. Table 4 reports the average stopping stage in R1 and R5 in each age group.

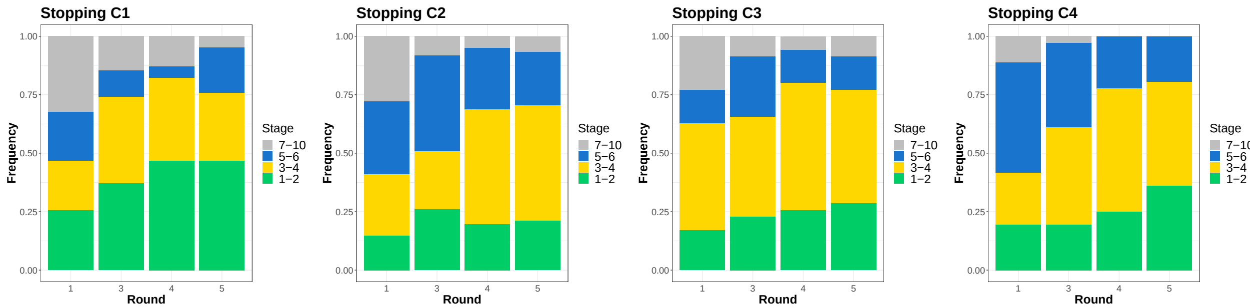

Figure 4: Evolution of the stopping strategy by age group

	C1	C2	C3	C4
Round 1	4.95	5.05	4.43	4.58
	(0.34)	(0.26)	(0.37)	(0.31)
Round 5	3.15	3.79	3.49	3.11
	(0.32)	(0.38)	(0.43)	(0.41)
difference	1.80***	1.26***	0.94*	1.47***

(st. errors in parenthesis); * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4: Average stopping stage in R1 and R5

As we can see from Figure 4, there is a noticeable tendency in all age groups to stop earlier as the experiment progresses. The differences between the average stopping stage in the first and last round reported in Table 4 are significant in all age groups (all p-values < 0.05). By the time they reach the fifth round, participants stop between one and two rounds earlier than they did in the first round.

The effect is most pronounced in the youngest population, who stopped the latest in R1, followed by the control group. This change in behavior over rounds is consistent with existing findings in adult populations (McKelvey and Palfrey, 1992; Fey et al., 1996). It is also quite natural. Participants whose rivals stop before they do receive feedback that incentivizes them to stop earlier in the next round. By contrast, those who stop before their rivals do not know the counterfactual of delaying stopping. This makes them less likely to postpone stopping in future rounds. Such asymmetry of incentives for future behavior between "winners" and "losers" results in progressive –though not full– unraveling.

To further investigate the evolution of play within each age group, we run OLS regressions of the stopping stage as a function of round (R1, R3, R4, R5) and gender, separately for each age group. We use R3 as the default round, to cleanly determine the changes in behavior for the entire length of the experiment. The results are presented in Table 5.

	Stages before stop			
	C1	**C2**	**C3**	**C4**
R1	1.241***	0.980**	0.429	0.683*
	(0.365)	(0.295)	(0.421)	(0.282)
R4	-0.739*	-0.330	-0.581	-0.393
	(0.286)	(0.240)	(0.324)	(0.252)
R5	-0.641*	-0.283	-0.533	-0.848**
	(0.307)	(0.262)	(0.356)	(0.324)
Male	-1.543***	-0.231	-0.324	-0.451
	(0.395)	(0.319)	(0.430)	(0.362)
const.	4.682***	4.164	4.213***	4.159***
	(0.375)	(0.249)	(0.353)	(0.320)
Adj. R^2	0.196	0.072	0.026	0.098
# obs.	248	244	140	140

(clustered st. errors in parenthesis); * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: OLS Regressions of evolution in stopping stage in each age group

The regressions confirm that the stage at which participants stop decreases over the course of the experiment in all age groups. Once again, the change is most pronounced in **C1**, featuring significant differences between all rounds. In **C2**, the change is most abrupt early in the experiment. By contrast, changes are not significant in **C3**. An immediate implication is the corresponding decrease in payoffs for participants in all age groups as the game progresses. We also observe that the gender effect previously noted in R1 and R2, whereby males are more likely to stop earlier, persists but is significant only in **C1**.

