Cognitive skills, strategic sophistication, and life outcomes *†

Eduardo Fe ‡
David Gill §
Victoria Prowse ¶

This version: December 29, 2021

Abstract

We investigate how childhood cognitive skills affect strategic sophistication and adult outcomes. In particular, we emphasize the importance of childhood theory-of-mind as a cognitive skill. We collected experimental data from more than seven hundred children in a variety of strategic interactions. First, we find that theory-of-mind ability and cognitive ability both predict level-k behavior. Second, 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. Third, 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 measures of cognitive skill correspond to different cognitive processes in strategic situations that involve understanding intentions. Using the ALSPAC birth-cohort study, we find that childhood theory-of-mind and cognitive ability are both associated with enhanced adult social skills, higher educational participation, better educational attainment, and lower fertility in young adulthood. Finally, we provide evidence that school spending improves theory-of-mind in childhood.

Keywords: *Cognitive skills; theory-of-mind; cognitive ability; fluid intelligence; children; experiment; strategic sophistication; level-k; bounded rationality; non-equilibrium thinking; intentions; gift-exchange game; competitive game; strategic game; ALSPAC; social skills; adult outcomes; life outcomes; education; fertility; labor market; wages; employment; school spending; childhood intervention.* **JEL Classification**: *C91; D91; J24*

*We are grateful for helpful comments from seminar participants and in private conversations. We thank: the University of Strathclyde and the Purdue University Research Center in Economics for funding; Robin Dunbar for many useful discussions; and Annie Chiaroni, Debasmita Das, Mouli Modak, Benjamin Raymond, Lu Wang, Chen Wei, Daniel Woods and Junya Zhou for excellent research assistance. Re the experiment, we thank: school staff, children, and their parents; 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. Re ALSPAC: we are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

†This paper replaces an earlier version titled: "Cognitive skills and the development of strategic sophistication."

‡University of Manchester; eduardo.fe@manchester.ac.uk

§Purdue University; dgill.econ@gmail.com

¶Purdue University; vprowse@purdue.edu

1 Introduction

Cognitive skills vary widely across the population and ability gaps open up early in life. Thus, it is essential to understand how childhood cognitive skills influence economic behavior and life outcomes, and to learn more about how we can improve cognitive skills in children. In this paper, we emphasize the importance of childhood theory-of-mind as a cognitive skill.

Individuals with theory-of-mind impute mental states (such as beliefs, desires, intentions and emotions) to others and use theory-of-mind to help interpret others' behavior, and in practice psychologists measure the theory-of-mind ability of adults and children by their ability to correctly infer or understand others' mental states (we provide a detailed description of the concept of theory-of-mind from the psychology literature in Section 2.1). The standard economic framework assumes that agents have a perfect theory-of-mind: these agents understand how the beliefs and desires of others drive their behavior. In practice, understanding the thought processes of others is difficult. Our mind-reading capabilities are limited and heterogeneous: theory-of-mind is a skill that varies across the population.

Theory-of-mind has attracted enormous attention from psychologists and cognitive neuroscientists (a Google Scholar search for "theory of mind" returns nearly half a million results). Estes and Bartsch (2017) argue that, together with language, theory-of-mind accounts for superior human general intelligence, and conclude that "our greatest accomplishments must rest on our basic capacity to imagine and recognize the variations and vulnerabilities of human cognitive states." Despite this interest, we know little about how theory-of-mind affects economic behavior. Instead, psychologists have focused on the development of theory-of-mind ability through childhood (see Stone et al., 1998, for a summary; Henry et al., 2013's meta-analysis compares theory-of-mind ability in young and old adults) and on the relationship between theory-of-mind deficits and psychiatric disorders such as autism (e.g., Baron-Cohen et al., 1985).

We begin by developing a detailed conceptual framework in which childhood theory-of-mind and cognitive ability are skills that affect strategic sophistication, social skills in adulthood, and life outcomes such as educational attainment, fertility and labor market behavior. In this framework, theory-of-mind operates as a cognitive skill and not through preferences: we provide evidence that theory-of-mind is not, or only very modestly, related to pro-social preferences and behaviors such as altruism, empathy or reciprocity.

We then test hypotheses that flow from this conceptual framework using a rich birth-cohort study that follows thousands of participants from childhood into young adulthood, and using experimental data from more than seven hundred children in a variety of incentivized strategic interactions. We also provide evidence about the effect of primary (i.e., elementary) school spending on theory-of-mind ability. Since theory-of-mind undergoes significant development in early to middle childhood (e.g., Stone et al., 1998), primary-school spending might be of particular importance in helping children to build theory-of-mind skills.

As we note above, we measure theory-of-mind by the ability to correctly infer or understand the mental states of others. Cognitive ability encompasses fluid intelligence, which is the ability to use logical reasoning to solve new problems, and crystallized intelligence, which includes acquired knowledge and verbal skills (e.g., Carpenter et al., 1990). We collected experimental data from five schools in Spain and measured the theory-of-mind ability and cognitive ability of children ranging from age five to age twelve. Our birth-cohort data come from the Avon

Longitudinal Study of Parents and Children (ALSPAC), which measured the theory-of-mind ability and cognitive ability at age eight of children from the Avon region in the South West of England. In both cases, we find a low correlation between the theory-of-mind and cognitive ability test scores, which provides evidence that the psychometric tests capture different cognitive skills.

As we explain in our conceptual framework, strategic sophistication underpins the ability to succeed in life. Indeed, the importance of developing strategic skills has been recognized by educational programs that emphasize strategic ability, such as Accelium that develops thinking abilities and life skills through strategy skills.[1] Our experimental data shed light on how childhood cognitive skills influence strategic sophistication, providing three main results.

First, we find that theory-of-mind and cognitive ability both predict level-k behavior (see Crawford et al., 2013, for a survey on level-k thinking). In a competitive game designed to trigger level-k thinking, we find that children with higher theory-of-mind ability are more likely to exhibit level-1 behavior (which coincides with the best-response to the empirical distribution in our dataset). Similarly, we find that cognitive ability and age also predict level-1 behavior.

Second, we find that children respond to information about the cognitive ability of their opponent. In particular, older children are more likely to exhibit level-2 behavior when they face an opponent of high cognitive ability (rather than an opponent of low cognitive ability). This ability to adjust behavior to the cognitive skill of the opponent provides support for the emergence of a sophisticated strategic theory-of-mind in children as young as eight to twelve years old.

Third, we find that theory-of-mind and age strongly predict whether children respond to intentions in a bespoke gift-exchange game, while cognitive ability has no influence. In particular, we find that theory-of-mind ability, but not cognitive ability, helps children to direct reciprocity appropriately according to the intentions of the allocator. 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.

Next, we turn to the ALSPAC birth-cohort study. To analyze how childhood cognitive skills affect adult outcomes, we regress the outcome of interest on theory-of-mind ability and cognitive ability at age eight using a rich set of controls that includes parental and school characteristics.

As predicted by our conceptual framework, we find that theory-of-mind ability at age eight is associated with enhanced adult social skills, higher educational participation, better educational attainment, and lower fertility in early adulthood. We do not find an effect of theory-of-mind on the probability of being employed or on wages in participants' early twenties; however, we do find evidence that workers with better theory-of-mind have a comparative advantage in larger firms. The absence of an effect of theory-of-mind on the probability of employment should be seen in the context of record employment rates in the United Kingdom. Regarding wages, we conjecture that any effects of theory-of-mind are not yet apparent because the wage distribution in the United Kingdom is compressed in the early years following graduation.

We find similar effects of childhood cognitive ability on adult outcomes, with two main differences: the effects of cognitive ability on education are larger in magnitude than those of

[1] Accelium (previously called Mind Lab; www.accelium.com) is integrated into the school curriculum; to date, over four million students from around the world have participated in Accelium programs.

theory-of-mind; and cognitive ability is associated with higher wages.

Finally, we use the ALSPAC data to provide evidence that primary-school spending can improve theory-of-mind ability in childhood, with the positive effect of school spending concentrated among advantaged students (as measured by relative family income in childhood). The evidence comes from a difference-in-differences approach that exploits different trends over time in primary-school spending across English Local Education Authorities (LEAs) in the midst of a substantial increase in government spending on primary-school education. In contrast to the positive effects on theory-of-mind, we find no evidence that primary-school spending improves cognitive ability.

What is fundamentally new in our paper is a conceptual and empirical investigation of the role that childhood theory-of-mind plays as a cognitive skill that helps to explain life outcomes and the development of strategic sophistication in children. We are not aware of existing work that undertakes the same exercise, and our findings highlight the importance of theory-of-mind as a cognitive skill that is distinct from cognitive ability.

Strategic sophistication matters because people engage in strategic interactions in a wide range of real-world environments. Together, our novel conceptual framework and experimental results provide evidence that theory-of-mind plays an important role distinct from that of cognitive ability in the development of strategic sophistication in children. This novel evidence helps us to understand how theory-of-mind affects economic behavior and provides new insights about the relationship between cognitive skills and bounded rationality.

Turning to our analysis of the birth-cohort data from ALSPAC, we are not aware of any previous systematic study of the relationship between childhood theory-of-mind and adult outcomes. To the best of our knowledge, ALSPAC is the only birth-cohort study to have measured both theory-of-mind in children and adult outcomes. Our finding that childhood theory-of-mind predicts key adult outcomes, including educational attainment and social skills, advances our understanding of the importance of childhood cognitive skills. Furthermore, our conceptual framework together with our mediation analyses clarify the dynamic mechanisms by which theory-of-mind affects outcomes.

Finally, our novel finding that school spending can improve childhood theory-of-mind provides evidence about the returns to educational investments through the production of an economically relevant cognitive skill, theory-of-mind, that economists have, for the most part, neglected.

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):[2] our conceptual framework clarifies the mechanisms by which improvements in childhood cognitive skills might translate into life outcomes and highlights the importance of theory-of-mind; our empirical results help to quantify the importance of these skills; and our findings on the effects of school spending provide new evidence about how cognitive skills can be improved.

Second, we complement recent research on bounded rationality and non-equilibrium thinking (see the survey by Crawford et al., 2013). In particular, we add empirical foundations to new

[2] 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.

theoretical work that seeks to model bounds on depth of reasoning (Alaoui and Penta, 2016, 2018): in our conceptual framework, we link theory-of-mind and cognitive ability to the reasoning processes that determine the observed behavioral level in Alaoui and Penta (2016)'s model; and our experimental findings quantify the importance of cognitive skills for level-k behavior.

Third, our finding that primary-school spending increases theory-of-mind ability complements existing work in psychology that trains children to improve their theory-of-mind. These training interventions are designed specifically to improve theory-of-mind ability measured shortly after the intervention. Based on a meta-analysis of nearly fifty experiments, Hofmann et al. (2016) document that these training interventions prompt children to think about alternative perspectives or mental states, and conclude that these interventions are generally effective at increasing theory-of-mind.[3,4] Our findings also complement Lillard and Else-Quest (2006), who consider the effect of Montessori education on childhood theory-of-mind.

We now turn to related literature. We are not aware of any prior work that investigates how psychometric measures of theory-of-mind and cognitive ability predict the strategic behavior of children in competitive games, although a handful of papers study related questions (Steinbeis et al., 2012; Sher et al., 2014; Geng et al., 2015; Czermak et al., 2016).[5] By contrast, 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, 2021; Brocas and Carrillo, forthcoming).[6]

We are not 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. Our results on children complement Gill and Prowse (2016) who find that adults respond to information about their opponent's cognitive ability in a repeated beauty contest with feedback.[7,8]

Despite an extensive literature in psychology that investigates how children's understanding of intentions affects their moral judgments (e.g., Zelazo et al., 1996), 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 focus on autistic children.[9] By contrast, economists and psychologists have studied how the importance of intentions changes with age (Sutter, 2007; Güroğlu et al., 2009, 2011; Bereby-Meyer and Fiks, 2013, Gummerum and Chu, 2014; Bueno-Guerra et al., 2016; Sul et al., 2017).[10]

[3] See Web Appendix I.1 for details of such theory-of-mind interventions.

[4] Our finding that the effect of school spending on theory-of-mind is concentrated among advantaged students also complements previous research that has found lower theory-of-mind among children of low socio-economic status (SES) (see Charness et al., 2019, and the meta-analysis of Devine and Hughes, 2018).

[5] See Web Appendix I.2 for details.

[6] See Section 4.6 for a review of this literature, and see Web Appendix I.3 for a review of the literature on how children cooperate and coordinate in strategic games.

[7] See Web Appendix I.4 for a description of related work that studies how adults respond to information about their opponent's experience.

[8] In a sample of adults, Gill and Prowse (2016) focuses on the relationship between cognitive ability and strategic behavior in the beauty contest (a zero-sum undercutting game). Gill and Prowse (2016): (i) did not consider theory-of-mind or its relationship to cognitive ability (the study did not measure theory-of-mind); (ii) did not explore the development of strategic sophistication in childhood (all experimental subjects were adults); and (iii) did not study responses to intentions (the beauty contest game does not allow players to infer the intentions of others).

[9] 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.

[10] See Web Appendix I.5 for a discussion of how to measure intentions and reciprocity.

The focus of our experiment is on how observable cognitive skills influence strategic sophistication in children. Sutter et al. (2019)'s review of experiments on children in economics cites a small literature that studies rational choice in children in non-strategic settings (Harbaugh et al., 2001; Apesteguia et al., 2018; Barash et al., 2019); we describe that literature in Web Appendix I.6. Our findings complement existing research on the relationship between cognition and the behavior of adults in strategic settings (e.g.: Burnham et al., 2009; Brañas-Garza et al., 2012; Georganas et al., 2015; Gill and Prowse, 2016; Ridinger and McBride, 2017; Corgnet et al., 2018; Proto et al., 2019). For example, in a repeated beauty contest game, Gill and Prowse (2016) find that cognitive ability predicts how fast adults learn to play equilibrium, while Proto et al. (2019) study cooperation. 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 develops our conceptual framework; Section 3 describes the experimental design; Section 4 presents results from the competitive games; Section 5 presents results from the gift-exchange game; Section 6 presents results from the ALSPAC data; and Section 7 concludes. The Web Appendix includes further details.

2 Conceptual framework

2.1 Introduction to the conceptual framework

We start by developing a conceptual framework that helps us to understand how cognitive skills affect economic behavior and to place our empirical findings in a broader context. In this conceptual framework, childhood theory-of-mind and cognitive ability are skills that affect strategic sophistication, social skills in adulthood, and life outcomes such as educational attainment, fertility and labor market behavior. Theory-of-mind has attracted enormous attention from developmental and social psychologists, as well as from cognitive neuroscience; as noted by Kidd and Castano (2013): "the capacity to identify and understand others' subjective states is one of the most stunning products of human evolution." Despite this, theory-of-mind has attracted little interest from economists. Our conceptual framework is a first step in understanding how theory-of-mind affects a range of economic behaviors. Although economists have shown interest in cognitive ability, our conceptual framework also clarifies the different roles of theory-of-mind and cognitive ability in economic environments.

Premack and Woodruff (1978)'s study of chimpanzees introduced the concept of theory-of-mind by positing that "an individual has a theory of mind if he imputes mental states to himself and others." Using theory-of-mind to impute mental states to others requires setting aside one's own perspective (Baron-Cohen and Wheelwright, 2004). These mental states can include others' beliefs, desires, intentions and emotions (Baron-Cohen, 2000; Agnew et al., 2007), and people use theory-of-mind to interpret others' behavior through the prism of beliefs about their mental states (e.g., Robalino and Robson, 2012, Estes and Bartsch, 2017). In practice, psychologists measure the theory-of-mind ability of adults and children by their ability to correctly infer or understand the mental states of others (see, e.g., the variety of tests used by Kidd and Castano, 2013), and we follow that tradition. Agnew et al. (2007) nicely summarize the meaning of theory-of-mind in the psychology literature as "the knowledge that other animals have mental states which may differ from our own; and the ability to infer what these internal states may be."[11]

We classify theory-of-mind as a cognitive skill for two reasons. First, psychologists measure theory-of-mind ability by the ability to perform well on tasks that rely on cognitive processes. As we note above, these tasks measure the ability to correctly infer or understand the mental states of others. Stone et al. (1998) categorize theory-of-mind as a "cognitive function," while Saxe et al. (2004) call it a "special domain of cognition." According to Wimmer and Perner (1983), who developed the false-belief test of theory-of-mind, understanding another person's false beliefs is a "novel cognitive skill." Second, we use the term 'skill' to distinguish theory-of-mind from a preference: as we describe in the final paragraph of this section, the evidence suggests that theory-of-mind is not, or only very modestly, related to pro-social behaviors and preferences. However, we emphasize that by categorizing theory-of-mind as a cognitive skill, we do not intend to take any particular stance toward debates in the psychology and neuroscience literature about the specific cognitive processes that underpin theory-of-mind (we discuss this literature in Web Appendix I.9).

Although we categorize theory-of-mind as a cognitive skill, theory-of-mind differs from cognitive ability. As we note above, theory-of-mind tests measure the ability to correctly infer or understand the mental states of others; by contrast, tests of cognitive ability do not refer to the mental states of others, but instead measure fluid intelligence (logical reasoning and problem-solving ability) or crystallized intelligence (acquired knowledge and verbal skills) (see in particular Gray and Thompson, 2004, Box 1). Evidence from psychology and neuroscience supports this distinction between theory-of-mind and cognitive ability. Camerer et al. (2005) and Mahy et al. (2014) review the neuroscientific evidence for a specialized theory-of-mind module in the brain. From a behavioral perspective, Stone et al. (1998) emphasize: (i) the particular developmental sequence of theory-of-mind in childhood; and (ii) evidence that various developmental disorders affect theory-of-mind and cognitive ability differently (as we discuss in Web Appendix I.8, one part of this evidence comes from the false photograph test, which suggests that autistic children suffer from a specific theory-of-mind deficit as opposed to a lack of cognitive ability). Henry et al. (2013)'s meta-analysis finds that theory-of-mind ability deteriorates more rapidly in old age than performance on matched control tasks that measure cognitive ability. Finally, our own data (and data from the literature) show a low correlation between theory-of-mind ability and cognitive ability (see Section 3.2.3, Section 6.2.3 and footnote 28).

An important function of theory-of-mind is to help predict the behavior of others (Premack and Woodruff, 1978). However, we emphasize that people who have theory-of-mind can differ in predictive ability and in the ability to best-respond to the predicted behavior of others. Prediction requires cognitive ability alongside theory-of-mind, because prediction relies on an understanding of the incentives that others face and the ability to link incentives and mental attitudes to forecast behavior. Focusing on strategic interactions, in Section 2.2 we note that: (i) cognitive ability allows people to understand the strategic context in which they and others are choosing; (ii) such understanding helps to predict how others will choose; and (iii) cognitive ability also helps people to calculate best-responses to their predictions.

[11] In Web Appendix I.7 we provide further discussion of Premack and Woodruff (1978). According to Chorman and Baron-Cohen (1992), data from the false photograph test and the false drawing test show that autistic children's failure on the false belief test is due to a specific theory-of-mind deficit (i.e., an inability to represent the mental states of others), as opposed to a lack of cognitive ability; we discuss this evidence in Web Appendix I.8.

Before linking theory-of-mind and cognitive ability to specific behaviors and outcomes, we emphasize that in our conceptual framework theory-of-mind operates as a skill and not through preferences. Data from the existing literature and data from our gift-exchange game show that theory-of-mind is not, or only very modestly, related to pro-social preferences and behaviors such as altruism, empathy or reciprocity: we review this evidence in Web Appendix II. By contrast, depending on the environment, theory-of-mind skill can be used to deceive, manipulate, or take advantage of others (e.g., Talwar and Lee, 2008, Doenyas, 2017, DeAngelo and McCannon, 2017).[12] Finally, we note that the existing literature finds that theory-of-mind and cognitive ability show a substantial degree of persistence: we review this evidence in Web Appendix III.

2.2 Strategic sophistication

Strategic sophistication matters because people constantly engage in strategic interactions with others inside and outside the workplace.[13] Strategic skills can be used to cooperate successfully with others, for example by sustaining longterm relationships and reputation built on trust, or by helping people to work well together in teams. Strategic skills also have a darker side, for example when they allow the more strategically able to control negotiations or dominate competitions for jobs or other resources. In our experiment, we measure strategic ability in children up to age twelve; existing literature together with some of our results suggest that strategic sophistication emerges by age twelve (see Section 4.6), and so the strategic skills that we measure are relevant to behavior over the life course.

Strategic interactions are those in which each agent's payoff depends on her own choice together with that of others with whom she interacts. Following Costa-Gomes et al. (2001), we define strategic sophistication as the extent to which players in a strategic interaction analyze their environment as a game by taking the game structure and other players' incentives into account when deciding how to behave.[14] The cognitive reasoning that underpins strategic sophistication involves understanding the payoff structure of the interaction, forming beliefs about how others choose based on the incentives that they face, and choosing well given those beliefs. Costa-Gomes et al. (2001) note that traditional game theory takes strategic sophistication to be unlimited, while evolutionary game theory and adaptive learning models take strategic sophistication to be severely limited. Our approach follows recent work in behavioral economics that studies people who attempt to think strategically without reaching the unlimited strategic sophistication of traditional game theory (see Crawford et al., 2013, for a survey).[15]

[12] Talwar and Lee (2008) find that three- to eight-year-olds with better theory-of-mind lie more. Surveying the literature, Doenyas (2017) concludes that children with high theory-of-mind use this ability to either manipulate others or act pro-socially, depending on their underlying social preferences, leading to "nice ToM behaviors" or "nasty ToM behaviors." DeAngelo and McCannon (2017) find that adults with better theory-of-mind cooperate less and earn more in the prisoners' dilemma game.

[13] Goleman (1995, 1998) stresses the importance of emotional intelligence for success inside and outside the workplace. In Web Appendix I.10 we discuss the relationship between Goleman (1995)'s concept of emotional intelligence and strategic sophistication.

[14] Web Appendix I.11 provides the supporting quote from Costa-Gomes et al. (2001).

[15] In game theory, a "rational" agent is one who best-responds to her beliefs (which, however, need not be correct; e.g., Bernheim, 1984). Thus, a rational agent can be viewed as an expert in one particular aspect of strategic sophistication. However, strategically sophisticated agents do not have to believe that others are rational, and so there is no clear connection between strategic sophistication and common knowledge of rationality (Bernheim, 1984) or strong belief in rationality (Battigalli and Siniscalchi, 2002).

We study strategic sophistication in three ways. First, we study strategic sophistication in a competitive strategic interaction designed to trigger level-k thinking. According to the level-k model, in this setting strategically unsophisticated or "naive" (level-0) agents choose a salient and instinctive action that does not require strategic thinking, while more sophisticated agents form beliefs about the behavioral type of others (level-1 agents best-respond to others being the level-0 type, level-2 agents best-respond to others being the level-1 type, and so on).[16] Second, in the same setting, we study whether subjects exhibit strategic sophistication by using information about their opponent (namely about their opponent's cognitive ability) that can help to predict their opponent's behavior. Third, in a gift-exchange game we study whether subjects show strategic sophistication by taking into account the intentions of the subject that they are matched with, which requires subjects to understand how the other subject thinks about the game. In the remainder of this section we discuss how theory-of-mind and cognitive ability affect strategic sophistication.

Theory-of-mind helps to understand how others are thinking. In strategic interactions, theory-of-mind helps to understand how others perceive the strategic environment, how others think about how you might behave, how others think about what you think about how they will behave, and so on. This type of iterative thinking plays a crucial role in many game-theoretic models because it underpins belief formation and facilitates successful behavior in strategic interactions (e.g., see Robalino and Robson, 2012).

Cognitive ability helps to reason logically and systemize new information. In strategic situations, cognitive ability helps to understand the structure of the game, such as the rules and the payoff consequences of actions. Thus, cognitive ability allows people to understand the strategic context in which they and others are choosing, which in turns helps people to form beliefs about what others will do and to choose well given those beliefs. Just like iterative thinking, the ability to understand the structure of the game and to best-respond to beliefs plays an important role in game theory because success in strategic interactions depends on this ability.

We expect that theory-of-mind and cognitive ability both help to explain behavior in the competitive games that we study. As we explain in Section 4.1, the game structure is designed to trigger level-k non-equilibrium thinking. Alaoui and Penta (2016) develop a theoretical model to study behavior in this type of game, in which each step of reasoning incurs a cognitive cost, and each player has a "cognitive bound" given by the number of steps of reasoning that she is able to perform. Each step of reasoning requires both an understanding of the structure of the game and an awareness that the opponent could have engaged in one fewer step, and so we conjecture that the cost function depends on both cognitive ability and theory-of-mind. In Alaoui and Penta (2016)'s model, the observed behavioral level also depends on beliefs about the opponent's cognitive bound, and we conjecture that theory-of-mind helps to form these beliefs.[17] In Section 4.1 we build on this logic to outline how theory-of-mind ability and cognitive ability link to a sequence of steps of reasoning that helps to underpin level-k thinking in the competitive games.

[16] Other papers that also use the level-k or cognitive hierarchy model to study strategic sophistication include, e.g., Danz et al. (2012), Lindner (2014), Georganas et al. (2015) and Hortaçsu et al. (2019).

[17] Formally, behavior is determined by a player's own cognitive bound, together with her first- and higher-order beliefs about the cost functions (e.g., her belief about the opponent's cost function, and her belief about the opponent's belief about her own cost function).

By contrast, we expect that theory-of-mind plays a relatively more important role in the gift-exchange game, which is designed to understand how children respond to intentions. In that setting, the structure of the game is rather simple, limiting the role of cognitive ability. By contrast, we expect theory-of-mind to help receivers understand the subtle role of the forgone alternative in determining the intentions of the allocator. In particular, second-order beliefs matter: to understand the allocator's intentions, the receiver has to understand how the allocator thinks about the game and her option set. In sum, we predict that theory-of-mind helps children in our gift-exchange game to direct reciprocity appropriately according to the intentions of the allocator. In Section 5 we build on this logic to outline how theory-of-mind ability and cognitive ability link to the steps of reasoning that underpin the receiver's understanding of the allocator's intentions.

2.3 Social skills

Social skills matter because interpersonal relationships are crucial to labor market success (e.g., Conti et al., 2013) and to well-being outside of the workplace (e.g., Segrin and Taylor, 2007). Deming (2017a) argues that the labor market increasingly rewards social skills because high paying jobs have become more likely to require interpersonal skills alongside analytical skills. We define social skills as the ability to build and maintain positive interpersonal relationships, which in turn helps people to function effectively in social networks and teams. People with good social skills build and maintain good interpersonal relationships by, for example: (i) communicating effectively in social interactions; (ii) taking into account the viewpoint of others; (iii) resolving conflict; and (iv) understanding and abiding by conventional rules of social behavior.[18]

We expect that childhood theory-of-mind ability helps children to develop social skills that persist into adulthood. Children that understand better the mental states of others can build more durable social relationships by responding appropriately to the emotions and beliefs of others. For example, theory-of-mind can help children to avoid and resolve interpersonal conflict (e.g., Olson et al., 2011, find less peer aggression in young children with high theory-of-mind). Furthermore, theory-of-mind can allow children to please others by meeting their wants and expectations. And, as noted in Section 2.2, children with better theory-of-mind can direct behaviors such as reciprocity according to others' intentions or other appropriate social cues. As Stone et al. (1998) put it, theory-of-mind seems to be a "modular cognitive capacity that underlies humans' ability to engage in complex social interaction." Indeed, there is already evidence from cross-sectional studies that theory-of-mind correlates with friendship group size or peer popularity in childhood (e.g.:, Nowicki and Duke, 1994, Bosacki and Astington, 1999).[19]

The theoretical link between cognitive ability and social skills is less obvious, but we expect any effect to be positive. People with lower reasoning ability might struggle to understand the subtle rules and norms that underpin social interactions, limiting their ability to acquire social skills. Indeed, Bellanti and Bierman (2000) argue that children with delays in cognitive development have difficulty understanding the rules of group play. Furthermore, people who

[18]Our definition of social skills is closely related to Borghans et al. (2014) who "conceptualize people skills as the ability to effectively interact with or handle interactions with people." In the context of the labor market, Deming (2017a) conceptualizes social skills as the "ability to work with others" by trading tasks.

[19]The psychology literature suggests that theory-of-mind is not just a proxy for social skills because theory-of-mind plays a fundamental role in developing and underpinning social skills; see Web Appendix I.12.

struggle to acquire knowledge or verbal skills might be less interesting or useful to others, again limiting opportunities to develop social skills.

2.4 Education

We expect that theory-of-mind promotes learning and educational attainment. In particular, we conjecture that theory-of-mind operates through one direct and two indirect channels.

First, theory-of-mind might affect learning directly. Estes and Bartsch (2017) argue that readiness to learn hinges on a recognition of the internal content of teachers' minds and that theory-of-mind helps children to learn by allowing them to recognize the pedagogical motives of teachers. Ziv et al. (2008) find that five-year-olds begin to understand that teaching is intentional, and argue that this understanding helps learning. Meltzoff et al. (2009) report that when learning, young children do not slavishly duplicate adult behavior, but instead infer the goals and intentions of the adult.

Second, theory-of-mind might affect learning indirectly via its effect on social skills (see Section 2.3). Children who get on well with their teachers might be rewarded with more individual attention, and children who get on well with other children might learn more from their peers. Davis (2001) argues that education is fundamentally interpersonal in nature and that a good student-teacher relationship motivates learning.

Third, theory-of-mind might affect learning indirectly via children's level of attention in the classroom. Goodfellow and Nowicki (2009) and von Salisch et al. (2017) find that children with worse theory-of-mind are rated as less attentive by teachers. Linking to the first channel above, theory-of-mind deficits could lower attention by inducing boredom if the child struggles to understand the teacher's motives and intentions. Linking to the second channel, lower theory-of-mind could lead to children switching off in class if they do not get along well with their teacher.

The link between cognitive ability and educational attainment is clear, and so we are brief here. Children with higher fluid intelligence are better able to learn by reasoning logically, grasping abstract concepts, and systemizing new information. Children with higher crystallized intelligence have been able to acquire more useful knowledge and better verbal skills. All of these abilities link directly to higher educational attainment.

2.5 Fertility

We expect that theory-of-mind and cognitive ability reduce fertility via education. However, other effects are ambiguous, and so we have no strong prior about the total effect of theory-of-mind or cognitive ability on fertility.

First, we expect that theory-of-mind and cognitive ability reduce fertility in early adulthood via their effects on educational attainment (see Section 2.4). A number of studies find causal evidence that education delays motherhood, although the effect on fertility over the life cycle is inconclusive (see Fort et al., 2016, for a review of the literature, and for evidence from the United Kingdom that education reduces total fertility). Effects of education on fertility are likely to operate through a number of channels, such as access to contraception, marriage rates, and income and substitution effects.

The direct effect of theory-of-mind on fertility is ambiguous. People with better theory-of-mind might have greater insight into the minds of children. This better understanding of children could lead to a more accurate appreciation of the demands of parenthood, thus reducing fertility. On the other hand, understanding children better might increase the desire to raise and nurture offspring. Similarly, there is no clear direct effect of cognitive ability on fertility.

Finally, the indirect effects of theory-of-mind and cognitive ability on fertility via their effects on social skills (see Section 2.3) are also ambiguous. Better social skills might increase the probability of finding a high quality partner. On the other hand, better social skills might increase the value of leisure activities not related to child-rearing.

2.6 Labor market behavior

We expect that theory-of-mind and cognitive ability affect labor market behavior via their effects on strategic sophistication, social skills, educational attainment and fertility, which we discussed in Sections 2.2 to 2.5. The labor market advantages of education and disadvantages of child-bearing are clear, and so do not merit further discussion.

