I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11326

Cognitive Skills and the Development of Strategic Sophistication

Eduardo Fe
David Gill

FEBRUARY 2018

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11326

Cognitive Skills and the Development of Strategic Sophistication

Eduardo Fe
University of Strathclyde

David Gill
Purdue University and IZA

FEBRUARY 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Cognitive Skills and the Development of Strategic Sophistication*

In this paper we investigate how observable cognitive skills influence the development of strategic sophistication. To answer this question, we study experimentally how psychometric measures of theory-of-mind and cognitive ability (or 'fluid intelligence') work together with age to determine the strategic ability and level-*k* behavior of children in a variety of incentivized strategic interactions. We find that better theory-of-mind and cognitive ability predict strategic sophistication in competitive games. Furthermore, age and cognitive ability act in tandem as complements, while age and theory-of-mind operate independently. Older children respond to information about the cognitive ability of their opponent, which provides support for the emergence of a sophisticated strategic theory-of-mind. Finally, theory-of-mind and age strongly predict whether children respond to intentions in a gift-exchange game, while cognitive ability has no influence, suggesting that different psychometric measures of cognitive skill correspond to different cognitive processes in strategic situations that involve the understanding of intentions.

JEL Classification: C91, D91, J24

Keywords: cognitive skills, theory-of-mind, cognitive ability, fluid intelligence, strategic sophistication, age, children, experiment, level-k, bounded rationality, non-equilibrium thinking, intentions, gift-exchange game, competitive game, strategic game, strategic interaction

Corresponding author:
David Gill
Department of Economics
Purdue University
Rawls 4023
Krannert School of Management
403 W. State Street
West Lafayette, IN 47907
USA
E-mail: dgill.econ@gmail.com

* We thank: school staff, children, and their parents for making this research possible; the University of Strathclyde for funding; Robin Durban and Victoria Prowse for many useful discussions; Mouli Modak, Benjamin Raymond, Chen Wei and Daniel Woods for excellent research assistance; and Alonso Gutiérrez Morillo, Valentín Pastor (respectively Director and Inspector at Consejería de Educación de Cantabria) and Aurora Gil Álvarez (Dean of the College of Psychology of Cantabria) for institutional support.

1 Introduction

The development of strategic sophistication matters because people constantly engage in strategic interactions with others in their environment. As Binmore (2007) eloquently puts it: "A game is being played whenever people have anything to do with each other." Examples of strategic decisions include: how much effort to allocate to competitions for valuable prizes or outcomes; when to reciprocate the kind behavior of others and when to punish bad behavior; whom to form friendships with; when to maintain a reputation for honesty over lying or breaking promises; how to use communication to coordinate on mutually beneficial outcomes; how to induce cooperation in repeated interactions that involve short-term incentives to behave badly; and how to bargain effectively. Strategic decisions also extend to the realm of emotions such as pride, disappointment, guilt or anger.[1]

The importance of developing strategic skills has been recognized by educational programs that emphasize strategic ability. For example, the "Mind Lab" project is "an innovative in-school methodology for the development of thinking abilities and life skills through strategy games." Mind Lab is integrated into the school curriculum; to date, over four million grade-school-aged students from around the world have participated in Mind Lab programs. Mind Lab argues that its successful programs provide children with "the most important tools they need to succeed in their studies, careers, and personal lives" and "empowers them with skills and knowledge relevant to real-life situations."[2]

In this paper we aim to discover whether observable cognitive skills influence the development of strategic sophistication. To answer this question, we study how cognitive skills and age work together to determine the strategic sophistication of children in a variety of strategic interactions. Accounting for the role of cognitive skills in how strategic ability develops is important because success in strategic interactions often requires high-level reasoning, while at the same time cognitive skills vary greatly at the population level, and differences in cognitive skills emerge early in life. Furthermore, an understanding of how observable cognitive skills influence the development of strategic sophistication will facilitate the use of psychometric testing to predict future individual-level behavior and outcomes in real-life settings that involve strategic decision-making. Relatedly, providing quantitative estimates of the importance of cognitive skills in the development of strategic ability will help to support childhood interventions that aim to improve cognitive skills.[3]

To discover how observable cognitive skills influence the development of strategic sophistication, we designed an experiment to answer four related questions. First, do cognitive skills help children in competitive strategic games? Second, do children adjust their strategic behavior according to the cognitive skill of their opponent? Third, do cognitive skills predict how children respond to intentions in strategic situations? Fourth, do different cognitive skills affect the

[1] For example: how to influence others' beliefs about personal characteristics that cause pride (Ellingsen and Johannesson, 2008); whether to restrain ambition to avoid disappointment (Gill and Stone, 2010; Gill and Prowse, 2012); how to use guilt to improve effort in teams (Gill and Stone, 2014); and when to exhibit anger (Battigalli et al., 2017).

[2] All quotes are from https://www.mindlab-group.com. More details are available there.

[3] Heckman (2006) and Heckman and Kautz (2014) survey childhood interventions that aim to improve cognitive and non-cognitive skills. As emphasized by Cunha et al. (2010), cognitive skills are most malleable at a young age.

strategic behavior of children in different ways? In each case, we consider how cognitive skills and age act in tandem to determine how strategic sophistication develops.

We begin by using psychometric tests to measure the cognitive skills of 730 child subjects, who range in age from five to twelve years old. We measure two cognitive skills: theory-of-mind (the ability to understand the mental states of others) and cognitive ability (the ability to use logical reasoning to solve new problems, also called 'fluid intelligence'). We measure theory-of-mind using an Imposing Memory Task (IMT) designed for children (Liddle and Nettle, 2006); this test asks questions about the mental states of characters in a number of stories. We measure cognitive ability using an age-appropriate Raven's Progressive Matrices test (Raven et al., 2000); the Raven test consists of non-verbal multiple-choice questions that require subjects to identify the missing element that completes a visual pattern. We standardize test scores within age group in order to separate the effect of age from that of theory-of-mind and cognitive ability; the influence of psychometric test scores on strategic behavior therefore captures the effect of test scores within age group. The low correlation between our theory-of-mind and cognitive ability test scores provides evidence that our psychometric tests capture different cognitive skills.

We then study how our subjects' observable cognitive skills influence their behavior in an incentivized competitive game (a simplified variant of the 11-20 Money Request game originally developed by Arad and Rubinstein, 2012). Our interest lies in studying initial responses, as opposed to repeated-game effects or learning from experience. Therefore, across repetitions of the game, we rematched subjects and gave no feedback (Costa-Gomes et al., 2001, and others study initial responses of adults by rematching and suppressing feedback; for a discussion, see Crawford et al., 2013). We find that children who have a better theory-of-mind and children who are more cognitively able tend to be more strategically sophisticated, in the sense that they are more likely to best-respond to the empirical distribution of choices. Furthermore, age and cognitive ability work together as complements in the development of strategic sophistication, while age and theory-of-mind operate independently.

The structure of Arad and Rubinstein (2012)'s competitive game is designed to trigger non-equilibrium level-k thinking (Nagel, 1995; Stahl and Wilson, 1995; Crawford et al., 2013). We find that our subjects are more likely to exhibit level-1 behavior as their theory-of-mind and cognitive ability improve (and as they become older), while they are less likely to exhibit unsophisticated level-0 behavior as they become more cognitively able.

In one repetition of the competitive game, we gave subjects information about the cognitive ability of their opponent, with the aim of creating exogenous variation in beliefs about the opponent's strategic sophistication. Remarkably, we find that our subjects respond to this exogenous variation: older children who face a high-cognitive-ability opponent (rather than an opponent of low cognitive ability) are more likely to exhibit level-2 behavior. This ability of older children to adjust behavior to the cognitive skill of their opponent provides support for the emergence of a sophisticated strategic theory-of-mind in children as young as eight to twelve years old.

Finally, we use a bespoke gift-exchange game to investigate whether cognitive skills predict how children respond to intentions. Conditional on the allocator choosing to split a pie equally, the allocator's intentions, as perceived by the receiver, depend on the allocator's unchosen or forgone alternative (see Charness and Levine, 2007, for a discussion of games that use foregone options to vary intentions). In one treatment, the alternative is favorable to the allocator, and choosing an equal split is unambiguously generous; in the second treatment, the alternative is favorable to the receiver. The theory that underlies intentions-based reciprocity in strategic games (Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006) models reciprocal behavior as depending on beliefs about intentions.[4] We find that theory-of-mind and age strongly predict whether receivers respond to the allocator's intentions in our gift-exchange game, while cognitive ability has no influence. This striking result suggests that different psychometric measures of cognitive skill correspond to different cognitive processes in strategic situations that involve the understanding of intentions.

Our results complement a number of important literatures. First, we complement existing work on childhood interventions that aim to improve cognitive skills (see Heckman and Kautz, 2014, for a summary). In particular, our findings help to understand the mechanism by which improvements in childhood cognitive skills might translate into life outcomes: our results suggest that, at least in part, interventions that improve cognitive skills could improve life outcomes by enhancing the ability to perform well in domains that require good strategic decision-making. Second, we complement existing literature on bounded rationality and non-equilibrium thinking (see the survey by Crawford et al., 2013). In particular, by showing how level-k thinking develops in childhood, and how this development interacts with observable cognitive skills, we hope to add empirical foundations to new theoretical work that seeks to model bounds on depth of reasoning (Alaoui and Penta, 2016a; Alaoui and Penta, 2016b). Third, we complement the literature on evolutionary game theory (e.g., Weibull, 1997). The behavior of our subjects is likely to be driven, in part, by hardwired abilities and behaviors that start to manifest themselves in childhood. We hope that our empirical evidence about how strategic sophistication develops will help to inform evolutionary models of strategic behavior.

Our novel contribution is to investigate whether psychometric measures of cognitive ability and theory-of-mind predict the development of strategic sophistication in children. We are not aware of any papers that undertake the same exercise. The rest of this introduction places our work in the context of the existing literature, focusing first on competitive games and then on the response to intentions.

Economists and psychologists have begun to investigate how age influences strategic behavior in competitive games (Perner, 1979; Sher et al., 2014; Brosig-Koch et al., 2015; Czermak et al., 2016; Brocas and Carrillo, 2017).[5] However, we are not aware of any existing papers that investigate how psychometric measures of cognitive ability and theory-of-mind predict the strategic behavior of children in competitive games, although a handful of papers study related

[4]The theory builds on the framework of psychological games introduced by Geanakoplos et al. (1989).

[5]Perner (1979) finds that the ability to detect an opponent's dominant strategy improves with age. Sher et al. (2014) find that behavior tends to move toward equilibrium with age. Brosig-Koch et al. (2015) find that success against a computer opponent in a dominant-strategy race game improves with age. Czermak et al. (2016) find that the proportion of subjects playing Nash does not change with age. Using a sample of preschoolers (53 to 66 months old), Brocas and Carrillo (2017) study behavior in simple individual decision problems based on "Connect 4" and success in a related simple dominant-strategy race game against the experimenter without any incentives; solving the easier individual decision problem correlates with success in the game, but solving the harder problem does not. Perhaps unsurprisingly given the narrow age range, Brocas and Carrillo (2017) find no effect of age on success in the game.

questions (Steinbeis et al., 2012; Sher et al., 2014; Geng et al., 2015; Czermak et al., 2016).[6] Nor are we aware of existing work that studies how children respond to information about the cognitive ability (or theory-of-mind) of their opponent in strategic environments.

We study strategic behavior in competitive games, rather than cooperative behavior. A few papers investigate how cooperation or coordination changes with age in strategic games (Kagan and Madsen, 1971; Stingle and Cook, 1985; Fan, 2000; Harbaugh and Krause, 2000; Sally and Hill, 2006; Devetag et al., 2013; Grueneisen et al., 2015; Brocas et al., 2017), while a couple of papers investigate the role of theory-of-mind with a focus on autistic children (Sally and Hill, 2006; Li et al., 2014).[7]

We now turn to intentions. Despite an extensive literature in psychology that investigates how children's understanding of intentions affects their moral judgments,[8] the question of how children respond to intentions in strategic situations has received little attention. We are not aware of any existing papers that use a gift-exchange game to measure how children respond to intentions.[9] Sutter (2007) popularized the use of the ultimatum game with variation in forgone

[6]Steinbeis et al. (2012) do not directly measure strategic behavior in a competitive game, but they argue that the across-game difference between ultimatum game offers and dictator game allocations captures strategic behavior. Their focus is on the effect of age for their full sample of 174 subjects, finding an increase in the difference with age (as do Bereby-Meyer and Fiks, 2013). Cognitive ability was measured only for the twenty-eight children who also underwent FMRI testing; unsurprisingly given the small sample, Steinbeis et al. (2012) find no effect of cognitive ability on the difference. Sher et al. (2014) study children's behavior in an undercutting game (the "stickers" game, which has a unique pure-strategy symmetric Nash equilibrium) and a sender-receiver game, with a focus on the effect of age. The games were not played against another child, but against an experimenter who was not trying to win, and instead followed a non-equilibrium behavioral rule. Sher et al. (2014) do not measure cognitive ability (in our sense of abstract or logical reasoning ability), but they do measure working memory, finding that better working memory predicts choices closer to equilibrium in the stickers game but not in the sender-receiver game. Geng et al. (2015) study the behavior of adolescents rather than children, and find that their subjects do not play minimax in zero-sum games; since opponents do not play minimax, it is perhaps not surprising that they also find no correlation between cognitive ability and closeness to minimax play for adolescents (nor do they find a correlation with earnings). Finally, using a sample of children and adolescents (ages ten to seventeen), Czermak et al. (2016) study the consistency of choices and beliefs in the normal-form games from Costa-Gomes et al. (2001). Czermak et al. (2016) do not investigate subjects' performance in the games or their likelihood of best-responding to the empirical distribution of choices (the focus of our study), and nor do they measure cognitive ability; instead they find that subjects with better self-reported math grades are more likely to act consistently with their first-order beliefs, suggesting that math ability allows subjects to better calculate the best-response to given beliefs. Splitting their sample by gender, Czermak et al. (2016) also find an effect of self-reported math ability on the probability of being a strategic type for females, but not for males.

[7]Using bespoke games, Kagan and Madsen, 1971, find cooperation decreases with age, while Stingle and Cook, 1985, find the opposite. In a repeated prisoners' dilemma, Fan (2000) finds that cooperation increases with age, while Sally and Hill (2006) find no age effect. In a repeated public good game, Harbaugh and Krause (2000) find that contributions increase over rounds for younger children, but decline for older children. Devetag et al. (2013) and Grueneisen et al. (2015) find that older children better recognize focal points as coordination devices. Brocas et al. (2017) find that cooperation in a repeated alternating dictator game goes up with age. Sally and Hill (2006)'s children play against a confederate who does not maximize her own payoff; the paper reports a marginally significant positive correlation between theory-of-mind and cooperation for the whole sample (including autists). Using a similar design and sample, Li et al. (2014) find some evidence that theory-of-mind helps cooperation, but only for the autistic children.

