Acknowledgement

It feels as though only yesterday I was working on my undergraduate thesis. And now, I am writing the acknowledgements for my master's dissertation. The flowers blooming and fading on the lawn in front of my college, St Hugh's College, seem to be telling me that I am about to pass four seasons in Oxford.

Firstly, I would like to sincerely appreciate my supervisor, Prof. Lars-Erik Malmberg. Lars is a highly responsible and dedicated supervisor. He has provided tremendous support in statistical methods, research design, and even in guiding me towards my future DPhil application. But what I remember most are the small moments: him cycling past with a warm greeting, asking about my lunch while holding a carrot, or cheerfully saying "nihao". Thank you so much, Lars.

Secondly, I want to thank my parents Xingfang Zhou and Jianbo Mao, as well as other family members. Despite the seven or eight-hour time difference, they always keep an eye on how I'm doing. Their love and unwavering support stretch across half the globe — from the time widget on my dad's phone showing both Beijing and London time, to the carefully packed snacks my mum sends, one after another. I also need to thank Dr. Tingyu Li. Dr. Li was not only my supervisor during my undergraduate years, but has remained a lifelong guide on my journey in research.

Thirdly, Oxford is a real-life Hogwarts. Here, I've met many wonderful friends. We are all on different paths with different dreams, but each constantly striving upwards. My heartfelt thanks go to Zixi Jiang, Yue Yin, Junyan Yin, Weibin Li, Liusiyuan He, Renyang Liu and Zeyang Wang. This year has been one of the amazing years in my life, and I'm deeply grateful to have shared all its sweetness, bitterness, and everything in between with you. Special thanks to our "Western Development Group" — our photos always lifted my spirits. Wishing all my friends the very best in everything that lies ahead.

Fourthly, Thank you so much my lovely Department of Education and St Hugh's College. I trusted the department and St Hughs embraced me. CDE cohort has offered me encouragement and affirmation, and I'm especially grateful to Martina Kurillová.

Hughs is such a heaven — the library, JCR, MCR, the quads — all hold countless moments of joy. Maybe next time, it will be my turn to open my arms to Hughs.

Finally, I want to thank the brave me. On 2 December 2023, I gave up a very nice job, and applied for a master's at Oxford. I didn't know if it was the right choice, nor was I ready to face the consequences. But on 14 March 2024, Oxford welcomed me. That one brave decision rewrote the course of my life. I hope that in 2026, I'll return to Oxford for a DPhil.

Abstract

Given the integration of robots into educational contexts, understanding how children evaluate information from artificial agents is essential. This dissertation examines how children aged 4 to 6 selectively trust robots when receiving information across STEM (Science, Technology, Engineering, and Mathematics) and non-STEM domains. Building on the Theory of Artificial Mind (ToAM) which is an extension of the Theory of Mind (ToM), the association between ToM/ToAM and selective trust was investigated. Three research questions guided this study: (1) To what extent do children's selective trust in robots and humans differ? (2) How does the domain of testimony (non-STEM versus STEM) influence children's selective trust? (3) Do ToAM and ToM relate to children's selective trust in robot and human informants?

The study employed a conflicting informants paradigm wherein 107 Chinese children (M = 5.57 years, SD = 0.58, 44.86% girls) were randomly allocated into two between-subjects conditions: a *Nao-accurate* condition and a *Human-accurate* condition. Each child encountered four informant dyads (one human, one robot) who provided conflicting testimony across four domains: one non-STEM domain (object labelling) and three STEM domains (physical science, life science, mathematics). Standardised scales were used to assess children's ToM and ToAM abilities, with a focus on core components including desire, belief, knowledge, and emotion. Statistical analyses indicated that children consistently preferred accurate informants, irrespective of informant type, particularly in non-STEM domain. However, domain-specific analyses revealed nuanced preferences. Interestingly, within the physical science domain, children demonstrated notable uncertainty; they were inclined to nominate robots despite their inaccuracies yet endorsed information provided by humans and deemed humans as reliable source. Lastly, in the *Nao-accurate* condition, children's ToM and ToAM scores both negatively predicted selective trust, suggesting that children with stronger cognitive functions were more cautious about trusting robots. In contrast, no significant relationships emerged in the *Human-accurate* condition. Collectively, these findings deepen our understanding of how young children evaluate informational reliability in AI-integrated early STEM education and highlight their developing cognitive sophistication.