We noted above that strategic and social skills matter for workplace success. Deming (2017a) finds evidence that the labor market increasingly rewards teamwork. In particular, Deming (2017a) shows that labor market success increasingly requires the ability to negotiate complex and repeated interactions within teams, where workers trade specialized tasks among themselves. Success in this environment depends on strategic and social skills that are underpinned by theory-of-mind ability and cognitive ability. Interestingly, one implication is that workers with high theory-of-mind have a comparative advantage in bigger firms, where teams are larger and more complex.

Strategic sophistication further helps in the workplace by improving bargaining skills and bolstering the ability to succeed in repeated competitions, e.g., for promotions or bonuses. Social skills also improve workplace success by creating better relationships with line managers, subordinates and customers, alongside the effects on teamwork.

3 Experimental design

3.1 Overview of the experimental design

The experiment was conducted during the first half of 2016 at five schools in Santander (Spain). Web Appendix IV 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, our sample consists of 730 subjects (377 boys and 352 girls)[20]: Table 1 describes the distribution of subjects by school and age.

[20] One subject did not report their gender.

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.[21]

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. Web Appendix V.1 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. We collected data from the experimental games in sessions with an average size of ten children. We randomly allocated children to a session that included children only from the same school and academic year (but not necessarily from the same class). The sessions were run one-by-one in a room familiar to the children, with only children from that session present in the room. All sessions were conducted by the alphabetically-first author and supervised by a member of staff who was familiar to the children. Since the experimental data were collected from sessions, in all of our analysis we cluster standard errors at the session level. The matching procedures for each game are described in detail in Sections 3.3.2 and 3.4.[22]

[21] Our sample of 730 subjects excludes children who failed to complete the tests in phase 1 or 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.

[22] Web Appendix V.2 describes the experimental software and configuration of the portable lab. Web Appendix V.3 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.

3.2 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.

3.2.1 Theory-of-mind test: Imposing Memory Task

We describe the concept of theory-of-mind in Section 2.1: as we explain there, we measure the theory-of-mind ability of our subjects by their ability to correctly infer or understand the mental states of others. The psychology literature (e.g., Liddle and Nettle, 2006, Warrier and Baron-Cohen, 2018) distinguishes first-order from higher-order theory-of-mind: first-order theory-of-mind refers to "the ability to understand another person's mental state" (Warrier and Baron-Cohen, 2018), while higher-order theory-of-mind refers to the ability to reason recursively about the content of the mental states of others (e.g., understanding what Alice thinks 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.[23]

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. 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.[24] Table A.24 in Web Appendix XIII provides the distribution of IMT scores by age.

3.2.2 Cognitive ability test: Raven's Progressive Matrices

In the experiment, we measured cognitive ability using a test of 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., 2000b). The Raven test consists of non-verbal multiple-choice questions and is recognized as a leading measure of analytic or fluid intelligence (Carpenter et al., 1990; Gray

[23]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).

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

and Thompson, 2004).[25] The Raven test correlates strongly with general cognitive ability g.[26] 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 Coloured Progressive Matrices test, which consists of thirty-six questions split across three sets of increasing difficulty. For children aged eight years and older, we used the Standard Progressive Matrices test, which consists of sixty questions split into fives sets of increasing difficulty. Following convention (Raven et al., 2000b), no time limit was imposed.[27] Table A.25 in Web Appendix XIII provides the distribution of Raven test scores by age.

3.2.3 Age-standardization and correlation of test scores

We standardize psychometric test scores within age group to separate cleanly the effect of age from the effect of within-cohort variation in cognitive skills. 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 Pearson correlation coefficient between the age-standardized theory-of-mind test score and the age-standardized cognitive ability test score is 0.28, which provides evidence that our two tests capture different skills. The low correlation is consistent with existing evidence from the literature on children and adults.[28]

3.3 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. As described in Section 4.1, the game structure was designed by Arad and Rubinstein (2012) to trigger level-k thinking. To study the subjects' initial behavior in the game

[25] 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., 2018), with more accurate beliefs (Burks et al., 2009) and with success in the p-beauty contest for adult populations (Gill and Prowse, 2016).

[26] Gray and Thompson (2004, Box 1) explain that scores on various tests of cognitive ability "can be factor-analysed to give g, a single summary measure of cognitive ability." The circle diagram in Gray and Thompson (2004, Box 1) illustrates "how specific tasks correlate with g – strongly in the centre, weakly at the periphery," with the Raven test sitting right at the center of this circle.

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

[28] Summarizing this literature, Warrier and Baron-Cohen (2018) state that: "previous studies have identified a modest, positive correlation between different measures of theory of mind and cognition." A recent example is Coyle et al. (2018), who find correlations in the range of 0.20 to 0.25 between college entrance math scores and different types of theory-of-mind tests. The meta-study by Baker et al. (2014) of seventy-seven effect sizes finds a correlation of 0.24 between theory-of-mind and intelligence; when restricting attention to child samples, the correlation is 0.22. Similarly, the meta-study by Murphy and Hall (2011) finds a correlation of 0.22 between performance in emotion-recognition tasks and intelligence. Consistent with these findings, Clemmensen et al. (2016)'s recent study of 1,600 Danish eleven- and twelve-year-olds found a modest but statistically significant correlation between a measure of theory-of-mind and cognitive ability.

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 3.1 and Web Appendix V.1 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.

3.3.1 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:[29]

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.[30,31]

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. Web Appendix VI describes the unique symmetric mixed-strategy Nash equilibrium for risk-neutral players (Table A.2) and provides proofs.

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

Baseline game

Subjects played the 'Baseline game' five times. In each of the five repetitions 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.[32]

[29]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 (Web Appendix V.3) it became apparent that the youngest children did not understand the meaning of the word 'additional'. Thus we modified the wording of the game for the sake of clarity.

[30]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).

[31]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 Web Appendix IV for further details.

[32]Therefore, subjects were matched within school and not within session, and we explained to the subjects that often their opponent was not in the same room.

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.[33] Furthermore, each subject was told whether her opponent's Raven test score was above or below average for children of her age (subjects were never given information about their own performance in the Raven test).[34,35] For the purposes of presenting our experimental results, we describe above-average opponents as of 'high cognitive ability' and below-average opponents as of 'low cognitive ability'. By providing age-specific 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).

Order of play

Table A.26 in Web Appendix XIII describes the order of play in phase 2. As noted there, subjects also played the 1-6 Token Request game a single time against the computer, which chose numbers uniform randomly: this 'Computer game' is described in Web Appendix V.4, with results reported in Web Appendix XIII (Tables A.30 to A.35). Subjects played the games in one of two orders, X and Y. Half of the sessions from each school-academic-year pair were randomly allocated to each order, and the allocation to order was balanced in terms of characteristics.[36]

[33]Subjects were matched within session, and we explained to the subjects that their opponent was in the same room. On average, sessions included ten children. When there was an odd number of children in a session, one child was randomly matched with two others (the primary and secondary opponents). The action of the first child was used to compute the payoffs of the primary and secondary opponents, but only the action of the primary opponent was used to compute the payoff of the first child.

[34]The wording used to describe the opponent's cognitive ability was as follows: "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 A.3] whether your partner is in the high or low score group [show in each player's tablet]."

[35]Children were categorized according to whether their Raven test score was above or below the mean for children of the same age in the same school. The small number of children exactly at the mean were placed in the below-mean category.

[36]The smallest school provided only enough children 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 $p = 0.365$. 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.

3.4 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.[37] The currency in this game were 'super-tokens' worth four normal tokens.[38] In each pair, one child 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.[39]

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.[40] Of the 730 subjects, 345 were allocators, while 385 were receivers.[41]

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.[42] 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).[43]

[37] Just like for the Raven game in Phase 2, subjects were matched within session, and we explained to the subjects that their opponent was in the same room.

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

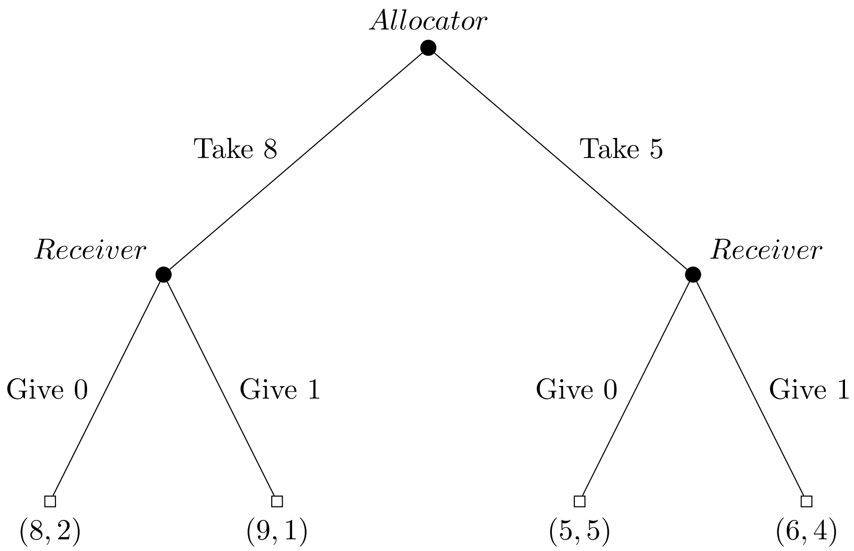
[39] Subjects were told only the total number of tokens that they accumulated in the experiment. Thus allocators were not told whether their matched receiver gave them a super-token.

[40] 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 36 also applies to the allocation to treatment. As noted in footnote 36, all of the children 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.

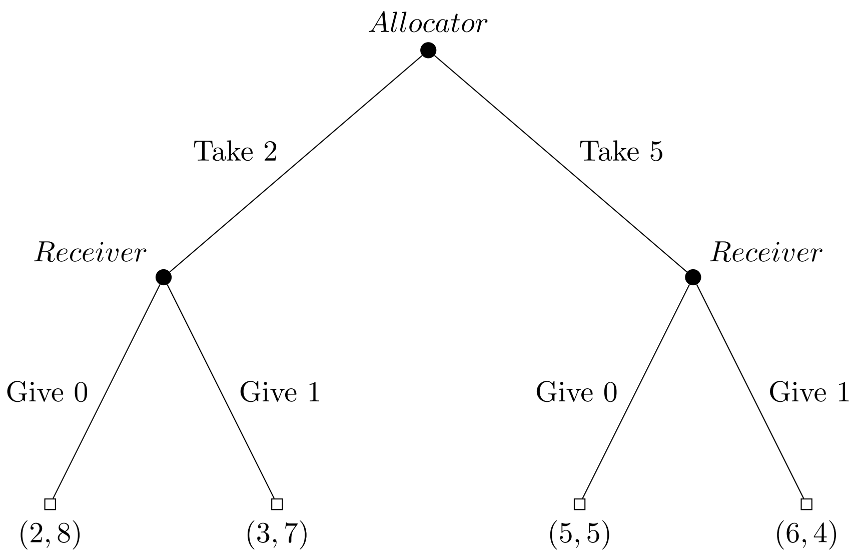
[41] On average, sessions included ten children. When there was an odd number of children in a session, the choice of one randomly selected allocator affected the payoffs of two receivers, but only the choice of the first of these receivers affected the payoff of this allocator, which explains why more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

[42] 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.

[43] 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.

4 Results from the competitive games

4.1 Using the level-k model to analyze behavior in the competitive games

We use the level-k model of non-equilibrium thinking (Nagel, 1995; Stahl and Wilson, 1995) to analyze strategic behavior in the 1-6 Token Request game.[44] The level-k model is based on cognitively simple iterated reasoning: in the level-k hierarchy of types, the level-0 type is strategically unsophisticated, the level-1 type best-responds to others being the unsophisticated level-0 type, the level-2 type best-responds to others being the level-1 type, and so on. We now describe why the level-k model of non-equilibrium thinking is well-suited to analyzing strategic behavior in our setting.

First, we study the subjects' initial behavior in the game without any repeated game effects or learning from experience (as explained in Section 3.3, we randomly rematched subjects and gave subjects no feedback about their performance). Subjects' thinking in initial responses to games tends to avoid fixed-point reasoning or iterated dominance; instead, the evidence supports the use of simpler level-k rules of thumb (Costa-Gomes and Crawford, 2006; Crawford et al., 2013).[45] Furthermore, the level-k model makes no assumption that the subjects know or can learn the empirical distribution of choices, and so the model is well-adapted to our setting without feedback. The fact that our subjects are children, whose strategic sophistication is likely to be less developed than that of adults, strengthens these considerations.

Second, Arad and Rubinstein (2012) designed the game structure to trigger level-k thinking:

- In the level-k model, the incentive of the level-$k > 0$ type to best-respond to others being the level-$(k-1)$ type depends on the assumption that the incentive to compete dominates any incentive to cooperate. The design of the game structure induces competition as a predominant incentive. In that sense, the game is "competitive."[46]

- By design, the game has no pure-strategy Nash equilibrium, and no strategy is weakly or strictly dominated; as a result, subjects cannot reason their way to a pure-strategy Nash equilibrium using iterated dominance or otherwise.

- The game structure is designed to induce a choice of 6 by the strategically unsophisticated level-0 type: this choice is instinctive, salient and guarantees a payoff of six in the absence of strategic thinking.

- The best-response calculation of each level-$k > 0$ type is straightforward: the level-1 type best-responds to the level-0 type's choice of 6 by choosing 5, the level-2 type best-responds

[44]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.

[45]Conceptually, level-k thinking is separate from iterated dominance. Costa-Gomes and Crawford (2006) develop variants of the p-beauty contest game that are designed to distinguish level-k thinking from iterated dominance reasoning, and find that level-k thinking predominates.

[46]First, only one of the two players can earn the ten-token bonus, and so the players are competing for this bonus. Second, given a belief about what the other player will do, undercutting the other player (instead of matching their choice) earns the bonus without changing the other player's payoff. In that sense, "the game does not call for social preferences" (Arad and Rubinstein, 2012, p. 3563). Third, there is no symmetric joint-payoff-maximizing profile of choices: maximizing joint payoffs requires one player to undercut the other to earn the bonus, and so coordination is not a salient incentive.

to the level-1 type's choice of 5 by choosing 4, and so on.[47]

- Finally, the behavior of level-$k > 0$ types is robust to a range of other assumptions about the behavior of the level-0 type (e.g., uniform randomization and any distribution in which 6 is the modal choice).

Before turning to the data, we now build on the conceptual framework in Section 2.2 to outline how theory-of-mind ability and cognitive ability link to a sequence of steps of reasoning that helps to underpin level-k thinking in the 1-6 Token Request game. The steps of reasoning are inspired by Sher et al. (2014) and Alaoui and Penta (2016).

1. Understand the structure of the game, e.g., the rules and payoff consequences of actions. This step mainly uses the player's cognitive ability.

2. Conceive of the opponent as a strategically unsophisticated level-0 type who chooses 6 (see the third and fifth bullets in the bullet list above). This step mainly uses first-order theory-of-mind (we introduced first- and higher-order theory-of-mind in Section 3.2.1).

3. Calculate that the choice of 5 is the level-1 best-response to a level-0 opponent. This step mainly uses cognitive ability.

4. Conceive of the opponent as a level-1 type who thinks her opponent is a level-0 type. This step mainly uses second-order theory-of-mind.

5. Eventually, hit a limit on higher-order theory-of-mind ability, and thus hit an upper limit on the level-k type of the opponent that the player is able to conceive of. Alternatively, the player could understand that this iterated reasoning process can continue indefinitely.

6. Form a belief about the actual level-k type of the opponent, among the levels that the player is able to conceive of. This step mainly uses first- and higher-order theory-of-mind.[48]

7. Calculate the best-response to this belief about the actual level-k type of the opponent, which determines the observed level-k type of the player herself. This step mainly uses cognitive ability.

4.2 Level-k behavior in the Baseline games

As explained in Section 4.1, we use the level-k model of non-equilibrium thinking to analyze strategic behavior in the 1-6 Token Request game. In Sections 2.2 and 4.1, we hypothesized that theory-of-mind and cognitive ability both help to explain level-k behavior: the results in this section confirm this prediction.

To measure how cognitive skills influence strategic behavior, we begin by studying how age, theory-of-mind and cognitive ability affect level-k behavior in the Baseline games, in which subjects shared the same information about the strategic sophistication of their opponent (in each

[47] 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).

[48] The reasoning process that underlies this step depends on beliefs about the strategic sophistication of the opponent, beliefs about the opponent's beliefs about the player's own strategic sophistication, and so on in a hierarchy of beliefs.

of the five repetitions of the Baseline game, the opponent was drawn randomly from the children of all ages from the same school). As explained in Section 3.2.3, we standardize psychometric test scores within age group to separate cleanly the effect of age from the effect of within-cohort variation in cognitive skills.

Figure 2 reports the distribution of choices in the Baseline games, while Table 2 provides 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 2 shows the correspondence between choices and levels explained in Section 4.1. Since subjects' choices vary from round to round, we allow their level to also vary across repetitions of the Baseline game.[49] The third row shows the expected payoff from each choice, given the empirical distribution of choices across the five Baseline games; the best-response to the empirical distribution is to choose 5.

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 2). 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 ten percent. Second, the subjects are less likely to be strategically unsophisticated level-0 types as they age and become more cognitively able (sixth column). Relative to the average, these changes represent decreases of seventeen 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).[50]

When we replicate the analysis using binary rather than continuous measures of cognitive skills, we find an effect of theory-of-mind ability on level-1 behavior that is statistically significant at the one-percent level. This binary analysis is informative since it reduces the influence of outliers in the distributions of skills. Table A.27 in Web Appendix XIII reports this analysis; the table shows that subjects whose theory-of-mind ability is above the median for their age are five percentage points more likely to be a level-1 type. This effect on level-1 behavior is larger in magnitude than that of cognitive ability, although Table A.27 shows that subjects whose cognitive ability is above the median for their age are also more likely to be level-2 types. Finally, unlike Table 2, Table A.27 finds that the effect of theory-of-mind ability on level-3 behavior is not significant at the ten-percent level. In Web Appendix X we provide evidence that our estimates in Table 2 and Table A.27 are robust to correcting for bias due to measurement error, with a modest change in precision in some cases.

The first column of Table 2 shows that the probability of choosing 1 decreases with age and with age-standardized 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.

Web Appendix XIII shows that the parameter estimates in Table 2 are stable when we use a

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

[50] 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.

linear probability model (Table A.28) and when we add controls for the demographics that we collected (gender and school; Table A.29).

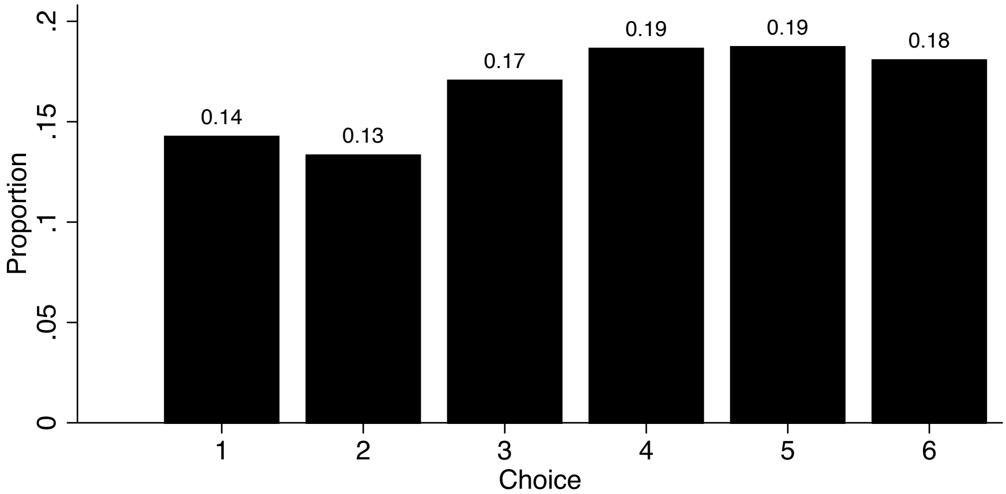


Figure 2: Distribution of choices in the Baseline games.

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***	-0.030***
	(0.007)	(0.003)	(0.004)	(0.004)	(0.004)	(0.006)
Theory-of-mind	-0.000	-0.011	-0.013*	0.007	0.017*	0.001
	(0.009)	(0.006)	(0.007)	(0.009)	(0.010)	(0.011)
Cognitive ability	-0.016*	-0.003	0.006	0.012	0.019**	-0.017*
	(0.009)	(0.006)	(0.007)	(0.008)	(0.009)	(0.010)
Subjects	730	730	730	730	730	730

Notes: Each column reports average marginal effects from a fractional logit. For each choice, the dependent variable is the fraction of times each subject made that choice in the five Baseline games (that is, the number of times each subject made that choice divided by five). The independent variables are age, age-standardized theory-of-mind ability and age-standardized cognitive ability (as explained in Section 3.2.3, we standardize test scores within age group). Section 4.1 explains the correspondence between choices and levels. We calculate the expected payoff (in tokens) to each choice given the empirical distribution of subjects' choices across the five Baseline games. Heteroskedasticity-robust standard errors, clustered at the session level, are shown in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table 2: Probability of choices in the Baseline games.

4.3 Dynamics of choices over repetitions of the Baseline game

As noted in Section 3.3, we study the subjects' initial behavior in the game without feedback, and so they cannot learn from experiencing payoffs or observing the play of their opponents. Nonetheless, our subjects might learn from introspection. In the absence of feedback, Weber (2003) continues to find some learning over repetitions of a competitive game, concluding that such learning comes from "experience with an environment and procedures and by repeatedly thinking about a game." This suggests that cognitive ability could improve learning from introspection in our game by helping subjects to understand better the rules of the game and the payoff consequences of actions as they have more opportunity to consider the strategic environment over time.

The dynamics of the mean choice provide some evidence of learning by introspection: the mean choice falls modestly from 3.85 in the first repetition of the Baseline game to 3.60 in the fifth repetition. Figure A.7 in Web Appendix XIV shows how the mean choice changes over time. Even though the mean choice falls, the best-response to the empirical distribution remains 5 in every repetition. Regressing choices on a linear time trend, the decline in the mean choice is statistically significant at the one-percent level.[51] However, when we add age, age-standardized theory-of-mind ability and age-standardized cognitive ability, and interact these characteristics with the time trend, none of the interactions is statistically significant at the ten-percent level.

The dynamics of level-k behavior from one repetition of the Baseline game to the next suggest a role for cognitive ability. In particular, we study the probability of moving away from level-0 behavior, the probability of moving to level-1 behavior, and the probability of moving to level-2 behavior.[52] In all three cases, age has a positive and statistically significant effect at the one-percent level. Age-standardized cognitive ability has a positive and statistically significant effect in the second and third cases (respectively, $p = 0.053$ and $p = 0.085$; in the first case the effect is positive but not significant at the ten-percent level). The effect of age-standardized theory-of-mind ability is never statistically significant ($p > 0.5$).

4.4 Other analyses of behavior in the Baseline games

Section 4.1 explained why the level-k model of non-equilibrium thinking is well-suited to analyzing strategic behavior in our setting. Here we briefly consider: (i) the probability that subjects best-responded to the empirical distribution of choices in the Baseline games; and (ii) deviations from the Nash equilibrium. However, since we study initial choices without feedback, subjects could not learn from experience about the empirical distribution of choices or adjust their behavior to observed play, and so it is unreasonable to expect them to be able to predict accurately how others play the game or to discover the mixed-strategy Nash equilibrium (Web Appendix VI describes the equilibrium; recall that, by design, the game has no pure-strategy Nash equilibrium and no dominated strategies, and subjects cannot discover the equilibrium using level-k thinking).

Table A.30 in Web Appendix XIII shows that age, age-standardized theory-of-mind ability

[51] Formally, we run an OLS regression of choices on the repetition number.

[52] In the first case, we take all observations of a choice of 6 (level-0 behavior) in the first to fourth repetitions of the Baseline game, construct an indicator of whether the subject chose 1 through 5 in the next repetition, and run a logistic regression of this indicator on our characteristics. The other two regressions are constructed similarly.

and age-standardized cognitive ability positively predict the probability of best-responding (Column 1).[53] Furthermore, there is some evidence that age and cognitive ability act in tandem as complements, while age and theory-of-mind ability operate independently (Column 2).[54] Tables A.33 and A.34 extend the analysis to payoffs: payoffs are noisy because they depend on the specific opponent that a subject is matched with, and so precision is lower. In an attempt to reduce this noise, Table A.35 instead uses expected payoffs given the empirical distribution of choices.

Table A.36 in Web Appendix XIII shows that the distribution of choices of older children is closer to the mixed-strategy Nash equilibrium than that of younger children. The same is true for more cognitively able children and children with better theory-of-mind (among peers of the same age), although the differences are not as quantitatively important. Across the five repetitions of the Baseline game, the deviation from Nash equilibrium increases slightly (the deviation metric defined in Table A.36 increases by 0.19 units); this increase is consistent with our expectation that our subjects cannot discover the Nash equilibrium.[55]

4.5 Response to information about opponent cognitive ability in Raven game

As described in Section 3.3.2, 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). That is, in the Raven game each subject was matched with another child from the same academic year and was told whether her opponent's Raven test score was above or below average for children of her age. For the purposes of presenting our experimental results, we describe above-average opponents as of 'high cognitive ability' and below-average opponents as of 'low cognitive ability'. We now study whether our child subjects responded to this information.

In Section 4.2 we found strong effects of age on level-k behavior. Since children 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.[56]

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 a low-cognitive-ability opponent, while the right-hand-side panel shows the distribution for older subjects who faced a high-cognitive-ability opponent. Older subjects whose opponent is of high cognitive ability are more likely to choose 4, which corresponds to level-2 behavior. Older subjects whose opponent is of low cognitive ability spread

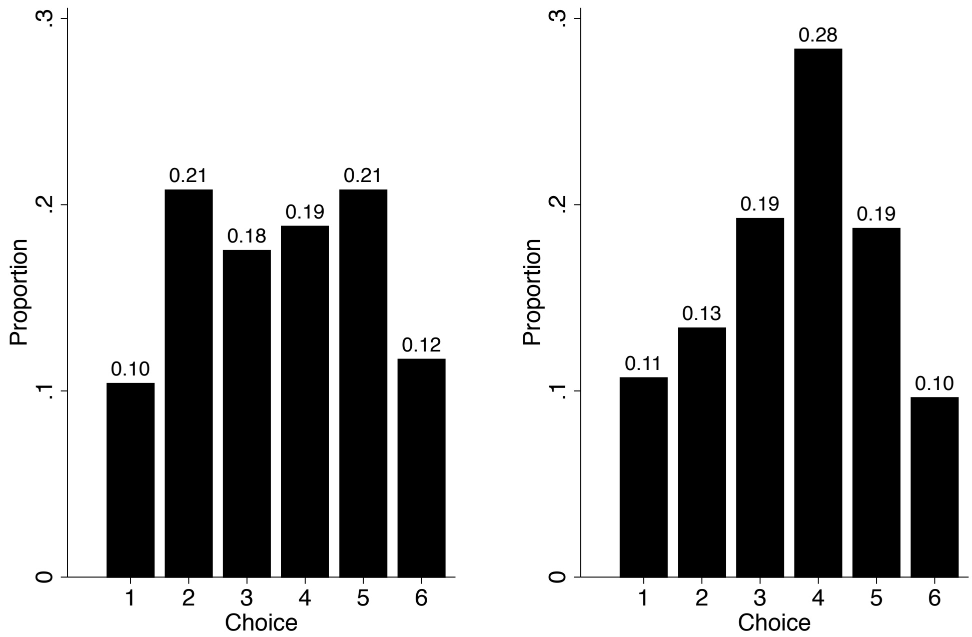
[53]Since the best-response is 5, these findings replicate our results on level-1 behavior from Section 4.2.

[54]The parameter estimates in Table A.30 are stable when we use a linear probability model (Table A.31) and when we add controls for the demographics that we collected (gender and school; Table A.32).

[55]Over the five repetitions, above-median age subjects and above-median cognitive ability subjects move away from the Nash more than do their below-median counterparts. Above-median theory-of-mind ability subjects move away from the Nash, while below-median subjects move toward the Nash. The differences in the change in our deviation metric are, respectively, 0.11, 0.03 and 0.25.

[56]Table 1 in Section 3.1 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).

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.



(a) Low cognitive ability opponent. (b) High cognitive ability opponent.

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 cognitive ability was below (above) the mean for children of her age (see Section 3.3.2). 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). We allocate the small number of subjects exactly at the median age to the above-median category.

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

Table 3 shows that the shift among older children toward level-2 behavior when facing a high-cognitive-ability opponent is statistically significant. The first row of results provides 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 above the median age, with two-sided tests of statistical significance. 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 level-2 behavior by about ten percentage points, with the effect statistically significant at the five-percent level.[57] With caution, we note that these results mesh with those from Section 4.2: in the five repetitions of the Baseline game we found that cognitive ability increases the likelihood of level-1 behavior; and here we find that subjects who face 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 children of any age from the same school.

[57] Older 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.

Choice	1	2	3	4	5	6
Level-k	Level-5	Level-4	Level-3	Level-2	Level-1	Level-0
Subjects above median age (341)						
High cognitive ability opponent	0.003 (0.033)	-0.074* (0.041)	0.017 (0.038)	0.095** (0.046)	-0.021 (0.039)	-0.021 (0.027)
Subjects below median age (343)						
High cognitive ability opponent	-0.019 (0.050)	0.046 (0.039)	-0.036 (0.044)	0.058 (0.042)	-0.036 (0.047)	-0.013 (0.051)

Notes: The table reports average marginal effects from multinomial logits. The first (second) row of results reports average marginal effects for subjects above (below) the median age. In each case, the dependent variable is a categorical indicator of subjects' choices in the Raven game, and the independent variable is an indicator taking value 1 when a subject was matched with an opponent whose cognitive ability was above the mean for children of her age (see Section 3.3.2). We have 341 (343) subjects above (below) the median age who were given information about the cognitive ability of their opponent; these numbers are each less than half of our sample because some subjects were matched with children for whom we had no cognitive ability score (see footnote 21). 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). We allocate the small number of subjects exactly at the median age to the above-median category. Section 4.1 explains the correspondence between choices and levels. Heteroskedasticity-robust standard errors, clustered at the session level, are shown in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table 3: Probability of choices in the Raven game.

The second row of results in Table 3 and Figure A.8 in Web Appendix XIV 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.

Table A.37 in Web Appendix XIII shows our results are robust when we add controls for the demographics that we collected (gender and school).