[8]This literature finds that intentions start to matter at a young age: see Zelazo et al. (1996) for an example and a review of the early literature. Using questions about stories, Feinfield et al. (1999) find that children start to develop the ability to distinguish intentions and outcomes from the age of four.

[9]Using a gift-exchange game, Charness and Levine (2007) find that intentions matter for adult subjects. Rather than using foregone payoffs like we do, they use uncertainty.

options to study whether responder age predicts responses to proposer intentions.[10] Sutter (2007) finds that intentions matter for children, teenagers and students, while later experiments find that the importance of intentions goes up with age.[11] We are not aware of any papers that investigate how children's theory-of-mind or cognitive ability predict how they respond to intentions in strategic situations, with the exception of Pelligra et al. (2015) who use the ultimatum game with a focus on autistic children.[12,13]

Our focus is on how observable cognitive skills influence the development of strategic sophistication in children. Some recent papers study the relationship between cognition and the behavior of adults in competitive games (e.g.: Burnham et al., 2009; Brañas-Garza et al., 2012; Gill and Prowse, 2016) or the cooperation of adults in strategic games (e.g., Proto et al., forthcoming). For example, in a repeated beauty contest game, Gill and Prowse (2016) find that cognitive ability predicts how fast adults learn to play equilibrium. We are not aware of papers that study whether the theory-of-mind or cognitive ability of adults predicts their response to intentions in strategic situations.

The paper proceeds as follows: Section 2 describes the experimental design; Section 3 presents results from the competitive games; Section 4 presents results from the gift-exchange game; and Section 5 concludes. The Supplementary Web Appendix includes the experimental instructions and further details.

[10] Although the ultimatum game is not well designed to measure reciprocity, it can be used to study whether the proposer's intentions affect the responder's rejection rate (see, e.g., Falk et al., 2003, with adult subjects). It is not well suited to measure reciprocity for two main reasons: first, proposers can try to influence rejection rates, and so proposals confound generosity and strategic incentives; second, by destroying the entire pie, rejection is socially costly, and so rejection rates confound preferences over efficiency and reciprocity.

[11] See Güroğlu et al. (2009), Güroğlu et al. (2011), Bereby-Meyer and Fiks (2013), Gummerum and Chu (2014) and Sul et al. (2017). Bueno-Guerra et al. (2016) find evidence consistent with Sutter (2007), but without reaching statistical significance.

[12] Autistic children tend to suffer from impaired theory-of-mind (Baron-Cohen et al., 1985). Using a sample of autistic and non-autistic children, Pelligra et al. (2015) find that intentions matter, except for the case of children who are autistic and fail a single-question second-order theory-of-mind test.

[13] Without studying the role of intentions, some papers link theory-of-mind and cognitive ability to simple reciprocity: using an ultimatum game followed by an unexpected dictator game, Schug et al. (2016) find that children's first-order theory-of-mind predicts simple reciprocity; using a sequential prisoners' dilemma, Burks et al. (2009) find that the cognitive ability of adults predicts simple reciprocity; and using a gift-exchange game, Filiz-Ozbay et al. (forthcoming) find that the cognitive ability of adult men predicts simple reciprocity.

2 Experimental design

2.1 Overview

The experiment was conducted during the first half of 2016 at five schools in Santander (Spain). Supplementary Web Appendix I provides the experimental instructions. The subjects were children between the ages of five and twelve inclusive whose parents provided written consent. Primary Institutional Review Board (IRB) approval was obtained from the University of Strathclyde. In total, 730 children (377 boys and 352 girls)[14] participated in the experiment: Table 1 describes the distribution of subjects by school and age.

School	Age								Total
	5	6	7	8	9	10	11	12	
1	26	36	50	27	31	45	33	5	253
2	4	9	13	20	11	20	16	3	96
3	34	37	28	41	30	30	7	2	209
4	7	7	8	2	7	11	9	0	51
5	13	14	11	21	21	19	19	3	121
Total	84	103	110	111	100	125	84	13	730

Notes: We define a subject's age to be her age in years on the date that the subject completed the cognitive ability and theory-of-mind tests. We selected subjects by academic year, not age, and so the sample includes some twelve-year-olds who were the older segment of the oldest academic-year cohort.

Table 1: Distribution of subjects by school and age.

The experiment was divided into three phases. In phase 1 subjects completed psychometric tests to measure theory-of-mind and cognitive ability. In phase 2 subjects repeatedly played an incentivized competitive game without feedback and with variation in the type of opponent. In phase 3 subjects played an incentivized gift-exchange game designed to measure how subjects respond to intentions. All subject interactions were anonymous. Phase 1 and phase 2 were separated by about eight weeks, while phase 3 immediately followed phase 2.[15]

At the beginning of phase 2 subjects were told how the games would be incentivized. Subjects accumulated tokens. Younger children exchanged tokens for stationery and educational supplies. Older children exchanged tokens for a voucher from either a local bookshop or an online retailer: each token was worth 0.1 euros. There was no show-up fee. Subjects earned an average of fifty-five tokens. Supplementary Web Appendix II provides further details about the payment protocols.

Subjects were given seven-inch electronic tablets to complete all the tasks. The experimental instructions were read aloud, with explanations supported by projections of decision screens. All sessions were conducted by the alphabetically-first author. Each session was supervised by one

[14] One subject did not report their gender.

[15] Our sample of 730 subjects excludes children who, due to absence, failed to complete the tests in phase 1 or participate in the games in phases 2 and 3. Sixteen subjects in our sample completed phase 1 after phases 2 and 3 because they were on a school trip on the day of phase 1.

member of staff who was familiar to the children.[16]

2.2 Phase 1: Psychometric tests

In phase 1 subjects completed a theory-of-mind test followed by a cognitive ability test. Following the convention in the psychometric literature, these tests were not incentivized.

2.2.1 Theory-of-mind test: Imposing Memory Task

Theory-of-mind refers to a person's ability to understand the mental states of other persons (e.g., Premack and Woodruff, 1978). Higher-order theory-of-mind refers to the ability to reason recursively about the content of the mental states of others (e.g., reasoning about what Alice believes about Bob's mental state). Developed by Kinderman et al. (1998), the Imposing Memory Task (IMT) measures first-order and higher-order theory-of-mind ability. In a typical IMT, the experimenter reads a series of stories. At the end of each story, subjects are asked binary-choice questions about the mental states of the characters in the stories. Questions that measure first-order theory-of-mind ask about a character's mental state (e.g., about what that character believes about an event or the location of an object, or what the character wants); questions that measure second-order theory-of-mind ask what a character believes about another character's mental state, and so on.[17]

Subjects in our experiment completed an IMT designed specifically for children (Liddle and Nettle, 2006). Each story includes two questions that measure theory-of-mind.[18] For children under the age of nine, we used the first and fourth stories from Liddle and Nettle (2006), which measure up to third-order theory-of-mind. For children aged nine years and older, we also used the second story, which measures up to fourth-order theory-of-mind.[19] Table SWA.1 in Supplementary Web Appendix VI provides the distribution of IMT scores by age.

2.2.2 Cognitive ability test: Raven's Progressive Matrices

We use the term 'cognitive ability' to mean analytic or fluid intelligence, which is "the ability to reason and solve problems involving new information, without relying extensively on an explicit base of declarative knowledge" (Carpenter et al., 1990). After the theory-of-mind test, subjects completed a Raven's Progressive Matrices test of cognitive ability (Raven et al., 2000). The Raven test consists of non-verbal multiple-choice questions and is recognized as a leading

[16] Supplementary Web Appendix III describes the experimental software and configuration of the portable lab. Supplementary Web Appendix IV describes three small pre-intervention pilots. We also collected information from the children about their social network and from teachers about the children's social competence (using the Social Competence questionnaire from Nettle and Liddle, 2008) that we have not yet analyzed. Finally, we attempted to collect a questionnaire from parents, but the completion rate was low.

[17] Scores in the IMT for adults have been shown to correlate with the personality trait agreeableness, as well as with scores on the Internal, Personal and Situational Attributions Questionnaire that measures the ability to separate agency from external causes (Kinderman et al., 1998; Nettle and Liddle, 2008).

[18] Some questions are not related to theory-of-mind but instead check understanding. Following the convention in the literature that uses the IMT to measure theory-of-mind, we exclude these questions when calculating IMT scores.

[19] Perner and Wimmer (1985) and Wellman et al. (2001) study the development of theory-of-mind in children.

measure of analytic or fluid intelligence (Carpenter et al., 1990; Gray and Thompson, 2004).[20] Specifically, the Raven test consists of a series of visual patterns with a missing element. In each case, subjects have to identify (among six or eight choices) the missing element that completes the pattern.

Subjects in our experiment completed Raven tests appropriate for their age. For children under the age of eight, we used the Colored Progressive Matrices test, which consists of thirty-six questions split across three sets of increasing difficulty (A, Ab and B). For children aged eight years and older, we used the Standard Progressive Matrices test, which consists of sixty questions split into fives sets (A to E) of increasing difficulty.[21] Following convention (Raven et al., 2000), no time limit was imposed.[22] Table SWA.2 in Supplementary Web Appendix VI provides the distribution of Raven test scores by age.

2.2.3 Age-standardization and correlation of test scores

To separate the effect of age from that of theory-of-mind and cognitive ability when analyzing behavior, we standardize psychometric test scores within age group. As described in the notes to Table 1, we define a subject's age to be her age in years on the date that the subject completed the tests. For each age group (five to twelve), we standardize both the theory-of-mind test scores and the cognitive ability test scores. Thus, the influence of psychometric test scores on behavior captures the effect of test scores within age group.

The correlation between the age-standardized theory-of-mind test score and the age-standardized cognitive ability test score is 0.24, which provides evidence that our two tests capture different skills.

2.3 Phase 2: Competitive games

In phase 2 subjects repeatedly played an incentivized competitive game that we call the '1-6 Token Request game'. This game is a simplified variant of Arad and Rubinstein (2012)'s 11-20 Money Request game. To study the subjects' initial behavior in the game without any repeated game effects or learning from experience, we randomly rematched subjects and gave subjects no feedback about their performance during the course of the experiment.

At the beginning of phase 2 subjects were told how the games would be incentivized using tokens (see the third paragraph of Section 2.1 and Supplementary Web Appendix II for details). The experimenter further explained that: (i) choices and partners would be anonymous; (ii) partners would change from game to game; and (iii) subjects would be told the total number of tokens that they accumulated at the end of the experiment (and so would receive no feedback about individual games). Next, the experimenter explained how subjects could make decisions using the computer interface. Finally, subjects repeatedly played the 1-6 Token Request game.

[20] According to Carpenter et al., 1990, Raven test scores capture the ability to use abstract reasoning and correlate highly with scores on other complex cognitive tasks. In economics, the Raven test has enjoyed recent popularity; for example, Raven test scores have been found to correlate positively with fewer Bayesian updating errors (Charness et al., forthcoming), with more accurate beliefs (Burks et al., 2009) and with success in the p-beauty contest for adult populations (Gill and Prowse, 2016).

[21] Sets A and B coincide across the two tests. Two more difficult Raven tests also exist: Standard Progressive Matrices Plus and Advanced Progressive Matrices.

[22] Phase 1 never lasted more than fifty-five minutes.

2.3.1 Rules of the 1-6 Token Request game

The rules of the game were read to the children as follows, and were repeated multiple times throughout the experiment:[23]

Your partner and you are going to ask for an amount of tokens. The amount must be between 1 and 6. I will give you the amount of tokens you ask for. However, I will give you 10 more tokens if you ask for exactly one token less than your partner. How many tokens are you going to ask for?

We restricted the strategy set compared to Arad and Rubinstein (2012)'s 11-20 Money Request game to aid the understanding of the game by our younger subjects.[24,25]

2.3.2 Nash equilibrium in the 1-6 Token Request game

Just as in Arad and Rubinstein (2012)'s 11-20 Money Request game, the 1-6 Token Request game has no pure-strategy Nash equilibrium. Table 2 describes the unique symmetric mixed-strategy Nash equilibrium for risk-neutral players. See Supplementary Web Appendix V for proofs.

Choice	1	2	3	4	5	6
Probability	0.0	0.0	0.4	0.3	0.2	0.1

Table 2: Nash equilibrium mixing distribution

2.3.3 Level-k model predictions in the 1-6 Token Request game

Following Arad and Rubinstein (2012) in the 11-20 Money Request game, we use the level-k model of non-equilibrium thinking (Nagel, 1995; Stahl and Wilson, 1995) to analyze behavior in our 1-6 Token Request game variant.[26] Level-0 types are non-strategic. Following the level-0 specification in Arad and Rubinstein (2012), we assume that level-0 types choose 6 in our setting; this choice is instinctive, salient and guarantees a payoff of six in the absence of strategic thinking. Level-1 types act as if they best-respond to the belief that their opponent is level-0 and so choose

[23] The original wording in Arad and Rubinstein (2012) states that a player will receive a number of "additional" experimental units if she asks for exactly one unit less than the other player. In one of the pre-intervention pilots (Supplementary Web Appendix IV) it became apparent that the youngest subjects did not understand the meaning of the word 'additional'. Thus we modified the wording of the game for the sake of clarity.

[24] There is evidence that children develop a basic understanding of counting before the age of four and a basic understanding of addition and subtraction before the age of five (Wynn, 1990; Bryant et al., 1999; Canobi et al., 2002)

[25] To clarify the rules of the game, subjects were presented with hypothetical scenarios. In each scenario, the children were asked how may tokens each player would receive and which player, if any, received the ten additional tokens. After hearing their responses, the experimenter further explained in detail the allocation of tokens in each scenario. See the experimental instructions in Supplementary Web Appendix I for further details.

[26] Gill and Prowse (2016) use the level-k framework to study the relationship between cognitive ability and strategic sophistication in adults in the p-beauty contest game. See Crawford et al. (2013) for a survey of applications of the level-k model.

5, level-2 types act as if they best-respond to the belief that their opponent is level-1 and so choose 4, and so on.[27]

The game structure was designed by Arad and Rubinstein (2012) to naturally trigger level-k reasoning and to allow simple identification of types from choices. We focus on three key aspects here. First, the level-0 choice is salient and does not require strategic reasoning, and the behavior of level-$k > 0$ types is robust to a range of other assumptions about the behavior of level-0 types (e.g., uniform randomization and any distribution in which 6 is the modal choice). Second, there are no other obvious and simple decision rules given the absence of dominated strategies and pure-strategy equilibria. Finally, the best-response calculation of each level-$k > 0$ type is straightforward.