4.2 Age effects

To better assess the determinants of stopping and identify potential changes as the experiment progresses, we conduct the same OLS regressions as in Table 2 for Rounds 3, 4 and 5. The results are presented in Table 6.

In Table 5, we emphasized changes in behavior between R1 and R3 and we noted that they were largest for children in elementary, followed by children in middle-school.

	Stages before stop			Payoff of winner		
	R3	R4	R5	R3	R4	R5
Age	-0.005	-0.001	-0.003	-0.716	-0.067	-0.367
	(0.006)	(0.004)	(0.004)	(0.620)	(0.545)	(0.528)
Male	-2.166*	-1.592*	-1.784*	-303.2*	-222.9*	-249.7*
	(0.839)	(0.777)	(0.762)	(117.5)	(108.7)	(106.6)
Age×Male	0.013*	0.009	0.010*	1.814*	1.242	1.463*
	(0.006)	(0.005)	(0.005)	(0.805)	(0.745)	(0.731)
Player2	0.804***	-0.348*	0.334*	42.6*	21.2	-23.3
	(0.150)	(0.143)	(0.139)	(21.0)	(20.0)	(19.5)
const.	2.705***	2.374***	2.353***	408.7***	292.4***	359.4***
	(0.647)	(0.568)	0.561)	(90.5)	(79.5)	(78.5)
Adj. R^2	0.184	0.085	0.067	0.061	0.042	0.043
# obs.	158	158	158	158	158	158

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 6: OLS of stopping stage and payoffs in R3, R4 and R5 in school-age population

Consistent with this result, Table 6 shows that the effect of age on delaying stopping has dissipated by Round 3, and remains non-significant in Rounds 4 and 5. Consequently, age effects on payoffs also dissipate. We also notice that the gender effect observed in Table 2 is still present but it is modulated by age: males delay less the time to stop but the effect is stronger in the youngest ones (significantly in R3 and R5).

Taken together, the analysis of the evolution of stopping behavior support our Hypothesis **H3** that participants stop earlier as the game progresses. Interestingly, the rate of change is different at different ages. By round 5, we observe convergence in behavior, with individuals stopping early but not immediately. This decrease in "cooperative passing" has detrimental payoff consequences in all age groups.

4.3 Reaction to empirical risk in the long games

As stages progress, a participant should realize that not stopping has both an increased risk (likelihood that the rival stops next) and a higher opportunity cost (difference between the current payoff of stopping and the payoff if the rival stops). We should therefore observe that participants are each time more likely to stop conditional on reaching a given stage. This pattern would indicate that participants both form logical beliefs about their partners and they themselves reason logically.

To test this hypothesis and reveal differences across ages, we determine for each age group and each stage t, the stopping *hazard rate* h_t. This is the probability that a participant stops in stage t, p_t, given that such stage has been reached. Formally, $h_t = \frac{p_t}{\sum_{i=t}^{T} p_i} \in [0, 1]$. Strategic decision-making predicts an increasing hazard rate ($h_{t+1} > h_t$). Figure 5 reports the data by age group. Since the number of observations decreases significantly as we move to later stages and since stopping after round 6 is rare, we present hazard rates only for the relevant range of stages 1 through 6. We also pool together Rounds 1, 3, 4 and 5 to increase statistical power, even though we are aware that behavior changes over rounds, and observations across rounds are collected from the same individuals.