We now consider how the response to information about the opponent's cognitive ability depends on subjects' own theory-of-mind. To do this, we categorize subjects as being of high or low theory-of-mind according to whether their own theory-of-mind ability is above or below the median for subjects of their age.[58]

As before, first we consider the older (above median age) subjects. Being matched with an opponent of high cognitive ability lowers the probability of choosing 6 (the level-0 choice) for older subjects with high theory-of-mind, but increases this probability for older subjects with low theory-of-mind; the difference according to own theory-of-mind is statistically significant at the ten-percent level ($p = 0.079$). Being matched with an opponent of high cognitive ability

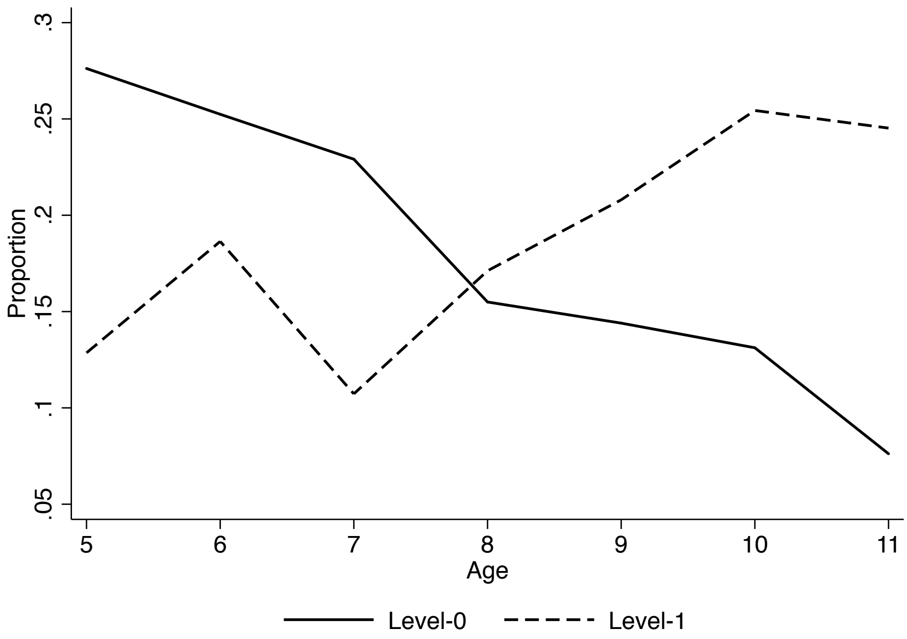
[58] We allocate the small number of subjects exactly at the median to the above-median category.

increases the probability of choosing 5 (the level-1 choice) for older subjects with high theory-of-mind, but reduces this probability for older subjects with low theory-of-mind; again the difference is statistically significant at the 10% level ($p = 0.059$). For choices 1-4, the differences according to own theory-of-mind are not statistically significant at the ten-percent level. Finally, for the younger subjects, these differences according to own theory-of-mind are not statistically significant at the ten-percent level for any of the six choices. When interpreting these results, the reader should keep in mind that we are cutting the data quite finely here, since we are cutting the data by own age, own theory-of-mind and opponent cognitive ability.

4.6 Age and strategic sophistication

In Section 4.2 we found that our subjects are less likely to exhibit strategically unsophisticated level-0 behavior as they become older. Here we delve deeper into this finding. Figure 4 shows how the proportion of level-0 behavior falls with age in the Baseline games. The figure also shows how, at the same time, the proportion of level-1 behavior increases with age. By age eleven the proportion of unsophisticated level-0 behavior falls to under eight percent. This proportion is close to the six percent of adult subjects who exhibit level-0 behavior in Arad and Rubinstein (2012)'s data from the 11-20 game, which shares its strategic structure with our game.

We conclude that our children display strategic sophistication by age eleven, in the sense that by age eleven the extent of unsophisticated level-0 behavior is low and similar to that for adults. This result is consistent with our finding from the Raven game that older subjects exhibit sophistication, in the sense that they shift their level-k behavior in response to information about the cognitive ability of their opponent (Section 4.5).



Notes: The figure shows the proportion of level-0 behavior (choices of 6) and level-1 behavior (choices of 5) in the five Baseline games by age. Section 4.1 explains the correspondence between choices and levels. As explained in the notes to Table 1, we have only thirteen twelve-year-old subjects, and so we exclude these subjects from this figure.

Figure 4: Age and strategic sophistication in the Baseline games.

Our results on age and strategic sophistication provide a unique insight into how unsophisticated level-0 behavior changes with age in a competitive game that is designed to trigger level-k non-equilibrium thinking, with no dominated strategies and no pure-strategy Nash equilibrium. By contrast, the existing literature focuses on dominance-solvable games of varying degrees of complexity. Nonetheless, our finding that children display strategic sophistication by age eleven is broadly consistent with that literature, which finds evidence of the emergence of strategic sophistication between the ages of seven and twelve (Perner, 1979; Sher et al., 2014; Brosig-Koch et al., 2015; Czermak et al., 2016; Brocas and Carrillo, 2021; Brocas and Carrillo, forthcoming).[59,60]

Figure A.9 in Web Appendix XIV repeats the analysis from Figure 4, splitting our sample by whether a subject's theory-of-mind ability is above or below the median for their age. Similarly, Figure A.10 splits the sample by whether a subject's cognitive ability is above or below the median for their age. For subjects aged ten or eleven, the proportion of level-1 behavior is statistically significantly higher among above-median theory-of-mind ability subjects ($p = 0.049$) and among above-median cognitive ability subjects ($p = 0.050$). However, for subjects aged ten or eleven, or for eleven-year-olds alone, the proportion of level-0 behavior among above-median theory-of-mind ability subjects or above-median cognitive ability subjects is not statistically significantly different (at the ten-percent level) from the proportion among the below-median subjects.[61] When interpreting these results, the reader should keep in mind that we are cutting the data quite finely here, since we are using only the data from the oldest subjects, and then cutting again by own theory-of-mind or cognitive ability.

5 Results from the Gift-exchange game

We now turn to the results from our Gift-exchange game, which was designed to understand how children respond to intentions. In Section 2.2, we hypothesized that theory-of-mind plays a more important role than cognitive ability in our gift-exchange setting. In fact, we find that only theory-of-mind predicts whether children respond to intentions: as we describe below, theory-

[59] Perner (1979) studies four- to ten-year-olds, and finds that the ability to detect an opponent's dominant strategy and best-respond to it increases throughout this age range. In an undercutting game, Sher et al. (2014) find that by age seven children play the Nash equilibrium; however, the equilibrium strategy profile generates the minimum payoff, and older children and adults move away from the equilibrium. In a dominant-strategy race game, Brosig-Koch et al. (2015) find that the main developmental step takes place between grades one and four for males, and between grades six and nine for females, although college students perform best. Czermak et al. (2016) study ten- to seventeen-year-olds, and find that the proportion of Nash equilibrium play does not vary with age (although female subjects become less likely to choose dominated strategies as they get older). Brocas and Carrillo (2021) study third- to eleven-graders, and find that the proportion of Nash equilibrium play increases up to around age twelve, and stabilizes thereafter. Finally, in a setting with simple games that have an "intuitive" equilibrium, Brocas and Carrillo (forthcoming) study four- to seven-year-olds, and find that the proportion of Nash equilibrium play increases throughout this age range, with the oldest children playing similarly to adults.

[60] We study strategic behavior in a competitive game, rather than cooperative behavior. In Web Appendix I.3, we review the literature that investigates how cooperation changes with age in strategic games. That literature does not find clear evidence of an age by which cooperative behavior emerges in childhood. For example, Brocas et al. (2017) find that cooperation increases throughout their age range from elementary school to university, and some studies find that cooperation does not increase with age (e.g., Sally and Hill, 2006).

[61] These results are based on fractional logits. The dependent variable is the fraction of times each subject chose 5 (for the level-1 results) or 6 (for the level-0 results) in the five Baseline games (that is, the number of times each subject made that choice divided by five). In each case, the independent variable is an indicator taking value 1 if the subject is above the relevant median for subjects of their age, with subjects exactly at the median allocated to the above-median category.

of-mind ability helps children to direct reciprocity appropriately according to the intentions of the allocator, while we find no effect of cognitive ability.

Section 3.4 describes the game, while Figure 1 presents the game tree. As we explain in Section 3.4, 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 3.4, choosing the even split in Treatment A is unambiguously generous since the 8/2 split is strictly dominant for a selfish allocator.

Before turning to the data, we now build on the conceptual framework in Section 2.2 to outline how theory-of-mind ability and cognitive ability link to the steps of reasoning that underpin the receiver's understanding of the allocator's intentions.

1. Understand the structure of the game, e.g., the rules and payoff consequences of actions. This step mainly uses the receiver's cognitive ability; however, the game structure is rather simple and so the demands on cognitive ability are limited.

2. Conceive of the allocator's perception of her choice set to understand how the allocator thinks about her chosen and foregone alternatives. This step mainly uses the receiver's first-order theory-of-mind (we introduced first- and higher-order theory-of-mind in Section 3.2.1).

3. Conceive of the allocator's social preferences to understand how the foregone alternative interacts with the allocator's generosity to determine her intentions. For instance, take the case when the allocator chooses the equal 5/5 split when she could instead have chosen the more favorable 8/2 split. To understand the allocator's generous intention, the receiver needs to understand: (i) how the allocator perceived her foregone alternative (from Step 2); together with (ii) an understanding that the allocator cares about the receiver's payoff. This step mainly uses the receiver's first-order theory-of-mind.

4. Potentially, engage in further reasoning about the allocator's motivation and social preferences that engages the receiver's higher-order theory-of-mind. For instance, the receiver might think that the allocator's social preferences depend on the allocator's beliefs about the receiver's own preferences or expectations.

	(1)	(2)	(3)
Treatment A	0.095	0.086	-0.126*
	(0.068)	(0.065)	(0.069)
Age		0.015	-0.000
		(0.013)	(0.014)
Theory-of-mind		0.009	-0.028
(T.o.M.)		(0.026)	(0.028)
Cognitive Ability		0.002	0.003
(Cogn. Ab.)		(0.029)	(0.038)
Treatment A. × Age			0.055**
			(0.024)
Treatment A. × T.o.M.			0.140***
			(0.050)
Treatment A. × Cogn. Ab.			0.008
			(0.061)
Subjects	207	207	207

Notes: We include the 207 subjects who were receivers and whose matched allocator chose the 5/5 split. Each column reports average marginal effects from a logit 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 3.2.3, we standardize test scores within age group). Heteroskedasticity-robust standard errors, clustered at the session level, are shown in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table 4: Effect of treatment on probability that receiver gives one super-token to allocator.

Table 4 studies whether subjects respond to intentions by measuring: (i) how the treatment affects the probability that the receiver gives one super-token back to the allocator; and (ii) how this treatment effect interacts with subject characteristics. The table reports average marginal effects, with two-sided tests of statistical significance, for the 207 receivers whose matched allocator chose the 5/5 split. Column (1) shows that receivers are about ten percentage points more likely to give a super-token back to the allocator in Treatment A, although the effect is not quite statistically significant. Thus, on average subjects respond to intentions. 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.[62] 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, the treatment affects the likelihood of giving back a super-token more strongly for older children and children whose theory-of-mind compares favorably to peers of the same age.

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 fourteen 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. In Web Appendix VII we provide evidence that our results in Table 4 on the effects of theory-of-mind and cognitive ability are robust to correcting the estimates for bias due to measurement error.

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 5 replicates the analysis from Table 4, 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 age-standardized theory-of-mind ability (or with age-standardized cognitive ability). Older subjects are more likely to exhibit simple reciprocity, with the effect significant at the five-percent level.

Web Appendix XIII shows that our results are robust when we use a linear probability model (Tables A.38 and A.39) and when we add controls for the demographics that we collected (gender and school; Tables A.40 and A.41).

[62] As explained in Section 3.2.3, we standardize psychometric test scores within age group to separate cleanly the effect of age from the effect of within-cohort variation in cognitive skills.

	(1)	(2)	(3)
Number of super-tokens received	0.040***	0.041***	0.016
(Super-tokens)	(0.009)	(0.009)	(0.016)
Age		0.007	-0.026
		(0.009)	(0.018)
Theory-of-mind		0.011	-0.027
(T.o.M.)		(0.018)	(0.037)
Cognitive ability		-0.004	-0.049
(Cogn. Ab.)		(0.020)	(0.039)
Super-tokens × Age			0.009**
			(0.004)
Super-tokens × T.o.M.			0.009
			(0.010)
Super-tokens × Cogn. Ab.			0.011
			(0.009)
Subjects	385	385	385

Notes: We include all 385 subjects who were receivers (see footnote 41). Each column reports average marginal effects from a logit 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 3.2.3, we standardize test scores within age group). Heteroskedasticity-robust standard errors, clustered at the session level, are shown in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table 5: Effect of number of tokens received on probability that receiver gives one super-token to allocator.

6 Analysis of longitudinal data from ALSPAC

6.1 Overview of ALSPAC data

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a rich birth-cohort study of children born in the early 1990s in the South West of England. Pregnant women resident in Avon with expected dates of delivery from April 1st 1991 to December 31st 1992 were invited to take part in the ALSPAC study. Boyd et al. (2013), Fraser et al. (2013) and Northstone et al. (2019) describe the ALSPAC birth-cohort data.[63,64]

Primary Institutional Review Board (IRB) approval to study data from ALSPAC was obtained from Purdue University. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Informed consent for the use of data collected via questionnaires and clinics was obtained by ALSPAC from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time.[65]

6.2 Psychometric tests

ALSPAC measured theory-of-mind ability and cognitive ability at mean age eight years and eight months when participants attended the "Focus @ 8" interview session.

6.2.1 Theory-of-mind test: Diagnostic Analysis of Nonverbal Accuracy

ALSPAC measured theory-of-mind ability at age eight using the Child Facial Expressions test from the Diagnostic Analysis of Nonverbal Accuracy (DANVA2-CF; see Nowicki and Duke, 1994, and Nowicki, 2015). We describe the concept of theory-of-mind in Section 2.1: as we explain there, psychologists measure theory-of-mind by the ability to correctly infer or understand the mental states of others (which can include others' beliefs, desires, intentions and emotions; Baron-Cohen, 2000). The DANVA2-CF measures theory-of-mind by the ability to recognize emotions in facial expressions. Specifically, the test uses twenty-four photographs of child faces, each showing one of four core emotions: happiness, sadness, anger or fear.[66] The ability of children to recognize these four core emotions measures their social-perceptual theory-of-mind because recognizing these emotions requires children to judge and interpret mental states (Tager-Flusberg and Sullivan, 2000). Web Appendix I.13 provides further details about how ALSPAC implemented the DANVA2-CF, while Web Appendix XI discusses the relationship between the DANVA2-CF and the IMT theory-of-mind test that we used in our experiments (Section 3.2.1).

[63]The total ALSPAC sample size is 15,454 pregnancies. This includes 913 eligible pregnancies that did not join the study initially, but were recruited in later phases (starting when the oldest children were around age seven).

[64]Please note that the study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool: http://www.bristol.ac.uk/alspac/researchers/our-data/. ALSPAC data were collected using REDCap (Harris et al., 2009, 2019).

[65]The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and Eduardo Fe, David Gill and Victoria Prowse will serve as guarantors for the contents of this paper. A comprehensive list of grants funding is available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf); this research was specifically funded by the UK Medical Research Council (MR/M006727/1; G0701503/85179) and the UK Medical Research Council and Wellcome (092731).

[66]The test produces a score out of twenty-four. Each emotion is shown six times, three times with high intensity, and three times with low intensity.

Psychologists often measure theory-of-mind by the ability to recognize emotions in photographs of human faces (Deming, 2017b, notes that there is a well-grounded theory of how theory-of-mind relates to emotion recognition in human faces). Kidd and Castano (2013), a recent and heavily cited paper on theory-of-mind in the journal *Science*, uses both the Reading the Mind in the Eyes Test (RMET) and the Facial Expressions test from DANVA2 to measure theory-of-mind. Like the widely used RMET test of theory-of-mind (Baron-Cohen et al., 1997), the Facial Expressions test from DANVA2 measures the ability to identify emotions from facial expressions in photographs. As noted by Kidd and Castano (2013, Supplementary Materials, p.4), the Facial Expressions test from DANVA2 has been "extensively validated with normal, clinical, adult, and child populations."[67,68]

We are not aware of any study that reports a correlation between performance on the IMT (see Section 3.2.1) and the DANVA2-CF. Similarly, we are not aware of any study that reports a correlation between the IMT and the RMET in children; however, Osterhaus et al. (2016) find that items from the IMT and the RMET load heavily on the same factor for children, and so "were found to be indicators of the same factor."[69]

6.2.2 Cognitive ability test: Wechsler Intelligence Scale for Children

ALSPAC measured cognitive ability at age eight using the Wechsler Intelligence Scale for Children (WISC-III; see Wechsler, 1991). As we note in Section 2, fluid intelligence is the ability to use logical reasoning to solve new problems, while crystallized intelligence includes acquired knowledge and verbal skills (e.g., Carpenter et al., 1990). WISC-III captures both fluid intelligence (Dickson et al., 2016) and crystallized intelligence (Smajlagić et al., 2018). The WISC-III Full Scale IQ score is based on ten subtests: information; similarities; arithmetic; vocabulary; comprehension; picture completion; coding; picture arrangement; block design; and object assembly.

6.2.3 Standardization and correlation of tests scores

We standardize the theory-of-mind tests scores and the cognitive ability tests scores. The Pearson correlation coefficient between the standardized theory-of-mind test score and the standardized cognitive ability test score is 0.21, which provides evidence that the two tests capture different skills. This low correlation is similar to the correlation in our experimental data, and

[67] The Facial Expressions test from DANVA2 is designed to measure "the identification of emotion in facial expressions" (Collins and Nowicki, 2001; see also Goodfellow and Nowicki, 2009). It is one of the tests in the Diagnostic Analysis of Nonverbal Accuracy (DANVA), which was "designed to assess individual differences in children's ability to accurately send and receive emotional information through nonverbal means" (Nowicki and Duke, 1994) (the ALSPAC database does not include data from other DANVA tests).

[68] Specifically, Kidd and Castano (2013) use the RMET and the Adult Facial Expressions test (DANVA2-AF), which is the same as the Child Facial Expressions test (DANVA2-CF), except that the photographs used in the test are of adults instead of children (Nowicki, 2015). The photographs used in the RMET show a part of the face that includes the eyes, while the photographs used in the Facial Expressions test from DANVA2 show the entire face.

[69] In adults, Lyons et al. (2010), Vonk et al. (2015), Ewing et al. (2016) and Vonk and Pitzen (2017) report correlations that range from 0.34 to 0.46 between the IMT and the RMET (all with $p < 0.001$). To put these figures in context, we note that the retest reliability of the RMET is around 0.7 for tests taken between one week and one month apart; in particular, Hallerbäck et al. (2009) and Lee et al. (2020) report correlations of 0.6 and 0.74, respectively, while Prevost et al. (2014) and Charernboon and Lerthattasilp (2017) report rank correlations of 0.7 and 0.95, respectively.

is consistent with existing evidence (see Section 3.2.3 and footnote 28).

6.3 Effect of theory-of-mind and cognitive ability on adult outcomes

6.3.1 Overview of approach and set of controls

We now study how theory-of-mind and cognitive ability at age eight affect adult outcomes. We measure social skills, fertility and labor market behavior at the latest age that is currently available in the ALSPAC dataset. As explained below, we measure educational participation at ages seventeen and twenty. In all cases, we regress the outcome of interest on childhood theory-of-mind ability and cognitive ability using a rich set of controls that is described in Web Appendix VIII. In each case, we use the participants for whom we observe the relevant outcome. The set of controls includes, among other variables: gender; mother's age at birth, marital status and parity; parental race, education, occupation, income and spending on childcare; the number of books owned and neighborhood quality in childhood; and primary (i.e., elementary) school characteristics. Given the large number of controls, we use linear probability models in the cases where the outcome is an indicator variable (rather than less transparent logits). Tables A.42 to A.45 in Web Appendix XIII show that our results on adult outcomes are robust to further including a control for locus of control. Tables A.46 to A.49 in Web Appendix XIII show that our results on the relationship between theory-of-mind and adult outcomes are robust when we include separate controls for the verbal and non-verbal components of cognitive ability. In Web Appendix IX we provide evidence that our estimates of the effects of theory-of-mind and cognitive ability on adult outcomes in Tables 6 to 9 are broadly stable after correcting for bias due to measurement error, with a modest loss of precision in some cases.

6.3.2 Social skills

In Section 2.3, we explained that social skills are crucial to labor market success and to well-being outside the workplace, and we hypothesized that childhood theory-of-mind ability helps children to develop social skills that persist into adulthood. We also hypothesized that any effect of cognitive ability on social skills would be positive, although the theoretical link is weaker. The results in Table 6 support these predictions.

We use two measures of social skills in adulthood: (i) the number of close friends; and (ii) an indicator of whether the participant has ten or more close friends. The regression results in Table 6 show positive effects of childhood theory-of-mind and cognitive ability on adult social skills that are statistically significant at the five-percent level or lower. A one-standard-deviation increase in theory-of-mind ability at age eight is associated with an increase of around 0.20 in the number of close friends and increase of around two percentage points in the probability of having ten or more close friends, both measured at age twenty-four. Similarly, a one-standard-deviation increase in cognitive ability at age eight is associated with an increase of around 0.28 in the number of close friends and an increase of around two percentage points in the probability of having ten or more close friends.

	Number of close friends	Ten or more close friends
Theory-of-mind	0.203**	0.021***
	(0.079)	(0.007)
Cognitive ability	0.284***	0.021***
	(0.090)	(0.008)
Participants	3,887	3,887
Mean age	24.49	24.49
Mean of dependent variable	6.76	0.16

Notes: Each column reports an OLS regression with a rich set of controls that is described in Web Appendix VIII. In the first column the dependent variable is the number of close friends that each participant reported when they attended the "Focus @ 24" interview session. In the second column the dependent variable is an indicator taking value 1 if a participant reported having ten or more close friends at the same interview session. The independent variables are standardized theory-of-mind ability and standardized cognitive ability measured at age eight (see Section 6.2). We derived the number of close friends from the participant's categorical report. The categories were: 0; 1; 2-4; 5-9; 10-14; 15-19; $\geq$ 20. We used interval regression to assign a value to each category that includes more than one value. Heteroskedasticity-robust standard errors are shown in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table 6: Close friends.

6.3.3 Education

In Section 2.4, we hypothesized that childhood theory-of-mind and cognitive ability promote learning and educational attainment. The results in Table 7 support these predictions, and in Web Appendix XII we quantify the importance of attention in the classroom as a dynamic mechanism that mediates the effect of childhood theory-of-mind on educational attainment.

We measure educational participation at ages seventeen and twenty, and we also have one measure of educational attainment. In England, compulsory schooling for the relevant cohorts ended at sixteen, and so our measure of educational participation at seventeen captures whether or not the participant chose to continue their secondary (i.e., high school) education beyond the compulsory period. The measure of educational participation at twenty captures whether the participant chose to continue with some form of tertiary education (e.g., college or vocational training) at the end of their secondary education. Finally, we measure whether the participant completed one or more A-levels; in England A-levels are subject-specific exams that determine college entry and that are generally taken at age eighteen.

The results in Table 7 show positive effects of childhood theory-of-mind and cognitive ability on education that are statistically significant at the five-percent level or lower, with the effects of cognitive ability larger in magnitude. A one-standard-deviation increase in theory-of-mind ability at age eight is associated with an increase of around one percentage point in the probability of educational participation at age seventeen, an increase of around two percentage points in educational participation at age twenty, and an increase of around two percentage points in the probability of completing one or more A-levels. For cognitive ability at age eight, the corresponding magnitudes are increases of two percentage points, nine percentage points, and nine percentage points, respectively.

	In education at 17	In education at 20	A-level qualification
Theory-of-mind	0.013**	0.020**	0.016**
	(0.006)	(0.009)	(0.008)
Cognitive ability	0.023***	0.094***	0.091***
	(0.007)	(0.009)	(0.008)
Participants	4,091	4,332	4,332
Mean age	17.76	20.95	20.95
Mean of dependent variable	0.88	0.55	0.73

Notes: Each column reports an OLS regression with a rich set of controls that is described in Web Appendix VIII. In the first column the dependent variable is an indicator taking value 1 if a participant reported being in education when they attended the "Teen Focus 4" interview session. In the second column the dependent variable is an indicator taking value 1 if a participant reported being in education or training that is not as part of a job when they answered the "It's All About You (20+)" questionnaire. In the third column the dependent variable is an indicator taking value 1 if a participant reported having at least one A-level or equivalent qualification when they answered the "It's All About You (20+)" questionnaire. The independent variables are standardized theory-of-mind ability and standardized cognitive ability measured at age eight (see Section 6.2). Heteroskedasticity-robust standard errors are shown in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table 7: Educational participation and attainment.

6.3.4 Fertility

In Section 2.5, we hypothesized that childhood theory-of-mind and cognitive ability both lower fertility in early adulthood, although we cautioned that our priors are not strong because some effects are theoretically ambiguous. The results in Table 8 support this prediction, and in Web Appendix XII we quantify the importance of educational participation as a dynamic mechanism that mediates the effect of childhood theory-of-mind on fertility in early adulthood.

We have two measures of fertility in early adulthood: (i) whether the participant is a parent; and (ii) the participant's total number of children. The results in Table 8 show negative effects of childhood theory-of-mind and cognitive ability on fertility that are statistically significant at the five-percent level or lower. A one-standard-deviation increase in theory-of-mind ability at age eight is associated with a reduction of around one percentage point in the probability of being a parent and a reduction of around 0.02 children, both measured at age twenty-five. This reduction of 0.02 children represents a decline of twelve percent relative to the average number of children among our participants. For cognitive ability at age eight, the corresponding magnitudes are reductions of two percentage points and 0.03 children.

	Parent	Number of children
Theory-of-mind	-0.012**	-0.022**
	(0.006)	(0.009)
Cognitive ability	-0.023***	-0.033***
	(0.006)	(0.010)
Participants	4,371	4,371
Mean age	25.76	25.76
Mean of dependent variable	0.12	0.18

Notes: Each column reports an OLS regression with a rich set of controls that is described in Web Appendix VIII. In the first column the dependent variable is an indicator taking value 1 if a participant reported being a parent when they answered the "Life @ 25+" questionnaire. In the second column the dependent variable is the number of children that each participant reported having in the same questionnaire. The independent variables are standardized theory-of-mind ability and standardized cognitive ability measured at age eight (see Section 6.2). In all cases "children" is defined to include biological, step, foster and adopted children. Heteroskedasticity-robust standard errors are shown in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table 8: Fertility.

6.3.5 Labor market behavior

In Section 2.6, we hypothesized that childhood theory-of-mind and cognitive ability both affect labor market behavior. In Table 9 we find no statistically significant effect of theory-of-mind on the probability of being in employment at age twenty-four or on wages at age twenty-three. Interestingly, we do find evidence that supports our hypothesis that workers with better theory-of-mind have a comparative advantage in larger firms: a one-standard deviation increase in theory-of-mind ability at age eight is associated with an increase of two percentage points at age twenty-three in the probability of being employed in a workplace with twenty-five or more employees ($p = 0.079$).

The absence of an effect of theory-of-mind on the probability of employment should be seen in the context of record employment rates in the United Kingdom in recent years.[70] Regarding wages, we conjecture that any effects of theory-of-mind are not yet apparent in participants' early twenties because the wage distribution in the United Kingdom is compressed in the early years following graduation.[71]

Turning to cognitive ability, we find no statistically significant effect on the probability of employment, but we do find a positive effect of childhood cognitive ability on wages that is statistically significant at the five-percent level: a one-standard deviation increase in cognitive

[70] Based on OECD records for the United Kingdom that start in 1984, the employment rate in 2016 (around the time that our measure of employment was collected by ALSPAC) was at a record high of 82.6% for individuals aged 25-54 (OECD, 2019).

[71] Blundell et al. (2016) find that the ratio of the median wage of first-degree holders to the median wage of high school graduates in the United Kingdom is around 1.2 for individuals at age twenty-three; this ratio increases to around 1.6 at age forty.

ability at age eight is associated with an increase of two percent in wages at age twenty-three. Throughout Section 6.3, we find effects of cognitive ability that are somewhat larger than those of theory-of-mind, and so it makes sense that cognitive ability should start to affect wages earlier in people's careers. Finally, we also find that childhood cognitive ability increases the probability of being employed in a workplace with twenty-five or more employees.

	Employed	Log hourly wage	Large workplace
Theory-of-mind	0.005	-0.009	0.020*
	(0.008)	(0.008)	(0.011)
Cognitive ability	0.005	0.022**	0.029**
	(0.009)	(0.009)	(0.012)
Participants	3,884	2,999	3,129
Mean age	24.49	23.89	23.89
Mean of dependent variable	0.81	2.05	0.60

Notes: Each column reports an OLS regression with a rich set of controls that is described in Web Appendix VIII. In the first column the dependent variable is an indicator taking value 1 if a participant reported being in employment and not in education when they attended the "Focus @ 24" interview session. In the second column the dependent variable is the log of the hourly wage that each participant reported when they answered the "Me @ 23+" questionnaire. In the third column the dependent variable is an indicator taking value 1 if a participant reported that twenty-five or more people were employed at their place of work when they answered the "Me @ 23+" questionnaire. The second and third columns use only participants who reported being in employment in the "Me @ 23+" questionnaire. The independent variables are standardized theory-of-mind ability and standardized cognitive ability measured at age eight (see Section 6.2). We derived the hourly wage from the participant's categorical report of monthly take-home pay (after tax and national insurance) and from their report of average hours worked per week. To reduce the influence of outliers, we winsorized the hourly wage at the 99th percentile. The categories for monthly take-home pay were: £1–499; £500–999; £1000–1,499; £1,500–1,999; £2,000–2,499; £2,500–2,999; $\geq$ £3,000. We used interval regression to assign a value to each category. Heteroskedasticity-robust standard errors are shown in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table 9: Employment, wages, and workplace size.

6.4 Effect of school spending on theory-of-mind

Finally, we provide evidence that school spending can improve theory-of-mind ability in childhood. The evidence comes from a difference-in-differences approach that uses data from primary-school spending in England.

Starting in the late 1990s, there was a substantial increase in government spending on primary (i.e., elementary) school education in England: in the ten-year period from 1996-97 to 2006-07, annual primary-school spending per pupil increased from around £2,300 to around £4,300 (in 2019-20 prices; see Farquharson and Sibieta, 2019). In England, each school is funded by a regional Local Education Authority (LEA). LEAs received different increases in funding from the central government that depended on the characteristics of the pupils in the LEA.

We use a difference-in-differences approach that exploits different trends over time in school spending across LEAs to estimate how primary-school spending affected childhood theory-of-mind. The ALSPAC dataset includes participants from three consecutive school-year cohorts and schools from four LEAs. Specifically, we regress standardized theory-of-mind at mean age eight years and eight months (see Section 6.2) on the cumulative per-pupil primary-school spending of the participant's LEA when the participant was in Year 1, Year 2 and Year 3 (children were aged five, six, and seven at the beginning of Year 1, Year 2, and Year 3, respectively). We report the effect of £300 of cumulative per-pupil spending over three years (in 2003-04 prices), since £300 is approximately one standard deviation of the identifying variation in school spending in our data.[72] In the regression, we include LEA fixed effects and cohort fixed effects, along with a rich set controls that is summarized in Section 6.3.1 and described in Web Appendix VIII. The cohort fixed effects absorb trends over time in school spending and in theory-of-mind that are common to all LEAs. The LEA fixed effects control for systematic differences in school spending and in theory-of-mind across LEAs. The estimated effect of school spending is therefore identified from variation across LEAs in the time path of spending.[73]

Table 10 provides evidence that primary-school spending increased theory-of-mind ability. The first column includes participants who attended a primary school in one of our four LEAs: we find that a £300 (in 2003-04 prices) increase in cumulative per-pupil spending over three years increases theory-of-mind ability at age eight by around 0.13 of a standard deviation ($p = 0.059$).

To help interpret our finding that school spending improved theory-of-mind on average: (i) we first describe some of the features of the increase in primary-school spending; and (ii) we then use evidence from psychology to propose mechanisms by which school spending could have improved theory-of-mind. The increase in English primary-school expenditure over the relevant period was spent as follows: forty-four percent on teachers; twenty percent on educational support, administrative and clerical staff; nine percent on books and equipment; and twenty-seven percent on other goods and services.[74] Furthermore, around the relevant time period there was a push by the United Kingdom government to provide extra assistance to school children, and this extra assistance was a mix of support inside and outside of the classroom provided by class teachers, support teachers and learning support assistants (Croll and Moses, 2003). Building on existing evidence from psychology, we propose three specific mechanisms which can help to explain our finding that school spending improved theory-of-mind ability: the first operates through reading and literacy; the second is based on conversational discourse; and the third links to classroom activities.