2.3.4 Variation in opponents in the 1-6 Token Request game

Baseline game

Subjects played the 'Baseline game' five times. In every repetition of the Baseline game, subjects played the 1-6 Token Request game against a randomly selected anonymous opponent from the same school, who could be of any age between five and twelve (random rematching). This was explained to the subjects, and thus all subjects shared the same information about the strategic sophistication of their opponent.

Raven game

Subjects played the 'Raven game' a single time. In the Raven game, subjects played the 1-6 Token Request game against a randomly selected anonymous opponent from the same academic year, and this was explained to the subjects. Furthermore, each subject was told whether her opponent's age-standardized Raven test score was in the top half of scores or the bottom half of scores of subjects of her age (subjects were never given information about their own performance in the Raven test).[28] By providing age-standardized information about the cognitive ability of opponents, we aimed to create exogenous variation in beliefs about the strategic sophistication of opponents (in a repeated beauty contest game with feedback, Gill and Prowse, 2016, find that Raven test scores predict how fast adults learn to play equilibrium, and that adults respond to information about their opponent's cognitive ability during the learning process).

Computer game

Subjects played the 'Computer game' a single time. In the Computer game, subjects played the 1-6 Token request game against the computer, which chose numbers uniform randomly. To convey the concept of uniform randomization to subjects, the experimenter explained that the computer did not understand the rules of the game and did not have any preference for particular numbers. Subjects were told that, instead, the computer would select the number of tokens at random, similarly to "rolling a die." Thus, all subjects were told the distribution from which their opponent selected its strategy, and so the opponent's behavior was predictable. For

[27] A level-6 type would cycle back to choosing 6, but we restrict attention to levels 0 to 5: empirically in adult populations the vast majority of subjects are found to be level-0 to level-3 types (Crawford et al., 2013).

[28] The small number of subjects exactly at the median were allocated to the top half.

a risk-neutral subject, the best-response to the computer's uniform randomization is to choose 5.[29]

2.3.5 Order of play and matching procedures

Table 3 describes the order of play in phase 2. As explained above, each subject played the 1-6 Token Request game seven times: each subject played the Baseline game five times, the Computer game a single time and the Raven game a single time. To control for order effects, subjects played the games in one of two orders, X and Y. Games 1, 4 and 7 were Baselines games for all subjects. Across games 2 and 3 we balanced the order of the Baseline game and Computer game. Across games 5 and 6 we balanced the order of the Baseline game and Raven game.

We randomly allocated subjects to sessions that included subjects only from the same school and academic year (but not necessarily from the same class). Half of the sessions from each school-year pair were randomly allocated to each order.[30] In the Raven game we randomly matched subjects from the same session; on average, sessions included ten subjects. In every repetition of the Baseline game we randomly matched subjects from the same school and order.[31]

Game	Order X	Order Y
1	Baseline game	Baseline game
2	Computer game	Baseline game
3	Baseline game	Computer game
4	Baseline game	Baseline game
5	Raven game	Baseline game
6	Baseline game	Raven game
7	Baseline game	Baseline game

Table 3: Order of games in phase 2.

[29]The payoff from choosing 6 is 6. The expected payoff from choosing $x < 6$ is $x + 10/6$, which is maximized at $x = 5$, giving $40/6 > 6$. A strongly risk-averse subject will choose 6. For any degree of risk aversion, $x < 5$ is never a best-response.

[30]The smallest school provided only enough subjects for one session per academic year: we allocated all these sessions to the same randomly chosen order (X). The allocation to order was balanced in terms of age, age-standardized theory-of-mind ability, age-standardized cognitive ability, gender and school. A χ-squared test of the joint null that the differences in the means of the characteristics between order X and order Y all equal zero gives $p = 0.368$. The χ-squared test is based on the results of a probit regression of an indicator for order X on an intercept and the five characteristics.

[31]When there was an odd number of subjects in a session (for the Raven game) or in a school-order pair (for the Baseline games), one subject was randomly matched with two others (the primary and secondary opponents). The actions of the first subject were used to compute the payoffs of the primary and secondary opponents, but only the actions of the primary opponent were used to compute the payoffs of the first subject.

2.4 Phase 3: Gift-exchange game

In phase 3 subjects played the 'Gift-exchange game' a single time. Subjects were matched with a randomly selected anonymous partner from the same academic year, and this was explained to the subjects.[32] The currency in this game were 'super-tokens' worth four normal tokens.[33] In each pair, one subject took the role of the 'allocator', while the other took the role of the 'receiver'. The allocator chose between two ways of splitting a pie of ten super-tokens. After finding out the split chosen by the allocator, the receiver chose whether or not to give one super-token back to the allocator.[34]

Figure 1 shows the game tree. In Treatment A, the allocator chose between taking 8 super-tokens (and thus giving 2 to the receiver) or taking 5 super-tokens (and giving 5 to the receiver). In Treatment B, the allocator chose between taking 2 super-tokens (and thus giving 8 to the receiver) or taking 5 super-tokens (and giving 5 to the receiver). In total, we have 390 subjects in Treatment A and 340 subjects in Treatment B.[35] Of the 730 subjects, 345 were allocators, while 385 were receivers.[36]

The game is designed to measure how subjects respond to intentions. The theory that underlies intentions-based reciprocity in strategic games (Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006) models reciprocal behavior as depending on beliefs about intentions, and the game structure is inspired by Falk and Fischbacher (2006)'s survey evidence that the perceived intentions of an allocator depend on the set of unchosen or foregone alternatives.[37] In particular, across treatments we will compare the decision of receivers after receiving 5 super-tokens (right-hand-side decision node in Figure 1). At that node, in the two treatments the receiver has been given the same number of super-tokens and her choice has the same distributional consequences. However, receivers who like to reciprocate and who take into account the intentions of the allocator are more likely to give back a super-token in Treatment A, since the allocator in that treatment sacrificed the option of taking 8 super-tokens in order to select the even split, while in Treatment B the even split is more advantageous to the allocator than her alternative option of taking 2 super-tokens. Importantly, a selfish allocator has a strictly dominant strategy (in Treatment A taking 8 super-tokens, and in Treatment B taking 5 super-tokens), and so choosing the even split in Treatment A is unambiguously generous (in the sense that the allocator is giving up money for sure).[38]

[32]Phases 2 and 3 shared the same sessions. As described in Section 2.3.5, we randomly allocated subjects to sessions that included subjects only from the same school and academic year (but not necessarily from the same class). Just like for the Raven game in Phase 2, we randomly matched subjects from the same session.

[33]Super-tokens were given a different color to distinguish them from normal tokens.

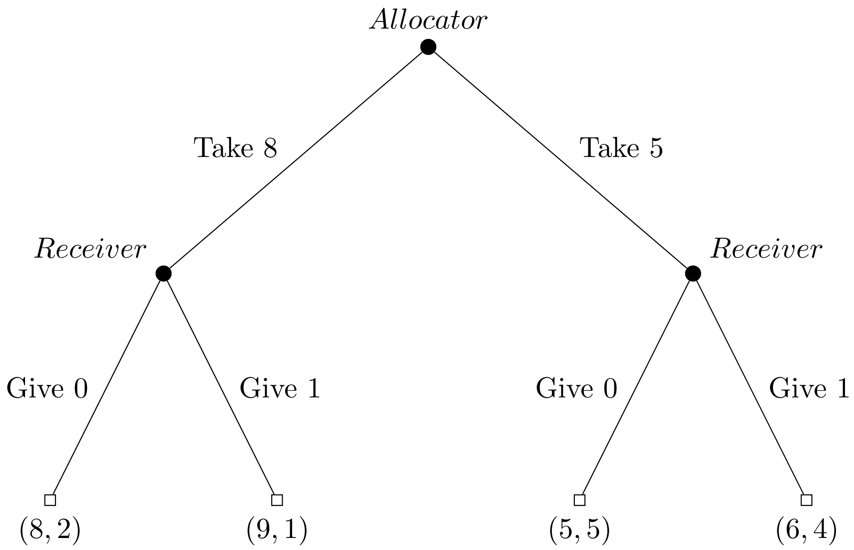
[34]As explained at the beginning of phase 2, subjects were told only the total number of tokens that they accumulated in the experiment. Thus allocators were not told whether their matched receiver gave them a super-token.

[35]The allocation to Treatment A or B in the Gift-exchange game and to order X or Y in Phase 2 was the same, and thus the balance test with respect to the allocation to order reported in footnote 30 also applies to the allocation to treatment. As noted in Section 2.3.5, all of the subjects in the smallest school were allocated to the same randomly chosen order (X) / treatment (A), which explains why we have more subjects in Treatment A.

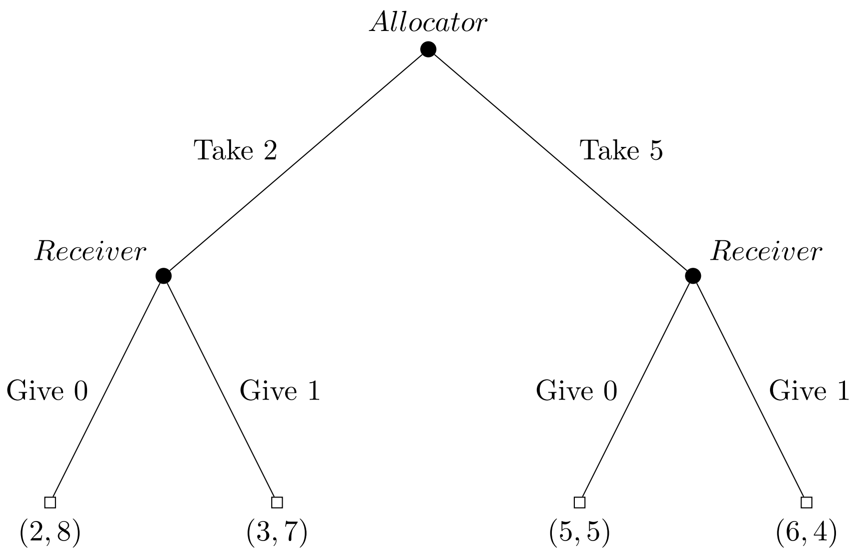
[36]On average, sessions included ten subjects. When there was an odd number of subjects in a session, the choice of one randomly selected allocator affected the payoff of two receivers, but only the choice of the first of these receivers affected the payoff of this allocator, which explains why more subjects were receivers than allocators.

[37]Fehr et al. (1998) popularized gift-exchange games more generally. See Charness and Levine (2007) for a discussion of methods used to measure the role of intentions in strategic games.

[38]By contrast, in the ultimatum game a selfish proposer needs to consider how the receiver will respond to his offer, and thus apparently generous offers can be strategic.



(a) Treatment A



(b) Treatment B

Figure 1: Gift-exchange game

3 Results from the competitive games

3.1 Strategic sophistication and best-responding

To measure how cognitive skills influence the development of the strategic sophistication, we start by studying how age, theory-of-mind and cognitive ability interact to predict the probability that our child subjects best-respond to the empirical distribution of choices in the competitive games. We include data from the Baseline games and the Computer game, in which subjects shared the same information about their opponent (in each of the five repetitions of the Baseline game, the opponent was known to be another child drawn randomly from the subjects of all ages from the same school; in the Computer game the opponent was a computer known to randomize uniformly). Importantly, the subjects could not learn from experience about the empirical distribution of choices: in order to study the subjects' initial behavior in a new strategic environment, the subjects did not receive any feedback during the experiment.

Table 4 presents the results from a linear probability model, with two-sided tests of statistical significance. The dependent variable is the probability of best-responding to the empirical distribution of choices in the games. We pool choices from the five Baseline games to calculate a single empirical distribution; the distribution for the Computer game is known to be uniform. The independent variables are age, age-standardized theory-of-mind ability and age-standardized cognitive ability. As explained in Section 2.2.3, we standardize test scores within age group to separate the effect of age from that of theory-of-mind and cognitive ability, and so the influence of psychometric test scores on behavior captures the effect of variation in test scores within age group. To help interpretation, we make the youngest age group (five-year-olds) the omitted category. The first two columns pool the data from the Baseline games and the Computer game, while the next four columns break the data down according to game type. As we will see in Table 5 below, the best-response in the Baseline games is to choose 5. The best-response in the Computer game is also 5 (see Section 2.3.4 above)

Column 1 in Table 4 shows that strategic ability improves with age. Furthermore, within an age group, children who have a better theory-of-mind and children who are more cognitively able tend to be more strategically sophisticated. Specifically, the probability of best-responding increases by about two percentage points with each additional year of life. Furthermore, among children of the same age, a one-standard-deviation increase in age-standardized theory-of-mind ability raises the probability of best-responding by about two percentage points. Similarly, a one-standard-deviation increase in age-standardized cognitive ability also raises the probability of best-responding by about two percentage points. Relative to the average best-response rate of nineteen percent, these changes increase the probability of best-responding by about ten percent.

The second column in Table 4 interacts age with our psychometric measures of cognitive skill. We find that as children get older, marginal changes in cognitive ability matter more for strategic sophistication, while theory-of-mind does not interact significantly with age. For example, among eight-year-olds a one-standard-deviation increase in age-standardized cognitive ability raises the probability of best-responding by 1.4 percentage points, while the corresponding increase for eleven-year-olds is 5.3 percentage points.[39] We conclude that age and cognitive ability act in tandem as complements in the development of strategic sophistication. In contrast, age and theory-of-mind operate independently in their effects on strategic ability.

	Pooled		Baseline games		Computer game	
	(1)	(2)	(3)	(4)	(5)	(6)
Age	0.020***	0.020***	0.021***	0.021***	0.017**	0.017**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.008)	(0.008)
Theory-of-mind	0.016**	0.005	0.017*	0.001	0.012	0.024
(T.o.M.)	(0.008)	(0.013)	(0.009)	(0.014)	(0.016)	(0.027)
Cognitive ability	0.018**	-0.025	0.018**	-0.013	0.015	-0.083***
(Cogn. Ab.)	(0.008)	(0.016)	(0.009)	(0.017)	(0.016)	(0.030)
T.o.M. × Age		0.003		0.005		-0.004
		(0.004)		(0.004)		(0.008)
Cogn. Ab. × Age		0.013***		0.010**		0.031***
		(0.004)		(0.004)		(0.008)
Intercept	0.126***	0.128***	0.122***	0.123***	0.149***	0.151***
	(0.014)	(0.014)	(0.015)	(0.015)	(0.027)	(0.027)
Subjects	730	730	730	730	730	730
Subject-round obs.	4,380	4,380	3,650	3,650	730	730

Notes: All columns report OLS regressions where the dependent variable is an indicator taking value 1 if a subject best-responded to the empirical distribution of choices in the games (we pool all choices from the five Baseline games to calculate a single empirical distribution; the distribution for the Computer game is known to be uniform). The independent variables are age, age-standardized theory-of-mind ability and age-standardized cognitive ability (as explained in Section 2.2.3, we standardize test scores within age group). To help interpretation, we make the youngest age group the omitted category (by coding five-year-olds as age zero). Heteroskedasticity-robust standard errors, clustered at the subject level, are shown in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table 4: Probability of best-responding.