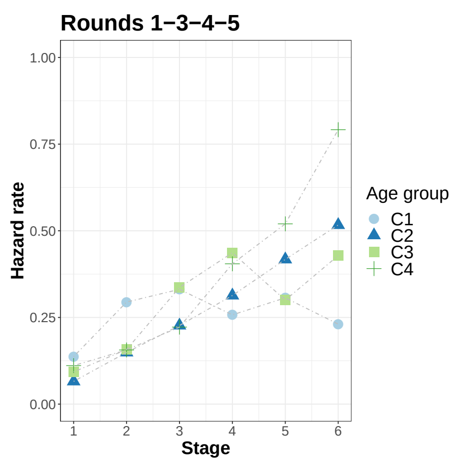

Figure 5: Stopping hazard rate in long games by age group

As typical in experiments with adults, individuals in **C4** display an increasing hazard rate, with a steep (0.124) and very significantly different from zero (p = 0.008) estimated slope of the time series. Hazard rates also increase steadily in **C2** and **C3**, although the slope is statistically significant in the former (0.089, p = 0.009) but not in the latter (0.067, p = 0.132). By contrast, our youngest participants (**C1**) exhibit a constant hazard rate (with an estimated slope of 0.004, p = 0.999). Behavior of **C2**, **C3** and **C4** is consistent with a belief that the risk of stopping is higher the more stages have passed. By contrast, participants in **C1** play in a memoryless fashion. Their behavior cannot be reconciled with a theory where rivals are perceived as minimally strategic, and they trade-off the expected costs and benefits of passing at each stage.[15]

The result reinforces the findings in section 3, where we argued that many of our

[15]The fact that the opportunity cost of passing increases with the number of stages implies that even if rivals use a constant hazard rate, it is still an optimal best response to use an increasing hazard rate.

youngest participants do not stop because they do not anticipate the strategic incentives of others and rather focus on their potential increasing payoffs.

4.4 Factors of behavior in the long games

We have noted that behavior converges with repeated play and that age is not a predictor of behavior from round 3 on. Here, we hypothesize that heterogeneity within age groups may underlie this observation: while age per se does not tell how a participant is likely to behave, their reasoning abilities may. We retain the participants who played first in R1 and R2 and whose behavior was categorized as *SS*, *CS* and *CC* depending on their initial choice in both games (we omitted the 2 participants classified as *SC*). We then run an OLS regression of the stopping stages of these participants in the long games R3 through R5, whenever they were given the opportunity to stop and acted upon it. We use *CC* as the benchmark category. It is worth noting that the number of observations drops significantly compared to previous regression analyses. The result is reported in Table 7.

	Stages before stop		
	R3	R4	R5
SS	-1.389*	-1.895**	-1.219
	(0.635)	(0.616)	(0.712)
CS	-0.117	-0.451	-0.828*
	(0.388)	(0.379)	(0.393)
Male	-0.146	-0.682	-0.700
	(0.375)	(0.355)	(0.383)
const.	3.963***	3.973***	4.391***
	(0.312)	(0.255)	(0.290)
Adj. R^2	0.027	0.116	0.105
# obs.	94	107	88

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 7: OLS of stopping stage in long games on factors of behavior

Compared to participants who behaved as if they were motivated by rewards in R1 and R2, both those who played at equilibrium and those who applied ToM reasoning continue to stop earlier in all games. These differences are significant in R3 and R4 for *SS* participants and in R5 for *CS* participants. It indicates that tendencies observed at the beginning of the experiment persist for its entire duration.

4.5 Summary

There is a decrease in the stopping stage over the course of the experiment. Behavior converges to a large extent. Given that choice in the first round was more noticeably different in **C1**, the changes are more abrupt in that group. Still, we observe that the underpinnings of behavior continue to differ. While age groups **C2**, **C3** and **C4** respond to empirical risk in a logical fashion by decreasing the conditional probability of passing as the game progresses, young participants in **C1** do not account for the increasing incentives to stop within a round. Furthermore, despite the general tendency of all participants to learn to stop earlier, initial tendencies to act purely logically, to respond to rewards or to apply ToM abilities are still present in later rounds.

5 Conclusion

Behavior in strategic games is driven by cognitive ability, theory of mind, individual motivation and incentives. In the centipede game, each decision to pass is a gamble that may lead to higher rewards and participants might be affected by these prospects. While equilibrium theory prescribes stopping at each stage, numerous experiments show that adults delay stopping. With our payoffs, passing can occur for two main reasons: inability to perform logical inferences or decision to best respond to the empirical risk and take a measured chance. Our study sheds light on these motives and on how they develop with age.