1. According to the psychology literature, reading and literacy are crucial for theory-of-mind development. For example, Hofmann et al. (2016)'s survey of training interventions to improve childhood theory-of-mind reports that many of these interventions are designed around storytelling (we discuss this survey in the introduction); similarly, Kidd and Castano (2013) provide evidence that reading literary fiction improves theory-of-mind. Based

[72] After calculating the change in cumulative per-pupil spending from the first cohort to the third cohort for each LEA, we find that £300 is approximately one standard deviation of the changes.

[73] Mobility between LEAs is low, which suggests that pupils did not select into LEAs based on the level of school spending. ALSPAC collected information about participants' LEA in Years 3, 4 and 6 of primary school: 0.28% of participants changed LEA between Years 3 and 4 and 0.94% changed LEA between Years 4 and 6.

[74] We calculated these figures using data on English LEA primary-school based spending between 1997-98 and 2000-01 from DfES (1999, 2002).

on this literature, we suggest that an important mechanism by which school spending likely improved theory-of-mind is through better reading and literacy skills. Indeed, according to Croll and Moses (2003), the push to provide extra assistance to school children described in the paragraph above was directed toward supporting reading and literacy (rather than numeracy or other areas of the curriculum).

2. de Rosnay and Hughes (2006)'s literature review concludes that conversational interactions are of fundamental importance to the development of theory-of-mind. By increasing the provision of support teachers and learning support assistants to help teachers inside and outside the classroom, school spending boosted the opportunities for children to engage in theory-of-mind improving conversational interactions at school.

3. Hofmann et al. (2016)'s survey of training interventions to improve childhood theory-of-mind highlights the importance of activities like socio-dramatic play and imaginative role playing. We suggest that the increase in support teachers and learning support assistants provided teachers with some opportunity to pivot away from traditional classroom academics toward more "fun" learning and play activities based on interpersonal interaction and role playing that helped to improve theory-of-mind.

	All	Disadvantaged participants	Advantaged participants
School spending per pupil (£300 cumulative over three years)	0.128* (0.067)	-0.041 (0.105)	0.267** (0.120)
Participants	4,549	1,999	2,199

Notes: Data on LEA primary-school based spending are from DfES (2004). The third paragraph of Section 6.4 describes the empirical specification and strategy. The specification also includes school fixed effects, which absorb the LEA fixed effects because all schools in the ALSPAC dataset remain in the same LEA. The first column includes participants who completed the theory-of-mind test at age eight (see Section 6.2) and attended a publicly-funded primary school in one of the four LEAs in the ALSPAC dataset; we allocate each participant to a school based on the earliest time that we observe the participant attending one of these schools. The second (third) column includes only disadvantaged (advantaged) participants: the fifth paragraph of Section 6.4 and footnote 75 explain how we define these categories based on family income in childhood; family income is described in Web Appendix VIII. Our sample includes a partial school-year cohort of participants born between February 1991 and August 1991, a full school-year cohort born between September 1991 and August 1992, and a partial school-year cohort born between September 1992 and January 1993. To ensure that we estimate the effect of school spending from differences between comparable cohorts, we construct an indicator taking value 1 for participants born between February and August, and in our regressions we include this indicator variable interacted with each of the fixed effects and controls. Heteroskedasticity-robust standard errors, clustered at the school level, are shown in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table 10: Difference-in-differences estimates of the effect of school spending on theory-of-mind.

Next, we provide evidence that the effect of school spending on theory-of-mind was stronger for advantaged participants, as measured by family income in childhood. We call a participant 'advantaged' if her family income in childhood was at or above the median for the participants in her primary school; we call her 'disadvantaged' if her family income was below the median.[75] In the second column of Table 10, we find no statistically significant effect of school spending on theory-of-mind for disadvantaged participants ($p = 0.694$). Instead, in the third column we find that the positive effect of school spending on theory-of-mind ability is concentrated among the advantaged students ($p = 0.028$). Table A.50 in Web Appendix XIII shows that our results are robust when we allocate participants at the median of childhood family income to the disadvantaged category.

At first blush, it might seem surprising that school spending improved theory-of-mind for advantaged participants but had no effect for disadvantaged participants. To help interpret these findings: (i) we first describe a stream of literature which reports that universal interventions often benefit the advantaged more and discusses the ethical implications for policy design (e.g., Ceci and Papierno, 2005); and (ii) we then build on recent evidence from economics and psychology to propose three specific mechanisms that can help to explain our findings.

Our finding that school spending was more effective for advantaged children is consonant with exciting new work in economics by Barcellos et al. (2021) and Huillery et al. (2021). Leveraging a change in years of compulsory schooling in the United Kingdom, Barcellos et al. (2021) show that returns to schooling are higher for advantaged children; furthermore, genetic data from children and their parents suggest that parental and environmental characteristics are responsible for these higher returns (rather than characteristics of the children themselves). Huillery et al. (2021) study a large-scale mindset intervention in French schools designed to increase locus of control and the perceived return to effort, and find that the effect of the intervention on academic achievement was substantially stronger for advantaged students. Furthermore, the small effects for disadvantaged students found by Huillery et al. (2021) only appeared in the fourth year of the sustained intervention, which raises the possibility that sustained increases in school spending do eventually start to have some effect on the theory-of-mind of disadvantaged children in the ALSPAC sample, but in a timeframe beyond the scope of our study.

Ceci and Papierno (2005) survey broader evidence showing that many universal interventions and policies widen initial disparities by benefitting advantaged groups more, and they go on to discuss the thorny ethical implications for policy design. For example, Ceci and Papierno (2005) describe how the Sesame Street educational television series and the availability of computer technology in schools both widened pre-existing gaps between advantaged and disadvantaged children. More recent examples include Mascini and Braster (2017), who use OECD PISA data to argue that choice and competition in education benefit high-SES children more. Indeed, the effect extends beyond impacts on educational attainment: for example, in a mental-health context, Kavanagh et al. (2009)'s meta-study provides some evidence that school-based mental-health interventions designed to prevent depression are more effective for higher-SES children.

Why might the theory-of-mind of advantaged children benefit more from school spending? Building on existing evidence from economics and psychology, we propose three specific mechanisms which can help to explain our finding that school spending was more effective for advantaged children: the first operates through biased beliefs and motivation; the second is based on differences in utilization; and the third links to better executive function among the advantaged.

[75] We use school-level medians to ensure that a participant's categorization is orthogonal to spending in her particular school. The total of advantaged and disadvantaged participants is lower than the number of participants in the first column of Table 10 because we do not have data on family income for all participants.

1. Guyon and Huillery (2021) find that the children of low-SES parents have downward-biased beliefs about their academic ability and potential, which then feed into lower educational aspirations. Because self-confidence affects motivation (e.g., Bénabou and Tirole, 2002), this type of excess pessimism among the disadvantaged could adversely affect learning and individual development in the classroom, which in turn could limit for disadvantaged children the impact of school spending on the evolution of cognitive skills such as theory-of-mind.

2. In their survey of the evidence that many interventions widen disparities, Ceci and Papierno (2005) describe the "utilization benefit" that advantaged groups gain when they are better able to access or utilize the intervention's benefits. In our context, high-SES parents might have been more able and willing to ensure that their children reaped the benefits of the school spending increase. For example, we noted above that part of the relevant school spending increase was directed toward extra assistance outside of the classroom: high-SES parents might have been more pro-active in securing such extra assistance for their children when needed.

3. Evidence from psychology suggests that executive function mediates changes in theory-of-mind. In particular, as discussed in Hofmann et al. (2016)'s survey, Benson et al. (2013) find that theory-of-mind training is more effective for children with better executive function, which they define as "the processes that underlie goal directed behavior, including self-regulation, planning, working memory, response inhibition, and resistance to interference." Furthermore, psychologists have also found that SES positively predicts executive function (e.g., Lawson et al., 2018, Rosen et al., 2019). Bringing these findings together, we suggest that school spending increased theory-of-mind more effectively for the advantaged participants partly because of their better executive function.

We find no statistically significant effects of primary-school spending on cognitive ability at age eight. When we replicate the analysis using cognitive ability instead of theory-of-mind, the equivalents of the regressions reported in the first to third columns of Table 10 give p-values of 0.605, 0.361 and 0.815, respectively. Across all our participants, we find a statistically insignificant effect of school spending on cognitive ability that is in fact negative (but small).

Since the ALSPAC dataset includes only three cohorts and four LEAs, we consider our evidence on the effects of school spending to be suggestive rather than conclusive. We encourage researchers to attempt to replicate our results by measuring the effects of similar interventions on theory-of-mind and cognitive ability when they run such interventions directly in the field.

7 Conclusion

In this paper, we have emphasized the importance of childhood theory-of-mind as a cognitive skill that underpins strategic ability, as well as success in social, educational and workplace interactions. We find that theory-of-mind operates as a skill separate from cognitive ability, and we provide evidence that primary-school spending can help to improve theory-of-mind ability in childhood. We hope that our findings illustrate the fundamental importance of theory-of-mind for economic behavior. Since teamwork and soft skills are becoming ever more critical to success in life (Deming, 2017a), we expect that the importance of theory-of-mind will only increase over time.

Estes and Bartsch (2017) define theory-of-mind as "the universal propensity of humans to understand and explain their own and others' behavior in terms of internal mental states and processes such as beliefs, desires, goals, and intentions" and note that theory-of-mind appears "rapidly and dependably early in development across all cultures without explicit instruction and even in the most disadvantaged and deficient environments." Yet, economists have shown little interest in studying theory-of-mind.

We hope that our findings will spur a new research agenda in economics that takes seriously how people understand the mental processes of others, that aims to learn more about how theory-of-mind skills drive behavior, and that aspires to understand how to improve theory-of-mind ability. For example, game theorists should model explicitly how people come to form beliefs about others' mental states in strategic situations,[76] researchers studying cognitive and non-cognitive skills should include theory-of-mind in the set of key skills that they examine, and researchers designing early-life interventions should study systematically the effects of their interventions on theory-of-mind ability.

Recent theoretical work on bounded rationality and non-equilibrium thinking seeks to model bounds on depth of reasoning (Alaoui and Penta, 2016, 2018). By developing specific hypotheses that link theory-of-mind and cognitive ability to level-k behavior, and by confirming experimentally that these cognitive skills matter for level-k thinking, we hope that our findings will add useful foundations to future theoretical work that models explicitly how cognitive skills help to drive the beliefs and cognitive processes that determine depth of reasoning in games.

We find that theory-of-mind and cognitive ability affect strategic sophistication. Alongside research on how to improve cognitive skills, future studies should also investigate how to improve strategic skills directly. For example, Ashraf et al. (forthcoming) find that training the strategic skills of adolescent girls in developing countries helps them to negotiate better educational resources from their parents. Furthermore, this type of training should be tailored to the person's specific level of theory-of-mind skill and cognitive ability. It is also 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 theory-of-mind and cognitive ability in non-strategic settings.

The ALSPAC birth-cohort study follows children born in the early 1990s into young adulthood. Following these participants further through their lives will, over time, allow us and others to build a comprehensive picture of the roles of theory-of-mind and cognitive ability in driving behavior throughout the life cycle.

[76] Initial work in this direction studies the evolution of the ability to understand others' preferences (Robalino and Robson, 2016), while Gauer and Kuzmics (2017) and Kimbrough et al. (2017) study learning about others' preferences in games.

References (for main text and Web Appendix)

Agnew, Z.K., **Bhakoo, K.K.**, and **Puri, B.K.** (2007). The human mirror system: A motor resonance theory of mind-reading. *Brain Research Reviews*, 54(2): 286–293

Agranov, M., **Potamites, E.**, **Schotter, A.**, and **Tergiman, C.** (2012). Beliefs and endogenous cognitive levels: An experimental study. *Games and Economic Behavior*, 75(2): 449–463

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

Alaoui, L. and **Penta, A.** (2018). Cost-benefit analysis in reasoning. *Universitat Pompeu Fabra, Economics Working Paper 1621*

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

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

Artinger, F., **Exadaktylos, F.**, **Koppel, H.**, and **Sääksvuori, L.** (2014). In others' shoes: Do individual differences in empathy and theory of mind shape social preferences? *PloS One*, 9(4): e92844

Ashraf, N., **Bau, N.**, **Low, C.**, and **McGinn, K.** (forthcoming). Negotiating a better future: How interpersonal skills facilitate inter-generational investment. *Quarterly Journal of Economics*

Baker, C.A., **Peterson, E.**, **Pulos, S.**, and **Kirkland, R.A.** (2014). Eyes and IQ: A meta-analysis of the relationship between intelligence and "Reading the Mind in the Eyes". *Intelligence*, 44: 78–92

Banerjee, R., **Watling, D.**, and **Caputi, M.** (2011). Peer relations and the understanding of faux pas: Longitudinal evidence for bidirectional associations. *Child Development*, 82(6): 1887–1905

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

Barcellos, S.H., **Carvalho, L.**, and **Turley, P.** (2021). The effect of education on the relationship between genetics, early-life disadvantages, and later-life SES. *NBER Working Paper 28750*

Baron-Cohen, S. (2000). Theory of mind and autism: a fifteen year review. In S. Baron-Cohen, H. Tager-Flusberg, and D.J. Cohen, editors, *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience, 2nd edition*, 3–20. Oxford University Press

Baron-Cohen, S., **Jolliffe, T.**, **Mortimore, C.**, and **Robertson, M.** (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry*, 38(7): 813–822

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

Baron-Cohen, S. and **Wheelwright, S.** (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, 34(2): 163–175

Bastiaansen, J.A., **Thioux, M.**, and **Keysers, C.** (2009). Evidence for mirror systems in emotions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1528): 2391–2404

Battigalli, P. and **Siniscalchi, M.** (2002). Strong belief and forward induction reasoning. *Journal of Economic Theory*, 106(2): 356–391

Beldoch, M. (1964). Sensitivity to expression of emotional meaning in three modes of communication. In J.R. Davitz, editor, *The Communication of Emotional Meaning*, 31–42. McGraw-Hill

Bellanti, C.J. and **Bierman, K.L.** (2000). Disentangling the impact of low cognitive ability and inattention on social behavior and peer relationships. *Journal of Clinical Child Psychology*, 29(1): 66–75

Bénabou, R. and **Tirole, J.** (2002). Self-confidence and personal motivation. *Quarterly Journal of Economics*, 117(3): 871–915

Benson, J.E., **Sabbagh, M.A.**, **Carlson, S.M.**, and **Zelazo, P.D.** (2013). Individual differences in executive functioning predict preschoolers' improvement from theory-of-mind training. *Developmental Psychology*, 49(9): 1615–1627

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

Bernheim, B.D. (1984). Rationalizable strategic behavior. *Econometrica*, 52(4): 1007–1028

Białecka-Pikul, M., **Szpak, M.**, **Zubek, J.**, **Bosacki, S.**, and **Kołodziejczyk, A.** (forthcoming). The psychological self and advanced theory of mind in adolescence. *Self and Identity*

Bloom, B.S. (1964). *Stability and Change in Human Characteristics*. New York: John Wiley

Blundell, R., **Green, D.A.**, and **Jin, W.** (2016). The UK wage premium puzzle: How did a large increase in university graduates leave the education premium unchanged? *Institute for Fiscal Studies Working Paper W16/01*

Borghans, L., **Ter Weel, B.**, and **Weinberg, B.A.** (2014). People skills and the labor-market outcomes of underrepresented groups. *Industrial and Labor Relations Review*, 67(2): 287–334

Bosacki, S. and **Astington, J.W.** (1999). Theory of mind in preadolescence: Relations between social understanding and social competence. *Social Development*, 8(2): 237–255

Boyd, A., **Golding, J.**, **Macleod, J.**, **Lawlor, D.A.**, **Fraser, A.**, **Henderson, J.**, **Molloy, L.**, **Ness, A.**, **Ring, S.**, and **Davey Smith, G.** (2013). Cohort Profile: The 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology*, 42: 111–127

Braakmann, N. (2011). Female education and fertility – Evidence from changes in British compulsory schooling laws. *Mimeo, Newcastle University*

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

Breslau, J., **Miller, E.**, **Breslau, N.**, **Bohnert, K.**, **Lucia, V.**, and **Schweitzer, J.** (2009). The impact of early behavior disturbances on academic achievement in high school. *Pediatrics*, 123(6): 1472–1476

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

Brocas, I. and **Carrillo, J.D.** (forthcoming). Iterative dominance in young children: Experimental evidence in simple two-person games. *Journal of Economic Behavior & Organization*

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

Bruins, M., **Duffy, J.A.**, **Keane, M.P.**, and **Smith, A.A.** (2018). Generalized indirect inference for discrete choice models. *Journal of Econometrics*, 205(1): 177–203

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

Camerer, C., **Loewenstein, G.**, and **Prelec, D.** (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature*, 43(1): 9–64

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

Ceci, S.J. and **Papierno, P.B.** (2005). The rhetoric and reality of gap closing: When the "have-nots" gain but the "haves" gain even more. *American Psychologist*, 60(2): 149–160

Charernboon, T. and **Lerthattasilp, T.** (2017). The Reading the Mind in the Eyes Test: Validity and reliability of the Thai version. *Cognitive and Behavioral Neurology*, 30(3): 98–101

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

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

Charness, G., **Rustichini, A.**, and **Van de Ven, J.** (2018). Self-confidence and strategic behavior. *Experimental Economics*, 21(1): 72–98

Chorman, T. and **Baron-Cohen, S.** (1992). Understanding drawing and beliefs: a further test of the meta-representation theory of mind. *Journal of Child Psychology and Psychiatry*, 33: 1105–1112

Chung, Y.S., **Barch, D.**, and **Strube, M.** (2014). A meta-analysis of mentalizing impairments in adults with schizophrenia and autism spectrum disorder. *Schizophrenia Bulletin*, 40(3): 602–616

Clemmensen, L., **van Os, J.**, **Drukker, M.**, **Munkholm, A.**, **Rimvall, M.**, **Vaever, M.**, **Rask, C.**, **Bartels-Velthuis, A.**, **Skovgaard, A.**, and **Jeppesen, P.** (2016). Psychotic experiences and hyper-theory-of-mind in preadolescence – a birth cohort study. *Psychological Medicine*, 46(1): 87–101

Collins, M. and **Nowicki, S.** (2001). African American children's ability to identify emotion in facial expressions and tones of voice of European Americans. *Journal of Genetic Psychology*, 162(3): 334–346

Conti, G., **Galeotti, A.**, **Mueller, G.**, and **Pudney, S.** (2013). Popularity. *Journal of Human Resources*, 48(4): 1072–1094

Corgnet, B., **Desantis, M.**, and **Porter, D.** (2018). What makes a good trader? On the role of intuition and reflection on trader performance. *Journal of Finance*, 73(3): 1113–1137

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

Costa-Gomes, M.A. and **Crawford, V.P.** (2006). Cognition and behavior in two-person guessing games: An experimental study. *American Economic Review*, 96(5): 1737–1768

Cowell, J.M., **Lee, K.**, **Malcolm-Smith, S.**, **Selcuk, B.**, **Zhou, X.**, and **Decety, J.** (2017). The development of generosity and moral cognition across five cultures. *Developmental Science*, 20(4): e12403

Cowell, J.M., **Samek, A.**, **List, J.**, and **Decety, J.** (2015). The curious relation between theory of mind and sharing in preschool age children. *PLoS One*, 10(2): e0117947

Coyle, T.R., **Elpers, K.E.**, **Gonzalez, M.C.**, **Freeman, J.**, and **Baggio, J.A.** (2018). General intelligence (g), ACT scores, and theory of mind: (ACT)g predicts limited variance among theory of mind tests. *Intelligence*, 71: 85–91

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

Croll, P. and **Moses, D.** (2003). Special educational needs across two decades: Survey evidence from English primary schools. *British Educational Research Journal*, 29(5): 731–747

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

Danz, D.N., **Fehr, D.**, and **Kübler, D.** (2012). Information and beliefs in a repeated normal-form game. *Experimental Economics*, 15(4): 622–640

Davis, H.A. (2001). The quality and impact of relationships between elementary school students and teachers. *Contemporary Educational Psychology*, 26(4): 431–453

de Rosnay, M. and **Hughes, C.** (2006). Conversation and theory of mind: Do children talk their way to socio-cognitive understanding? *British Journal of Developmental Psychology*, 24(1): 7–37

DeAngelo, G. and **McCannon, B.C.** (2017). Theory of mind predicts cooperative behavior. *Economics Letters*, 155: 1–4

Deming, D.J. (2017a). The growing importance of social skills in the labor market. *Quarterly Journal of Economics*, 132(4): 1593–1640

Deming, D.J. (2017b). The value of soft skills in the labor market. *NBER Reporter*, 4: 7–11

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

Devine, R.T. and **Hughes, C.** (2018). Family correlates of false belief understanding in early childhood: A meta-analysis. *Child Development*, 89(3): 971–987

DfES (1999). Education and training expenditure since 1989–90. *National Statistics Bulletin B10 - 1999 (July 1999), Department for Education and Skills*

DfES (2002). Education and training expenditure since 1992–93. *National Statistics Bulletin B06 - 2002 (September 2002), Department for Education and Skills*

DfES (2004). Education and training expenditure since 1994–95. *National Statistics Bulletin Bweb03 - 2004 (October 2004–Revised May 2005), Department for Education and Skills* Statistical Appendix: https://webarchive.nationalarchives.gov.uk/20130323081649///media.education.gov.uk/assets/files/xls/bweb032004sbppxls.xls

Dickson, M., **Gregg, P.**, and **Robinson, H.** (2016). Early, late or never? When does parental education impact child outcomes? *Economic Journal*, 126(596): F184–F231

Doenyas, C. (2017). Self versus other oriented social motivation, not lack of empathic or moral ability, explains behavioral outcomes in children with high theory of mind abilities. *Motivation and Emotion*, 41(6): 683–697

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

Duncan, G.J., **Dowsett, C.J.**, **Claessens, A.**, **Magnuson, K.**, **Huston, A.C.**, **Klebanov, P.**, **Pagani, L.S.**, **Feinstein, L.**, **Engel, M.**, **Brooks-Gunn, J.**, **Sexton, H.**, and **Duckworth, K.** (2007). School readiness and later achievement. *Developmental Psychology*, 43(6): 1428–1446

Edele, A., **Dziobek, I.**, and **Keller, M.** (2013). Explaining altruistic sharing in the dictator game: The role of affective empathy, cognitive empathy, and justice sensitivity. *Learning and Individual differences*, 24: 96–102

Engel, D., **Woolley, A.W.**, **Jing, L.X.**, **Chabris, C.F.**, and **Malone, T.W.** (2014). Reading the mind in the eyes or reading between the lines? Theory of mind predicts collective intelligence equally well online and face-to-face. *PloS One*, 9(12): e115212

Estes, D. and **Bartsch, K.** (2017). Theory of mind: A foundational component of human general intelligence. *Behavioral and Brain Sciences*, 40: e201

Ewing, D., **Zeigler-Hill, V.**, and **Vonk, J.** (2016). Spitefulness and deficits in the social–perceptual and social–cognitive components of Theory of Mind. *Personality and Individual Differences*, 91: 7–13

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

Farquharson, C. and **Sibieta, L.** (2019). *2019 Annual Report on Education Spending in England: Schools*. Institute for Fiscal Studies

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

Ferguson, F.J. and **Austin, E.J.** (2010). Associations of trait and ability emotional intelligence with performance on Theory of Mind tasks in an adult sample. *Personality and Individual Differences*, 49(5): 414–418

Fernández-Abascal, E.G., **Cabello, R.**, **Fernández-Berrocal, P.**, and **Baron-Cohen, S.** (2013). Test-retest reliability of the 'Reading the Mind in the Eyes' test: A one-year follow-up study. *Molecular Autism*, 4(33): 1–6

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*

Fort, M., **Schneeweis, N.**, and **Winter-Ebmer, R.** (2016). Is education always reducing fertility? Evidence from compulsory schooling reforms. *Economic Journal*, 126(595): 1823–1855

Fraser, A., **Macdonald-Wallis, C.**, **Tilling, K.**, **Boyd, A.**, **Golding, J.**, **Davey Smith, G.**, **Henderson, J.**, **Macleod, J.**, **Molloy, L.**, **Ness, A.**, **Ring, S.**, **Nelson, S.M.**, and **Lawlor, D.A.** (2013). Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International Journal of Epidemiology*, 42: 97–110

Gale, C.R., **Cooper, R.**, **Craig, L.**, **Elliott, J.**, **Kuh, D.**, **Richards, M.**, **Starr, J.M.**, **Whalley, L.J.**, and **Deary, I.J.** (2012). Cognitive function in childhood and lifetime cognitive change in relation to mental wellbeing in four cohorts of older people. *PLoS One*, 7(9): e44860

Gallese, V. and **Goldman, A.** (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12): 493–501

Gallese, V., **Keysers, C.**, and **Rizzolatti, G.** (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, 8(9): 396–403

Gauer, F. and **Kuzmics, C.** (2017). Cognitive empathy in conflict situations. *SSRN Working Paper 2715160*

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

Georganas, S., **Healy, P.J.**, and **Weber, R.A.** (2015). On the persistence of strategic sophistication. *Journal of Economic Theory*, 159: 369–400

Geruso, M. and **Royer, H.** (2018). The impact of education on family formation: Quasi-experimental evidence from the UK. *NBER Working Paper 24332*

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

Gillen, B., **Snowberg, E.**, and **Yariv, L.** (2019). Experimenting with measurement error: Techniques with applications to the Caltech cohort study. *Journal of Political Economy*, 127(4): 1826–1863

Goldman, A.I. and **Sripada, C.S.** (2005). Simulationist models of face-based emotion recognition. *Cognition*, 94(3): 193–213

Goleman, D. (1995). *Emotional Intelligence*. New York: Bantam

Goleman, D. (1998). What makes a leader? *Harvard Business Review*, Nov.-Dec.: 93–102

Goleman, D. and **Boyatzis, R.** (2017). Emotional intelligence has 12 elements. Which do you need to work on? *Harvard Business Review*, 84(2): 1–5

Goodfellow, S. and **Nowicki, S.** (2009). Social adjustment, academic adjustment, and the ability to identify emotion in facial expressions of 7-year-old children. *Journal of Genetic Psychology*, 170(3): 234–243

Gow, A.J., **Johnson, W.**, **Pattie, A.**, **Brett, C.E.**, **Roberts, B.**, **Starr, J.M.**, and **Deary, I.J.** (2011). Stability and change in intelligence from age 11 to ages 70, 79, and 87: The Lothian Birth Cohorts of 1921 and 1936. *Psychology and Aging*, 26(1): 232–240

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

Gross, K. and **Wronski, J.** (forthcoming). Helping the homeless: The role of empathy, race and deservingness in motivating policy support and charitable giving. *Political Behavior*

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

Guajardo, N.R. and **Watson, A.C.** (2002). Narrative discourse and theory of mind development. *The Journal of Genetic Psychology*, 163(3): 305–325

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

Guyon, N. and **Huillery, E.** (2021). Biased aspirations and social inequality at school: Evidence from French teenagers. *Economic Journal*, 131(634): 745–796

Hallerbäck, M.U., **Lugnegård, T.**, **Hjärthag, F.**, and **Gillberg, C.** (2009). The Reading the Mind in the Eyes Test: Test–retest reliability of a Swedish version. *Cognitive Neuropsychiatry*, 14(2): 127–143

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

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

Harris, P.A., **Taylor, R.**, **Minor, B.L.**, **Elliott, V.**, **Fernandez, M.**, **O'Neal, L.**, **McLeod, L.**, **Delacqua, G.**, **Delacqua, F.**, **Kirby, J.**, and **Duda, S.N.** (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics*, 95: 103208

Harris, P.A., **Taylor, R.**, **Thielke, R.**, **Payne, J.**, **Gonzalez, N.**, and **Conde, J.G.** (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42(2): 377–381

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

Henry, J.D., **Phillips, L.H.**, **Ruffman, T.**, and **Bailey, P.E.** (2013). A meta-analytic review of age differences in theory of mind. *Psychology and Aging*, 28(3): 826–839

Hofmann, S.G., **Doan, S.N.**, **Sprung, M.**, **Wilson, A.**, **Ebesutani, C.**, **Andrews, L.A.**, **Curtiss, J.**, and **Harris, P.L.** (2016). Training children's theory-of-mind: A meta-analysis of controlled studies. *Cognition*, 150: 200–212

Hortaçsu, A., **Luco, F.**, **Puller, S.L.**, and **Zhu, D.** (2019). Does strategic ability affect efficiency? Evidence from electricity markets. *American Economic Review*, 109(12): 4302–4342

Huillery, E., **Bouguen, A.**, **Charpentier, A.**, **Algan, Y.**, and **Chevallier, C.** (2021). The role of mindset in education: A large-scale field experiment in disadvantaged schools. *Mimeo, Santa Clara University*

Jensen, A.R. (1969). How much can we boost IQ and scholastic achievement. *Harvard Educational Review*, 39(1): 1–123

Jensen, A.R. (1980). *Bias in Mental Testing*. New York: Free Press

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

Kavanagh, J., **Oliver, S.**, **Lorenc, T.**, **Caird, J.**, **Tucker, H.**, **Harden, A.**, **Greaves, A.**, **Thomas, J.**, and **Oakley, A.** (2009). School-based cognitive-behavioural interventions: A systematic review of effects and inequalities. *Health Sociology Review*, 18(1): 61–78

Khorashad, B.S., **Baron-Cohen, S.**, **Roshan, G.M.**, **Kazemian, M.**, **Khazai, L.**, **Aghili, Z.**, **Talaei, A.**, and **Afkhamizadeh, M.** (2015). The "Reading the Mind in the Eyes" test: Investigation of psychometric properties and test–retest reliability of the Persian version. *Journal of Autism and Developmental Disorders*, 45(9): 2651–2666

Kidd, D.C. and **Castano, E.** (2013). Reading literary fiction improves theory of mind. *Science*, 342(6156): 377–380

Kimbrough, E.O., **Robalino, N.**, and **Robson, A.J.** (2017). Applying "theory of mind": Theory and experiments. *Games and Economic Behavior*, 106: 209–226

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

Kogut, T., **Slovic, P.**, and **Västfjäll, D.** (2016). The effect of recipient identifiability and neediness on children's sharing behavior. *Journal of Behavioral Decision Making*, 29(4): 353–362

Lang, H., **DeAngelo, G.**, and **Bongard, M.** (2018). Theory of mind and general intelligence in dictator and ultimatum games. *Games*, 9(2): 16

Lawson, G.M., **Hook, C.J.**, and **Farah, M.J.** (2018). A meta-analysis of the relationship between socioeconomic status and executive function performance among children. *Developmental Science*, 21(2): e12529

Le Coq, C. and **Sturluson, J.T.** (2012). Does opponents' experience matter? Experimental evidence from a quantity precommitment game. *Journal of Economic Behavior & Organization*, 84(1): 265–277

Lecce, S., **Bianco, F.**, **Devine, R.T.**, and **Hughes, C.** (2017). Relations between theory of mind and executive function in middle childhood: A short-term longitudinal study. *Journal of Experimental Child Psychology*, 163: 69–86

Lee, H.R., **Nam, G.**, and **Hur, J.W.** (2020). Development and validation of the Korean version of the Reading the Mind in the Eyes Test. *PloS One*, 15(8): e0238309

Leekam, S.R. and **Perner, J.** (1991). Does the autistic child have a metarepresentational deficit? *Cognition*, 40(3): 203–218

Leslie, A.M. and **Thaiss, L.** (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition*, 43(3): 225–251

Leuner, B. (1966). Emotional intelligence and emancipation. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 15: 193–203

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

Lillard, A. and **Else-Quest, N.** (2006). Evaluating Montessori education. *Science*, 313(5795): 1893–1894

Lindner, F. (2014). Decision time and steps of reasoning in a competitive market entry game. *Economics Letters*, 122(1): 7–11