When we look at the data from the five repetitions of the Baseline game (Columns 3 and 4 in Table 4), the results are virtually identical to those for the pooled data (Columns 1 and 2). This is perhaps not surprising given that most of our data come from the Baseline games.

[39] The derivative of the probability of best-responding with respect to cognitive ability in Column 2 is given by $-0.025 + 0.013 \times$ (age - 5). As explained in the notes to Table 4, we have coded five-year-olds as age zero in these regressions to make the youngest age group the omitted category.

Finally, when we compare the data from the Computer game to that from the Baseline games, the effect of theory-of-mind is weaker in the Computer game (0.012 in Column 5 vs. 0.017 in Column 3). Although the difference is not statistically significant, this suggests that when playing against a human opponent whose behavior is difficult to predict, theory-of-mind matters more than when playing against a computer whose behavior is predictable. Cognitive ability, on the other hand, interacts statistically significantly with age in both cases.

Table SWA.3 in Supplementary Web Appendix VI shows that the parameter estimates in Table 4 are stable when we add controls for the demographics that we collected (gender and school).

Payoffs are a noisy measure of strategic sophistication. Payoffs depend on the behavior of the specific opponents that a subject is matched with, and so a subject who best-responds to the empirical distribution of choices is not guaranteed a high payoff. For completeness, Tables SWA.4 and SWA.5 in Supplementary Web Appendix VI show how tokens earned in the games vary with age, theory-of-mind and cognitive ability, even though we think that the best-response regressions in Table 4 provide a more accurate picture of how these characteristics influence strategic ability. The noisiness of payoff data reduces the precision of our estimates. Nonetheless, two broad patterns that emerged when studying the probability of best-responding in Table 4 carry over when evaluating payoffs. First, payoffs increase with age. Second, age and cognitive ability act as complements in determining payoffs, while age and and theory-of-mind act independently.

3.2 Choices, level-k behavior, and deviation from Nash equilibrium

Next we study how age and our psychometric measures of cognitive skill affect choices, level-k behavior and deviation from Nash equilibrium in the Baseline games, in which subjects shared the same information about the strategic sophistication of their human opponent (in each of the five repetitions of the Baseline game, the opponent was another child drawn randomly from the subjects of all ages from the same school).

3.2.1 Choices and level-k behavior

Figure 2 reports the distribution of choices in the Baseline games, while the fourth to sixth rows of Table 5 present the results from a multinomial choice logit that provide estimates of the marginal effects of age, age-standardized theory-of-mind ability and age-standardized cognitive ability on the probability of each of the six choices, with two-sided tests of statistical significance. The second row of Table 5 shows the correspondence between choices and levels explained in Section 2.3.3. The third row shows the expected payoff from each choice, given the empirical distribution of choices across the five Baseline games.

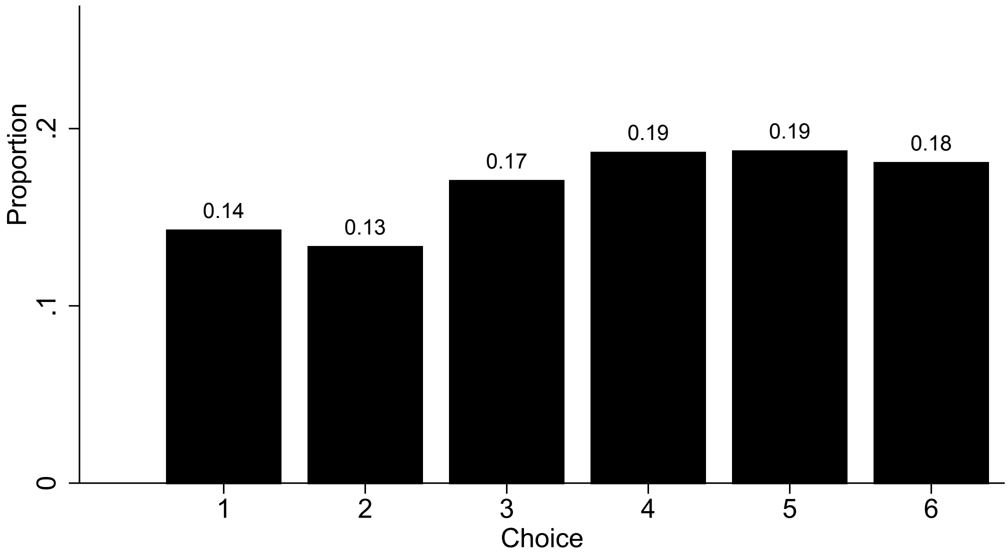


Figure 2: Distribution of choices in the Baseline games

Choice	1	2	3	4	5	6
Level-k	Level-5	Level-4	Level-3	Level-2	Level-1	Level-0
Expected payoff	2.332	3.704	4.863	5.871	6.805	6.000
Age	-0.022***	0.001	0.003	0.027***	0.020***	-0.029***
	(0.005)	(0.003)	(0.004)	(0.003)	(0.004)	(0.005)
Theory-of-mind	0.000	-0.010	-0.014*	0.006	0.017*	0.001
	(0.009)	(0.007)	(0.008)	(0.008)	(0.009)	(0.010)
Cognitive ability	-0.016*	-0.004	0.006	0.012	0.019**	-0.017*
	(0.009)	(0.007)	(0.007)	(0.007)	(0.009)	(0.010)
Subjects				730		
Subject-round obs.				3,650		

Notes: The table reports average marginal effects from a multinomial logit. The dependent variable is a categorical indicator of subjects' choices in the five Baseline games. The independent variables are age, age-standardized theory-of-mind ability and age-standardized cognitive ability (as explained in Section 2.2.3, we standardize test scores within age group). Section 2.3.3 explains the correspondence between choices and levels. We calculate the expected payoff to each choice given the empirical distribution of choices across the five Baseline games. Heteroskedasticity-robust standard errors, clustered at the subject level, are shown in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table 5: Probability of choices in the Baseline games.

The third row of Table 5 shows that the best-response to the empirical distribution is to choose 5. Thus, the fifth column of the table replicates the results from Column 3 of Table 4: the probability of best-responding increases with age, and within-age-group with both theory-of-mind ability and cognitive ability.

We now turn to the level-k model of non-equilibrium thinking. Recall from Section 2.3.3 that the game structure is designed to naturally trigger level-k reasoning. Non-strategic level-0 types choose 6; level-1 types act as if their opponent is a strategically unsophisticated level-0 type and so choose 5; level-2 types act as if their opponent is a level-1 type and so choose 4; and so on. Since subjects' choices vary from round to round, we allow their level to also vary across rounds.[40]

Three main results on level-k thinking emerge. First, our child subjects are more likely to exhibit level-1 behavior as they get older, and as their theory-of-mind ability and cognitive ability improve compared to their peers of the same age (fifth column of Table 5). Relative to the average shown in Figure 2, an additional year of life, a one-standard-deviation increase in age-standardized theory-of-mind ability and a one-standard-deviation increase in age-standardized cognitive ability all increase the probability of level-1 behavior by about about ten percent. Second, the subjects are less likely to be unsophisticated level-0 types as they age and become more cognitively able (sixth column). Relative to the average, these changes represent decreases of sixteen and nine percent respectively. Third, older children are more likely to be level-2 types who act as if their opponent is a level-1 type (fourth column). We note that, in expectation, level-2 types earn less than level-1 types: the proportion of level-1 types in our sample is not large enough to make level-2 behavior optimal.

The first column of Table 5 shows that the probability of choosing 1 decreases with age and with cognitive ability. This choice returns a low payoff; furthermore, the literature finds that for adult populations almost nobody thinks at a level higher than level-3 (Crawford et al., 2013). Thus, we interpret this decrease as a reduction in random play by children who do not understand well the strategic environment, which happens to mimic level-5 behavior.[41]

Table SWA.6 in Supplementary Web Appendix VI shows that the parameter estimates in Table 5 are stable when we add controls for the demographics that we collected (gender and school).

3.2.2 Deviation from Nash equilibrium

So far in this section we have studied how age and cognitive skills influence individual choices in the Baseline games. Now we investigate whether these same characteristics move the distribution of choices in the Baseline games toward the Nash equilibrium prediction (which we reported in Section 2.3.2).

The notes to Table 6 explains the construction of our deviation metric. In short, we first split our subjects according to whether they are above or below median age, median age-standardized theory-of-mind ability or median age-standardized cognitive ability, and for each sub-sample we

[40] A more complicated alternative could assume that each subject follows a fixed level but chooses with noise.

[41] Theory-of-mind appears to reduce the likelihood of level-3 behavior (third column), but we find no statistically significant effects on level-3 behavior when we add controls (see Table SWA.6 in Supplementary Web Appendix VI).

then calculate the average deviation from Nash equilibrium behavior across the five Baseline games.

Table 6 shows how the characteristics change the distance to the Nash equilibrium prediction. We can see that the distribution of choices of older children is substantially closer to Nash equilibrium than that of younger children. The same is true for more cognitively able children and children with better theory-of-mind, although the differences are not as quantitatively important. Since we only have five observations per sub-sample (one for each repetition of the Baseline game), statistical tests are not appropriate.

	Below median	Above median	Difference
Age	0.89	0.61	0.28
Theory-of-mind	0.76	0.72	0.04
Cognitive ability	0.77	0.69	0.08

Notes: For each sub-sample, we calculate the deviation metric $\frac{1}{5}\sum_{g=1}^{5}\sum_{c=1}^{6}|f^g(c)-f^*(c)|$, where $g \in \{1, 2, \ldots, 5\}$ denotes a Baseline game, $c \in \{1, 2, \ldots, 6\}$ denotes a choice, $f^*(c)$ is the Nash equilibrium density (reported in Section 2.3.2), and $f^g(c)$ is the empirical density in Baseline game g for that sub-sample. In other words, for each sub-sample, for each game we first calculate the sum of absolute deviations between the observed frequency of a choice in that game and the Nash equilibrium frequency, and then average across the five Baseline games. The sub-samples are based on age, age-standardized theory-of-mind ability and age-standardized cognitive ability (as explained in Section 2.2.3, we standardize test scores within age group). To minimize the number of subjects at the median age, we classify subjects to be above or below the median age using birth date (rather than age in years). The small number of subjects exactly at the median of age, age-standardized theory-of-mind ability or age-standardized cognitive ability were allocated to the above-median category.

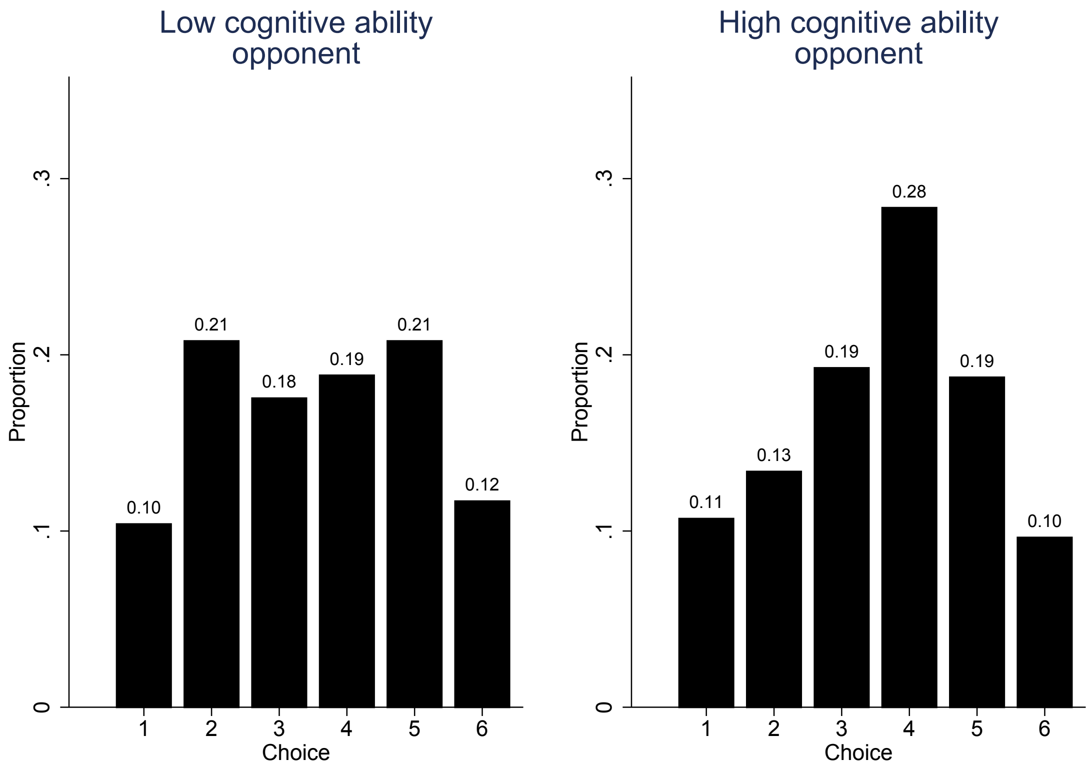
Table 6: Deviations from Nash equilibrium in the Baseline games.

3.3 Response to information about opponent cognitive ability

As described in Section 2.3.4, in the Raven game we aimed to create exogenous variation in beliefs about the strategic sophistication of opponents by giving subjects information about the cognitive ability of their opponent (but not about their own ability). Specifically, in the Raven game each subject was matched with another subject from the same academic year and was told whether her opponent's age-standardized Raven test score was in the top half or the bottom half of scores of subjects of her age. Here we describe subjects in the top half of the distribution as of 'high cognitive ability' and subjects in the bottom half as of 'low cognitive ability', but we did not use these subjective labels in the experiment. We now study whether our child subjects responded to this information.

In Section 3.1 we found that the positive effect of a subject's own cognitive ability on her capacity to best-respond is stronger for older children. Since subjects were matched by academic year in the Raven game, this finding leads us to first investigate whether older subjects in the Raven game responded to information about the cognitive ability of their (also older) opponent. In particular, we first look at the behavior of subjects whose age was above the median.

Figure 3 shows that older subjects shift their behavior when they are told that they face an (also older) opponent of high cognitive ability. The left-hand-side panel shows the distribution of choices for older subjects who faced an opponent from the bottom half of the age-standardized distribution of Raven test scores, while the right-hand-side panel shows the distribution for older subjects who faced an opponent from the top half of the distribution. Older subjects whose opponent is of high cognitive ability are more likely to choose 4. Older subjects whose opponent is of low cognitive ability spread their choices more uniformly over 2 to 5, which suggests that they are less certain about how their opponent will behave. This ability to adjust behavior to the characteristics of the opponent provides behavioral evidence of the emergence of a sophisticated strategic theory-of-mind in children as young as eight to twelve years old.[42]



Notes: The figure shows the distribution of choices in the Raven game for subjects above the median age. The left-hand-side (right-hand-side) panel shows the distribution for subjects matched with an opponent whose age-standardized cognitive ability was below (above) the median (as explained in Section 2.2.3, we standardize test scores within age group). To minimize the number of subjects at the median age, we classify subjects to be above or below the median age using birth date (rather than age in years). The small number of subjects exactly at the median of age or age-standardized cognitive ability were allocated to the above-median category.