We find that logical abilities develop gradually, leading to a decrease in stopping stage with age. Interestingly, while young participants are least likely to act logically, those who do, tend to over-estimate the ability of their peers to behave similarly. As a consequence, they apply the logic blindly. With age, participants learn to anticipate what others may do and best respond to their beliefs. Starting in middle school, students who reason logically know that the unraveling argument should not be applied blindly. The behavior of middle- and high-schoolers is in line with the literature on the centipede that documents strong deviations from backward inductions even after experience. It is also consistent with Levitt et al. (2011) who show no correlation between early stopping in the centipede game and ability to backward induct in more challenging, dominance-solvable games.

The intuitive way to approach the game for participants with limited cognition is as a series of gambles leading to increasing rewards, which is what many young participants do. The inclination of young children to take high risks has been widely documented in the literature (Paulsen et al., 2012; Sutter et al., 2019) and it likely results from their difficulty to simultaneously process rewards and probabilities (Brocas et al., 2019). This leads them to pay disproportionally more attention to rewards. In our case, it makes them not only willing to take more risks than older participants, but also willing to stop at each stage with the same conditional probability. Overall, passing in young children results from cognitive limitations, while it follows a calculated empirical risk in older participants.

While the observed behavior of participants may be captured by a risk preference argument, we believe that the underlying mechanisms that lead to choice not only are complex but they also change with age. Delayed stopping and constant hazard rate by many young participants is only consistent with salience effects and limited cognition. By contrast, delayed stopping and increasing hazard rate by teenagers more likely reflects impulsivity or underdeveloped risk avoidance mechanisms. Also, the studies referenced above have typically found that female are more risk averse than males, which would predict that females would stop earlier than males in our game. This is not supported by our data. It may be due to the fact that the task does not explicitly describes probabilities and that behavior in our strategic task relates to gender-related differences in competitiveness (Niederle and Vesterlund, 2011). These observations taken together suggest that differences in stopping are driven more by differences in cognition and ToM abilities than by differences in intrinsic risk attitudes.

Middle- and high-schoolers understand well the unraveling logic of the short game. And yet, there is heterogeneity within age: 21% of very young participants understand it while 28% of highly educated adults do not. This is consistent with results obtained in other paradigms involving iterative reasoning (Brocas and Carrillo, 2021). It is also interesting to notice that cognitively equipped young children are not able to *not apply* their skill, while their older peers are, which underscores the importance of ToM. This faculty is particularly important in our game, since once the first player passes in stage 1, backward induction ceases to be the relevant concept.

To our knowledge, this is the first experimental study where earnings monotonically decrease over a large age span (8 to 16 years old, and up to 23 years old if we include the control group). Young children obtain larger earnings because they are lured by the high rewards in late stages. As logical abilities and ToM develop, participants take fewer chances, resulting in earlier stopping and therefore lower earnings. Larger deviations by less cognitively developed participants may seem unsurprising. And yet, we expected that more sophisticated individuals would be able to replicate the behavior of their less sophisticated peers and obtain (at least) their same payoff in the long game. This is what typically happens in developmental studies, but not what we find in this paper.

Experience heavily disciplines behavior, and by the fifth round participants in all grades end up making very similar choices on aggregate. The fact that young children rapidly adapt their decisions indicates that they use feedback as an input in their choices across rounds, and apply inductive logic to what they observe. It would be interesting to refine those ideas and see whether children learn to respond to empirical risk, to backward induct or both. It would also be interesting to link individual learning to feedback, as we expect children to learn differently if their first rival stops early or takes some chances.

As noted earlier, the centipede game is largely dependent on beliefs. It is therefore an ideal game to mix participants who differ in cognitive abilities but also in abilities to read other players' minds. This exercise is instructive because it tells us about the capacity of people to integrate relevant information about potential rivals and to fine tune their behavior to this information. Examples include Palacios-Huerta and Volij (2009) who mixed undergraduate students and professional chess players in the centipede paradigm and Proto et al. (2022) who mixed players with different intelligence levels in the repeated prisoner's dilemma. An interesting alley for future research would be to mix participants from different grades. On the one hand, it would allow us to measure the capacity of young children to act in a more sophisticated manner in the presence of older children. On the other hand, it would shed light on the capacity of older children to take advantage of rivals who are (presumably) less cognitively able.

References

Robert J Aumann. Irrationality in game theory. *Economic analysis of markets and games*, pages 214–227, 1992.