Liu, B., **Huang, Z.**, **Xu, G.**, **Jin, Y.**, **Chen, Y.**, **Li, X.**, **Wang, Q.**, **Song, S.**, and **Jing, J.** (2016). Altruistic sharing behavior in children: Role of theory of mind and inhibitory control. *Journal of Experimental Child Psychology*, 141: 222–228

Livingston, L.A., **Colvert, E.**, **the Social Relationships Study Team**, **Bolton, P.**, and **Happé, F.** (2019). Good social skills despite poor theory of mind: exploring compensation in autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, 60(1): 102–110

Lu, H., **Su, Y.**, and **Wang, Q.** (2008). Talking about others facilitates theory of mind in Chinese preschoolers. *Developmental Psychology*, 44(6): 1726–1736

Lyons, M., **Caldwell, T.**, and **Shultz, S.** (2010). Mind-reading and manipulation – Is Machiavellianism related to theory of mind? *Journal of Evolutionary Psychology*, 8(3): 261–274

Mahy, C.E., **Moses, L.J.**, and **Pfeifer, J.H.** (2014). How and where: Theory-of-mind in the brain. *Developmental Cognitive Neuroscience*, 9: 68–81

Mascini, P. and **Braster, S.** (2017). Choice and competition in education: Do they advance performance, voice and equality? *Public Administration*, 95(2): 482–497

Mayer, J.D., **Salovey, P.**, **Caruso, D.R.**, and **Cherkasskiy, L.** (2011). Emotional intelligence. In R.J. Sternberg and S.B. Kaufman, editors, *The Cambridge Handbook of Intelligence*, 528–549. Cambridge University Press

Meltzoff, A.N., **Kuhl, P.K.**, **Movellan, J.**, and **Sejnowski, T.J.** (2009). Foundations for a new science of learning. *Science*, 325(5938): 284–288

Murphy, B.A. and **Lilienfeld, S.O.** (2019). Are self-report cognitive empathy ratings valid proxies for cognitive empathy ability? Negligible meta-analytic relations with behavioral task performance. *Psychological Assessment*, 31(8): 1062–1072

Murphy, N.A. and **Hall, J.A.** (2011). Intelligence and interpersonal sensitivity: A meta-analysis. *Intelligence*, 39(1): 54–63

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

Northstone, K., **Lewcock, M.**, **Groom, A.**, **Boyd, A.**, **Macleod, J.**, **Timpson, N.**, and **Wells, N.** (2019). The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019. *Wellcome Open Research*, 4(51). https://doi.org/10.12688/wellcomeopenres.15132.1

Nowicki, S. (2015). Manual for the receptive tests of the Diagnostic Analysis of Nonverbal Accuracy 2 (DANVA2). *Research Gate Publication 309121252*

Nowicki, S. and **Duke, M.P.** (1994). Individual differences in the nonverbal communication of affect: The Diagnostic Analysis of Nonverbal Accuracy Scale. *Journal of Nonverbal Behavior*, 18(1): 9–35

Nowicki, S., **Gregory, S.**, **Iles-Caven, Y.**, **Ellis, G.**, and **Golding, J.** (2018). Early home-life antecedents of children's locus of control. *Frontiers in Psychology*, 9: 2032

Nowicki, S. and **Strickland, B.R.** (1973). A locus of control scale for children. *Journal of Consulting and Clinical Psychology*, 40(1): 148–154

OECD (2019). Employment database - Employment indicators: Employment rate, by gender and detailed age group (annual). *Organisation for Economic Co-operation and Development* http://www.oecd.org/employment/emp/employmentdatabase-employment.htm

Olson, S.L., **Lopez-Duran, N.**, **Lunkenheimer, E.S.**, **Chang, H.**, and **Sameroff, A.J.** (2011). Individual differences in the development of early peer aggression: Integrating contributions of self-regulation, theory of mind, and parenting. *Development and Psychopathology*, 23(1): 253–266

ONS (2000). *Standard Occupational Classification 2000 Volume 1: Structure and Description of the Unit Groups*. Office for National Statistics and the Government Economic Service

Osterhaus, C., **Koerber, S.**, and **Sodian, B.** (2016). Scaling of advanced theory-of-mind tasks. *Child Development*, 87(6): 1971–1991

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

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

Peterson, C.C. and **Wellman, H.M.** (2019). Longitudinal theory of mind (ToM) development from preschool to adolescence with and without ToM delay. *Child Development*, 90(6): 1917–1934

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

Prevost, M., **Carrier, M.E.**, **Chowne, G.**, **Zelkowitz, P.**, **Joseph, L.**, and **Gold, I.** (2014). The Reading the Mind in the Eyes test: validation of a French version and exploration of cultural variations in a multi-ethnic city. *Cognitive Neuropsychiatry*, 19(3): 189–204

Prifitera, A., **Weiss, L.G.**, and **Saklofske, D.H.** (1998). The WISC-III in context. In A. Prifitera and D.H. Saklofske, editors, *WISC-III Clinical Use and Interpretation*, 1–38. The Psychological Corporation

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

Qu, L., **Shen, P.**, **Chee, Y.Y.**, and **Chen, L.** (2015). Teachers' theory-of-mind coaching and children's executive function predict the training effect of sociodramatic play on children's theory of mind. *Social Development*, 24(4): 716–733

Raven, J., **Raven, J.C.**, and **Court, J.H.** (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales: Section 2 (Coloured Progressive Matrices)*. San Antonio: Pearson

Raven, J., **Raven, J.C.**, and **Court, J.H.** (2000a). *Manual for Raven's Progressive Matrices and Vocabulary Scales: Section 3 (Standard Progressive Matrices)*. San Antonio: Pearson

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

Realo, A., **Allik, J.**, **Nõlvak, A.**, **Valk, R.**, **Ruus, T.**, **Schmidt, M.**, and **Eilola, T.** (2003). Mind-reading ability: Beliefs and performance. *Journal of Research in Personality*, 37(5): 420–445

Ridinger, G. and **McBride, M.** (2017). Theory-of-mind ability and cooperation. *Mimeo, UC Irvine*

Robalino, N. and **Robson, A.** (2012). The economic approach to 'theory of mind'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1599): 2224–2233

Robalino, N. and **Robson, A.** (2016). The evolution of strategic sophistication. *American Economic Review*, 106(4): 1046–72

Rochat, P., **Dias, M.D.**, **Liping, G.**, **Broesch, T.**, **Passos-Ferreira, C.**, **Winning, A.**, and **Berg, B.** (2009). Fairness in distributive justice by 3-and 5-year-olds across seven cultures. *Journal of Cross-Cultural Psychology*, 40(3): 416–442

Rosen, M.L., **Amso, D.**, and **McLaughlin, K.A.** (2019). The role of the visual association cortex in scaffolding prefrontal cortex development: A novel mechanism linking socioeconomic status and executive function. *Developmental Cognitive Neuroscience*, 39: 100699

Rütgen, M., **Seidel, E.M.**, **Riečanskỳ, I.**, and **Lamm, C.** (2015a). Reduction of empathy for pain by placebo analgesia suggests functional equivalence of empathy and first-hand emotion experience. *Journal of Neuroscience*, 35(23): 8938–8947

Rütgen, M., **Seidel, E.M.**, **Silani, G.**, **Riečanskỳ, I.**, **Hummer, A.**, **Windischberger, C.**, **Petrovic, P.**, and **Lamm, C.** (2015b). Placebo analgesia and its opioidergic regulation suggest that empathy for pain is grounded in self pain. *Proceedings of the National Academy of Sciences*, 112(41): E5638–E5646

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

Salovey, P. and **Mayer, J.D.** (1990). Emotional intelligence. *Imagination, Cognition and Personality*, 9(3): 185–211

Saxe, R., **Carey, S.**, and **Kanwisher, N.** (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology*, 55: 87–124

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

Segrin, C. and **Taylor, M.** (2007). Positive interpersonal relationships mediate the association between social skills and psychological well-being. *Personality and Individual Differences*, 43(4): 637–646

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

Silles, M.A. (2011). The effect of schooling on teenage childbearing: Evidence using changes in compulsory education laws. *Journal of Population Economics*, 24(2): 761–777

Smajlagić, D., **Kvarme Jacobsen, K.**, **Myrum, C.**, **Haavik, J.**, **Johansson, S.**, and **Zayats, T.** (2018). Moderating effect of mode of delivery on the genetics of intelligence: Explorative genome-wide analyses in ALSPAC. *Brain and Behavior*, 8(12): e01144

Sobel, M.E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 13: 290–312

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

Stone, V.E., **Baron-Cohen, S.**, and **Knight, R.T.** (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, 10(5): 640–656

Stylianou, M.S. (2007). *Does Executive Functioning Training Improve Mentalising Ability?* Ph.D. thesis, University of Liverpool

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

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

Tager-Flusberg, H. and **Sullivan, K.** (2000). A componential view of theory of mind: evidence from Williams syndrome. *Cognition*, 76(1): 59–90

Talwar, V. and **Lee, K.** (2008). Social and cognitive correlates of children's lying behavior. *Child Development*, 79(4): 866–881

Trentacosta, C.J. and **Izard, C.E.** (2007). Kindergarten children's emotion competence as a predictor of their academic competence in first grade. *Emotion*, 7(1): 77–88

Trentacosta, C.J., **Izard, C.E.**, **Mostow, A.J.**, and **Fine, S.E.** (2006). Children's emotional competence and attentional competence in early elementary school. *School Psychology Quarterly*, 21(2): 148–170

von Salisch, M., **Denham, S.A.**, and **Koch, T.** (2017). Emotion knowledge and attention problems in young children: A cross-lagged panel study on the direction of effects. *Journal of Abnormal Child Psychology*, 45(1): 45–56

Vonk, J. and **Pitzen, J.** (2017). Believing in other minds: Accurate mentalizing does not predict religiosity. *Personality and Individual Differences*, 115: 70–76

Vonk, J., **Zeigler-Hill, V.**, **Ewing, D.**, **Mercer, S.**, and **Noser, A.E.** (2015). Mindreading in the dark: Dark personality features and theory of mind. *Personality and Individual Differences*, 87: 50–54

Warrier, V. and **Baron-Cohen, S.** (2018). Genetic contribution to 'theory of mind' in adolescence. *Scientific Reports*, 8(3465): 1–9

Warrier, V., **Grasby, K.L.**, **Uzefovsky, F.**, **Toro, R.**, **Smith, P.**, **Chakrabarti, B.**, **Khadake, J.**, **Mawbey-Adamson, E.**, **Litterman, N.**, **Hottenga, J.J.**, **Lubke, G.**, **Boomsma, D.I.**, **Martin, N.G.**, **Hatemi, P.K.**, **Medland, S.E.**, **Hinds, D.A.**, **Bourgeron, T.**, and **Baron-Cohen, S.** (2018). Genome-wide meta-analysis of cognitive empathy: Heritability, and correlates with sex, neuropsychiatric conditions and cognition. *Molecular Psychiatry*, 23(6): 1402–1409

Weber, R.A. (2003). 'Learning' with no feedback in a competitive guessing game. *Games and Economic Behavior*, 44(1): 134–144

Wechsler, D. (1991). *WISC-III: Wechsler Intelligence Scale for Children: Manual.* The Psychological Corporation

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

Wilson, T. (2017). Compulsory education and teenage motherhood. *Stirling Economics Discussion Paper 2017-01*

Wimmer, H. and **Perner, J.** (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1): 103–128

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

Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. *Cognition*, 35(1): 41–68

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

Ziv, M., **Solomon, A.**, and **Frye, D.** (2008). Young children's recognition of the intentionality of teaching. *Child Development*, 79(5): 1237–1256

Web Appendix

(Intended for Online Publication)

Web Appendix I

More details regarding related literature

Web Appendix I.1 Details referred to in footnote 3

Examples include social discourse focusing on events related to mental states from stories, training children to talk about others through storytelling, and socio-dramatic play based on a story (Guajardo and Watson, 2002; Lu et al., 2008; Qu et al., 2015). More recently, Charness et al. (2019) find that theory-of-mind can be activated in low SES children by stimulating them to think about the interaction with another child, while Ashraf et al. (forthcoming) improve negotiating skills in adolescents partly by training them to understand others' perspectives. Among adults, Kidd and Castano (2013) find that reading literary fiction improves theory-of-mind.

Web Appendix I.2 Details referred to in footnote 5

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 across-game 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 across-game 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). 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.

Web Appendix I.3 Details referred to in footnote 6

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). 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.

Web Appendix I.4 Details referred to in footnote 7

In related work, Palacios-Huerta and Volij (2009), Agranov et al. (2012) and Le Coq and Sturluson (2012) find that adults respond to information about their opponent's experience, which is also potentially informative about the opponent's sophistication (see Gill and Prowse, 2016, for a detailed discussion). For example, in a one-shot beauty contest, Agranov et al. (2012) find that inexperienced undergraduates choose lower numbers that are closer to Nash equilibrium when they play against graduates who have some experience of the game.

Web Appendix I.5 Details referred to in footnote 10

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. Sutter (2007) popularized the use of the ultimatum game with variation in forgone options to study whether responder age predicts responses to proposer intentions. 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. 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.

Web Appendix I.6 Details referred to in the penultimate paragraph of the introduction

Harbaugh et al. (2001) study violations of the Generalized Axiom of Revealed Preference (GARP) using choices among consumption bundles, finding that seven-year-old children violate GARP more frequently than eleven-year-olds and adults (who violate GARP at about the same rate); the paper finds no statistically significant relationship between math ability and GARP violations. Apesteguia et al. (2018) study the behavior of fourth-grade school children in a belief updating problem based on urn selection, finding that children predicted by teachers to attend the Gymnasium (the higher selective school track in Austria) performed better, but that children were not able to improve their performance when observing others. Barash et al. (2019) study the behavior of subjects from kindergarten to university in a Bayesian updating problem, finding that as children become older, they start to use inference-based reasoning instead of simple heuristic rules.

Web Appendix I.7 Further discussion of Premack and Woodruff (1978) referred to in footnote 11

Premack and Woodruff (1978) study theory-of-mind in chimpanzees. They limit attention to whether chimpanzees impute mental states to others because the accuracy of inferences about mental states is difficult to determine experimentally in chimpanzees. However, they note that the accuracy of theory-of-mind inferences is an important question that deserves further study. In their own words: "We will not be concerned at this time with whether the chimpanzee's theory is a good or complete one, whether he infers every mental state we infer and does so accurately... These questions are not out of order, but they are, for the moment at least, too difficult to deal with experimentally. It will be sufficient now to consider whether or not the chimpanzee imputes mental states to others at all. If we succeed with that claim, we may later seek to determine how accurate and complete his inferences are." (Premack and Woodruff, 1978, p.515)

Web Appendix I.8 Discussion of the evidence from the false photograph test referred to in footnote 11

In the false belief test (Wimmer and Perner, 1983), an object is moved from one location to another while a third party is absent, and the subject is then asked where the third party will search for the object. In the false photograph test (Zaitchik, 1990), the experimenter takes a photograph of an object, the object is moved, and finally the subject is asked where in the photograph the object will appear. The false photograph test replicates the false belief test in terms of representational difficulty, but without requiring any representation of mental states (the false drawing test (Chorman and Baron-Cohen, 1992) serves the same purpose). Autistic children generally fail the false belief test, but pass the false photograph or false drawing test (Leekam and Perner, 1991; Chorman and Baron-Cohen, 1992; Leslie and Thaiss, 1992). According to Chorman and Baron-Cohen (1992), these data show that autistic children's failure on the false belief test is due to a specific theory-of-mind deficit (i.e., an inability to represent the mental states of others), as opposed to a lack of cognitive ability (i.e., a general difficulty with understanding both mental and non-mental representations).

Web Appendix I.9 Details referred to in the third paragraph of Section 2.1

Different research groups emphasize different aspects of the cognitive processes that might help to underpin theory-of-mind ability. Autism researchers emphasize that autistic children lack the specific ability to represent the mental states of others, as opposed to a lack of cognitive ability (in Web Appendix I.8 we discuss evidence for this claim from the false photograph test). According to the "simulation-theory" view of theory-of-mind, people attribute mental states to others by duplicating or mimicking the mental processes of the target in their own minds (Gallese and Goldman, 1998). This view is motivated by neuroscientific evidence that empathetic affect sharing for disgust and pain relies on neural processes that are also activated when experiencing the affect first-hand (Bastiaansen et al., 2009, surveys the evidence for these "mirror systems"; Rütgen et al., 2015a,b, provide recent examples), while Gallese and Goldman (1998) and Gallese et al. (2004) posit that the activity of "mirror neurons" helps to create in the observer a mental state that matches that of the target. According to the "theory-theory" view of theory-of-mind, people attribute mental states to others using commonsense ("folk psychology") theoretical reasoning based on tacitly known causal laws together with information about the target (the description here follows Gallese and Goldman, 1998). Goldman and Sripada (2005) review the proponents of each theory across developmental psychology, philosophy and neuroscience, and also discuss some hybrid models.

Web Appendix I.10 Details referred to in footnote 13

Goleman (1995)'s concept of emotional intelligence incorporates a broad range of soft skills or character skills such as: self-control, zeal, persistence, and self-awareness; and the ability to motivate oneself, to delay gratification, to regulate one's moods, to empathize, to hope, to read another's innermost feelings, and to handle relationships smoothly. As Goleman (1995) notes, his work builds on Salovey and Mayer (1990), and "there is an old-fashioned word for the body of skills that emotional intelligence represents: *character*." Goleman (1998) emphasizes the importance of emotional intelligence for effective leadership, while Goleman and Boyatzis (2017) categorize emotional intelligence skills into four key domains: self-awareness, self-management, social awareness, and relationship management. As noted above, one of the many soft skills included in Goleman (1995)'s concept of emotional intelligence is the ability to read others' feelings. Thus, emotional intelligence incorporates aspects of theory-of-mind, and so emotional intelligence can influence strategic sophistication through the effect of theory-of-mind on strategic sophistication discussed in Section 2.2. More broadly, emotional intelligence will help determine success in strategic interactions that depend on relationship management and leadership skills.

Beldoch (1964) is recognized as the first to use the term "emotional intelligence." However, Beldoch (1964) uses the term only once and does not provide a definition. Beldoch (1964) uses the term "emotional sensitivity" to refer to the ability to identify emotions (from spoken language, art and music) and views emotional sensitivity as an aspect of emotional intelligence. Leuner (1966) also uses the term "emotional intelligence;" the article is in German, but Mayer et al. (2011) provide a useful summary: "In a prefeminist German article on motherhood, the author speculated that women might reject their roles as housewives and mothers due to a lack of emotional intelligence (Leuner, 1966). (We note that Leuner proposed LSD as a treatment for such women!)"

Web Appendix I.11 Quote referred to in footnote 14

"Many unresolved questions about strategic behavior concern the extent to which it reflects players' analyses of their environment as a game, taking its structure and other players' incentives into account. This notion, which we call *strategic sophistication*, is the main difference between the behavioral assumptions of traditional noncooperative and cooperative game theory, which take it to be unlimited, and evolutionary game theory and adaptive learning models, which take it to be nonexistent or severely limited." (Costa-Gomes et al., 2001, p.1193)

Web Appendix I.12 Details referred to in footnote 19

The psychology literature suggests that theory-of-mind is not just a proxy for social skills because theory-of-mind plays a fundamental role in developing and underpinning social skills. In Section 2.3 we define what we mean by social skills and we describe some of the mechanisms by which childhood theory-of-mind helps children to develop these social skills, and below we list a number of quotes from the psychology literature that emphasize how social skills depend on theory-of-mind.

- " "Theory of mind," the ability to make inferences about others' mental states, seems to be a modular cognitive capacity that underlies humans' ability to engage in complex social interaction" (Stone et al., 1998)
- "Understanding others' mental states is a crucial skill that enables the complex social relationships that characterize human societies" and "disengagement of ToM has been linked to the breakdown of positive interpersonal and intergroup relationships" (Kidd and Castano, 2013)
- "Theory of mind, or the ability to make inferences about the mental states of other people, is thought to be the proximate mechanism underlying humans' ability to function in complex, collaborative social networks" (Liddle and Nettle, 2006)
- "In conclusion, ToM is an essential prerequisite for the development and maintenance of close interpersonal relationships" and "in the context of normal adult aging, age-related reductions in ToM mediate a substantial decline in social participation" (Henry et al., 2013)
- "it has been widely proposed that difficulties in inferring others' mental states, or 'theory of mind' (ToM), are responsible for the reduced social skills observed in ASD" (Livingston et al., 2019)

Web Appendix I.13 Details referred to in the first paragraph of Section 6.2.1

The following two quotes from the ALSPAC documentation provide details about how the DANVA2-CF was implemented.[77]

- "The faces subtest comprises 24 photos of child faces, with each face showing one of four emotions: happiness, sadness, anger or fear. The photos are presented to the child for two seconds each and he or she must respond as to whether the person in the photo is happy, sad, angry or afraid."
- "The full 24-item faces subtest was administered. The tester explained the procedure to the child and read out the four words, Happy; Sad; Angry and Fearful and ensured that the child understood their meaning. The child was placed in front of a computer screen on which each face was shown for approximately two seconds. As each picture appeared the child was asked to state which emotion the face was showing. Note, that the pictures were of either high or low intensity (i.e. the emotion displayed was easier to identify or a little harder)."

For each photograph, the subject scores one point if she correctly identifies the corresponding emotion. Thus, the test produces a score out of twenty-four. ALSPAC followed the recommendation in the DANVA2 manual (Nowicki, 2015) by showing each photograph for two seconds.

[77] See "Focus @ 8" (March 2018) in the ALSPAC data dictionary (www.bristol.ac.uk/alspac/researchers/access).

Web Appendix II

Theory-of-mind and altruism, empathy, reciprocity, and openness

In Section 2 we note that data from the existing literature and data from our gift-exchange game show that theory-of-mind is not, or only very modestly, related to pro-social preferences and behaviors such as altruism, empathy or reciprocity. Here we review that evidence, together with the evidence on the relationship between theory-of-mind and openness.[78]

Altruism

First, we use our experimental data to provide novel evidence on the relationship between theory-of-mind and altruism. In particular, we study the behavior of the allocators in our gift-exchange game. From the perspective of the allocator, our gift-exchange game is similar to a mini-dictator game. We find that subjects with higher theory-of-mind ability (compared to peers of the same age) are slightly less likely to offer the generous split of the ten super-tokens, although the effect is not statistically significant at the ten-percent level.[79]

Second, to study further the relationship between theory-of-mind and altruism, we survey existing evidence on the relationship between theory-of-mind and sharing in the dictator game. Among three- to five-year-olds, Rochat et al. (2009) find no relationship when controlling for age, while Cowell et al. (2015) find a strong negative relationship. Among three- to eight-year-olds, Kogut et al. (2016) find no relationship, unless the recipient is identified as needy. Among three- to eleven-year olds, Liu et al. (2016) find no relationship. Among five- to twelve-year olds, Cowell et al. (2017) find a small positive relationship. Among university students, Edele et al. (2013) and Artinger et al. (2014) find no relationship, while Lang et al. (2018) find a positive relationship. Among adults, Gross and Wronski (forthcoming) find no relationship, unless the recipient and donor share the same race or the recipient is identified as deserving.

Empathy

We first note that the term "cognitive empathy" is synonymous with theory-of-mind, while "affective empathy" is defined as "feeling an appropriate emotion triggered by seeing/learning of another's emotion" (Baron-Cohen and Wheelwright, 2004). Based on a meta-analysis of sixty-six studies with over eleven thousand subjects, Murphy and Lilienfeld (2019) find a modest positive correlation of 0.12 between theory-of-mind and affective empathy.

Reciprocity

First, we remind the reader that we find no statistically significant relationship between theory-of-mind and reciprocity using the data from our gift-exchange game, in the sense that theory-of-mind does not predict whether receivers respond to the number of tokens given to them when

[78] We exclude studies based on clinically atypical samples (e.g., autistic children).

[79] In Treatment A (B), the generous split is 5/5 (2/8). In each case, we run a logistic regression of an indicator for whether the allocator offered the generous split on age, age-standardized theory-of-mind ability and age-standardized cognitive ability, with and without controls for the demographics that we collected (gender and school). In all cases, the estimated average marginal effect is negative but not statistically significant at the ten-percent level.

deciding whether to give a super-token back to the allocator (see Table 5, and the associated discussion in the penultimate paragraph of Section 5).[80]

Second, we survey the limited evidence from the existing literature on the relationship between theory-of-mind and reciprocity. This existing literature is small and inconclusive. Among university students, Ridinger and McBride (2017) find that theory-of-mind correlates positively with the probability of cooperating after observing first-mover cooperation in a sequential prisoners' dilemma game; however, after controlling for beliefs the correlation is not statistically significant, and so Ridinger and McBride (2017) conclude that the effect of theory-of-mind operates through beliefs rather than preferences. Among preschoolers, and using an ultimatum game followed by an unexpected dictator game, Schug et al. (2016) find that theory-of-mind correlates positively with reciprocity.

Openness

We found eight correlations between theory-of-mind and openness reported in the literature (Realo et al., 2003; Nettle and Liddle, 2008; Ferguson and Austin, 2010; Engel et al., 2014; Ewing et al., 2016; all of these studies use adult subjects). Of those eight correlations, five were not statistically significant, while three were positive and statistically significant.

[80] Instead, we find that theory-of-mind predicts whether receivers condition their behavior on the allocator's forgone alternative, and so theory-of-mind helps children to direct reciprocity appropriately according to the intentions of the allocator.

Web Appendix III

Persistence of cognitive skills

In Section 2 we note that the existing literature finds that theory-of-mind and cognitive ability show a substantial degree of persistence. Here we review the evidence.

Persistence of theory-of-mind

In Table A.1 we survey existing literature on the persistence of theory-of-mind. The table notes explain how we selected the studies.

Study	Sample size	Initial age (years)	Final age (years)	Persistence (correlation)
Warrier et al. (2018)	259	37.0	39.0	0.47
Fernández-Abascal et al. (2013)	358	34.2	35.2	0.63
Khorashad et al. (2015)	44	25.8	26.8	0.74
Białecka-Pikul et al. (forthcoming)	96	16.2	17.2	0.63
Białecka-Pikul et al. (forthcoming)	105	13.2	14.2	0.74
Lecce et al. (2017)	113	9.3	10.3	0.44
Banerjee et al. (2011)	138	9.0	11.0	0.29
Peterson and Wellman (2019)	37	7.3	8.5	0.74
Banerjee et al. (2011)	72	6.0	8.0	0.45
von Salisch et al. (2017)	265	5.1	6.2	0.41

Notes: We restrict attention to studies that measure theory-of-mind of the same individuals at different ages using the same test over a period of one year or longer (when a study reports persistence over a period of time and over intervals of that period, we report results over the entire period). To align with the age of the youngest participants in our experiment, we further restrict attention to findings for individuals aged five years and older. Finally, we exclude studies based on clinically atypical samples (e.g., autistic children).

Table A.1: Persistence of theory-of-mind.

Persistence of cognitive ability

Researchers have been studying the persistence of cognitive ability in childhood for several decades. Bloom (1964, Chapter 3) surveys the literature and concludes that cognitive ability is variable in early childhood but becomes stable at around age eight (see Jensen, 1969, and Jensen, 1980, for further discussion). To illustrate the persistence of cognitive ability over the life course, we focus on studies that measure cognitive ability in both childhood and adulthood.

In particular, a handful of recent studies have tracked cognitive ability from childhood through to middle and old age. Drawing on samples from four British cohort surveys, Gale et al. (2012) find a correlation between cognitive ability at age eleven and later in life (age fifty-one to seventy-nine) of between 0.43 and 0.53. Based on the Lothian Birth Cohorts of 1921 and 1936, Gow et al. (2011) find correlations between cognitive ability at age eleven and cognitive ability at age seventy, seventy-nine and eighty-seven of 0.67, 0.66 and 0.51, respectively.

Web Appendix IV

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.

Web Appendix IV.1 Instructions for Phase 1

Hello everyone! My name is [**experimenter**] and I am a teacher[81] 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.

[81] In Spanish we used the word "profesor", which refers to both a teacher and a 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 3.2.1**]

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 3.2.2**.]

Web Appendix IV.2 Instructions for Phase 2

{*As described in Section 3.3.2, 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 A.1**]. Please do not touch anything until I tell you to do so.

Figure A.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.[82]

[82] 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)[83]

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 A.2**].

Figure A.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.

[83]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.[84] So, it is going to choose the number of tokens at random in each game.[85] 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.

[84] In Spanish we used the colloquial expression "le dan todos los números igual."
[85] 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 A.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 A.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.**]

Web Appendix IV.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 A.4**]. These super-tokens are yellow and receiving one super-token is the same as receiving four tokens.

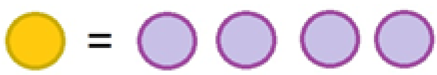


Figure A.4: Equivalence between tokens and super-tokens.

{*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 A.5**], while those of you who are lions continue to see the spinner.[86]

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

Figure A.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 A.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.**]

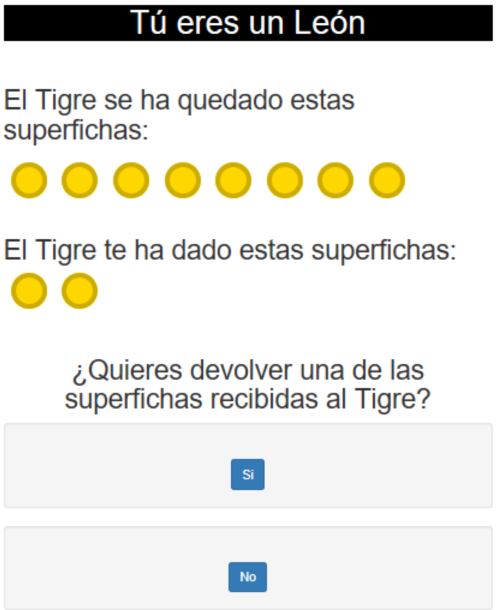


Figure A.6: Receiver (Lion) decision screen.

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 partner changed in each game. It is likely that your 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.**]

Web Appendix V

Further details about the experiment

Web Appendix V.1 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.

Web Appendix V.2 Software & configuration of 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.

Web Appendix V.3 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.

Web Appendix V.4 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 a risk-neutral subject, the best-response to the computer's uniform randomization is to choose 5. 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.

Web Appendix V.5 Verbal cognitive ability

In our experiment, we measured cognitive ability using only Raven's Progressive Matrices test, a leading measure of analytic or fluid intelligence (see Section 3.2.2). Here, we explain why verbal cognitive ability is unlikely to be an important omitted driver of behavior in our experiment that studies strategic sophistication. First, the Raven test correlates strongly with general cognitive ability g (see footnote 26). Second, in our conceptual framework (see Section 2.2) we describe in detail the cognitive reasoning that underpins strategic sophistication in our simple games: this reasoning depends heavily on theory-of-mind and fluid intelligence, with little additional role for verbal cognitive ability. Third, we carefully designed our experiment to enhance comprehension (by: keeping the structure of the games simple, reading out and repeating instructions, using hypothetical scenarios, and providing detailed explanations supported by projections of decision screens onto a screen/wall), further reducing any residual role of verbal cognitive ability.

Web Appendix VI

Nash equilibrium in the 1-6 Token Request game

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

Table A.2: Nash equilibrium mixing distribution.

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.