Contents

List of figures

List of tables

Abbreviations

AI: Artificial Intelligence

BERA: British Educational Research Association

BIC: Bayesian Information Criterion

CRI: Child-Robot Interaction

CUREC: The Central University Research Ethics Committee

DREC: Departmental Research Ethics Committee

DVA: Digital Voice Assistant

EYFS: Early Years Foundation Stage

GLM: Generalised Linear Model

GLMM: Generalised Linear Mixed Model

GVI: Generalised Variance Inflation

LMM: Linear Mixed Model

MSA: Mental State Attribution

NGSS: Next Generation Science Standards

ORA: Oxford Research Archive

RQ: Research Question

STEM: Science, Technology, Engineering, and Mathematics

ToM: Theory of Mind

ToAM: Theory of Artificial Mind

Chapter 1 Introduction

Moxie, a child-centred robot designed to support the development of social and emotional skills, was recently shut down due to funding failures. Many children reportedly felt as though they were losing a friend and even wrote farewell letters to Moxie (Notopoulos, 2024). A wise and enduring Chinese proverb from a millennium-old classic states: "*As the world changes, so do events; as events change, so must our responses.*" In today's digital age, since technological but intelligent agents become increasingly integrated into children's lives, understanding children's perceptions of robots becomes essential (Breazeal et al., 2016; Brink & Wellman, 2020). Spektor-Precel and Mioduser (2015) introduced the concept of Theory of Artificial Mind (ToAM). A recent study by Mao et al. (2025) systematically examined preschoolers' development of ToAM, suggesting a clear developmental progression. Notably, by the age of six, children exhibited near adult-level understanding of artificial minds.

ToAM can be viewed as an extension of Theory of Mind (ToM) in the age of Artificial Intelligence (AI). While ToM refers to the cognitive ability to attribute mental states to other people and to recognise that these states may differ from one's own (Premack & Woodruff, 1978), ToAM involves applying this capacity to artificial agents, such as robots or digital systems (Spektor-Precel & Mioduser, 2015). A key component of both ToM and ToAM is knowledge access. According to Vygotsky's sociocultural theory, which emphasises the role of language and other symbolic systems in children's learning and development within social contexts (Lantolf, 2000), children acquire knowledge through interactions with more knowledgeable others. By around age four, children begin to recognise that others may hold knowledge states different from their own (Wellman & Liu, 2004). Indeed, prior research has shown that even three-year-olds can selectively trust an accurate human informant over an inaccurate one (Koenig et al., 2004; Koenig & Harris, 2005). However, emerging research highlighted children aged 3 to 6 also attribute knowledge state to humanoid robots (Mao et al., 2025). This indicated that children may perceive robots as possessing specialised knowledge or information that is not readily accessible to themselves. Accordingly, important questions arose: Can children learn from robots? More specifically: do children trust and learn from robots in a way that mirrors their trust in humans?

The answer to the first question appears to be positive. Social robots are capable of supporting learning through interactive engagement (Belpaeme et al., 2018) and have been successfully used to teach various subjects, including English, science, and geometry (Rosanda & Istenic Starcic, 2020). In contrast, the answer to the second question is more complex. To better understand the social learning process, researchers have investigated the concept of selective trust (i.e., epistemic trust), which refers to children's ability to evaluate and choose whom to trust based on specific cues or contextual factors (Koenig et al., 2004). In the child-robot interactions (CRIs) preschoolers often apply their understanding of human social behaviour to their interactions with robots, using this framework to judge how they engage with, and how much they trust, the information robots provide (Geiskkovitch et al., 2019). Even three-year-olds have demonstrated the ability to selectively trust an accurate robot over an inaccurate one, showing patterns similar to how they evaluate human informants (Brink & Wellman, 2020).

STEM education, which refers to learning science, technology, engineering, and mathematics (STEM), is understood as a process initiated by teachers who assess and support students' conceptual development, inquiry skills, and connections to real-world applications (Allen et al., 2016). Due to the interdisciplinary nature of STEM learning, young children, particularly those in preschool, often find it difficult to grasp abstract concepts through direct observation or personal experience. Previous explorations have shown that robots can serve as effective instructional tools in STEM education by increasing motivation and making complex ideas more accessible (Budiharto et al., 2017). Robotic teachers have been shown to support primary school students in learning mathematics and programming (Mertzani & Drigas, 2023), while preschoolers have also benefited from geometric thinking tasks supported by robotic social systems (Keren & Fridin, 2014). More broadly, children aged 4 to 15 who engage with robot-supported programming activities demonstrate improvements in problem-solving, critical thinking, and coding skills, along with sustained motivation (Mertzani & Drigas, 2023).