Elchanan Ben-Porath. Rationality, nash equilibrium and backwards induction in perfect-information games. *The Review of Economic Studies*, 64(1):23–46, 1997.

Isabelle Brocas and Juan D Carrillo. Introduction to special issue "understanding cognition and decision-making by children." studying decision-making in children: challenges and opportunities. *Journal of Economic Behavior & Organization*, 179:777–783, 2020a.

Isabelle Brocas and Juan D Carrillo. Iterative dominance in young children: Experimental evidence in simple two-person games. *Journal of Economic Behavior & Organization*, 179:623–637, 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. Developmental meritocracy. USC Working Paper, 2022.

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.

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*, 179:653–666, 2020.

Simon Czermak, Francesco Feri, Daniela Glätzle-Rützler, and Matthias Sutter. How strategic are children and adolescents? experimental evidence from normal-form games. *Journal of Economic Behavior & Organization*, 2016.

Eduardo Fe, David Gill, and Victoria Prowse. Cognitive skills, strategic sophistication, and life outcomes. *Journal of Political Economy*, 130(10):2643–2704, 2022.

Mark Fey, Richard D McKelvey, and Thomas R Palfrey. An experimental study of constant-sum centipede games. *International Journal of Game Theory*, 25(3):269–287, 1996.

David Gill and Victoria Prowse. Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis. *Journal of Political Economy*, 124(6):1619–1676, 2016.

William T Harbaugh and Kate Krause. Children's altruism in public good and dictator experiments. *Economic Inquiry*, 38(1):95–109, 2000.

Eva M Krockow, Andrew M Colman, and Briony D Pulford. Cooperation in repeated interactions: a systematic review of centipede game experiments, 1992–2016. *European Review of Social Psychology*, 27(1):231–282, 2016.

Steven D Levitt, John A List, and Sally E Sadoff. Checkmate: Exploring backward induction among chess players. *American Economic Review*, 101(2):975–90, 2011.

Rosa Li. Flexing dual-systems models: How variable cognitive control in children informs our understanding of risk-taking across development. *Developmental cognitive neuroscience*, 27:91–98, 2017.

John A List, Ragan Petrie, and Anya Samek. How experiments with children inform economics. Technical report, National Bureau of Economic Research, 2021.

Richard D McKelvey and Thomas R Palfrey. An experimental study of the centipede game. *Econometrica: Journal of the Econometric Society*, pages 803–836, 1992.

Patricia H Miller. *Theories of developmental psychology.* Macmillan, 2002.

J Keith Murnighan and Michael Scott Saxon. Ultimatum bargaining by children and adults. *Journal of Economic Psychology*, 19(4):415–445, 1998.

Rosemarie Nagel and Fang Fang Tang. Experimental results on the centipede game in normal form: an investigation on learning. *Journal of Mathematical psychology*, 42(2-3):356–384, 1998.

Muriel Niederle and Lise Vesterlund. Gender and competition. *Annu. Rev. Econ.*, 3(1):601–630, 2011.

Ignacio Palacios-Huerta and Oscar Volij. Field centipedes. *American Economic Review*, 99(4):1619–35, 2009.

James E Parco, Amnon Rapoport, and William E Stein. Effects of financial incentives on the breakdown of mutual trust. *Psychological Science*, 13(3):292–297, 2002.

David J Paulsen, Michael L Platt, Scott A Huettel, and Elizabeth M Brannon. From risk-seeking to risk-averse: the development of economic risk preference from childhood to adulthood. *Frontiers in psychology*, 3:313, 2012.

Eugenio Proto, Aldo Rustichini, and Andis Sofianos. Intelligence, personality, and gains from cooperation in repeated interactions. *Journal of Political Economy*, 127(3):1351–1390, 2019.

Eugenio Proto, Aldo Rustichini, and Andis Sofianos. Intelligence, errors, and cooperation in repeated interactions. *The Review of Economic Studies*, 89(5):2723–2767, 2022.

Amnon Rapoport, William E Stein, James E Parco, and Thomas E Nicholas. Equilibrium play and adaptive learning in a three-person centipede game. *Games and Economic Behavior*, 43(2):239–265, 2003.

Philip J Reny. Rationality in extensive-form games. *Journal of Economic perspectives*, 6 (4):103–118, 1992.