Web Appendix VII

Robustness to measurement error in the gift-exchange game

Web Appendix VII.1 Introduction

In this appendix we provide evidence that our results in Table 4 (Section 5) on theory-of-mind and cognitive ability are robust to correcting the estimates for bias due to measurement error. Gillen et al. (2019) note that measurement error in an experimentally measured control variable can bias the estimate of an effect of interest. To correct for such bias due to measurement error, Gillen et al. (2019) propose using repeated measures of the same experimentally measured control variable.[87] In our experiment, subjects completed our measures of cognitive ability and theory-of-mind (i.e., the Raven test and the IMT) a single time, and so we do not have access to repeated measures. However, Gillen et al. (2019) use simulations to estimate the extent of the bias arising from different degrees of measurement error in an experimentally measured control variable when the variable is measured a single time. The literature provides us with information about the extent of measurement error in our measures of cognitive ability and theory-of-mind (from subjects who have taken the Raven test or IMT multiple times). Using this information about the extent of measurement error, we run Gillen et al. (2019)-style simulations to correct our estimates for bias due to this measurement error, and we find that our results in Table 4 on the effects of theory-of-mind and cognitive ability are robust to this correction.

Web Appendix VII.2 provides an overview of the bias-correction procedure, Web Appendix VII.3 describes the details of the bias-correction procedure, and Web Appendix VII.4 reports the bias-corrected estimates.

Web Appendix VII.2 Overview of the bias-correction procedure

First, we construct a data generation process (DGP) that explicitly includes noise in the measurement of cognitive ability and theory-of-mind, using information from the literature about the extent of this noise. Second, we estimate the parameters of this DGP by: (i) simulating many samples from the DGP (for a given vector of parameters); and (ii) then choosing the parameter vector that gives the best match between the average of the simulated samples (for the chosen parameter vector) and our experimental sample. Finally, we use these estimated parameters of the DGP to correct our estimates in Table 4 for the bias introduced by the measurement error. We conduct this exercise three times, first including error only in the measurement of cognitive ability, second including error only in the measurement of theory-of-mind, and third including error in the measurement of both cognitive ability and theory-of-mind.

[87] Gillen et al. (2019) write: "At the heart of our approach is the combination of duplicate elicitations (usually two) of behavioral proxies and methods from the econometrics literature, particularly the instrumental-variables approach to errors in variables."

Web Appendix VII.3 Details of the bias-correction procedure

The bias-correction procedure starts with the following specification of the DGP for receiver behavior in the gift-exchange game:

$$\begin{aligned} Y &= \mathbf{1}[\beta_0 + \beta_1 \text{TreatmentA} + \beta_2 \text{Age} + \beta_3 T^* + \beta_4 C^* + \\ &\quad \beta_5(\text{TreatmentA} \times \text{Age}) + \beta_6(\text{TreatmentA} \times T^*) + \beta_7(\text{TreatmentA} \times C^*) + \epsilon > 0], \end{aligned}$$

where Y takes value 1 if the receiver gives one super-token back to the allocator, TreatmentA is an indicator for being in Treatment A, Age is age in years, T^* denotes theory-of-mind, C^* denotes cognitive ability, and ϵ denotes an error with a standard logistic distribution. T^* and C^* follow a bivariate normal distribution with means of zero, variances of f_T and f_C, and a correlation of ρ. Finally, the variables $\{\text{TreatmentA}, \text{Age}, \epsilon\}$ are mutually independent and independent of $\{T^*, C^*\}$, and all variables are independent over subjects.

We now introduce measurement error. In particular, we assume that measured theory-of mind $T = T^* + u$, where the noise $u \sim N(0, 1 - f_T)$. Similarly, we assume that measured cognitive ability $C = C^* + v$, where the noise $v \sim N(0, 1 - f_C)$. Furthermore $\{u, v\}$ are mutually independent, independent of $\{\text{TreatmentA}, \text{Age}, \epsilon, T^*, C^*\}$, and independent over subjects. Together, our assumptions ensure that T and C have means of zero and variances of one at every age; thus we match our analysis in the main text, where we work with age-standardized measures of theory-of-mind and cognitive ability.

Note that $1 - f_T$ represents the proportion of the variance in measured theory-of-mind that is due to measurement error. We conclude from the literature that measurement error accounts for approximately 10% of the variance of measured theory-of-mind when using the IMT, and so we set $1 - f_T = 0.1$. Similarly, we conclude from the literature that measurement error accounts for approximately 20% of the variance of measured cognitive ability when using the Raven test, and so we set $1 - f_C = 0.2$.[88]

We use indirect inference to estimate the parameters of this DGP. In particular, we estimate the parameters of the DGP by: (i) simulating many samples from the DGP (for a given vector of parameters); and (ii) then choosing the parameter vector that gives the best match between the average of the simulated samples (for the chosen parameter vector) and our experimental sample. Each time we simulate samples from the DGP for given parameters, we simulate 100 samples of size $N = 207$ (to match the N from the experimental sample) using the values of age and treatment in the experimental sample (the values of T^* and C^* are generated by the bivariate distribution described above). When we estimate ρ, we match the average correlation between measured cognitive ability and measured theory-of-mind in the simulated samples with the correlation between the same variables in our experimental sample. When we estimate $\boldsymbol{\beta}$,

[88] We use the fact that the proportion of the variance of scores on a skill test that is due to measurement error is one minus the test-retest correlation between scores from two attempts on the test (assuming, as do Gillen et al., 2019, that measurement error is orthogonal across the two attempts and the variance of the measurement error is constant across attempts). The Raven manual (Raven et al., 2000a) reports that the test-retest correlation of the Standard Progressive Matrices version of the Raven test is 0.9 when the test is taken twice one week apart and 0.8 when the test is taken twice one month apart (retest reliabilities for the Coloured Progressive Matrices are similar; Raven et al., 1998); thus, we make the conservative assumption that $1 - f_C = 0.2$. Stylianou (2007) provides the only evidence that we are aware of about the retest reliability of the IMT: from Stylianou (2007)'s data, the test-retest correlation of the IMT is 0.94 for children aged ten to thirteen and 0.95 for adolescents aged thirteen to sixteen, and so we make the conservative assumption that $1 - f_T = 0.1$.

we match summary measures of choice behavior in our experimental sample with the average of the same summary measures in the simulated samples (using a quadratic distance metric of dissimilarity).[89]

Finally, we use these estimated parameters of the DGP to correct our estimates in Table 4 for the bias introduced by the measurement error. In particular, we calculate average marginal effects for each sample that we simulate from the DGP given the estimated $\boldsymbol{\beta}$ and ρ, and we then average these marginal effects over the simulated samples.

We conduct this exercise three times, first including error only in the measurement of cognitive ability (i.e., setting $1 - f_T = 0$), second including error only in the measurement of theory-of-mind (i.e., setting $1 - f_C = 0$), and third including error in the measurement of both cognitive ability and theory-of-mind. Tables A.3 to A.5 in Web Appendix VII.4 report the bias-corrected estimates.

[89] We follow Bruins et al. (2018) by using in each sample the coefficient estimates from a linear probability model to summarize choice behavior from a non-linear DGP with a binary outcome variable, and by using Bruins et al. (2018)'s generalized indirect inference approach to solve the optimization problem, which combines a gradient-based optimizer with a procedure to smooth the objective function. The explanatory variables in the linear probability model are treatment, age, measured theory-of-mind, measured cognitive ability, and the interaction of the treatment with age, measured theory-of-mind and measured cognitive ability.

Web Appendix VII.4 Tables of results

	(1)	(2)	(3)
Treatment A	0.095	0.082	-0.123*
	(0.068)	(0.060)	(0.063)
Age		0.013	-0.002
		(0.021)	(0.017)
Theory-of-mind		0.009	-0.026
(T.o.M.)		(0.026)	(0.031)
Cognitive Ability		0.002	0.003
(Cogn. Ab.)		(0.032)	(0.039)
Treatment A. × Age			0.051
			(0.037)
Treatment A. × T.o.M.			0.136**
			(0.065)
Treatment A. × Cogn. Ab.			0.005
			(0.073)
Subjects	207	207	207

Notes: See the notes to Table 4. Column (3) reports bias-corrected marginal effects following the procedure outlined in Web Appendix VII.2 and described in Web Appendix VII.3 (setting $1 - f_T = 0$). Column (2) reports bias-corrected marginal effects from the same procedure, but where interaction terms are omitted from the DGP and estimation. Column (1) repeats the first column from Table 4. Standard errors in Columns (2) and (3) were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the experimental sample for each replication, and then calculated the standard deviation of the bias-corrected marginal effects across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.3: Effect of treatment on probability that receiver gives one super-token to allocator: Robustness (measurement error in cognitive ability).

	(1)	(2)	(3)
Treatment A	0.095	0.082	-0.126**
	(0.068)	(0.060)	(0.063)
Age		0.013	0.001
		(0.021)	(0.017)
Theory-of-mind		0.009	-0.026
(T.o.M.)		(0.026)	(0.033)
Cognitive Ability		0.001	0.000
(Cogn. Ab.)		(0.029)	(0.034)
Treatment A. × Age			0.054
			(0.037)
Treatment A. × T.o.M.			0.151**
			(0.071)
Treatment A. × Cogn. Ab.			-0.002
			(0.061)
Subjects	207	207	207

Notes: See the notes to Table 4. Column (3) reports bias-corrected marginal effects following the procedure outlined in Web Appendix VII.2 and described in Web Appendix VII.3 (setting $1 - f_C = 0$). Column (2) reports bias-corrected marginal effects from the same procedure, but where interaction terms are omitted from the DGP and estimation. Column (1) repeats the first column from Table 4. Standard errors in Columns (2) and (3) were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the experimental sample for each replication, and then calculated the standard deviation of the bias-corrected marginal effects across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.4: Effect of treatment on probability that receiver gives one super-token to allocator: Robustness (measurement error in theory-of-mind).

	(1)	(2)	(3)
Treatment A	0.095	0.082	-0.125**
	(0.068)	(0.060)	(0.064)
Age		0.013	0.001
		(0.021)	(0.017)
Theory-of-mind		0.009	-0.026
(T.o.M.)		(0.028)	(0.034)
Cognitive Ability		0.002	0.002
(Cogn. Ab.)		(0.032)	(0.040)
Treatment A. × Age			0.052
			(0.037)
Treatment A. × T.o.M.			0.154**
			(0.072)
Treatment A. × Cogn. Ab.			-0.004
			(0.074)
Subjects	207	207	207

Notes: See the notes to Table 4. Column (3) reports bias-corrected marginal effects following the procedure outlined in Web Appendix VII.2 and described in Web Appendix VII.3. Column (2) reports bias-corrected marginal effects from the same procedure, but where interaction terms are omitted from the DGP and estimation. Column (1) repeats the first column from Table 4. Standard errors in Columns (2) and (3) were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the experimental sample for each replication, and then calculated the standard deviation of the bias-corrected marginal effects across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.5: Effect of treatment on probability that receiver gives one super-token to allocator: Robustness (measurement error in cognitive ability and theory-of-mind).

Web Appendix VIII

Description of controls used in analysis of ALSPAC data

For presentational purposes, we divide the controls into four groups: controls for characteristics of the participant's birth; controls for characteristics of the participant's home environment; controls for parental characteristics; and controls for school characteristics.

The controls for characteristics of the participant's birth are as follows:

- An indicator for being born female (from birth records).
- ALSPAC's preferred birthweight (combines obstetric records, birth records and ALSPAC's own measurements).
- ALSPAC's best estimate of gestational age at birth (from obstetric records).
- Maternal age at birth (calculated by ALSPAC from obstetric records).
- An indicator for the participant resulting from a multiple pregnancy (from obstetric records).
- An indicator for membership of the ALSPAC core sample (i.e., the sample recruited around the time of the participant's birth).

We construct controls for the characteristics of the participant's home environment using information collected from questionnaires that the mothers of ALSPAC participants were asked to complete during the prenatal period and in the early years of the ALSPAC participant's life, up to and around the time that theory-of-mind and cognitive ability were measured at mean age eight years and eight months (see Section 6.2).[90,91]

The controls for characteristics of the home environment are as follows:

- Mother's marital status.
- Family income per week.
- Amount spent on food per week
- Amount spent on childcare per week.
- Mother's home ownership status.
- Mother living in public housing.
- Mother's employment status.
- Mother's partner's employment status.
- Number of children living in the mother's household.
- Parity (i.e., the number of previous pregnancies that resulted in either a livebirth or stillbirth), winsorized at the 99th percentile.
- Mother having ever breast fed the participant.
- Number of cigarettes the mother smokes per day.

[90]Specifically, we use information from prenatal questionnaires called "Your Environment", "Having a Baby" and "Your Pregnancy" and from further questionnaires that were administered to mothers of the ALSPAC participants when the participants reached the followings ages (questionnaire names in parentheses): 8 weeks (Me and My Baby); 6 months (My Son/Daughter); 8 months (Looking after the Baby); 18 months (Boy/Girl Toddler Questionnaire); 21 months (Caring for a Toddler); 30 months (My Study Son/Daughter); 33 months (Your Health, Events and Feelings); 42 months (My Study Son/Daughter's Health and Behaviour); 47 months (Mother's New Questionnaire); 61 months (Study Mother's Questionnaire); 73 months (Mother's Lifestyle); 85 months (Mother and Home); 97 months (Mother and Family); and 110 months (Mother of a 9 Year Old).

[91]When a characteristic was measured multiple times, we sought to use all available measures, and we averaged across available measures at the participant level (before averaging, we de-meaned questionnaire-by-questionnaire to adjust for participant age when the questionnaire was administered).

- Number of cigarettes the mother's partner smokes per day.
- Number of books owned by the participant.
- ALSPAC's Neighbourhood Quality Index.

The controls for parental characteristics are as follows:[92]

- A set of controls based on indicators for the mother's highest educational qualification (the qualification categories are: no qualifications; GCSE grade D–G, CSE or vocational qualification; GCSE grade A–C or O-level; A-level; and university degree).
- An equivalent set of controls for the mother's partner's highest educational qualification.
- An indicator for the mother being white.
- An indicator for the mother's partner being white.
- A set of controls based on indicators for the mother's occupational group (the occupational groups are the nine major groups in the Standard Occupational Classification 2000 (ONS, 2000)).
- An equivalent set of controls for the mother's partners' occupational group.

We construct controls for the characteristics of the participant's school using information collected from questionnaires that head teachers of schools of participants were asked to complete.[93]

The controls for school characteristics are as follows:

- Percentage of children at the participant's school receiving free school meals, winsorized at the 1st and 99th percentiles.
- Percentage of children at the participant's school with special educational needs, winsorized at the 1st and 99th percentiles.
- Number of children at the participant's school, winsorized at the 1st and 99th percentiles.
- An indicator for the participant attending a publicly-funded school.

In Tables A.42 to A.45 in Web Appendix XIII we further include a control for locus of control. ALSPAC measured locus of control at age eight (in the "Focus @ 8" interview session) using a 12-item version of Nowicki and Strickland (1973)'s Children's Nowicki Strickland Internal External scale (CNSIE). Nowicki et al. (2018) describe the development of this 12-item version and use the ALSPAC data; following Nowicki et al. (2018) we use the variable in the ALSPAC dataset that measures the CNSIE score for children who answered all 12 questions.

In all cases, when a control is missing for a participant we include an indicator for the control being missing, and when a measure of a characteristic includes interval information we use interval regression to assign a value to each interval.

[92] We construct controls for parental characteristics using information from questionnaires that we also used for the characteristics of the home environment. For each parental characteristic, we use the first available observation of that characteristic for each parent.

[93] We use information from the "About Your School" questionnaire, administered when a participant was in Year 3, and the "Questionnaire for the Head Teacher", administered when the participant was in Year 6 (children were aged seven at the beginning of Year 3 and ten at the beginning of Year 6). When a school is in one of the four Local Education Authorities (LEAs) in the ALSPAC data, we have access to anonymized school identifiers, which allow us to create an average measure of that school's characteristics. We assign each participant to the first publicly-funded school in one of the four LEAs that we observe them attending; if the participant never attended such a school, she is assigned to a privately-funded school in one of the four LEAs or to a school outside these LEAs.

Web Appendix IX

Robustness to measurement error in analysis of ALSPAC data

Web Appendix IX.1 Introduction

In Web Appendix VII we describe how we run Gillen et al. (2019)-style simulations to correct estimates from our gift-exchange game for bias due to measurement error in cognitive ability and theory-of-mind. In this appendix we use the same approach to correct our estimates in Tables 6 to 9 (Section 6.3) for bias due to measurement error. Using information from the literature about the extent of measurement error in our measures of cognitive ability and theory-of-mind (from subjects who have taken the WISC-III or DANVA2-CF multiple times), we find that our estimates of the effects of theory-of-mind and cognitive ability on adult outcomes in Tables 6 to 9 are broadly stable after correcting for bias due to measurement error, with a modest loss of precision in some cases.

Web Appendix IX.2 provides an overview of the bias-correction procedure, Web Appendix IX.3 describes the details of the bias-correction procedure, and Web Appendix IX.4 reports the bias-corrected estimates.

Web Appendix IX.2 Overview of the bias-correction procedure

First, we construct a data generation process (DGP) that explicitly includes noise in the measurement of cognitive ability and theory-of-mind, using information from the literature about the extent of this noise. Second, we use indirect inference to estimate the parameters of this DGP by: (i) simulating many samples from the DGP (for a given vector of parameters); and (ii) then choosing the parameter vector that gives the best match between the average of the simulated samples (for the chosen parameter vector) and the estimation sample from the ALSPAC data. Finally, we use these estimated parameters of the DGP to correct our estimates in Tables 6 to 9 for the bias introduced by the measurement error. For each adult outcome that we study, we conduct this exercise three times, first including error only in the measurement of cognitive ability, second including error only in the measurement of theory-of-mind, and third including error in the measurement of both cognitive ability and theory-of-mind.

Web Appendix IX.3 Details of the bias-correction procedure

The bias-correction procedure is the same as the procedure that we used to correct our gift-exchange game estimates for bias due to measurement error (see Web Appendix VII.3), apart from the following differences.

First, given that we estimated linear models in Tables 6 to 9, we specify a linear DGP for each of the adult outcomes that we study.[94] Furthermore, since we are no longer in the setting of the gift-exchange game experiment, the DGP does not include variables that involve Age or Treatment A.

[94] We include a normally distributed error with a variance parameter in the DGP, and so our indirect inference procedure estimates this variance parameter.

Second, when using the estimation sample from the ALSPAC data in the indirect inference procedure, we residualize the outcome, measured theory-of-mind and measured cognitive ability with respect to the controls (see Web Appendix VIII for a description of the controls); differencing in this way removes the need to explicitly model the controls when performing the bias correction while still accounting for the correlation of the controls with the outcome, theory-of-mind and cognitive ability.[95]

Third, we conclude from the literature that measurement error accounts for approximately 20% of the variance of measured theory-of-mind when using the DANVA2-CF, and so here we set $1 - f_T = 0.2$. Similarly, we conclude from the literature that measurement error accounts for approximately 10% of the variance of measured cognitive ability when using the WISC-III, and so here we set $1 - f_C = 0.1$.[96]

Tables A.6 to A.17 in Web Appendix IX.4 report the bias-corrected estimates.

[95] As in Web Appendix VII, the indirect inference procedure uses a linear regression model (note that the residualized outcome is not binary, and the same is true of the outcome from the linear DGP).

[96] As in Web Appendix VII.3, we use the fact that the proportion of the variance of scores on a skill test that is due to measurement error is one minus the test-retest correlation between scores from two attempts on the test (assuming, as do Gillen et al., 2019, that measurement error is orthogonal across the two attempts and the variance of the measurement error is constant across attempts). The WISC-III manual (Wechsler, 1991) reports that the test-retest correlation of the WISC-III Full Scale IQ score is 0.94 for school children; thus, we make the conservative assumption that $1 - f_C = 0.1$. The DANVA2 manual (Nowicki, 2015) evidences that the test-retest correlation of the DANVA2-CF is 0.79 for school children; thus, we assume that $1 - f_T = 0.2$.

Web Appendix IX.4 Tables of results

	Number of close friends	Ten or more close friends
Theory-of-mind	0.188**	0.020***
	(0.081)	(0.007)
Cognitive ability	0.329***	0.025***
	(0.102)	(0.009)
Participants	3,887	3,887
Mean age	24.49	24.49
Mean of dependent variable	6.76	0.16

Notes: See the notes to Table 6. This table reports bias-corrected estimates following the procedure outlined in Web Appendix IX.2 and described in Web Appendix IX.3 (setting $1 - f_T = 0$). Standard errors were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the estimation sample (from the ALSPAC data) for each replication, and then calculated the standard deviation of the bias-corrected estimates across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.6: Close friends: Robustness (measurement error in cognitive ability).

	In education at 17	In education at 20	A-level qualification
Theory-of-mind	0.014**	0.017*	0.013*
	(0.006)	(0.009)	(0.008)
Cognitive ability	0.026***	0.109***	0.106***
	(0.007)	(0.010)	(0.009)
Participants	4,091	4,332	4,332
Mean age	17.76	20.95	20.95
Mean of dependent variable	0.88	0.55	0.73

Notes: See the notes to Table 7. This table reports bias-corrected estimates following the procedure outlined in Web Appendix IX.2 and described in Web Appendix IX.3 (setting $1 - f_T = 0$). Standard errors were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the estimation sample (from the ALSPAC data) for each replication, and then calculated the standard deviation of the bias-corrected estimates across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.7: Educational participation and attainment:
Robustness (measurement error in cognitive ability).

	Parent	Number of children
Theory-of-mind	-0.012**	-0.023**
	(0.006)	(0.009)
Cognitive ability	-0.028***	-0.040***
	(0.007)	(0.011)
Participants	4,371	4,371
Mean age	25.76	25.76
Mean of dependent variable	0.12	0.18

Notes: See the notes to Table 8. This table reports bias-corrected estimates following the procedure outlined in Web Appendix IX.2 and described in Web Appendix IX.3 (setting $1 - f_T = 0$). Standard errors were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the estimation sample (from the ALSPAC data) for each replication, and then calculated the standard deviation of the bias-corrected estimates across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.8: Fertility: Robustness (measurement error in cognitive ability).

	Employed	Log hourly wage	Large workplace
Theory-of-mind	0.005	-0.008	0.018
	(0.008)	(0.008)	(0.011)
Cognitive ability	0.006	0.027**	0.031**
	(0.010)	(0.011)	(0.014)
Participants	3,884	2,999	3,129
Mean age	24.49	23.89	23.89
Mean of dependent variable	0.81	2.05	0.60

Notes: See the notes to Table 9. This table reports bias-corrected estimates following the procedure outlined in Web Appendix IX.2 and described in Web Appendix IX.3 (setting $1 - f_T = 0$). Standard errors were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the estimation sample (from the ALSPAC data) for each replication, and then calculated the standard deviation of the bias-corrected estimates across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.9: Employment, wages, and workplace size:
Robustness (measurement error in cognitive ability).

	Number of close friends	Ten or more close friends
Theory-of-mind	0.253**	0.026***
	(0.103)	(0.009)
Cognitive ability	0.287***	0.022***
	(0.089)	(0.008)
Participants	3,887	3,887
Mean age	24.49	24.49
Mean of dependent variable	6.76	0.16

Notes: See the notes to Table 6. This table reports bias-corrected estimates following the procedure outlined in Web Appendix IX.2 and described in Web Appendix IX.3 (setting $1 - f_C = 0$). Standard errors were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the estimation sample (from the ALSPAC data) for each replication, and then calculated the standard deviation of the bias-corrected estimates across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.10: Close friends: Robustness (measurement error in theory-of-mind).

	In education at 17	In education at 20	A-level qualification
Theory-of-mind	0.018**	0.024**	0.019*
	(0.008)	(0.011)	(0.010)
Cognitive ability	0.023***	0.096***	0.093***
	(0.007)	(0.009)	(0.008)
Participants	4,091	4,332	4,332
Mean age	17.76	20.95	20.95
Mean of dependent variable	0.88	0.55	0.73

Notes: See the notes to Table 7. This table reports bias-corrected estimates following the procedure outlined in Web Appendix IX.2 and described in Web Appendix IX.3 (setting $1 - f_C = 0$). Standard errors were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the estimation sample (from the ALSPAC data) for each replication, and then calculated the standard deviation of the bias-corrected estimates across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.11: Educational participation and attainment:
Robustness (measurement error in theory-of-mind).

	Parent	Number of children
Theory-of-mind	-0.016**	-0.030**
	(0.007)	(0.012)
Cognitive ability	-0.025***	-0.036***
	(0.006)	(0.010)
Participants	4,371	4,371
Mean age	25.76	25.76
Mean of dependent variable	0.12	0.18

Notes: See the notes to Table 8. This table reports bias-corrected estimates following the procedure outlined in Web Appendix IX.2 and described in Web Appendix IX.3 (setting $1 - f_C = 0$). Standard errors were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the estimation sample (from the ALSPAC data) for each replication, and then calculated the standard deviation of the bias-corrected estimates across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.12: Fertility: Robustness (measurement error in theory-of-mind).

	Employed	Log hourly wage	Large workplace
Theory-of-mind	0.007	-0.011	0.023
	(0.010)	(0.011)	(0.014)
Cognitive ability	0.005	0.024**	0.028**
	(0.009)	(0.009)	(0.012)
Participants	3,884	2,999	3,129
Mean age	24.49	23.89	23.89
Mean of dependent variable	0.81	2.05	0.60

Notes: See the notes to Table 9. This table reports bias-corrected estimates following the procedure outlined in Web Appendix IX.2 and described in Web Appendix IX.3 (setting $1 - f_C = 0$). Standard errors were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the estimation sample (from the ALSPAC data) for each replication, and then calculated the standard deviation of the bias-corrected estimates across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.13: Employment, wages, and workplace size:
Robustness (measurement error in theory-of-mind).

	Number of close friends	Ten or more close friends
Theory-of-mind	0.243**	0.025***
	(0.104)	(0.009)
Cognitive ability	0.327***	0.024***
	(0.102)	(0.009)
Participants	3,887	3,887
Mean age	24.49	24.49
Mean of dependent variable	6.76	0.16

Notes: See the notes to Table 6. This table reports bias-corrected estimates following the procedure outlined in Web Appendix IX.2 and described in Web Appendix IX.3. Standard errors were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the estimation sample (from the ALSPAC data) for each replication, and then calculated the standard deviation of the bias-corrected estimates across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.14: Close friends:
Robustness (measurement error in cognitive ability and theory-of-mind).

	In education at 17	In education at 20	A-level qualification
Theory-of-mind	0.017**	0.021*	0.016
	(0.008)	(0.011)	(0.010)
Cognitive ability	0.025***	0.109***	0.106***
	(0.007)	(0.010)	(0.009)
Participants	4,091	4,332	4,332
Mean age	17.76	20.95	20.95
Mean of dependent variable	0.88	0.55	0.73

Notes: See the notes to Table 7. This table reports bias-corrected estimates following the procedure outlined in Web Appendix IX.2 and described in Web Appendix IX.3. Standard errors were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the estimation sample (from the ALSPAC data) for each replication, and then calculated the standard deviation of the bias-corrected estimates across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.15: Educational participation and attainment:
Robustness (measurement error in cognitive ability and theory-of-mind).

	Parent	Number of children
Theory-of-mind	-0.015**	-0.029**
	(0.007)	(0.012)
Cognitive ability	-0.028***	-0.040***
	(0.007)	(0.011)
Participants	4,371	4,371
Mean age	25.76	25.76
Mean of dependent variable	0.12	0.18

Notes: See the notes to Table 8. This table reports bias-corrected estimates following the procedure outlined in Web Appendix IX.2 and described in Web Appendix IX.3. Standard errors were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the estimation sample (from the ALSPAC data) for each replication, and then calculated the standard deviation of the bias-corrected estimates across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.16: Fertility: Robustness (measurement error in cognitive ability and theory-of-mind).

	Employed	Log hourly wage	Large workplace
Theory-of-mind	0.006	-0.011	0.022
	(0.010)	(0.011)	(0.014)
Cognitive ability	0.006	0.027**	0.031**
	(0.010)	(0.011)	(0.014)
Participants	3,884	2,999	3,129
Mean age	24.49	23.89	23.89
Mean of dependent variable	0.81	2.05	0.60

Notes: See the notes to Table 9. This table reports bias-corrected estimates following the procedure outlined in Web Appendix IX.2 and described in Web Appendix IX.3. Standard errors were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the estimation sample (from the ALSPAC data) for each replication, and then calculated the standard deviation of the bias-corrected estimates across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.17: Employment, wages, and workplace size:
Robustness (measurement error in cognitive ability and theory-of-mind).

Web Appendix X

Robustness to measurement error in the Baseline games

Web Appendix X.1 Introduction

In Web Appendix VII we describe how we run Gillen et al. (2019)-style simulations to provide evidence that the results from our gift-exchange game are robust to correcting for bias due to measurement error in cognitive ability and theory-of-mind. In this appendix, we use the same approach to correct our estimates in Table 2 (Section 4.2) for bias due to measurement error. We also correct our estimates in the associated Table A.27 (Web Appendix XIII), which replicates Table 2 using binary rather than continuous measures of cognitive skills. Using information from the literature about the extent of measurement error in our measures of cognitive ability and theory-of-mind (from subjects who have taken the Raven test or the IMT multiple times; see Web Appendix VII.3), we find that our estimates of the effects of theory-of-mind and cognitive ability in Table 2 and Table A.27 are robust to correcting for bias due to measurement error, with a modest change in precision in some cases. We note in particular that in the specification with binary measures of cognitive skills (which reduces the influence of outliers on the distribution of skills), the effect of theory-of-mind on level-1 behavior remains statistically significant at the one-percent level after correcting for bias.

Web Appendix X.2 provides an overview of the bias-correction procedure, Web Appendix X.3 describes the details of the bias-correction procedure, and Web Appendix X.4 reports the bias-corrected estimates.

Web Appendix X.2 Overview of the bias-correction procedure

First, we construct a data generation process (DGP) that explicitly includes noise in the measurement of cognitive ability and theory-of-mind, using information from the literature about the extent of this noise. Starting with Table 2, we use indirect inference to estimate the parameters of this DGP by: (i) simulating many samples from the DGP (for a given vector of parameters); and (ii) then choosing the parameter vector that gives the best match between the average of the simulated samples (for the chosen parameter vector) and our experimental sample. We then use these estimated parameters of the DGP to correct our estimates in Table 2 for the bias introduced by the measurement error. We conduct this exercise three times, first including error only in the measurement of theory-of-mind, second including error only in the measurement of cognitive ability, and third including error in the measurement of both cognitive ability and theory-of-mind. Finally, we repeat the procedure to correct the estimates in Table A.27.

Web Appendix X.3 Details of the bias-correction procedure

The bias-correction procedure is the same as the procedure that we used to correct our gift-exchange game estimates for bias due to measurement error (see Web Appendix VII.3), apart from the following differences.

First, the outcomes here are fractions, and so we replace the DGP for a binary outcome variable with the following DGP for the fraction of rounds Y_j that an individual chooses number $j \in \{1, 2, ..., 6\}$ over the five repetitions $r \in \{1, 2, ..., 5\}$ of the Baseline game:

$$\begin{aligned} Y_j &= \frac{1}{5}\sum_{r=1}^{5} \mathbf{1}[u_{j,r} > \max_{k \in \{1,2,...,6\} \setminus j} u_{k,r}], \\ u_{j,r} &= \beta_0^j + \beta_1^j \text{Age} + \beta_2^j T^* + \beta_3^j C^* + \varepsilon_{j,r}, \end{aligned}$$

where $\varepsilon_{j,r}$ denote independently-distributed errors from the standard extreme value type-1 distribution (here we use the extreme value distribution because Table 2 and Table A.27 are based on fractional logits in a multinomial setting; Age, T^* and C^* are defined in Web Appendix VII.3). Second, each time we simulate samples from the DGP for given parameters, we simulate samples of size $N = 730$ (to match the N from the experimental sample).[97]

Tables A.18 to A.23 in Web Appendix X.4 report the bias-corrected estimates.