Figure 3: Distribution of choices in the Raven game for older subjects.

Table 7 shows that the shift toward choosing 4 when facing a high-cognitive-ability opponent is statistically significant. The table presents the results from a multinomial choice logit that provide estimates of the effect of being matched with a high-cognitive-ability opponent (instead of a low-cognitive-ability opponent) on the probability of each of the six choices for subjects

[42]Table 1 in Section 2.1 shows the distribution of ages; the median lies within the eight-year-old category. To minimize the number of subjects at the median age, we classify subjects to be above or below the median age using birth date (rather than age in years).

above the median age, with two-sided tests of statistical significance. We include specifications with and without controls. The results show that when an older subject is told that her (also older) opponent is of high cognitive ability, on average she increases her probability of choosing 4 by about ten percentage points, with the effect statistically significant at the five-percent level. Subjects are also less likely to choose 2, which reflects the shift in Figure 3 from a uniform distribution over choices 2 to 5 in the left-hand-side panel to a distribution centered around 4 in the right-hand-side panel.

Subjects above median age						
Choice	1	2	3	4	5	6
Level-k	Level-5	Level-4	Level-3	Level-2	Level-1	Level-0
No controls						
High cognitive ability opponent	0.003 (0.033)	-0.073* (0.040)	0.016 (0.042)	0.096** (0.046)	-0.021 (0.043)	-0.021 (0.033)
With controls						
High cognitive ability opponent	0.000 (0.032)	-0.071* (0.039)	0.010 (0.043)	0.115** (0.046)	-0.021 (0.043)	-0.033 (0.032)
Subjects				341		
Subject-round obs.				341		

Notes: The table reports average marginal effects from a multinomial logit for subjects above the median age. The dependent variable is a categorical indicator of subjects' choices in the Raven game. The independent variable is an indicator taking value 1 when a subject was matched with an opponent whose age-standardized cognitive ability was above the median (as explained in Section 2.2.3, we standardize test scores within age group). The model with controls further includes age, age-standardized theory-of-mind ability, age-standardized cognitive ability, gender and school. We have 341 subjects above the median age for our whole sample and who were given information about the cognitive ability of their opponent; this is less than half of our sample because some subjects were matched with children for whom we had no cognitive ability score (see footnote 15). To minimize the number of subjects at the median age, we classify subjects to be above or below the median age using birth date (rather than age in years). The small number of subjects exactly at the median of age or age-standardized cognitive ability were allocated to the above-median category. Section 2.3.3 explains the correspondence between choices and levels. Heteroskedasticity-robust standard errors are shown in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table 7: Probability of choices in the Raven game for older subjects.

Using the prism of the level-k model of non-equilibrium thinking (see Section 2.3.3), Table 7 shows that older subjects are more likely to exhibit level-2 behavior when they are told that they face an (also older) opponent of high cognitive ability. With caution, we note that these results mesh with those from Section 3.2: in the five repetitions of the Baseline game we found that age and cognitive ability act together to increase the likelihood of choosing 5, which corresponds to level-1 behavior; and here we find that subjects who face older and more cognitively able opponents are more likely to act as level-2 types, who best-respond to level-1 behavior. We use caution in noting this correspondence because in the Raven game subjects were matched by academic year, but in the five repetitions of the Baseline game they were matched randomly with subjects of any age from the same school.

Table SWA.7 in Supplementary Web Appendix VI and Figure SWA.7 in Supplementary Web Appendix VII repeat the analysis for subjects below the median age. For the younger children, we find no statistically significant change in behavior according to the cognitive ability of their (also younger) opponent. This is consistent with the finding in Section 3.1 that the positive effect of a subject's own cognitive ability on her capacity to best-respond is weaker for younger children.

4 Results from the Gift-exchange game

In this section we turn to the results from the Gift-exchange game. Section 2.4 describes the game, while Figure 1 presents the game tree. As we explain in Section 2.4, the game was designed to measure how subjects respond to intentions. To recap, the allocator chooses between two ways of splitting a pie of ten super-tokens (a super-token is worth four normal tokens), and the receiver decides whether or not to give a super-token back to the allocator after observing the chosen split. In Treatment A, the allocator chooses between an 8/2 split and a 5/5 split. In Treatment B, the allocator chooses between a 2/8 split and a 5/5 split.

To identify cleanly whether subjects respond to intentions, we compare across treatments the decision of receivers after the allocator chooses a 5/5 split (that is, the allocator takes 5 super-tokens and gives the other 5 to the receiver). Conditional on this split, in the two treatments the receiver has been given the same number of super-tokens and her choice has the same distributional consequences. However, if the receiver likes to reciprocate and cares about the intentions of the allocator, she is more likely to give back a super-token in Treatment A, since the allocator in that treatment gave up the option of taking 8 super-tokens in order to select the even split, while in Treatment B the even split benefits the allocator. As we explain in Section 2.4, choosing the even split in Treatment A is unambiguously generous since the 8/2 split is strictly dominant for a selfish allocator.

	(1)	(2)	(3)
Treatment A	0.095	0.088	-0.135
	(0.068)	(0.068)	(0.100)
Age		0.015	0.000
		(0.014)	(0.017)
Theory-of-mind		0.009	-0.029
(T.o.M.)		(0.028)	(0.032)
Cognitive Ability		0.002	0.003
(Cogn. Ab.)		(0.031)	(0.036)
Treatment A. × Age			0.062**
			(0.029)
Treatment A. × T.o.M.			0.153**
			(0.060)
Treatment A. × Cogn. Ab.			0.006
			(0.064)
Intercept	0.160***	0.111**	0.164***
	(0.030)	(0.056)	(0.062)
Subjects	207	207	207
Subject-round obs.	207	207	207

Notes: The sample consists of the 207 receivers whose matched allocator chose the 5/5 split. All columns report OLS regressions where the dependent variable is an indicator taking value 1 if a receiver gave one super-token back to the allocator. The independent variable in Column (1) is an indicator for Treatment A. In Columns (2) and (3) we add age, age-standardized theory-of-mind ability and age-standardized cognitive ability (as explained in Section 2.2.3, we standardize test scores within age group). To help interpretation, we make the youngest age group the omitted category (by coding five-year-olds as age zero). Heteroskedasticity-robust standard errors are shown in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table 8: Effect of treatment on probability that receiver gives one super-token to allocator

Table 8 presents the results from a linear probability model, with two-sided tests of statistical significance. We study the 207 receivers whose matched allocator chose the 5/5 split. The dependent variable is the probability that the receiver gives one super-token back to the allocator. In Column (1) we regress this probability on an indicator for Treatment A: conditional on the 5/5 split, receivers are about ten percentage points more likely to reciprocate in Treatment A, although the effect is not quite statistically significant. When we add age, age-standardized theory-of-mind ability and age-standardized cognitive ability as independent variables in Column (2), we see that these characteristics do not predict whether the receiver gives back a super-token.[43] However, in Column (3) we find that age and age-standardized theory-of-mind ability interact statistically significantly with the treatment. Thus, conditional on a 5/5 split, older children and children whose theory-of-mind compares favorably to peers of the same age are more likely to reciprocate in Treatment A than in Treatment B.

This result shows that age and theory-of-mind predict whether receivers respond to the allocator's intentions in our gift-exchange game. Each additional year of life increases the effect of Treatment A on the probability of giving back one super-token by about six percentage points, while a one-standard-deviation increase in age-standardized theory-of-mind ability increases the effect of Treatment A by about fifteen percentage points. These effects are large compared to the average impact of Treatment A of ten percentage points from Column (1). Despite these findings for age and theory-of-mind, cognitive ability has no effect: the interaction of Treatment A and age-standardized cognitive ability is quantitatively small and far from statistical significance. These striking results suggest that different psychometric measures of cognitive skill correspond to different cognitive processes in strategic situations that involve the understanding of intentions.

Finally, we check whether subjects exhibit unconditional ('simple') reciprocity. That is, we study whether receivers respond to the number of tokens given to them by their matched allocator when we do not condition on treatment and so disregard the allocator's option set. Table 9 replicates the regressions from Table 8, but replacing the indicator for Treatment A with the number of super-tokens given to the receiver (2 in the 8/2 split, 5 in the 5/5 split, and 8 in the 2/8 split) and using all the receivers. Column (1) shows that receivers are indeed more likely to reciprocate when they are given more tokens. However, Column (3) shows that this simple reciprocity does not interact statistically significantly with theory-of-mind (or with cognitive ability). Older subjects are more likely to exhibit simple reciprocity, with the effect significant at the ten-percent level.

Tables SWA.8 and SWA.9 in Supplementary Web Appendix VI show our results are robust when we add controls for the demographics that we collected (gender and school).

[43] As explained in Section 2.2.3, we standardize test scores within age group to separate the effect of age from that of theory-of-mind and cognitive ability, and so the influence of test scores on behavior captures the effect of variation in test scores within age group. To help interpretation, we make the youngest age group (five-year-olds) the omitted category.

	(1)	(2)	(3)
Number of super-tokens received	0.040***	0.041***	0.015
(Super-tokens)	(0.011)	(0.011)	(0.018)
Age		0.007	-0.027
		(0.009)	(0.019)
Theory-of-mind		0.010	-0.026
(T.o.M.)		(0.018)	(0.034)
Cognitive ability		-0.003	-0.049
(Cogn. Ab.)		(0.021)	(0.042)
Super-tokens × Age			0.009*
			(0.005)
Super-tokens × T.o.M.			0.009
			(0.009)
Super-tokens × Cogn. Ab.			0.012
			(0.011)
Intercept	-0.014	-0.038	0.064
	(0.038)	(0.050)	(0.070)
Subjects	385	385	385
Subject-round obs.	385	385	385

Notes: The sample consists of all 385 receivers (see footnote 36). All columns report OLS regressions where the dependent variable is an indicator taking value 1 if a receiver gave one super-token back to the allocator. The independent variable in Column (1) is the number of super-tokens given to the receiver (2 in the 8/2 split, 5 in the 5/5 split, and 8 in the 2/8 split). In Columns (2) and (3) we add age, age-standardized theory-of-mind ability and age-standardized cognitive ability (as explained in Section 2.2.3, we standardize test scores within age group). To help interpretation, we make the youngest age group the omitted category (by coding five-year-olds as age zero). Heteroskedasticity-robust standard errors are shown in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table 9: Effect of number of tokens received on
probability that receiver gives one super-token to allocator

5 Conclusion

Cognitive skills vary widely across the population. To better grasp how cognitive skills translate into life outcomes, we need to understand more about the mechanisms by which cognitive skills influence economic behavior and success. Strategic sophistication is an important mediator in this relationship: people constantly make strategic decisions (from allocating effort in competitions to reciprocating kind and bad behavior) where successful outcomes often require high-level reasoning. Since ability gaps open early in life, it is particularly important to understand how strategic sophistication develops in children, and to learn how the development of strategic sophistication depends on observable cognitive skills.

In this article we study how the development of strategic sophistication depends on cognitive ability and theory-of-mind. Our analysis shows that children who have better cognitive skills tend to be more strategically sophisticated in competitive environments. Specifically, we show that age and cognitive ability work together as complements in the development of strategic sophistication, whereas age and theory-of-mind act independently. Strikingly, children as young as eight to twelve years old respond to variation in information about their opponent's strategic sophistication, which provides support for the emergence of a sophisticated strategic theory-of-mind. Furthermore, different cognitive skills influence strategic sophistication in distinct ways. For example, whereas theory-of-mind and age strongly predict children's reactions to intentions in a gift-exchange game, cognitive ability does not.

Our results suggest that interventions designed to promote cognitive skills could also improve life outcomes by enhancing good decision-making in situations that require strategic thinking. These insights will help to support childhood interventions that aim to improve cognitive skills. Furthermore, understanding the distinct effects of different cognitive skills on the development of strategic sophistication will help to better target these policies. Such understanding will also facilitate the use of psychometric testing to predict individuals' behavior and outcomes in real-life environments that involve strategic decision-making. Finally by showing how non-equilibrium level-k thinking develops in childhood, and how this development interacts with observable cognitive skills, we hope to add empirical foundations to new theoretical work that seeks to model bounds on depth of reasoning.

Future research could investigate how specific early-life interventions aimed at improving cognitive skills translate into strategic sophistication in both competitive and collaborative settings. Similarly, it is important to understand the extent to which spurring the type of high-level reasoning involved in strategic decision-making can feed back into the development of cognitive skills in non-strategic settings. The latter relationship is especially crucial in early-life settings when cognitive skills are most malleable. Finally, in future work we hope to investigate the role of non-cognitive and social skills in the development of strategic sophistication.