Robert W Rosenthal. Games of perfect information, predatory pricing and the chain-store paradox. *Journal of Economic theory*, 25(1):92–100, 1981.

Edward B Royzman, Kimberly Wright Cassidy, and Jonathan Baron. "i know, you know": Epistemic egocentrism in children and adults. *Review of General Psychology*, 7(1):38–65, 2003.

Itai Sher, Melissa Koenig, and Aldo Rustichini. Children's strategic theory of mind. *Proceedings of the National Academy of Sciences*, 111(37):13307–13312, 2014.

Matthias Sutter, Claudia Zoller, and Daniela Glätzle-Rützler. Economic behavior of children and adolescents–a first survey of experimental economics results. *European Economic Review*, 111:98–121, 2019.

Appendix A. Instructions: LILA (6-10) and USC

GOING DOWN THE STREET GAME

In this game, the computer will pair you with one other student. One of you will be BLUE and the other will be ORANGE. The computer decides who is BLUE and who is ORANGE. In this game, BLUE and ORANGE are walking down a street. On that street, there are blue and orange houses (see Figure 6 for the slides).

[SLIDE 1]

The first house is blue, the second one is orange, the third one is blue again, and so on. If BLUE and ORANGE pass by a BLUE house, BLUE decides whether to stop the game or to continue. If they pass by an orange house, ORANGE decides whether to stop the game or to continue. Because the first house is blue, BLUE always makes the first choice. Once a player stops, the game is over.

When a player decides to stop the game, both BLUE and ORANGE get points. But how many points?

[SLIDE 2]

When BLUE stops, BLUE gets more points the farther down the street he stops: 100, 240, 380, etc... In this game, 1 point equals 1 cent, so 1 dollar, 2.40 dollars, 3.80 dollars, etc.

[SLIDE 3]

When ORANGE stops, ORANGE also gets more points the farther down the street he stops: 170, 310, 450, etc.

[SLIDE 4]

Finally, when BLUE stops ORANGE gets only 50 points and when ORANGE stops BLUE gets only 50 points.

[SLIDE 5]

Putting all together, these are the points each person gets depending on who stops where.

[SLIDE 6]

For instance, imagine they reach the fifth house. This is just an example.

- Whose turn is it to decide? [answer: BLUE]
- If BLUE stops, how many points does he get? [answer: 380]
- How many points does ORANGE get? [answer: 50]
- If BLUE continues and ORANGE stops, how many points does ORANGE get? [answer: 450]
- And BLUE? [answer: 50]. Is it clear?

Now, let's look at what you will see on your tablet at the beginning of the game. If you are BLUE, your screen looks like this.

[SLIDE 7]

At the top of the screen, it says you are BLUE. You can see the whole street with the blue and orange houses. You and ORANGE are in front of the first blue house, which is highlighted in grey. Since the house is BLUE, it is your turn to choose. You decide to stop the game by clicking STOP and then OK or to continue by clicking the green arrow and then OK. If you stop the points you and ORANGE get are highlighted in grey.

If you are ORANGE, your screen looks like this.

[SLIDE 8]

It says you are ORANGE. It is the same as what BLUE saw except that you cannot make any choice, since you are in a BLUE house. If BLUE decides to stop, the points you and BLUE get are highlighted in grey. If BLUE decides to continue, you will both go down the street, reach the next house, which is now ORANGE and it will be your turn to choose. In that case, BLUE will see this screen.

[SLIDE 9]

and ORANGE will see this screen

[SLIDE 10]

Is the game clear? OK, let's play then. Remember you are matched with someone from this room, but you will not know with whom you are playing, and it is not the point to know.

Start CENTIPEDE GAME 1

[At the end] We are going to play another round of this game, except that the street is much shorter this time, with only 4 houses. You will be playing with a different person than before. Are you ready?

Start CENTIPEDE GAME 2

Ok, we are going to play a few more times, again with different partners each time. It is the same game as the first time, with a long street. You will also change color each time (one time blue, one time orange).