[97]As described in footnote 89, in Web Appendix VII.3 we followed Bruins et al. (2018) by using in each sample the coefficient estimates from a linear probability model to summarize choice behavior from a non-linear DGP; here we extend that approach by using the coefficient estimates from linear regression models of the fractional outcomes Y_j to summarize behavior in each sample. In relation to Table 2, the explanatory variables in each linear regression model are age, measured theory-of-mind and measured cognitive ability. In relation to Table A.27, the explanatory variables in each linear regression model are age, an indicator for above-median measured theory-of-mind, and an indicator for above-median measured cognitive ability.

Web Appendix X.4 Tables of results

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***
	(0.003)	(0.003)	(0.004)	(0.005)	(0.006)	(0.001)
Theory-of-mind	0.000	-0.011	-0.014	0.006	0.018*	0.000
	(0.009)	(0.006)	(0.009)	(0.009)	(0.011)	(0.004)
Cognitive ability	-0.017*	-0.003	0.008	0.011	0.018*	-0.017***
	(0.009)	(0.006)	(0.008)	(0.008)	(0.010)	(0.005)
Subjects	730	730	730	730	730	730

Notes: See the notes to Table 2. This table reports bias-corrected marginal effects following the procedure outlined in Web Appendix X.2 and described in Web Appendix X.3 (setting $1 - f_C = 0$). Standard errors were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the experimental sample for each replication, and then calculated the standard deviation of the bias-corrected marginal effects across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.18: Probability of choices in the Baseline games:
Robustness (measurement error in theory-of-mind).

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	-0.021***	0.000	0.001	0.026***	0.023***	-0.030***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)	(0.002)
Above median theory-of-mind	0.002	-0.021	-0.033*	-0.005	0.060***	-0.003
	(0.022)	(0.015)	(0.018)	(0.017)	(0.019)	(0.023)
Above median cognitive ability	-0.043**	0.006	-0.006	0.036**	0.030*	-0.023
	(0.018)	(0.013)	(0.015)	(0.014)	(0.016)	(0.019)
Subjects	730	730	730	730	730	730

Notes: See the notes to Table A.27. This table reports bias-corrected marginal effects following the procedure outlined in Web Appendix X.2 and described in Web Appendix X.3 (setting $1 - f_C = 0$). Standard errors were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the experimental sample for each replication, and then calculated the standard deviation of the bias-corrected marginal effects across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.19: Probability of choices in the Baseline games (binary measures of cognitive skills): Robustness (measurement error in theory-of-mind).

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	-0.021***	0.001	0.002	0.027***	0.021***	-0.029***
	(0.003)	(0.003)	(0.004)	(0.005)	(0.006)	(0.001)
Theory-of-mind	0.001	-0.010	-0.012	0.005	0.015	0.001
	(0.009)	(0.007)	(0.009)	(0.009)	(0.010)	(0.005)
Cognitive ability	-0.019**	-0.004	0.007	0.012	0.021**	-0.017***
	(0.010)	(0.006)	(0.008)	(0.008)	(0.011)	(0.005)
Subjects	730	730	730	730	730	730

Notes: See the notes to Table 2. This table reports bias-corrected marginal effects following the procedure outlined in Web Appendix X.2 and described in Web Appendix X.3 (setting $1 - f_T = 0$). Standard errors were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the experimental sample for each replication, and then calculated the standard deviation of the bias-corrected marginal effects across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.20: Probability of choices in the Baseline games: Robustness (measurement error in cognitive ability).

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	-0.020***	-0.001	0.000	0.028***	0.022***	-0.029***
	(0.004)	(0.003)	(0.005)	(0.005)	(0.006)	(0.002)
Above median theory-of-mind	0.006	-0.020	-0.028	-0.007	0.054***	-0.005
	(0.021)	(0.015)	(0.018)	(0.016)	(0.018)	(0.023)
Above median cognitive ability	-0.045**	0.003	-0.008	0.039**	0.036*	-0.025
	(0.020)	(0.015)	(0.018)	(0.016)	(0.019)	(0.019)
Subjects	730	730	730	730	730	730

Notes: See the notes to Table A.27. This table reports bias-corrected marginal effects following the procedure outlined in Web Appendix X.2 and described in Web Appendix X.3 (setting $1 - f_T = 0$). Standard errors were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the experimental sample for each replication, and then calculated the standard deviation of the bias-corrected marginal effects across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.21: Probability of choices in the Baseline games (binary measures of cognitive skills): Robustness (measurement error in cognitive ability).

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	-0.021***	0.001	0.002	0.026***	0.021***	-0.029***
	(0.003)	(0.004)	(0.004)	(0.005)	(0.006)	(0.001)
Theory-of-mind	0.001	-0.010	-0.014	0.006	0.017	0.000
	(0.009)	(0.007)	(0.010)	(0.009)	(0.011)	(0.004)
Cognitive ability	-0.019*	-0.004	0.007	0.012	0.021*	-0.017***
	(0.010)	(0.006)	(0.008)	(0.009)	(0.011)	(0.005)
Subjects	730	730	730	730	730	730

Notes: See the notes to Table 2. This table reports bias-corrected marginal effects following the procedure outlined in Web Appendix X.2 and described in Web Appendix X.3. Standard errors were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the experimental sample for each replication, and then calculated the standard deviation of the bias-corrected marginal effects across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.22: Probability of choices in the Baseline games:
Robustness (measurement error in cognitive ability and theory-of-mind).

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	-0.020***	-0.000	0.000	0.028***	0.022***	-0.029***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)	(0.002)
Above median theory-of-mind	0.006	-0.021	-0.031*	-0.006	0.056***	-0.004
	(0.022)	(0.016)	(0.019)	(0.016)	(0.019)	(0.023)
Above median cognitive ability	-0.046**	0.003	-0.007	0.040**	0.035*	-0.024
	(0.021)	(0.015)	(0.018)	(0.016)	(0.019)	(0.019)
Subjects	730	730	730	730	730	730

Notes: See the notes to Table A.27. This table reports bias-corrected marginal effects following the procedure outlined in Web Appendix X.2 and described in Web Appendix X.3. Standard errors were obtained by bootstrapping; in particular, we repeated the bias-correction procedure 5,000 times, using a different bootstrap sample from the experimental sample for each replication, and then calculated the standard deviation of the bias-corrected marginal effects across the 5,000 bootstrap replications. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.23: Probability of choices in the Baseline games (binary measures of cognitive skills): Robustness (measurement error in cognitive ability and theory-of-mind).

Web Appendix XI

Relationship between measures of theory-of-mind

Web Appendix XI.1 Discussion

To begin, we summarize our discussion in Section 6.2.1 of the DANVA2-CF that was used by the ALSPAC birth-cohort study to measure theory-of-mind:

- Psychologists measure theory-of-mind by the ability to correctly infer or understand the mental states of others, which can include others' beliefs, desires, intentions and emotions.
- ALSPAC measured theory-of-mind ability at age eight using the DANVA2-CF.
- Similarly to the Reading the Mind in the Eyes Test (RMET), the DANVA2-CF measures theory-of-mind by the ability to recognize emotions in photographs of faces.
- Psychologists often measure theory-of-mind using tests based on recognizing emotions in photographs. For example, Kidd and Castano (2013), a heavily cited paper on theory-of-mind in the journal *Science*, uses both the RMET and DANVA2 to measure theory-of-mind.
- Theory-of-mind tests based on recognizing emotions in photographs such as the RMET and the DANVA2-CF are "social-perceptual" tests of theory-of-mind.

In our experiments, we measured theory-of-mind using the IMT:

- We describe the IMT in Section 3.2.1.
- The IMT is a "social-cognitive" test of theory-of-mind. In a social-cognitive test of theory-of-mind such as the IMT, subjects infer others' mental states from information provided in stories or scenarios, rather than inferring them from perceptual cues as in social-perceptual tests of theory-of-mind such as the RMET and the DANVA2-CF.
- According to Ewing et al. (2016)'s succinct description, social-perceptual tests of theory-of-mind measure "the ability to identify the mental states of others using perceptual cues such as facial expressions, gestures, or bodily movements," while social-cognitive tests of theory-of-mind measure "the ability to reason about the mental states of other individuals without relying on perceptual cues."
- In our experimental game-theoretic setting, a social-cognitive test of theory-of-mind such as the IMT was most appropriate because experimental subjects in games like the ones we study cannot use perceptual cues to infer the mental states of those they interact with, but must instead reason about their mental states without using perceptual cues.

- Alongside social-perceptual tests of theory-of-mind such as the RMET and the DANVA2-CF, social-cognitive tests are a popular way to measure theory-of-mind. For example, the false-belief test (Wimmer and Perner, 1983) is widely used in psychology to measure theory-of-mind.

More on the relationship between social-perceptual and social-cognitive tests of theory-of-mind:

- Osterhaus et al. (2016) find that items from the IMT and the RMET load heavily on the same factor for children, and so "were found to be indicators of the same factor."
- Chung et al. (2014)'s meta-study finds that autistic subjects show similarly-sized deficits in social-perceptual and social-cognitive tests of theory-of-mind.
- Tager-Flusberg and Sullivan (2000) introduced the distinction between the social-perceptual and social-cognitive components of theory-of-mind. According to Tager-Flusberg and Sullivan (2000), the social-perceptual and social-cognitive components of theory-of-mind are closely related. As they say: "our capacity to make inferences about other people almost always involves both components." Furthermore, they argue that: (i) the social-cognitive component builds on social-perceptual knowledge; and (ii) the social-perceptual component itself entails some cognitive processing, and so tests of social-perceptual theory-of-mind partly capture the social-cognitive component.

Web Appendix XI.2 Empirical evidence

In Web Appendix XI.1 above, we discuss the relationship between the IMT and tests of theory-of-mind that are based on recognizing emotions in photographs such as the RMET and the DANVA2-CF. Existing empirical evidence provides further information about this relationship. Specifically, we focus on existing evidence about the relationship between the IMT and the RMET, since we are not aware of any study that reports a correlation between scores on the IMT and the DANVA2-CF.

First, we report correlations from the literature between scores on the IMT and the RMET for adults. In the context of relevant retest reliabilities, these correlations are sizable: the correlations between scores on the IMT and the RMET range from 0.34 to 0.46 (all with $p < 0.001$), while the retest reliability of the RMET is around 0.7 (see footnote 69 in Section 6.2.1 for more details including references).

Second, Osterhaus et al. (2016)'s factor analysis provides empirical evidence about the relationship between the IMT and the RMET for children:[98]

- In particular, Osterhaus et al. (2016) find that items from the IMT and the RMET load heavily on the same factor for children, and so "were found to be indicators of the same factor." Osterhaus et al. (2016) further argue that this factor that they identify is evidence of a single underlying ability, which they call "social reasoning" and which they suggest is "the core of AToM [advanced theory-of-mind]."
- In more detail, Osterhaus et al. (2016) conduct two separate studies, each with over 400 children from grades 2–4. Osterhaus et al. (2016) conclude that the studies provide strong and consistent support for a three-factor model in which items from the IMT and the RMET load heavily on the first factor (and load little on the other two factors). Quantitatively, across the two studies, the items from the IMT load on the first factor with coefficients averaging 0.48 (all coefficients have $p < 0.05$), and the items from the RMET load on the same factor with coefficients averaging 0.49 (again all coefficients have $p < 0.05$). To help interpret these figures, note that all items and the factor are scaled to have variances of one.

Finally, we explain why we rely on existing empirical evidence. As described in Section 3.2.1, we selected a version of the IMT that was designed specifically for children. Furthermore, theory-of-mind develops throughout childhood and adolescence (e.g., Stone et al., 1998). For these reasons, scores on the specific version of the IMT that we used in our experiments, but collected from a college population, would not provide useful evidence. Furthermore, collecting data from school children during the COVID pandemic is not feasible (e.g., due to IRB constraints).

[98] Osterhaus et al. (2016) do not report the raw correlation between scores on the IMT and the RMET, and we are not aware of any study that reports this correlation for children.

Web Appendix XII

Dynamic mechanisms

Web Appendix XII.1 Overview

Exploiting dynamic models that build on our conceptual framework in Section 2, in this appendix we quantify the importance of dynamic mechanisms that mediate the effect of childhood theory-of-mind on adult outcomes. We conduct two mediation analyses; in each mediation analysis we include one important theoretically-motivated dynamic mechanism whose empirical importance we quantify with the help of evidence from the existing literature. The conceptual framework in Section 2 discusses other direct and indirect mechanisms that also help to explain the total effect of childhood theory-of-mind on adult outcomes.

In our first mediation analysis, we quantify the importance of attention in the classroom as a dynamic mechanism that mediates the effect of theory-of-mind at age eight on academic attainment. This mediation analysis draws on the literature for estimates of the effect of childhood theory-of-mind on subsequent attention in the classroom and for estimates of the effect of attention in the classroom on later academic achievement. As described in detail in Web Appendix XII.2, our mediation analysis suggests that the dynamic effect of childhood theory-of-mind on subsequent attention in the classroom mediates 55% of the total effect of theory-of-mind at age eight on academic attainment that we report in Table 7 in Section 6.3.3.

In our second mediation analysis, we quantify the importance of educational participation (i.e., years of education) as a dynamic mechanism that mediates the effect of theory-of-mind at age eight on fertility in early adulthood. This mediation analysis uses our results from Table 7 to estimate the effect of childhood theory-of-mind on years of education, and the analysis draws on the literature that exploits a change in the compulsory schooling age in the UK in the 1970s for estimates of the effect of an extra year of education on fertility in early adulthood. As described in detail in Web Appendix XII.3, our mediation analysis suggests that the dynamic effect of childhood theory-of-mind on educational participation mediates 27% of the total effect of theory-of-mind at age eight on fertility in early adulthood that we report in Table 8 in Section 6.3.4.

Web Appendix XII.2 Mediation analysis 1: Relationship between childhood theory-of-mind and academic attainment

In our first mediation analysis, we quantify the importance of attention in the classroom as a dynamic mechanism that mediates the effect of theory-of-mind at age eight on academic attainment. The mediation analysis exploits a dynamic model in which childhood theory-of-mind increases subsequent attention in the classroom, which in turn facilitates learning and later academic achievement. This dynamic model builds on our conceptual framework in Section 2 in which childhood theory-of-mind improves attention in the classroom by: (i) increasing engagement and reducing boredom because theory-of-mind allows the child to better understand the pedagogical motives and intentions of teachers; and (ii) improving the child's relationship with teachers through the effect of theory-of-mind on social skills. Relatedly, von Salisch et al. (2017)

show that a poor understanding of the emotions of social peers adds cognitive load to children's working memory and can cause social problems that further absorb attention.

The mediation analysis uses the product-method approach (Sobel, 1982). First, we multiply the effect of childhood theory-of-mind on subsequent attention in the classroom by the effect of attention in the classroom on later academic achievement (with all quantities in standard deviations). Dividing this product by the total effect of theory-of-mind at age eight on academic attainment gives the proportion of the total effect that is mediated by the dynamic effect of childhood-theory of mind on subsequent attention in the classroom.

We draw on the literature for estimates of the effect of childhood theory-of-mind on subsequent attention in the classroom and for estimates of the effect of attention in the classroom on later academic achievement. The total effect of theory-of-mind at age eight on academic attainment comes from the third column of Table 7 in Section 6.3.3. Multiplying the effect of a one-standard-deviation increase in childhood theory-of-mind on subsequent attention in the classroom (0.156 of a standard deviation)[99] by the effect of a one-standard-deviation increase in attention in the classroom on later academic achievement (0.126 of a standard deviation),[100] and then dividing by the total effect of a one-standard-deviation increase in theory-of-mind at age eight on academic attainment (0.036 of a standard deviation),[101] gives 0.545. That is, we estimate that the dynamic effect of childhood theory-of-mind on subsequent attention in the classroom mediates 54.5% of the total effect of theory-of-mind at eight on academic attainment.

[99] We found three papers that provide estimates of the longitudinal relationship between theory-of-mind and attention in the classroom at a later age (Trentacosta et al., 2006; Trentacosta and Izard, 2007; von Salisch et al., 2017). Trentacosta et al. (2006) find that a one-standard-deviation increase in theory-of-mind in the fall is associated with a 0.11 standard-deviation increase in teacher-rated attention in the following spring (controlling for verbal ability and teacher-rated attention in the fall; mean age of 7.4 years in the fall). Using path-analytic modeling, Trentacosta and Izard (2007) find that a one-standard-deviation increase in theory-of-mind in kindergarten is associated with a 0.14 standard-deviation increase in teacher-rated attention in first grade (controlling for attention (not teacher-rated) and verbal ability in kindergarten); we note that the authors also explore a better-fitting alternative model which does not include the path from theory-of-mind to attention. von Salisch et al. (2017) find that a one-standard-deviation increase in theory-of-mind at age five is associated with a 0.22 standard-deviation increase in teacher-rated attention one year later (controlling for language ability and teacher-rated attention at age five), while attention at age five does not predict theory-of-mind at age six. The effect size of 0.156 comes from averaging these three estimates of the effect of theory-of-mind on subsequent attention in the classroom.

[100] For estimates of the effect of attention in the classroom on later academic achievement, we use Duncan et al. (2007)'s comprehensive study based on three longitudinal datasets (which produce consistent results). We also use estimates from Breslau et al. (2009), who extend Duncan et al. (2007)'s analysis to consider academic achievement at later ages. Using the ECLSK, NICHD, and MLEPS datasets, Duncan et al. (2007) regress several measures of academic achievement in grades 3 and 5 on teacher-rated attention around the time of school entry, with controls for socio-emotional skills and prior cognitive or academic achievement, and with controls for non-teacher-rated attention (NICHD only) or hyperactivity (MLEPS only) around the time of school entry. The results show that a one-standard-deviation increase in teacher-rated attention around the time of school entry is associated with an increase in later academic achievement (averaged over achievement measures) of 0.10, 0.14, and 0.17 of a standard deviation in, respectively, the ECLSK, NICHD, and MLEPS. Breslau et al. (2009) regress academic achievement at age 16 on teacher-rated attention at age six, with controls for cognitive ability at age six and family background. Again averaging over achievement measures, the results show that a one-standard-deviation increase in teacher-rated attention at age six is associated with an increase in academic achievement at age sixteen of 0.11 of a standard deviation (the findings are robust to adding controls for other behavioral problems). The effect size of 0.126 comes from averaging these four estimates of the effect of attention in the classroom on later academic achievement.

[101] We derive this effect size from the numbers reported in the third column of Table 7 in Section 6.3.3

Web Appendix XII.3 Mediation analysis 2: Relationship between childhood theory-of-mind and fertility in early adulthood

In our second mediation analysis, we quantify the importance of educational participation (i.e., years of education) as a dynamic mechanism that mediates the effect of theory-of-mind at age eight on fertility in early adulthood. The mediation analysis exploits a dynamic model in which childhood theory-of-mind increases the number of years spent in education, which in turn reduces fertility. This dynamic model builds on our conceptual framework in Section 2 in which: (i) childhood theory-of-mind increases educational participation and attainment (e.g., via the effect of theory-of-mind on attention studied in Web Appendix XII.2); and (ii) educational participation reduces fertility in early adulthood (e.g., through delayed marriage or a higher opportunity cost of child-rearing).

The mediation analysis uses the product-method approach (Sobel, 1982). First, we multiply the effect of theory-of-mind at age eight on years of education by the effect of an extra year of education on the probability of being a parent by early adulthood. Dividing this product by the total effect of theory-of-mind at age eight on the probability of being a parent by early adulthood gives the proportion of the total effect that is mediated by the dynamic effect of childhood theory-of-mind on educational participation.

The effect of theory-of-mind at age eight on years of education comes from the first two columns of Table 7 in Section 6.3.3. We draw on the literature that exploits a change in the compulsory schooling age in the UK in the 1970s for estimates of the effect of an extra year of education on the probability of being a parent by early adulthood. The total effect of theory-of-mind at age eight on the probability of being a parent by early adulthood comes from the first column of Table 8 in Section 6.3.4. Multiplying the effect of a one-standard-deviation increase in theory-of-mind at age eight on years of education (0.086 years)[102] by the effect of an extra year of education on the probability of being a parent by early adulthood (−0.118 of a standard deviation),[103] and then dividing by the total effect of a one-standard-deviation increase in theory-of-mind at age eight on the probability of being a parent by early adulthood (−0.037 of a standard deviation),[104] gives 0.274. That is, we estimate that the dynamic effect of childhood theory-of-mind on educational participation mediates 27.4% of the total effect of theory-of-mind at age eight on fertility in early adulthood.

[102] We derive this effect size from the numbers reported in the first two columns of Table 7 in Section 6.3.3, assuming that individuals not in education at 17 left education at 16 (schooling was compulsory to age 16 for the individuals in ALSPAC), individuals in education at 17 but not at 20 left education at 18 (high school beyond the compulsory age of 16 lasted for two more years until the age of 18), and individuals in education at 20 left education at 21 (undergraduate degrees in the UK generally last for three years).

[103] For estimates of the effect of years of education on fertility in early adulthood, we restrict attention to UK data to match our data from ALSPAC, and we further restrict attention to studies that identify the size of the relationship using the change in compulsory schooling age in the 1970s and include results for individuals age 20 or older; we found two studies that provide such estimates (Silles, 2011, and Wilson, 2017; using the same policy variation, Braakmann, 2011, Fort et al., 2016, and Geruso and Royer, 2018, estimate the effect of education on completed fertility). Silles (2011) and Wilson (2017) find that an extra year of education decreases the probability of being a parent by age 20 by, respectively, 0.133 and 0.103 of a standard deviation. The effect size of −0.118 comes from averaging these two estimates of the effect of years of education on fertility in early adulthood.

[104] We derive this effect size from the numbers reported in the first column of Table 8 in Section 6.3.4.

Web Appendix XII.4 How theory-of-mind and cognitive ability interact with each other in the mediation analyses

Web Appendix XII.4.1 Mediation analysis 1

In Web Appendix XII.2, we quantified the importance of attention in the classroom as a dynamic mechanism that mediates the effect of childhood theory-of-mind on academic attainment. Here, we explore whether theory-of-mind and cognitive ability act as complements or substitutes in this dynamic process. In summary, we conclude that: (i) theory-of-mind and cognitive ability are substitutes in the production of academic attainment; (ii) but this substitutability does not operate through the attention mediator.

Using our ALSPAC sample, we extend our analysis of academic attainment in the third column of Table 7 in Section 6.3.3 by adding the interaction between theory-of-mind and cognitive ability. We estimate a negative coefficient on the additional interaction, and so theory-of-mind and cognitive ability are substitutes, in the sense that additional cognitive ability (theory-of-mind) reduces the positive effect of theory-of-mind (cognitive ability) on academic attainment.[105] This substitutability between theory-of-mind and cognitive ability in the production of academic attainment does not appear to operate through the attention mediator: the literature that studies the effect of childhood theory-of-mind on subsequent attention in the classroom consistently finds that theory-of-mind predicts attention while cognitive ability does not.[106]

Web Appendix XII.4.2 Mediation analysis 2

In Web Appendix XII.3, we quantified the importance of years of education as a dynamic mechanism that mediates the effect of childhood theory-of-mind on fertility in early adulthood. Here, we explore whether theory-of-mind and cognitive ability act as complements or substitutes in this dynamic process. In summary, we find that: (i) theory-of-mind and cognitive ability are substitutes in their overall effect on fertility and in their effect on the education mediator; (ii) and this substitutability is stronger for fertility than for the education mediator.

Using our ALSPAC sample, we extend our analysis of fertility in early adulthood in the first column of Table 8 in Section 6.3.4 by adding the interaction between theory-of-mind and cognitive ability. We estimate a positive coefficient on the additional interaction, and so theory-of-mind and cognitive ability are substitutes, in the sense that additional cognitive ability (theory-of-mind) reduces in absolute terms the negative effect of theory-of-mind (cognitive ability) on fertility in early adulthood.[107]

Again using our ALSPAC sample, we further estimate how years of education depend on

[105]The coefficient on the interaction term is -0.021 ($p = 0.002$), while the coefficients on theory-of-mind and cognitive ability are, respectively, 0.018 ($p = 0.025$) and 0.090 ($p < 0.001$).

[106]Footnote 99 summarizes the literature that studies the longitudinal relationship between childhood theory-of-mind and subsequent attention in the classroom. Trentacosta et al. (2006) and von Salisch et al. (2017) both find that the effect of theory-of-mind on subsequent attention (reported in footnote 99) is strongly statistically significant, while the effect of cognitive ability on subsequent attention is smaller and not statistically significant (these studies use verbal ability or language ability as the measure of cognitive ability). Trentacosta and Izard (2007), the third study summarized in footnote 99, is not relevant here because it does not study the simultaneous effects of theory-of-mind and cognitive ability on subsequent attention.

[107]The coefficient on the interaction term is 0.011 ($p = 0.035$), while the coefficients on theory-of-mind and cognitive ability are, respectively, -0.012 ($p = 0.038$) and -0.022 ($p < 0.001$).

childhood theory-of-mind, childhood cognitive ability, and the interaction between theory-of-mind and cognitive ability. We estimate a negative coefficient on the interaction term, and so theory-of-mind and cognitive ability are substitutes, in the sense that additional cognitive ability (theory-of-mind) reduces the positive effect of theory-of-mind (cognitive ability) on years of education.[108]

To recap, we have found that childhood theory-of-mind and cognitive ability are substitutes in their overall effect on fertility and in their effect on the education mediator. Finally, we show that this substitutability is stronger for fertility than for the education mediator: dividing the relevant interaction coefficient by the standard deviation of the corresponding outcome variable gives (in absolute terms) 0.034 for fertility and 0.025 for years of education.

[108]The coefficient on the interaction term is -0.047 ($p = 0.079$), while the coefficients on theory-of-mind and cognitive ability are, respectively, 0.088 ($p = 0.004$) and 0.323 ($p < 0.001$). In more detail, we extend our analysis of educational participation in the first two columns of Table 7 in Section 6.3.3 by adding the interaction between theory-of-mind and cognitive ability. Based on the estimation results, we calculate the effects of theory-of-mind, cognitive ability, and their interaction on years of education (making the same assumptions as described in footnote 102 and obtaining standard errors using the delta method).

Web Appendix XII.5 How theory-of-mind and cognitive ability complement each other in the development of cognitive skills

In this appendix, we consider how theory-of-mind and cognitive ability complement each other in the development of cognitive skills during childhood. In particular, we draw on estimates from the literature to explore the dynamic interplay between theory-of-mind and cognitive ability.

First, we find that both theory-of-mind and cognitive ability display strong own-skill dynamic dependence. A one-standard-deviation increase in theory-of-mind at time t is associated with an increase in theory-of-mind at time $t+1$ of 0.47 of a standard deviation, controlling for cognitive ability at time t. A one-standard-deviation increase in cognitive ability at time t is associated with an increase in cognitive ability at time $t+1$ of 0.56 of a standard deviation, controlling for theory-of-mind at time t.[109]

Second, we find positive cross-skill dynamic dependencies in the development of theory-of-mind and cognitive ability that are weaker than the own-skill dynamic dependencies described above. A one-standard-deviation increase in cognitive ability at time t is associated with an increase in theory-of-mind at time $t+1$ of 0.19 of a standard deviation, controlling for theory-of-mind at time t. A one-standard-deviation increase in theory-of-mind at time t is associated with an increase in cognitive ability at time $t+1$ of 0.13 of a standard deviation, controlling for cognitive ability at time t.[110]

To provide context for these findings, we note that theory-of-mind increased substantially over the relevant periods in all of the studies described in footnotes 109 and 110.

[109] In Table A.1 in Web Appendix III, we surveyed existing literature on the persistence of theory-of-mind. These studies provide four estimates of the association between theory-of-mind at time t and theory-of-mind at time $t+1$ in childhood, controlling for cognitive ability at time t; 0.47 comes from averaging estimates of 0.45 and 0.36 from Lecce et al. (2017), 0.35 from von Salisch et al. (2017) and 0.74 from Peterson and Wellman (2019). The studies further provide two estimates of the association between cognitive ability at time t and cognitive ability at time $t+1$ in childhood, controlling for theory-of-mind at time t; 0.56 comes from averaging estimates of 0.76 and 0.36 from Lecce et al. (2017). All of these studies use verbal ability or language ability as the measure of cognitive ability. Peterson and Wellman (2019)'s sample includes deaf and autism spectrum disorder children, but Peterson and Wellman (2019) conclude that the results are invariant to their inclusion.

[110] See footnote 109. The studies in Table A.1 in Web Appendix III provide four estimates of the association between cognitive ability at time t and theory-of-mind at time $t+1$ in childhood, controlling for theory-of-mind at time t; 0.19 comes from averaging estimates of 0.24 and 0.22 from Lecce et al. (2017), 0.17 from von Salisch et al. (2017) and 0.13 from Peterson and Wellman (2019). The studies further provide two estimates of the association between theory-of-mind at time t and cognitive ability at time $t+1$ in childhood, controlling for cognitive ability at time t; 0.13 comes from averaging estimates of 0.10 and 0.15 from Lecce et al. (2017).

Web Appendix XIII

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	0.96	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 A.24: 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 A.25: Distribution of scores in the Raven's Progressive Matrices test.

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 A.26: Order of games in phase 2.

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	0.003	0.027***	0.021***	-0.030***
	(0.007)	(0.003)	(0.004)	(0.004)	(0.004)	(0.006)
Above median theory-of-mind	-0.000	-0.017	-0.026	-0.003	0.053***	-0.011
	(0.022)	(0.014)	(0.017)	(0.017)	(0.017)	(0.026)
Above median cognitive ability	-0.045**	0.005	-0.006	0.033**	0.029*	-0.016
	(0.019)	(0.012)	(0.015)	(0.014)	(0.017)	(0.021)
Subjects	730	730	730	730	730	730

Notes: See the notes to Table 2. Here above median theory-of-mind (cognitive ability) is an indicator taking value 1 when a subject's theory-of-mind ability (cognitive ability) is above the median for subjects of their age. We allocate the small number of subjects exactly at the median to the above-median category.

Table A.27: Probability of choices in the Baseline games: Binary measures of cognitive skills.

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***	-0.030***
	(0.007)	(0.003)	(0.004)	(0.004)	(0.004)	(0.006)
Theory-of-mind	-0.000	-0.011	-0.013*	0.007	0.017*	0.001
	(0.009)	(0.007)	(0.008)	(0.008)	(0.010)	(0.011)
Cognitive ability	-0.016	-0.003	0.006	0.012	0.018**	-0.017
	(0.010)	(0.006)	(0.007)	(0.008)	(0.009)	(0.010)
Intercept	0.212***	0.129***	0.160***	0.102***	0.123***	0.273***
	(0.027)	(0.013)	(0.014)	(0.013)	(0.013)	(0.027)
Subjects	730	730	730	730	730	730

Notes: See the notes to Table 2. Here we report OLS regressions with the same dependent and independent variables, where we recode age five (the youngest age) as 0, age six as 1, and so on.

Table A.28: Probability of choices in the Baseline games: Robustness (linear probability model).

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	-0.024***	0.001	0.004	0.027***	0.020***	-0.030***
	(0.006)	(0.003)	(0.004)	(0.004)	(0.003)	(0.006)
Theory-of-mind	0.004	-0.012*	-0.013*	0.006	0.017*	-0.001
	(0.008)	(0.007)	(0.007)	(0.009)	(0.010)	(0.011)
Cognitive ability	-0.013	-0.003	0.005	0.010	0.016*	-0.014
	(0.009)	(0.006)	(0.008)	(0.008)	(0.010)	(0.010)
Subjects	729	729	729	729	729	729

Notes: See the notes to Table 2. Here we add controls for subjects' gender and school. One subject is excluded because they did not report their gender.

Table A.29: Probability of choices in the Baseline games: Robustness (demographic controls).