References

Alaoui, L. and **Penta, A.** (2016a). Cost-benefit analysis in reasoning. *Mimeo, University of Wisconsin*

Alaoui, L. and **Penta, A.** (2016b). Endogenous depth of reasoning. *Review of Economic Studies*, 83(4): 1297–1333

Arad, A. and **Rubinstein, A.** (2012). The 11-20 Money Request game: A level-k reasoning study. *American Economic Review*, 102(7): 3561–3573

Baron-Cohen, S., **Leslie, A.M.**, and **Frith, U.** (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21(1): 37–46

Battigalli, P., **Dufwenberg, M.**, and **Smith, A.** (2017). Frustration and anger in games. *IGIER Working Paper 539, Bocconi University*

Bereby-Meyer, Y. and **Fiks, S.** (2013). Changes in negative reciprocity as a function of age. *Journal of Behavioral Decision Making*, 26(4): 397–403

Binmore, K. (2007). *Playing for Real: A Text on Game Theory.* Oxford University Press

Brañas-Garza, P., **García-Muño, T.**, and **Hernán, R.** (2012). Cognitive effort in the beauty contest game. *Journal of Economic Behavior & Organization*, 83(2): 254–260

Brocas, I. and **Carrillo, J.D.** (2017). The determinants of strategic thinking in preschool children. *Mimeo, University of Southern California*

Brocas, I., **Carrillo, J.D.**, and **Kodaverdian, N.** (2017). Altruism and strategic giving in children and adolescents. *Mimeo, University of Southern California*

Brosig-Koch, J., **Heinrich, T.**, and **Helbach, C.** (2015). Exploring the capability to reason backwards: An experimental study with children, adolescents, and young adults. *European Economic Review*, 74: 286–302

Bryant, P., **Christie, C.**, and **Rendu, A.** (1999). Children's understanding of the relation between addition and subtraction: Inversion, identity and decomposition. *Journal of Experimental Child Pychology*, 74: 194–212

Bueno-Guerra, N., **Leiva, D.**, **Colell, M.**, and **Call, J.** (2016). Do sex and age affect strategic behavior and inequity aversion in children? *Journal of Experimental Child Psychology*, 150: 285–300

Burks, S., **Carpenter, J.**, **Goette, L.**, and **Rustichini, A.** (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences*, 106(19): 7745–7750

Burnham, T., **Cesarini, D.**, **Johannesson, M.**, **Lichtenstein, P.**, and **Wallace, B.** (2009). Higher cognitive ability is associated with lower entries in a p-beauty contest. *Journal of Economic Behavior & Organization*, 72(1): 171–175

Canobi, K., **Reeve, R.**, and **Pattison, P.** (2002). Young children's understanding of addition concepts. *Educational Psychology*, 22(5): 513–532

Carpenter, P., **Just, M.A.**, and **Shell, P.** (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review*, 97(3): 404–413

Charness, G. and **Levine, D.I.** (2007). Intention and stochastic outcomes: An experimental study. *Economic Journal*, 117(522): 1051–1072

Charness, G., **Rustichini, A.**, and **Van de Ven, J.** (forthcoming). Self-confidence and strategic behavior. *Experimental Economics*

Costa-Gomes, M., **Crawford, V.P.**, and **Broseta, B.** (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, 69(5): 1193–1235

Crawford, V., **Costa-Gomes, M.**, and **Iriberri, N.** (2013). Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature*, 51(1): 5–62

Cunha, F., **Heckman, J.J.**, and **Schennach, S.M.** (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3): 883–931

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

Devetag, G., **Di Guida, S.**, and **Montinari, N.** (2013). Theory of mind in social sciences: An experiment on strategic thinking in children. *Methode-Analytic Perspectives*, 2: 67–92

Dufwenberg, M. and **Kirchsteiger, G.** (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, 47(2): 268–298

Ellingsen, T. and **Johannesson, M.** (2008). Pride and prejudice: The human side of incentive theory. *American Economic Review*, 98(3): 990–1008

Falk, A., **Fehr, E.**, and **Fischbacher, U.** (2003). On the nature of fair behavior. *Economic Inquiry*, 41(1): 20–26

Falk, A. and **Fischbacher, U.** (2006). A theory of reciprocity. *Games and Economic Behavior*, 54(2): 293–315

Fan, C.P. (2000). Teaching children cooperation — An application of experimental game theory. *Journal of Economic Behavior & Organization*, 41(3): 191–209

Fehr, E., **Kirchsteiger, G.**, and **Riedl, A.** (1998). Gift exchange and reciprocity in competitive experimental markets. *European Economic Review*, 42(1): 1–34

Feinfield, K.A., **Lee, P.P.**, **Flavell, E.R.**, **Green, F.L.**, and **Flavell, J.H.** (1999). Young children's understanding of intention. *Cognitive Development*, 14(3): 463–486

Filiz-Ozbay, E., **Ham, J.C.**, **Kagel, J.H.**, and **Ozbay, E.Y.** (forthcoming). The role of cognitive ability, personality traits and gender in gift exchange outcomes. *Experimental Economics*

Geanakoplos, J., **Pearce, D.**, and **Stacchetti, E.** (1989). Psychological games and sequential rationality. *Games and Economic Behavior*, 1(1): 60–79

Geng, S., **Peng, Y.**, **Shachat, J.**, and **Zhong, H.** (2015). Adolescents, cognitive ability, and minimax play. *Economics Letters*, 128: 54–58

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

Gill, D. and **Prowse, V.** (2016). Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis. *Journal of Political Economy*, 126(4): 1619–1676

Gill, D. and **Stone, R.** (2010). Fairness and desert in tournaments. *Games and Economic Behavior*, 69(2): 346–364

Gill, D. and **Stone, R.** (2014). Desert and inequity aversion in teams. *Journal of Public Economics*, 123: 42–54

Gray, J. and **Thompson, P.** (2004). Neurobiology of intelligence: Science and ethics. *Nature Reviews Neuroscience*, 5: 471–482

Grueneisen, S., **Wyman, E.**, and **Tomasello, M.** (2015). Children use salience to solve coordination problems. *Developmental Science*, 18(3): 495–501

Gummerum, M. and **Chu, M.T.** (2014). Outcomes and intentions in children's, adolescents', and adults' second- and third-party punishment behavior. *Cognition*, 133(1): 97–103

Güroğlu, B., **van den Bos, W.**, and **Crone, E.A.** (2009). Fairness considerations: Increasing understanding of intentionality during adolescence. *Journal of Experimental Child Psychology*, 104(4): 398–409

Güroğlu, B., **van den Bos, W.**, **van Dijk, E.**, **Rombouts, S.A.**, and **Crone, E.A.** (2011). Dissociable brain networks involved in development of fairness considerations: Understanding intentionality behind unfairness. *Neuroimage*, 57(2): 634–641

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

Heckman, J.J. and **Kautz, T.** (2014). Fostering and measuring skills: Interventions that improve character and cognition. In J.J. Heckman, J.E. Humphries, and T. Kautz, editors, *The Myth of Achievement Tests: The GED and the Role of Character in American Life*. Chicago: University of Chicago Press

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

Kagan, S. and **Madsen, M.C.** (1971). Cooperation and competition of Mexican, Mexican-American, and Anglo-American children of two ages under four instructional sets. *Developmental Psychology*, 5(1): 32–39

Kinderman, P., **Dunbar, R.**, and **Bentall, R.** (1998). Theory-of-mind deficits and causal attribution. *British Journal of Psychology*, 89: 191–204

Li, J., **Zhu, L.**, **Liu, J.**, and **Li, X.** (2014). Social and non-social deficits in children with high-functioning autism and their cooperative behaviors. *Research in Autism Spectrum Disorders*, 8(12): 1657–1671

Liddle, B. and **Nettle, D.** (2006). Higher-order theory of mind and social competence in school-age children. *Journal of Cultural and Evolutionary Psychology*, 4: 231–246

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review*, 85(5): 1313–1326

Nettle, D. and **Liddle, B.** (2008). Agreeableness is related to social-cognitive, but not social-perceptual, theory of mind. *European Journal of Personality*, 22: 323–335

Pelligra, V., **Isoni, A.**, **Fadda, R.**, and **Doneddu, G.** (2015). Theory of mind, perceived intentions and reciprocal behaviour: Evidence from individuals with Autism Spectrum Disorder. *Journal of Economic Psychology*, 49: 95–107

Perner, J. (1979). Young children's preoccupation with their own payoffs in strategic analysis of 2×2 games. *Developmental Psychology*, 15(2): 204–213

Perner, J. and **Wimmer, H.** (1985). "John *thinks* that Mary *thinks* that..." attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, 39(3): 437–471

Premack, D. and **Woodruff, G.** (1978). Does the chimpazee have a theory of mind? *Behavioural and Brain Sciences*, 1(4): 515–526

Proto, E., **Rustichini, A.**, and **Sofianos, A.** (forthcoming). Intelligence, personality and gains from cooperation in repeated interactions. *Journal of Political Economy*

Raven, J., **Raven, J.C.**, and **Court, J.H.** (2000). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. San Antonio: Pearson

Sally, D. and **Hill, E.** (2006). The development of interpersonal strategy: Autism, theory-of-mind, cooperation and fairness. *Journal of Economic Psychology*, 27(1): 73–97

Schug, J., **Takagishi, H.**, **Benech, C.**, and **Okada, H.** (2016). The development of theory of mind and positive and negative reciprocity in preschool children. *Frontiers in Psychology*, 7(888): 1–9

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

Stahl, D. and **Wilson, P.** (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1): 218–254

Steinbeis, N., **Bernhardt, B.C.**, and **Singer, T.** (2012). Impulse control and underlying functions of the left DLPFC mediate age-related and age-independent individual differences in strategic social behavior. *Neuron*, 73(5): 1040–1051

Stingle, S.F. and **Cook, H.** (1985). Age and sex differences in the cooperative and noncooperative behavior of pairs of American children. *Journal of Psychology*, 119(4): 335–345

Sul, S., **Güroğlu, B.**, **Crone, E.A.**, and **Chang, L.J.** (2017). Medial prefrontal cortical thinning mediates shifts in other-regarding preferences during adolescence. *Scientific Reports*, 7(8510): 1–10

Sutter, M. (2007). Outcomes versus intentions: On the nature of fair behavior and its development with age. *Journal of Economic Psychology*, 28(1): 69–78

Weibull, J.W. (1997). *Evolutionary Game Theory*. MIT press

Wellman, H.M., **Cross, D.**, and **Watson, J.** (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72(3): 655–684

Wynn, K. (1990). Children's understanding of counting. *Cognition*, 36(2): 155–193

Zelazo, P.D., **Helwig, C.C.**, and **Lau, A.** (1996). Intention, act, and outcome in behavioral prediction and moral judgment. *Child Development*, 67(5): 2478–2492

Supplementary Web Appendix

(Intended for Online Publication)

Supplementary Web Appendix I

Experimental instructions

The experimental instructions are translated from the original Spanish. In the original Spanish instructions we use the neutral form "lo" to refer to the third person. In this English translation we use the feminine gender throughout.

The experimental instructions were read aloud, with explanations supported by projections of decisions screens onto a screen/wall.

Supplementary Web Appendix I.1 Instructions for Phase 1

Hello everyone! My name is [**experimenter**] and I am a teacher[44] just like [**supervisor from school staff**]. Who knows why have we brought you here today? The reason is because we are going to play some games with you. Who likes games?

Today, we are going to complete some puzzles so that I can learn a little about you. For this I have given you a tablet like this one [**show**]. Who has used a tablet like this before? On this side of the tablet there are two buttons [**show**]. Please do not press them. If there is any problem with the tablet at any stage, or there is something that you do not understand, raise your hand. The tablet is personal, so you cannot swap it with another person.

What you do in each game is anonymous, which means that I will not know what you have done in the games. Your classmates must not know what you have done either, so do not discuss it or say it aloud. I am going to be very strict: if someone says what he is doing aloud, I will ask that person to return to class.

Let's start. Put the tablets on [**location**].

I am now going to read some stories. Pay attention. At the end of each story we will complete some exercises. In each exercise, I am going to read two sentences about the story: option A and option B. Only one of the options is true, and you have to decide which one is true. You will see the options written on your tablet, but you will not see the story, as in the example on the screen/wall [**point at the projection on the screen/wall**]. You will also see two green buttons. One has the letter A [**show**], and the other the letter B [**show**].

If you think that the true sentence is sentence A, then press on the button that has the letter A; and if you think that the true sentence is sentence B, then press on the button that has the letter B.

[44] In Spanish we used the word "profesor", which refers to both a teacher and an university professor/lecturer.

When you have pressed one of the buttons, you will be able to see the option you have chosen on your tablet, here [**show**]. If you make a mistake, you can change the option by pressing the other button. To submit the answer, press the blue button with the white arrow [**show**]. I'm going to read the stories only once. So you have to pay close attention.

Let's see an example. I'm going to read a story. Listen!

"Pedro went to the kitchen because he wanted a glass of water. When Pedro got the glass of water, it fell on the floor."

I'm going to read the first pair of sentences:

A. Pedro wanted an apple.

B. Pedro wanted a glass of water.

Which one is the right choice? Yes, the right choice is B. Look at the screen. Which button should you press now? When you press that button you will see the letter B here [**show**]. Another example regarding the story about Pedro:

A. Pedro went to the kitchen.

B. Pedro went to the bathroom.

Which one is the right choice? Yes, the right choice is A. Look at the screen. Which button should you press now? When you press that button you will see the letter A here. Now, take your tablets. Do not press any of the buttons.

I'm going to read the first story...

[**Subjects complete an Imposing Memory Task designed specifically for children, as described in Section 2.2.1**]

This is the last test of the day. As you can see [**point at the projection of the first Raven's Progressive Matrix (A1)**], at the top there is a figure with a missing piece. Each of the pieces below is of the right size to fit the gap, but not all complete the figure. Piece number 1 [**show**] is not the right one. Pieces number 2 and 3 [**show**] do not work either: they fill the gap, but neither is the right piece. What about piece number 6? It has the required pattern [**show**], but it does not totally cover the gap.

Using your tablet, press on the right piece. [**Check subjects' responses. If there were any wrong choices, repeat the explanation again**.] Yes, piece number 4 is the right one. Now, on your tablet, you can see the number of the piece that you have selected [**show**].

Next you will see a series of figures with a missing piece, similarly to this one [**point at the projection of the first Raven's Progressive Matrix (A1)**]. For each figure you have to decide which of the pieces at the bottom is the one that completes the figure at the top. When you discover it, press with your finger on the piece. At that time the tablet will show the number of the piece that you have selected. To move to the next image, press on the blue button with a white arrow at the bottom of the screen. [**Check that all the subjects can see the second Raven's Progressive Matrix (A2) on their tablets**.]

The problems are easy at the beginning, and gradually become more difficult. There is no catch. If you pay attention to the way the easy ones are solved, the later ones will be less difficult. Work on your own, and do not skip any exercise. You cannot go back either. You have all the time you need. Continue now until the end. When you finish, put the tablet [**location**] and raise your hand.

[**Subjects complete an age-appropriate Raven's Progressive Matrices test, as described in Section 2.2.2**.]

Supplementary Web Appendix I.2 Instructions for Phase 2

{*As described in Section 2.3.5, some subjects completed the games according to Order X and others according to Order Y. The instructions here are for Order X.*}

Today you are going to participate in a few games. In each game you will be paired with another player who will be called 'your partner'.

In general, your partner will change from one game to another and will be chosen at random by the computer, so only the computer knows who your partner is in each game. Neither you nor I will know who your partner is. In each game, your partner could be in this room, but it is also very likely that your partner is not in this room. In fact, your partner may not have participated in the games yet.

In each game you will receive an amount of tokens. The amount of tokens you receive in each game will depend on the decisions you take, but it will also depend on the decisions that your partner takes. When all the students in the school who are participating in the project have played, I will calculate the total amount of tokens that you have received in the games.

> {*Four lower academic years*} The total amount of tokens you have received can be exchanged for things like these [**show**]. For example, you can exchange one token for a colored pencil. The rest of the things have a value in tokens. You take these things home: they are yours.
>
> {*Three higher academic years*} Then, I will give you a voucher with which you can buy whatever you want in either [**local bookshop**] or [**online retailer**]. The value of your voucher will depend on the amount of tokens you have received. For each token you receive the value of your voucher will increase by 10 cents of a Euro.

The more games there are, the more tokens you can receive and...

{*Four lower academic years*} ...the more things you can take home.