Start CENTIPEDE GAME 3 (3 rounds)

We are done. In the next page you will see how many points you got in the "Going Down the Street" Game. These are worth 1 cent. You don't need to record it. The computer takes care of it. We are going to ask you some final questions. Please fill the questionnaire. We will then tell you how much money you made and you will receive your amazon e-giftcard in your email. Thanks for participating.

Start QUESTIONNAIRE

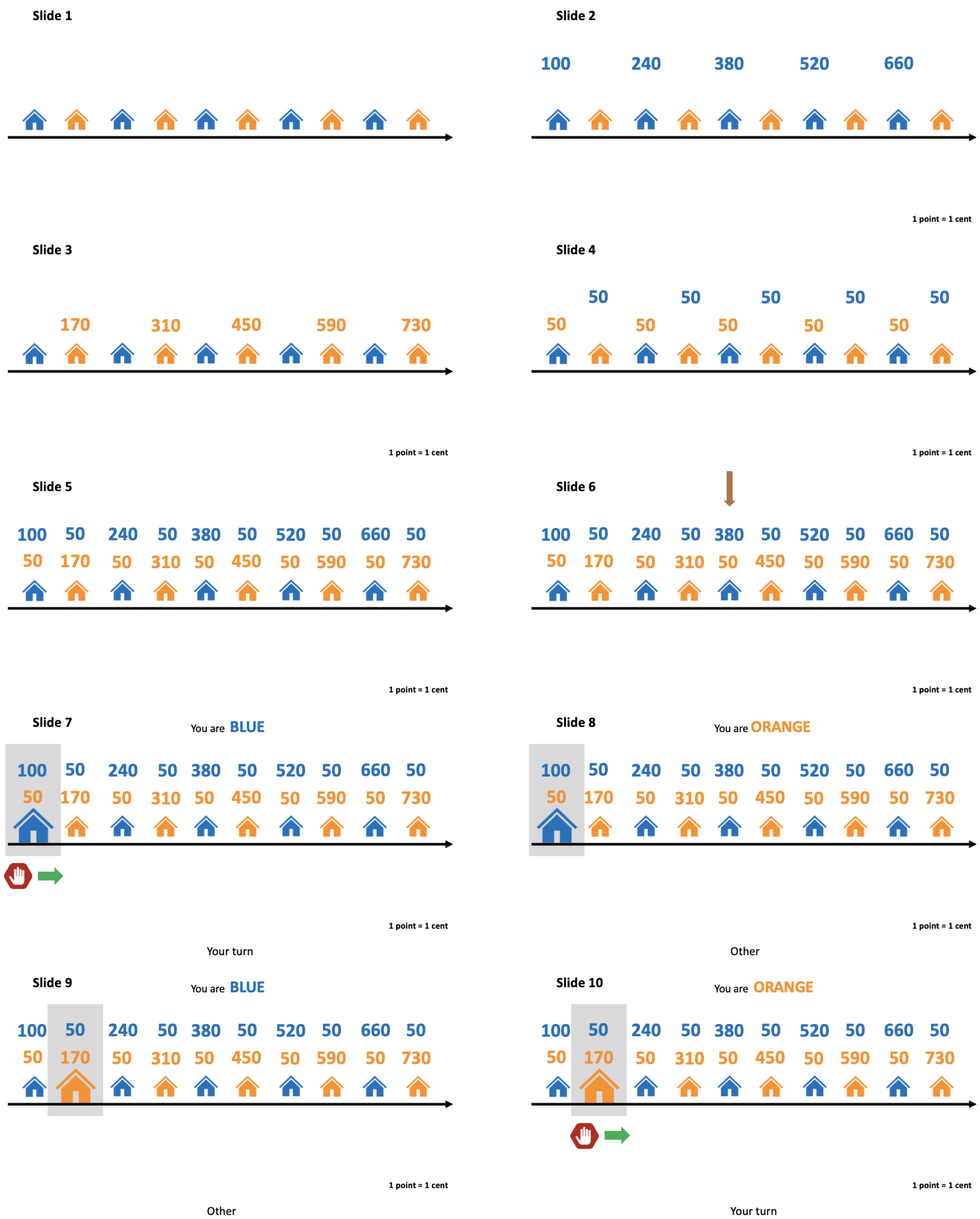

Figure 6: Slides projected on screen for instructions