	Baseline games		Computer game	
	(1)	(2)	(3)	(4)
Age	0.020***	0.019***	0.017*	0.016*
	(0.004)	(0.004)	(0.009)	(0.008)
Theory-of-mind	0.017*	0.004	0.013	0.026
(T.o.M.)	(0.010)	(0.017)	(0.015)	(0.027)
Cognitive ability	0.019**	-0.012	0.013	-0.089***
(Cogn. Ab.)	(0.009)	(0.019)	(0.017)	(0.032)
T.o.M.× Age		0.004		-0.004
		(0.005)		(0.006)
Cogn. Ab. × Age		0.009*		0.032***
		(0.005)		(0.008)
Subjects	730	730	730	730

Notes: Columns 1 and 2 each report average marginal effects from a fractional logit: the dependent variable is the fraction of times each subject best-responded to the empirical distribution of subjects' choices in the five Baseline games (that is, the number of times each subject best-responded divided by five); we pool all choices from the five Baseline games to calculate a single empirical distribution (to which the best-response is 5). Columns 3 and 4 each report average marginal effects from a logit: the dependent variable is an indicator taking value 1 if a subject best-responded to the computer's known uniform randomization in the Computer game (again the best-response is 5). In all columns the independent variables are age, age-standardized theory-of-mind ability and age-standardized cognitive ability (as explained in Section 3.2.3, we standardize test scores within age group). Heteroskedasticity-robust standard errors, clustered at the session level, are shown in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.30: Probability of best-responding.

	Baseline games		Computer game	
	(1)	(2)	(3)	(4)
Age	0.020***	0.020***	0.017*	0.017**
	(0.004)	(0.003)	(0.009)	(0.008)
Theory-of-mind	0.017*	0.002	0.012	0.023
(T.o.M.)	(0.010)	(0.014)	(0.014)	(0.023)
Cognitive ability	0.018**	-0.012	0.013	-0.083***
(Cogn. Ab.)	(0.009)	(0.016)	(0.016)	(0.031)
T.o.M.× Age		0.005		-0.004
		(0.005)		(0.006)
Cogn. Ab. × Age		0.010**		0.031***
		(0.005)		(0.008)
Intercept	0.123***	0.123***	0.150***	0.150***
	(0.013)	(0.013)	(0.034)	(0.034)
Subjects	730	730	730	730

Notes: See the notes to Table A.30. Here we report OLS regressions with the same dependent and independent variables, where we recode age five (the youngest age) as 0, age six as 1, and so on.

Table A.31: Probability of best-responding: Robustness (linear probability model).

	Baseline games		Computer game	
	(1)	(2)	(3)	(4)
Age	0.020***	0.019***	0.013*	0.012
	(0.003)	(0.003)	(0.008)	(0.008)
Theory-of-mind	0.017*	0.003	0.010	0.024
(T.o.M.)	(0.010)	(0.017)	(0.014)	(0.026)
Cognitive ability	0.016*	-0.010	0.014	-0.083***
(Cogn. Ab.)	(0.010)	(0.018)	(0.017)	(0.031)
T.o.M.× Age		0.004		-0.004
		(0.005)		(0.006)
Cogn. Ab. × Age		0.008		0.030***
		(0.005)		(0.008)
Subjects	729	729	729	729

Notes: See the notes to Table A.30. Here we add controls for subjects' gender and school. One subject is excluded because they did not report their gender.

Table A.32: Probability of best-responding: Robustness (demographic controls).

	Baseline games		Computer game	
	(1)	(2)	(3)	(4)
Age	0.118***	0.118***	-0.054	-0.054
	(0.041)	(0.039)	(0.074)	(0.070)
Theory-of-mind (T.o.M.)	0.111	0.087	-0.136	-0.155
	(0.078)	(0.119)	(0.133)	(0.202)
Cognitive ability (Cogn. Ab.)	-0.040	-0.277*	0.075	-0.471
	(0.077)	(0.160)	(0.153)	(0.294)
T.o.M.× Age		0.007		0.004
		(0.036)		(0.060)
Cogn. Ab. × Age		0.076*		0.176**
		(0.039)		(0.070)
Intercept	4.731***	4.731***	5.150***	5.150***
	(0.143)	(0.140)	(0.243)	(0.235)
Subjects	730	730	730	730

Notes: Each column reports an OLS regression. In Columns 1 and 2 the dependent variable is the average number of tokens that each subject earned in the five Baseline games (that is, the total number of tokens each subject earned in the Baseline games divided by five). In Columns 3 and 4 the dependent variable is the number of tokens that each subject earned in the Computer game. In all columns the independent variables are age, age-standardized theory-of-mind ability and age-standardized cognitive ability (as explained in Section 3.2.3, we standardize test scores within age group). We recode age five (the youngest age) as 0, age six as 1, and so on. Heteroskedasticity-robust standard errors, clustered at the session level, are shown in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.33: Payoffs.

	Baseline games		Computer game	
	(1)	(2)	(3)	(4)
Age	0.126***	0.125***	-0.042	-0.044
	(0.036)	(0.035)	(0.073)	(0.070)
Theory-of-mind	0.090	0.081	-0.172	-0.191
(T.o.M.)	(0.074)	(0.114)	(0.136)	(0.211)
Cognitive ability	-0.061	-0.264*	0.032	-0.449
(Cogn. Ab.)	(0.080)	(0.158)	(0.150)	(0.293)
T.o.M.× Age		0.002		0.004
		(0.035)		(0.061)
Cogn. Ab. × Age		0.067*		0.158**
		(0.038)		(0.069)
Intercept	4.981***	4.969***	5.027***	4.999***
	(0.192)	(0.188)	(0.372)	(0.358)
Subjects	729	729	729	729

Notes: See the notes to Table A.33. Here we add controls for subjects' gender and school. One subject is excluded because they did not report their gender.

Table A.34: Payoffs: Robustness (demographic controls).

	Baseline games		Computer game	
	(1)	(2)	(3)	(4)
Age	0.088***	0.088***	0.068**	0.068**
	(0.028)	(0.027)	(0.030)	(0.028)
Theory-of-mind	0.053	-0.012	0.065	0.091
(T.o.M.)	(0.043)	(0.075)	(0.050)	(0.094)
Cognitive ability	0.073*	-0.056	0.026	-0.249**
(Cogn. Ab.)	(0.044)	(0.094)	(0.058)	(0.115)
T.o.M.× Age		0.020		-0.009
		(0.021)		(0.026)
Cogn. Ab. × Age		0.041*		0.089***
		(0.022)		(0.029)
Intercept	4.830***	4.830***	4.819***	4.819***
	(0.112)	(0.110)	(0.120)	(0.116)
Subjects	730	730	730	730

Notes: Each column reports an OLS regression. In Columns 1 and 2 the dependent variable is the average expected payoff (in tokens) of each subject's choices in the five Baseline games (that is, the sum across the five Baseline games of the subject's expected payoff in each game, given her choice in that game, divided by five); we calculate the expected payoff to each choice given the empirical distribution of choices across the five Baseline games. In Columns 3 and 4 the dependent variable is the expected payoff (in tokens) of each subject's choice in the Computer game; we calculate the expected payoff to each choice given the computer's known uniform randomization in the Computer game. In all columns the independent variables are age, age-standardized theory-of-mind ability and age-standardized cognitive ability (as explained in Section 3.2.3, we standardize test scores within age group). We recode age five (the youngest age) as 0, age six as 1, and so on. Heteroskedasticity-robust standard errors, clustered at the session level, are shown in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels (two-sided tests).

Table A.35: Expected payoffs.

	Below median	Above median	Difference
Age	0.89	0.61	0.28
Theory-of-mind	0.76	0.72	0.05
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 Table A.2 in Web Appendix VI), 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 split subjects according to whether they are above or below: (i) median age; (ii) median theory-of-mind ability for subjects of their age; or (iii) median cognitive ability for subjects of their age. In each case, we allocate the small number of subjects exactly at the median to the above-median 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). Since we have only five observations per sub-sample (one for each repetition of the Baseline game), statistical tests are not appropriate.

Table A.36: Deviations from Nash equilibrium 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
Subjects above median age (341)						
High cognitive ability opponent	0.000 (0.032)	-0.072* (0.040)	0.011 (0.037)	0.114*** (0.044)	-0.021 (0.040)	-0.034 (0.029)
Subjects below median age (342)						
High cognitive ability opponent	-0.021 (0.049)	0.045 (0.036)	-0.039 (0.044)	0.045 (0.042)	-0.024 (0.047)	-0.007 (0.050)

Notes: See the notes to Table 3. Here we add controls for subjects' gender and school, as well as for age, age-standardized theory-of-mind ability and age-standardized cognitive ability (as explained in Section 3.2.3, we standardize test scores within age group). One subject of below median age is excluded because that subject did not report their gender.

Table A.37: Probability of choices in the Raven game: Robustness (demographic controls).

	(1)	(2)	(3)
Treatment A	0.095	0.087	-0.128
	(0.068)	(0.067)	(0.086)
Age		0.015	-0.000
		(0.013)	(0.014)
Theory-of-mind (T.o.M.)		0.009	-0.028
		(0.026)	(0.030)
Cognitive Ability (Cogn. Ab.)		0.002	0.003
		(0.029)	(0.036)
Treatment A. × Age			0.060**
			(0.023)
Treatment A. × T.o.M.			0.153***
			(0.051)
Treatment A. × Cogn. Ab.			0.005
			(0.057)
Intercept	0.160***	0.112**	0.162***
	(0.028)	(0.051)	(0.051)
Subjects	207	207	207

Notes: See the notes to Table 4. Here we report OLS regressions with the same dependent and independent variables, where we recode age five (the youngest age) as 0, age six as 1, and so on.

Table A.38: Effect of treatment on probability that receiver gives one super-token to allocator: Robustness (linear probability model).

	(1)	(2)	(3)
Number of super-tokens received	0.040***	0.041***	0.016
(Super-tokens)	(0.009)	(0.009)	(0.016)
Age		0.007	-0.026
		(0.009)	(0.018)
Theory-of-mind		0.011	-0.027
(T.o.M.)		(0.018)	(0.037)
Cognitive ability		-0.004	-0.049
(Cogn. Ab.)		(0.020)	(0.039)
Super-tokens × Age			0.009**
			(0.004)
Super-tokens × T.o.M.			0.009
			(0.010)
Super-tokens × Cogn. Ab.			0.011
			(0.009)
Intercept	-0.014	-0.037	0.060
	(0.036)	(0.045)	(0.069)
Subjects	385	385	385

Notes: See the notes to Table 5. Here we report OLS regressions with the same dependent and independent variables, where we recode age five (the youngest age) as 0, age six as 1, and so on.

Table A.39: Effect of number of tokens received on probability that receiver gives one super-token to allocator: Robustness (linear probability model).

	(1)	(2)	(3)
Treatment A	0.087	0.081	-0.145**
	(0.075)	(0.072)	(0.064)
Age		0.014	-0.003
		(0.013)	(0.014)
Theory-of-mind		0.002	-0.033
(T.o.M.)		(0.026)	(0.027)
Cognitive Ability		0.002	-0.007
(Cogn. Ab.)		(0.028)	(0.034)
Treatment A. × Age			0.061***
			(0.023)
Treatment A. × T.o.M.			0.134***
			(0.049)
Treatment A. × Cogn. Ab.			0.025
			(0.054)
Subjects	207	207	207

Notes: See the notes to Table 4. Here we add controls for subjects' gender and school.

Table A.40: Effect of treatment on probability that receiver gives one super-token to allocator: Robustness (demographic controls).

	(1)	(2)	(3)
Number of super-tokens received	0.042***	0.043***	0.017
(Super-tokens)	(0.008)	(0.008)	(0.016)
Age		0.008	-0.034
		(0.009)	(0.027)
Theory-of-mind		0.013	-0.022
(T.o.M.)		(0.018)	(0.050)
Cognitive ability		-0.003	-0.073
(Cogn. Ab.)		(0.020)	(0.049)
Super-tokens × Age			0.009*
			(0.005)
Super-tokens × T.o.M.			0.007
			(0.010)
Super-tokens × Cogn. Ab.			0.015
			(0.009)
Subjects	384	384	384

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 A.41: Effect of number of tokens received on probability that receiver gives one super-token to allocator: Robustness (demographic controls).

	Number of close friends	Ten or more close friends
Theory-of-mind	0.191**	0.020***
	(0.079)	(0.007)
Cognitive ability	0.244***	0.019**
	(0.091)	(0.008)
Participants	3,887	3,887
Mean age	24.49	24.49
Mean of dependent variable	6.76	0.16

Notes: See the notes to Table 6. Here we add a control for the participant's locus of control (see the penultimate paragraph of Web Appendix VIII).

Table A.42: Close friends: Robustness (with control for locus of control).

	In education at 17	In education at 20	A-level qualification
Theory-of-mind	0.012**	0.018**	0.015*
	(0.006)	(0.009)	(0.008)
Cognitive ability	0.023***	0.084***	0.085***
	(0.007)	(0.009)	(0.008)
Participants	4,091	4,332	4,332
Mean age	17.76	20.95	20.95
Mean of dependent variable	0.88	0.55	0.73

Notes: See the notes to Table 7. Here we add a control for the participant's locus of control (see the penultimate paragraph of Web Appendix VIII).

Table A.43: Educational participation and attainment:
Robustness (with control for locus of control).

	Parent	Number of children
Theory-of-mind	-0.011**	-0.022**
	(0.006)	(0.009)
Cognitive ability	-0.022***	-0.032***
	(0.006)	(0.010)
Participants	4,371	4,371
Mean age	25.76	25.76
Mean of dependent variable	0.12	0.18

Notes: See the notes to Table 8. Here we add a control for the participant's locus of control (see the penultimate paragraph of Web Appendix VIII).

Table A.44: Fertility: Robustness (with control for locus of control).

	Employed	Log hourly wage	Large workplace
Theory-of-mind	0.006	-0.010	0.020*
	(0.008)	(0.008)	(0.011)
Cognitive ability	0.008	0.019**	0.031**
	(0.009)	(0.010)	(0.012)
Participants	3,884	2,999	3,129
Mean age	24.49	23.89	23.89
Mean of dependent variable	0.81	2.05	0.60

Notes: See the notes to Table 9. Here we add a control for the participant's locus of control (see the penultimate paragraph of Web Appendix VIII).

Table A.45: Employment, wages, and workplace size:
Robustness (with control for locus of control).

	Number of close friends	Ten or more close friends
Theory-of-mind	0.203**	0.021***
	(0.079)	(0.007)
Participants	3,887	3,887
Mean age	24.49	24.49
Mean of dependent variable	6.76	0.16

Notes: See the notes to Table 6. The cognitive ability measure in Table 6 is the WISC-III Full Scale IQ score, which is based on the scores on ten subtests (see Section 6.2.2). Here, instead we include separate controls for the verbal and non-verbal components of cognitive ability. The first control is the Verbal Comprehension Index (VCI), which is based on the scores from the information, similarities, vocabulary, and comprehension subtests used in the WISC-III Full Scale IQ score (see Wechsler, 1991, and Prifitera et al., 1998, for more on the VCI). The second control is based on the scores from the remaining six subtests used in the WISC-III Full Scale IQ score. The correlation between the VCI and the second control is 0.55. The coefficients and standard errors reported here are not identical to those in Table 6 when rounded using more decimal places.

Table A.46: Close friends: Robustness
(with separate controls for verbal and non-verbal components of cognitive ability).

	In education at 17	In education at 20	A-level qualification
Theory-of-mind	0.012**	0.020**	0.016**
	(0.006)	(0.009)	(0.008)
Participants	4,091	4,332	4,332
Mean age	17.76	20.95	20.95
Mean of dependent variable	0.88	0.55	0.73

Notes: See the notes to Table 7. The cognitive ability measure in Table 7 is the WISC-III Full Scale IQ score, which is based on the scores on ten subtests (see Section 6.2.2). Here, instead we include separate controls for the verbal and non-verbal components of cognitive ability. The first control is the Verbal Comprehension Index (VCI), which is based on the scores from the information, similarities, vocabulary, and comprehension subtests used in the WISC-III Full Scale IQ score (see Wechsler, 1991, and Prifitera et al., 1998, for more on the VCI). The second control is based on the scores from the remaining six subtests used in the WISC-III Full Scale IQ score. The correlation between the VCI and the second control is 0.55. The coefficients and standard errors reported here are not identical to those in Table 7 when rounded using more decimal places.

Table A.47: Educational participation and attainment: Robustness
(with separate controls for verbal and non-verbal components of cognitive ability).

	Parent	Number of children
Theory-of-mind	-0.012**	-0.022**
	(0.006)	(0.009)
Participants	4,371	4,371
Mean age	25.76	25.76
Mean of dependent variable	0.12	0.18

Notes: See the notes to Table 8. The cognitive ability measure in Table 8 is the WISC-III Full Scale IQ score, which is based on the scores on ten subtests (see Section 6.2.2). Here, instead we include separate controls for the verbal and non-verbal components of cognitive ability. The first control is the Verbal Comprehension Index (VCI), which is based on the scores from the information, similarities, vocabulary, and comprehension subtests used in the WISC-III Full Scale IQ score (see Wechsler, 1991, and Prifitera et al., 1998, for more on the VCI). The second control is based on the scores from the remaining six subtests used in the WISC-III Full Scale IQ score. The correlation between the VCI and the second control is 0.55. The coefficients and standard errors reported here are not identical to those in Table 8 when rounded using more decimal places.

Table A.48: Fertility: Robustness
(with separate controls for verbal and non-verbal components of cognitive ability).

	Employed	Log hourly wage	Large workplace
Theory-of-mind	0.005	-0.009	0.020*
	(0.008)	(0.008)	(0.011)
Participants	3,884	2,999	3,129
Mean age	24.49	23.89	23.89
Mean of dependent variable	0.81	2.05	0.60

Notes: See the notes to Table 9. The cognitive ability measure in Table 9 is the WISC-III Full Scale IQ score, which is based on the scores on ten subtests (see Section 6.2.2). Here, instead we include separate controls for the verbal and non-verbal components of cognitive ability. The first control is the Verbal Comprehension Index (VCI), which is based on the scores from the information, similarities, vocabulary, and comprehension subtests used in the WISC-III Full Scale IQ score (see Wechsler, 1991, and Prifitera et al., 1998, for more on the VCI). The second control is based on the scores from the remaining six subtests used in the WISC-III Full Scale IQ score. The correlation between the VCI and the second control is 0.55. The coefficients and standard errors reported here are not identical to those in Table 9 when rounded using more decimal places.

Table A.49: Employment, wages, and workplace size: Robustness
(with separate controls for verbal and non-verbal components of cognitive ability).

	Disadvantaged participants	Advantaged participants
School spending per pupil (£300 cumulative over three years)	-0.058 (0.092)	0.351*** (0.126)
Participants	2,216	1,982

Notes: See the notes to Table 10. Here we allocate participants whose family income in childhood is at the median for the participants in their primary school to the disadvantaged category.

Table A.50: Difference-in-differences estimates of the effect of school spending on theory-of-mind: Robustness.

Web Appendix XIV

Additional figures

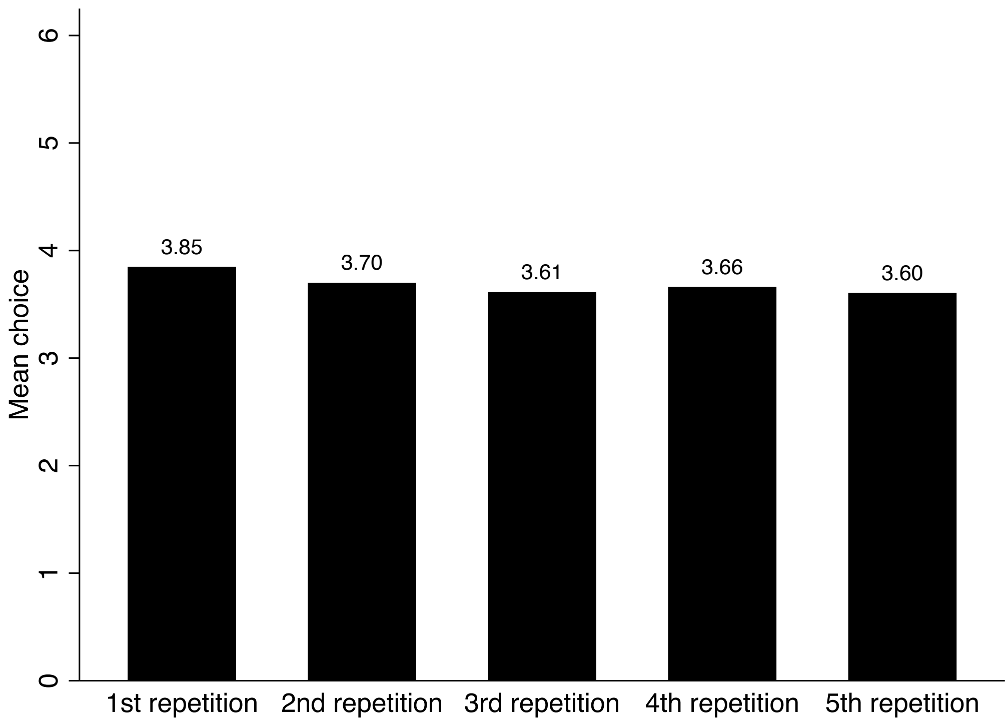
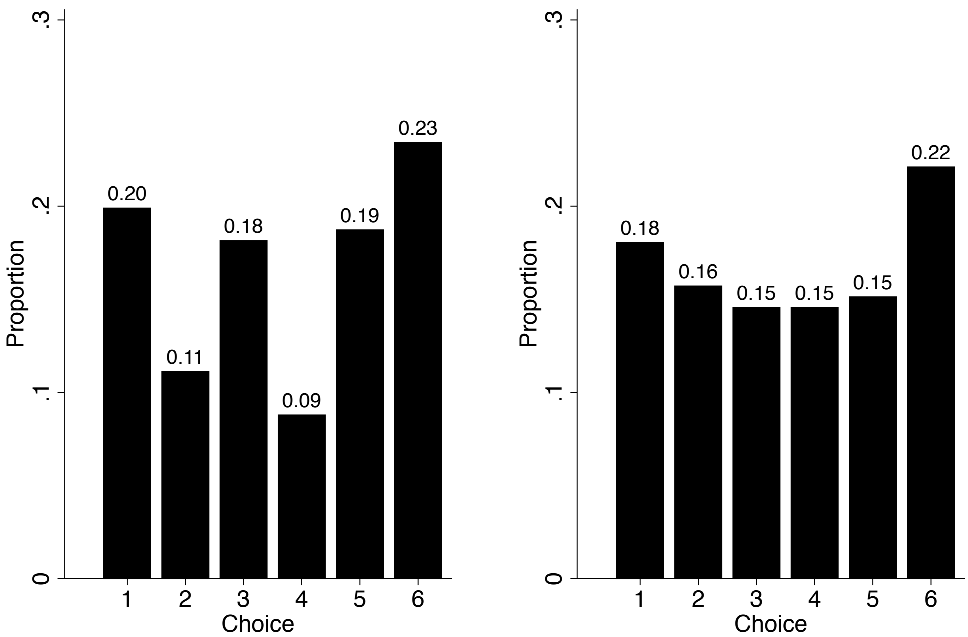


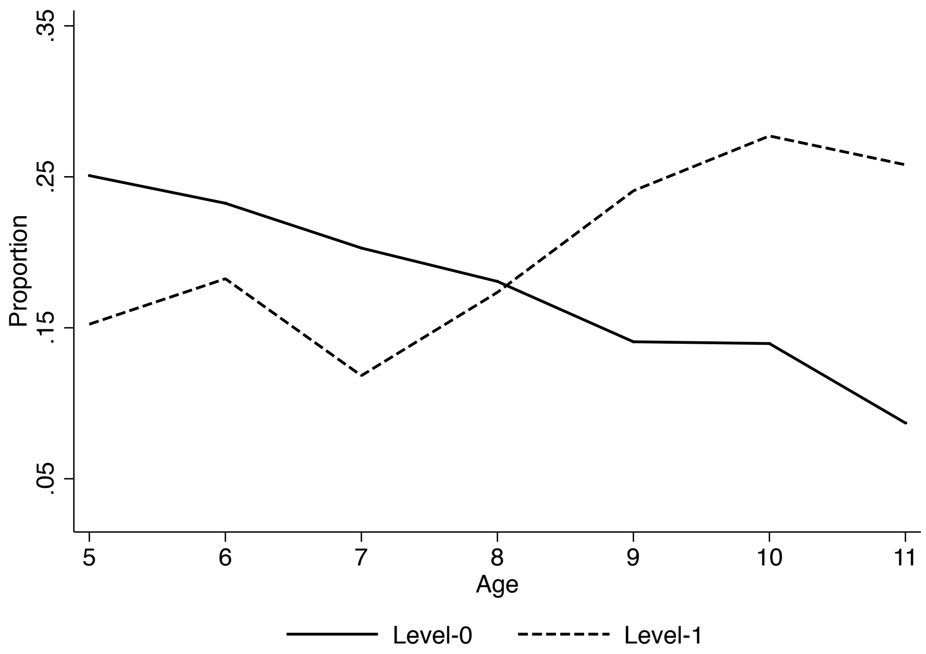
Figure A.7: Mean choices in the Baseline games.



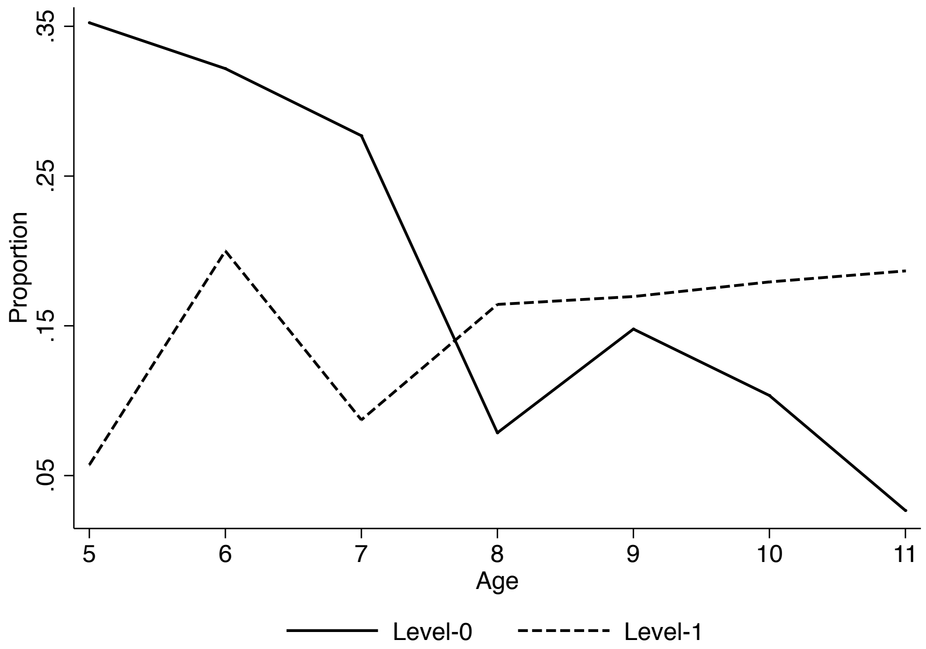
(a) Low cognitive ability opponent. (b) High cognitive ability opponent.

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

Figure A.8: Distribution of choices in the Raven game for younger subjects.



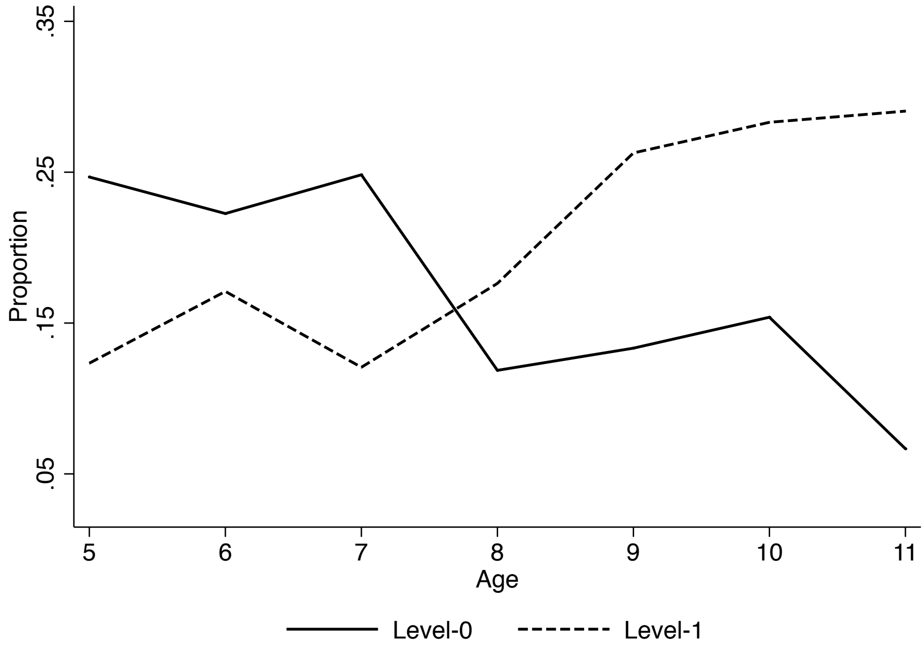
(a) Above median theory-of-mind ability subjects.



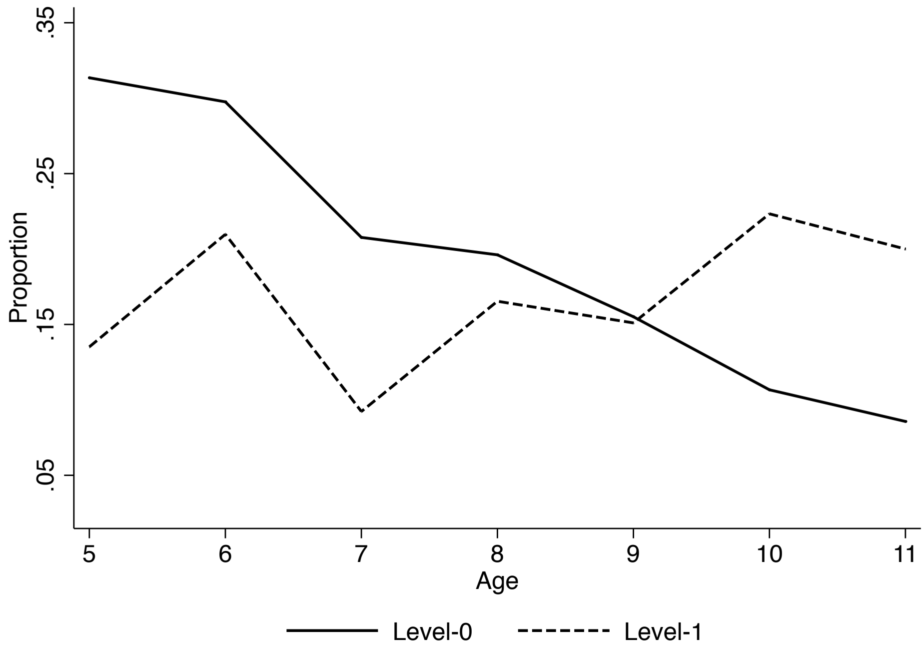
(b) Below median theory-of-mind ability subjects.

Notes: See the notes to Figure 4. The top (bottom) panel shows the proportion among subjects above (below) median theory-of-mind ability for subjects of their age. We allocate the small number of subjects exactly at the median to the above-median category.

Figure A.9: Age and strategic sophistication in the Baseline games: Split by theory-of-mind ability.



(a) Above median cognitive ability subjects.



(b) Below median cognitive ability subjects.

Notes: See the notes to Figure 4. The top (bottom) panel shows the proportion among subjects above (below) median cognitive ability for subjects of their age. We allocate the small number of subjects exactly at the median to the above-median category.

Figure A.10: Age and strategic sophistication in the Baseline games: Split by cognitive ability.