{*Three higher academic years*} ...the higher the value of your voucher will be.

Remember that your partner may have not participated in the games yet. That is why it is very important that you do not tell the other students in the school what the games are about or what you did in the games. If you do, other students will have time to think about what they will do when it is their turn to play. That will put you at a disadvantage and you could get many fewer tokens than you expected.

What you do in each game is anonymous, which means that I will not know what you have done in the games. Your classmates should not know what you have done either, so do not discuss it with each other or say it aloud. I will be very strict: if someone says what she is doing aloud, I will ask that person to leave the session and that person will not receive any tokens. If you have any questions at any stage, raise your hand. Pay close attention to the instructions of each game because the better you understand them, the more tokens you will receive in each game.

You can now see the decision screen of the first game in your tablet [**point at the projection of Figure SWA.1**]. Please do not touch anything until I tell you to do so.

Figure SWA.1: Decision screen in the 1-6 Token Request game.

The screen shows a chequered red and white tablecloth. Just under the tablecloth you can see 6 purple tokens, and just above the tablecloth there is a counter saying "I am going to ask for 0 tokens." The counter tells you how many tokens there are on the tablecloth. You can drag the tokens around the screen using your fingers.

The first thing I want you to do is put one token on the tablecloth [**check that subjects have done this**]. You will now see that the counter says "I am going to ask for 1 token." Now put another token on the tablecloth. What does the counter say now? [**Wait for responses.**] Now put the tokens that are on the tablecloth back to (more or less) where they were at the beginning, until the counter shows "I am going to ask for 0 tokens." This blue button ends the game, so you must only press it when you are sure about what you want to do.[45]

[45]The software did not allow children to submit a request of 0 tokens; in that case a pop-up window informed them that they had to choose a number of tokens between 1 and 6.

Game 1 ('Baseline' game)[46]

Now I am going to read the instructions for the first game. Listen carefully.

In this game your partner is another student from the school, randomly chosen by the computer, and who can be between 5 and 12 years old. Your partner might be in this room, but she might not be here. The rules of the game are the same for your partner.

Your partner and you are going to ask for an amount of tokens. The amount must be between 1 and 6. I will give you the amount of tokens you ask for. However, I will give you 10 more tokens if you ask for exactly one token less than your partner. How many tokens are you going to ask for?

Let's see if you have understood. How many tokens can you ask for? And how many tokens am I going to give you? When will I give you 10 tokens more than you asked for? Let's look at some examples. Suppose the amounts asked for by two paired players are these [**point at the projection of Figure SWA.2**].

Figure SWA.2: First example.

How many tokens has the first player asked for? And the second player? How many tokens does the first player receive? And the second player? Has any of the players asked for exactly one token less than the other player? Then the first player receives the two tokens she has asked for, and the second player receives the token that she has asked for and 10 more tokens, that is, 11 tokens.

[**Two further examples are given. In the second example, the first player asks for 4 tokens and the second player asks for 2 tokens. In the third example, they both ask for 3 tokens.**]

Using your fingers, put on the tablecloth the amount of tokens you want to ask for. When the counter shows the number of tokens you want to ask for, press the blue button.

[46] Game numbers and names were not read to the subjects.

Game 2 ('Computer' game)

Let's play the same game but remember that in general your partner changes in each game.

In this game your partner is this computer [**show**]. The computer does not understand the rules of the game. The computer is not interested in the tokens and it does not care about any particular number.[47] So, it is going to choose the number of tokens at random in each game.[48] What it is going to do is similar to rolling a die like this [**show**]. Before the die is rolled, nobody knows which number is going to come up and all numbers are equally likely. This means that a 1 can come up just the same as a 2, a 3, 4, 5, or 6. So you can imagine that the computer is going to roll a die to decide the amount of tokens it is going to ask for.

In this game your partner is this computer. The rules of the game are the same for your partner. Your partner and you are going to ask for an amount of tokens. The amount must be between 1 and 6. I will give you the amount of tokens you ask for. However, I will give you 10 more tokens if you ask for exactly one token less than your partner. How many tokens are you going to ask for?

Using your fingers, put on the tablecloth the amount of tokens you want to ask for. When the counter shows the number of tokens you want to ask for, press the blue button.

Game 3 ('Baseline' game)

Let's play the same game but remember that in general your partner changes in each game.

In this game your partner is another student from the school, randomly chosen by the computer, and who can be between 5 and 12 years old. Your partner might be in this room, but she might not be here. The rules of the game are the same for your partner.

Your partner and you are going to ask for an amount of tokens. The amount must be between 1 and 6. I will give you the amount of tokens you ask for. However, I will give you 10 more tokens if you ask for exactly one token less than your partner. How many tokens are you going to ask for?

Using your fingers, put on the tablecloth the amount of tokens you want to ask for. When the counter shows the number of tokens you want to ask for, press the blue button.

[47] In Spanish we used the colloquial expression "le dan todos los números igual."
[48] In Spanish we used the colloquial expressions "al azar (a suerte, a voleo)."

Game 4 (‘Baseline’ game).

Let’s play the same game but remember that in general your partner changes in each game.

In this game your partner is another student from the school, randomly chosen by the computer, and who can be between 5 and 12 years old. Your partner might be in this room, but she might not be here. The rules of the game are the same for your partner.

Proceed!

Game 5 (‘Raven’ game)

Let’s play the same game but remember that in general your partner changes in each game.

In this game your partner is a student who is in this room, chosen at random by the computer. The rules of the game are the same for your partner. This time I’m going to give you a piece of information about your partner. How many of you remember this figure [**point at the projection of the first Raven’s Progressive Matrix (A1)**]. We completed this exercise the last time I was here. The computer has calculated the total number of right answers that you and the students of your age got in this test. The computer has sorted the scores from lowest to highest and then the computer has separated the students into two groups of roughly the same size: the half of the students with the highest scores are in the ‘high score’ group and the half of the students with the lowest scores are in the ‘low score’ group.

For example, if the scores of students your age were these [**point at a projection showing the numbers 11, 12, 15 and 16**], then the students with scores 15 and 16 would be in the ‘high score’ group, and the students with scores 11 and 12 would be in the ‘low score’ group.

In your tablet, you can see here [**point at the projection of Figure SWA.3**] whether your partner is in the high or low score group [**show in each player’s tablet**].

Notes: Next to “Puntuación de tu pareja” (“Your partner’s score”), a subject’s decision screen states whether the subject’s partner is in the high or low cognitive ability group.

Figure SWA.3: Decision screen in the Raven game version of the 1-6 Token Request game.

I'm going to read the instructions.

In this game your partner is a student who is in this room, chosen at random by the computer. The rules of the game are the same for your partner. On your tablet you can see if your partner is in the high or low score group.

Your partner and you are going to ask for an amount of tokens. The amount must be between 1 and 6. I will give you the amount of tokens you ask for. However, I will give you 10 more tokens if you ask for exactly one token less than your partner. How many tokens are you going to ask for?

Using your fingers, put on the tablecloth the amount of tokens you want to ask for. When the counter shows the number of tokens you want to ask for, press the blue button.

Game 6 ('Baseline game).

[**Read same text as for Game 3 above.**]

Game 7 ('Baseline' game).

[**Read same text as for Game 4 above.**]

Supplementary Web Appendix I.3 Instructions for Phase 3

{*The instructions here are for Treatment A.*}

Let's play a different game but remember that in general your partner changes in each game.

In this game, you will receive an amount of super-tokens [**point at the projection of Figure SWA.4**]. These super-tokens are yellow and receiving one super-token is the same as receiving four tokens.

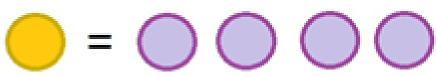


Figure SWA.4: Equivalence between tokens and super-tokens.

{*Four lower academic years*} So, for example, whereas you could exchange one token for one coloured pencil [**show**], you can exchange one super-token for four colored pencils [**show**].

{*Three higher academic years*} For each super-token you receive, the value of your voucher will increase by 40 cents of a Euro.

The computer has separated students in this room into two groups: tigers and lions. The computer has matched each tiger with a lion. You will know if you are a tiger or a lion, but neither you nor I will know who in this room is your partner.

Each tiger starts this game with 10 super-tokens, while each lion starts the game with no super-tokens. Each tiger has to decide how many super-tokens she is going to keep and how many super-tokens she is going to give to her lion partner. Each tiger has two options. A tiger can:

- Keep 8 super-tokens and give 2 super-tokens to her lion partner.
- Keep 5 super-tokens and give 5 super-tokens to her lion partner.

Once a tiger has made her decision, the computer will tell her lion partner what the tiger has done. In her tablet, the lion will see how many super-tokens the tiger has kept and how many super-tokens the tiger has given to her. Then, each lion has to make a decision. The lion can:

- Keep all the super-tokens that the tiger has given to her.
- Give back one of the super-tokens that the tiger has given to her.

Let us see if you have understood the game. With how many super-tokens does a tiger start the game? And a lion? How many super-tokens can a tiger keep? Once the tiger has made a decision, the computer will tell each lion what her partner tiger has done. Then, what can the lion do?

What you do in the game is anonymous, which means that I will not know what you have done in the games. Your classmates should not know what you have done either, so do not discuss it with each other or say it aloud. I will be very strict: if someone says what she is doing aloud, I will ask that person to leave the session and that person will not receive any tokens. If you have any questions at any stage, raise your hand.

Let's start the game.

Those of you who are tigers can now see this screen [**point at the projection of Figure SWA.5**], while those of you who are lions continue to see the spinner.[49]

[49]The tablets of Lions displayed a place-holder screen, but the Lions could also see the projection of a Tiger's decision screen.

Figure SWA.5: Allocator (Tiger) decision screen.

At the top of the screen it says "You are a Tiger." Below, there are two figures. Each figure contains a red and white chequered tablecloth with a number of super-tokens on top. In each figure, under the title 'Tiger' you can see the number of super-tokens the tiger would keep, and under the title 'Lion' you can see the number of super-tokens that the tiger would give to her lion partner.

For example, how many super-tokens would the tiger keep in this figure? [**Point at the top figure**.] And how many super-tokens would a tiger give to her lion partner?

[**Repeat previous paragraph for bottom figure**.]

If you are a tiger and want to keep 8 super-tokens and give 2 super-tokens to your lion partner, then you have to press this button [**show**]. If you are a tiger and want to keep 5 super-tokens and give 5 super-tokens to your lion partner, then you have to press this button [**show**].

Proceed!

If you are a lion, you now can see this screen [**point at the projection of Figure SWA.6**].

At the top of the figure, it says "You are a Lion." The screen shows the number of super-tokens that your tiger partner has kept, here [**show**], and the number of super-tokens that the tiger has given to you, here [**show**].

Here it says "Do you want to give back to the tiger one of the super-tokens that you received?" [**Show.**]

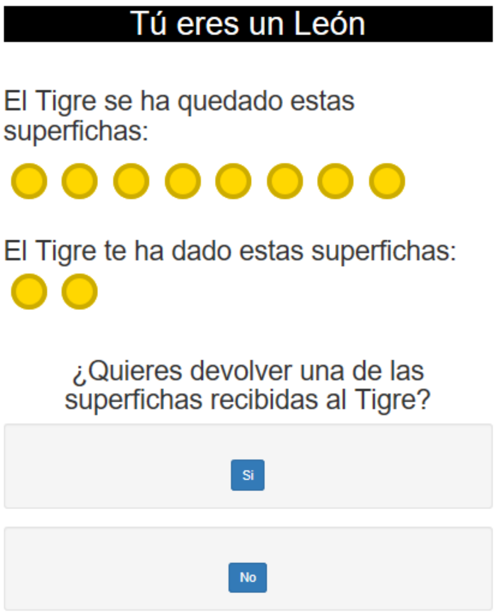


Figure SWA.6: Receiver (Lion) decision screen.

If you are a lion, and you want to give back to your tiger partner one of the super-tokens that she has given to you, then press the button 'Yes'.

If you are a lion, and you do not want to give back to your tiger partner one of the super-tokens that she has given to you, then press the button 'No'.

Proceed!

This is the end of the session. Thank you all for playing. When all the students who are participating in the project have played, I will let you know how many tokens in total you have received. Remember that the number of tokens you receive is not decided yet. In each game your partner has been chosen at random by the computer and, in general, your parter changed in each game. It is likely that you partner in some of the games has not played yet. For this reason, it is very important that you do not tell the other students in the school what the games are about or what you did in the games. If you do so, the other students will have time to think about what they are going to do when it is their turn to play. That puts you at a disadvantage and you may receive many fewer tokens than you expected.

[**An explanation of the procedure to exchange the tokens for items or vouchers followed. Students then went back to class.**]

Supplementary Web Appendix II

Payment protocols

Children in the four lower academic years exchanged their tokens for stationery and educational supplies. In each school, payoffs were computed and letters were produced to inform children of the number of tokens that they had earned. The letters included a menu with images of the items available for the children to choose from, together with the price, in tokens, of each item. Children could choose between graphite pencils, color pencils, color felt tips, erasers, pencil sharpeners, rulers, calculators, notebooks, pencil cases and Play-Doh. With the help of their teachers, children filled up the menus, stating the amounts of each item that they wanted.

Children in the three higher academic years exchanged their tokens for vouchers, with each token worth 0.1 euros. In each school, payoffs were computed and letters were produced to inform children of the value of the vouchers that they had earned. The children were free to choose whether the voucher would be provided by a well-known online retailer or a popular bookshop located in the city of Santander. Parental consent was sought from those children who opted for the online retailer. Once the letters were collected, vouchers from the retailers were obtained and distributed to the children.

Supplementary Web Appendix III

Software and configuration of the portable lab

The intervention relied on a portable lab consisting of twenty tablets, an access point and a server. The tablets were 7" Asus Zen Pad 16GB, running Android 5.0 (Lollipop) on an Intel Atom x3-C3200 Processor. The access point was a Belkin F9K1102UK N600 Dual Band Wireless Cable Router. The server was a Lenovo laptop computer with Windows 10 on an Intel Core i3-5005U CPU at 2.GHz and 8GB of installed memory. The access point was wired to the server using a standard Ethernet cable, while the communication between the tablets and the access point was configured by Computer Services at the University of Strathclyde.

The server had an installation of XAMPP v3.2.2., which is a free development environment containing, inter alia, the Apache HTML server, a database (MySQL), and PHP (v.5.6.3), a popular programming language that is typically used to control the communication between front-end HTML applications with a back-end sever and database. The main purpose of XAMPP is the development and testing of web applications; however, it can enable a standard laptop or desktop to operate as a server within a local network.

All the software used for the intervention was written by the alphabetically-first author using javaScript (including jQuery), PHP, HTML and CSS (including Bootstrap). In order to guarantee subject anonymity, we used Hashids (http://hashids.org), an open-source library that generates short, unique, non-sequential identifiers from numbers and enabled us to separate personal data from subjects' responses in the games and tests.

Supplementary Web Appendix IV

Pre-intervention pilots

Three small pilots were completed prior to the intervention.

In December 2014 forty-nine children aged sixteen and seventeen from a school in Santander, Spain, participated in a two-phase intervention similar to the first two phases described in this paper, but without any incentives. In the second phase, the children played four repetitions of the original 11-20 game (Arad and Rubinstein, 2012).

In March 2015 thirty children aged six and seven from a school in Manchester, United Kingdom, played two incentivized repetitions of a variant of the 11-20 game with a strategy set restricted to $\{1, 2, 3, 4\}$. In this pilot we tested the children's ability to subtract one from numbers up to ten and we discussed numeracy skills with the teacher responsible for the class. During the pilot, it became apparent that small children did not understand the meaning of the word 'additional', which resulted in an interruption while the experimenter was reading the instructions; we amended the instructions for the main experiment accordingly.

In January 2016 we tested our tablets and software using forty children aged between five and twelve from a school in Santander, Spain (this school was not used in the main experiment). We tested only phase 2 of our experiment (and so the Raven game was not included). We did not use any incentives.

Supplementary Web Appendix V

Nash equilibrium in the 1-6 Token Request game

Assume that utility increases monotonically in money. There are no pure-strategy Nash equilibria. If a player chooses $x \geq 3$, the best-response is $x-1$, to which the best-response is $x-2$. If a player chooses 2, the best-response is 1, to which the best-response is 6. If a player chooses 1, the best-response is 6, to which the best-response is 5.

Assume risk-neutral players. We search for symmetric mixed-strategy Nash equilibria. Let $p_x < 1$ denote the probability of choosing x. Let π_x denote the expected monetary payoff from choosing x. Expected payoffs are $\pi_6 = 6$ and $\pi_x = x + 10p_{x+1}$ for $x \in \{1, 2, \ldots, 5\}$. First, by definition, the players must be indifferent over all the strategies with $p_x > 0$, and $\sum p_x = 1$. Second, there can be no gaps in the distribution: if $p_x = 0$, then $\pi_x > \pi_{x-1}$, and so by induction $p_z = 0\ \forall z < x$. Thus $p_6 \in (0, 1)$, which implies $p_5 > 0$, and so:

$$\pi_5 = 5 + 10p_6 = 6 \implies p_6 = 0.1.$$

Next, $p_4 > 0$. If $p_4 = 0$, then $p_5 = 0.9$, which gives $\pi_4 = 13 > 6$, a contradiction. Thus:

$$\pi_4 = 4 + 10p_5 = 6 \implies p_5 = 0.2.$$

Next, $p_3 > 0$. If $p_3 = 0$, then $p_4 = 0.7$, which gives $\pi_3 = 10 > 6$, a contradiction. Thus:

$$\pi_3 = 3 + 10p_4 = 6 \implies p_4 = 0.3.$$

Next, $p_2 = 0$, which also implies that $p_1 = 0$. If $p_2 > 0$, then:

$$\pi_2 = 2 + 10p_3 = 6 \implies p_3 = 0.4 \implies \sum p_x > 1,$$

a contradiction. Thus, $p_3 = 0.4$ to give $\sum p_x = 1$; and $p_3 = 0.4$ ensures that there is no incentive to deviate to 2.

Supplementary Web Appendix VI

Additional tables

	Age							
	5	6	7	8	9	10	11	12
Mean score	2.08	2.26	2.85	3.14	4.50	5.02	5.21	5.00
Standard deviation	1.01	0.98	0.95	0.91	1.12	0.83	0.91	1.35
Minimum	0	0	0	0	1	3	3	2
Maximum	4	4	4	4	6	6	6	6
Number of questions	4	4	4	4	6	6	6	6
Number of subjects	84	103	110	111	100	125	84	13

Table SWA.1: Distribution of scores in the Imposing Memory Task.

	Age							
	5	6	7	8	9	10	11	12
Mean score	18.23	21.17	24.57	28.92	33.08	37.49	41.29	41.31
Standard deviation	5.36	5.71	5.66	7.70	8.64	7.65	6.76	9.71
Minimum	5	5	7	10	13	17	20	17
Maximum	28	32	35	45	53	57	53	50
Number of questions	36	36	36	60	60	60	60	60
Number of subjects	84	103	110	111	100	125	84	13

Table SWA.2: Distribution of scores in the Raven's Progressive Matrices test.

	Pooled		Baseline games		Computer game	
	(1)	(2)	(3)	(4)	(5)	(6)
Age	0.020***	0.019***	0.021***	0.020***	0.014*	0.013*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.008)	(0.008)
Theory-of-mind	0.015*	0.004	0.016*	0.000	0.011	0.021
(T.o.M.)	(0.008)	(0.013)	(0.009)	(0.014)	(0.016)	(0.027)
Cognitive ability	0.016*	-0.023	0.016*	-0.011	0.016	-0.079***
(Cogn. Ab.)	(0.009)	(0.016)	(0.009)	(0.017)	(0.017)	(0.029)
T.o.M. × Age		0.003		0.005		-0.004
		(0.004)		(0.004)		(0.008)
Cogn. Ab. × Age		0.012***		0.009*		0.031***
		(0.004)		(0.005)		(0.007)
Intercept	0.142***	0.142***	0.141***	0.140***	0.152***	0.149***
	(0.020)	(0.019)	(0.021)	(0.021)	(0.037)	(0.036)
Subjects	729	729	729	729	729	729
Subject-round obs.	4,374	4,374	3,645	3,645	729	729

Notes: See the notes to Table 4. Here we add controls for subjects' gender and school. One subject is excluded because they did not report their gender.

Table SWA.3: Probability of best-responding: Robustness.

	Pooled		Baseline games		Computer game	
	(1)	(2)	(3)	(4)	(5)	(6)
Age	0.090***	0.088***	0.118***	0.117***	-0.055	-0.058
	(0.033)	(0.032)	(0.036)	(0.036)	(0.065)	(0.064)
Theory-of-mind	0.066	0.040	0.108	0.084	-0.144	-0.179
(T.o.M.)	(0.066)	(0.117)	(0.072)	(0.127)	(0.137)	(0.239)
Cognitive ability	-0.020	-0.318**	-0.039	-0.285*	0.077	-0.487*
(Cogn. Ab.)	(0.074)	(0.153)	(0.078)	(0.153)	(0.133)	(0.268)
T.o.M. × Age		0.007		0.006		0.009
		(0.031)		(0.034)		(0.066)
Cogn. Ab. × Age		0.095**		0.078*		0.179***
		(0.038)		(0.040)		(0.064)
Intercept	4.798***	4.808***	4.727***	4.735***	5.154***	5.171***
	(0.122)	(0.122)	(0.131)	(0.131)	(0.246)	(0.245)
Subjects	730	730	730	730	730	730
Subject-round obs.	4,380	4,380	3,650	3,650	730	730

Notes: The regressions are the same as in Table 4, except that here the dependent variable is the number of tokens earned in each game. See the notes to Table 4.

Table SWA.4: Payoffs.

	Pooled		Baseline games		Computer game	
	(1)	(2)	(3)	(4)	(5)	(6)
Age	0.098***	0.096***	0.126***	0.124***	-0.044	-0.048
	(0.033)	(0.033)	(0.037)	(0.037)	(0.066)	(0.066)
Theory-of-mind	0.041	0.027	0.085	0.076	-0.180	-0.217
(T.o.M.)	(0.065)	(0.115)	(0.072)	(0.124)	(0.140)	(0.241)
Cognitive ability	-0.044	-0.303**	-0.060	-0.270*	0.033	-0.464*
(Cogn. Ab.)	(0.076)	(0.153)	(0.080)	(0.153)	(0.135)	(0.267)
T.o.M. × Age		0.003		0.002		0.010
		(0.030)		(0.034)		(0.066)
Cogn. Ab. × Age		0.084**		0.068*		0.161**
		(0.038)		(0.040)		(0.064)
Intercept	4.988***	4.981***	4.978***	4.973***	5.035***	5.022***
	(0.172)	(0.170)	(0.183)	(0.182)	(0.312)	(0.310)
Subjects	729	729	729	729	729	729
Subject-round obs.	4,374	4,374	3,645	3,645	729	729

Notes: See the notes to Table SWA.4. Here we add controls for subjects' gender and school. One subject is excluded because they did not report their gender.

Table SWA.5: Payoffs: Robustness.

Choice	1	2	3	4	5	6
Level-k	Level-5	Level-4	Level-3	Level-2	Level-1	Level-0
Expected payoff	2.332	3.704	4.863	5.871	6.805	6.000
Age	-0.023***	0.001	0.004	0.027***	0.020***	-0.029***
	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)
Theory-of-mind	0.004	-0.011	-0.014	0.005	0.016*	-0.001
	(0.009)	(0.007)	(0.008)	(0.008)	(0.009)	(0.010)
Cognitive ability	-0.013	-0.003	0.005	0.010	0.016*	-0.014
	(0.009)	(0.007)	(0.007)	(0.008)	(0.009)	(0.010)
Subjects				729		
Subject-round obs.				3,645		

Notes: See the notes to Table 5. Here we add controls for subjects' gender and school. One subject is excluded because they did not report their gender.

Table SWA.6: Probability of choices in the Baseline games: Robustness.

Subjects below median age						
Choice	1	2	3	4	5	6
Level-k	Level-5	Level-4	Level-3	Level-2	Level-1	Level-0
No controls						
High cognitive ability opponent	-0.017 (0.042)	0.046 (0.037)	-0.035 (0.040)	0.052 (0.035)	-0.035 (0.040)	-0.012 (0.045)
With controls						
High cognitive ability opponent	-0.020 (0.041)	0.045 (0.037)	-0.038 (0.040)	0.040 (0.035)	-0.024 (0.040)	-0.003 (0.044)
Subjects				344		
Subject-round obs.				344		

Notes: See the notes to Table 7. Here the subjects are those of below median age. One of the 344 subjects is excluded from the specification with controls because that subject did not report their gender.

Table SWA.7: Probability of choices in the Raven game for younger subjects.

	(1)	(2)	(3)
Treatment A	0.093	0.087	-0.180*
	(0.073)	(0.073)	(0.101)
Age		0.013	-0.004
		(0.015)	(0.017)
Theory-of-mind (T.o.M.)		0.003	-0.035
		(0.028)	(0.031)
Cognitive ability		0.002	-0.005
		(0.031)	(0.035)
Treatment A × Age			0.072**
			(0.029)
Treatment A × T.o.M.			0.153**
			(0.059)
Treatment A × Cogn. Ab.			0.026
			(0.062)
Intercept	0.061	0.020	0.065
	(0.046)	(0.072)	(0.075)
Subjects	207	207	207
Subject-round obs.	207	207	207

Notes: See the notes to Table 8. Here we add controls for subjects' gender and school.

Table SWA.8: Effect of treatment on probability that receiver gives one super-token to allocator: Robustness

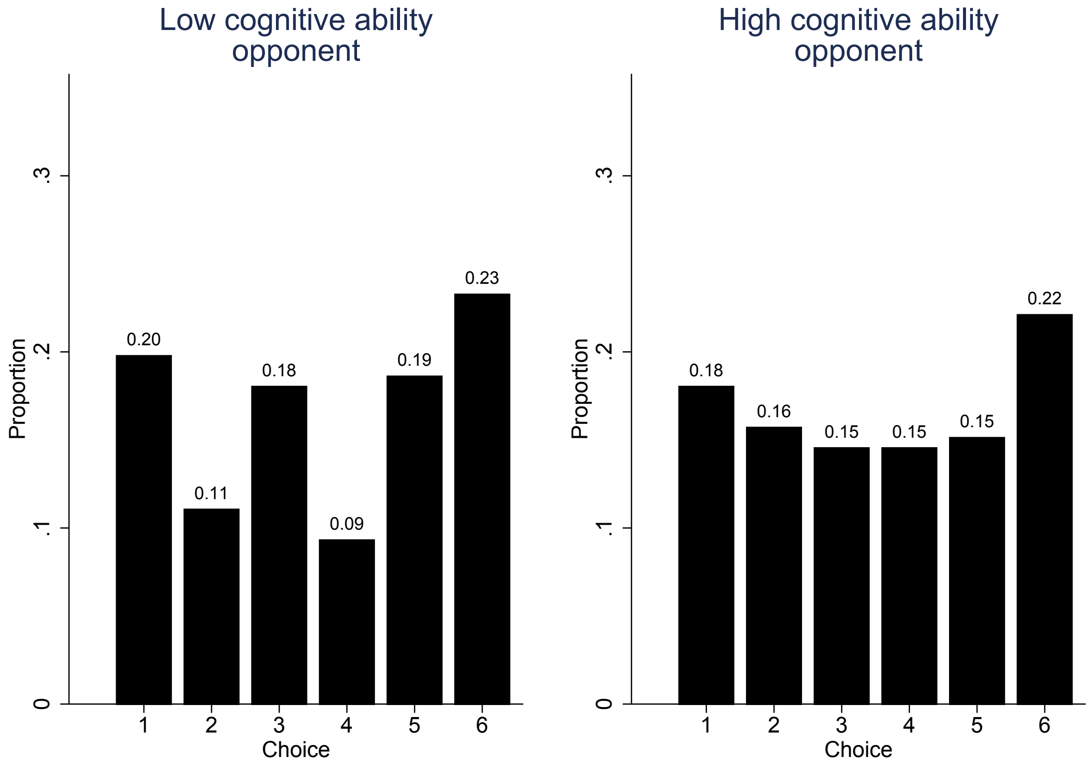
	(1)	(2)	(3)
Number of super-tokens received	0.043***	0.044***	0.017
(Super-tokens)	(0.010)	(0.010)	(0.018)
Age		0.007	-0.029
		(0.009)	(0.020)
Theory-of-mind		0.011	-0.021
(T.o.M.)		(0.018)	(0.035)
Cognitive ability		-0.002	-0.059
(Cogn. Ab.)		(0.021)	(0.041)
Super-tokens × Age			0.009*
			(0.005)
Super-tokens × T.o.M.			0.008
			(0.010)
Super-tokens × Cogn. Ab.			0.014
			(0.011)
Intercept	-0.084*	-0.104*	-0.006
	(0.048)	(0.056)	(0.069)
Subjects	384	384	384
Subject-round obs.	384	384	384

Notes: See the notes to Table 9. Here we add controls for subjects' gender and school. One subject is excluded because they did not report their gender.

Table SWA.9: Effect of number of tokens received on probability that receiver gives one super-token to allocator: Robustness

Supplementary Web Appendix VII

Additional figures



Notes: See the notes to Figure 3. Here the subjects are those of below median age.

Figure SWA.7: Distribution of choices in the Raven game for younger subjects.