Robots have been shown to be potentially powerful facilitators in STEM education across various age groups. However, limited research has examined how children learn from robot informants, and to date, no study has systematically manipulated robots to

deliver STEM-related testimony. This study addresses two significant gaps in the literature. First, it compares children's selective trust in a human versus a robot across both non-STEM and STEM domains. Second, given that knowledge access is a core element of both ToM and ToAM, this study explores the relationship between children's ToM and ToAM and selective trust in human and robotic informants. To the best of our knowledge, this is the first study to investigate how children learn from robot informants in the context of STEM-related testimony. In doing so, the study bridges cognitive development, social learning, STEM education, and early AI literacy, offering new insights into how young learners engage with artificial agents as knowledge sources and promoting AI-integrated STEM education.

The current chapter includes elementary background and an introduction of the study. Chapter 2 begins by defining and conceptualising the key concepts of the study, encompassing ToM, ToAM and selective trust. It also introduces underpinning theoretical frameworks and critically reviews existing literature, highlighting the relationships between these concepts and identifying meaningful empirical findings. Chapter 3 outlines the research objectives, methodology, and research questions, grounded in a clear identification of the research gap. The fourth chapter provides a detailed account of the methodology, including sampling procedures, materials, research design, ethical considerations, and analytical methods. Chapter 5 presents the results in relation to each research question, using descriptive and inferential analyses such as correlation matrices, logistic regression models, and generalised linear mixed models. Finally, Chapter 6 summarises and interprets the findings, discusses their theoretical and practical implications, and offers directions for future research along with a reflection on the study's limitations.

Chapter 2 Literature review

2.1 Introduction of the literature review

This chapter critically reviews the existing literature relevant to ToM, ToAM and children's selective trust in human and robotic informants, with particular attention to how such trust manifests across STEM and non-STEM domains.

Specifically, this review is structured into five main sections. It begins with an exploration of ToM and ToAM, tracing their developmental trajectories and cultural variations. Next, it examines the mechanisms of selective trust in human informants, detailing the cues children use to evaluate the credibility of information. The review then turns to selective trust in the context of STEM learning, followed by STEM education in diverse cultural backgrounds. Subsequently, there is a discussion of children's trust in robotic agents and a comparison between human and robot informants. Finally, the chapter explores how children's ToM and ToAM are empirically linked to their trust. Throughout, previous studies are reviewed by critically focusing on the advantages and disadvantages of cross-sectional, descriptive surveys, and experimental designs. Cross-sectional studies and surveys revealed cultural and developmental differences in ToM and ToAM. Experimental designs have been widely employed in selective trust research, enabling researchers to find potential causal relationships between conflicting information and informants.

In synthesising these diverse strands of research, this chapter identifies an underpinning theoretical framework that is grounded in previous literature yet especially tailored to support the scope of the current study. In parallel, research gaps are critically analysed at the conclusion of the chapter, providing the necessary foundation for formulating the research questions presented in the next chapter.

2.2 From ToM to ToAM

2.2.1 Theory of Mind (ToM)

Theory of Mind (ToM) refers to the ability to understand that others have beliefs, desires, intentions, and emotions that may be different from one's own. It enables children to understand and predict others' behaviour based on internal mental states (Premack &

Woodruff, 1978). Children's everyday conversations about people and minds are communicated through the use of terms like "think", "want", "feel", and "know", which are real-life expressions of the function of ToM. For instance, a child receives a candy from the teacher as a reward, but notices that the friend, who did not, looks sad. The child understands that the friend feels sad and may believe he was left out. In this situation, the child recognises that the friend has a different emotional experience (feeling sad) and a different belief (believing he was left out). Such an understanding of the other child would be indicative of ToM. Basically, the main stage of ToM development is preschool years. From basic theory exploration to social-emotional learning (Barlow et al., 2010) and future academic performance in the school years (Lecce et al., 2011), the worldwide discussion on ToM was supported by numerous evidence and nurtures thousands of children's cognitive development.

ToM consists of several key components, including desire, belief, knowledge, and emotion (Wellman & Liu, 2004) that emerge in a developmental sequence. Around age two, children begin to understand simple desires, and by age three, they start to grasp the concept of belief, although they still tend to explain behaviour primarily in terms of desires (Bartsch & Wellman, 1995). A three-year-old is more likely to say, "She *wants* (desire) some candies" than "She *knows* (belief) candies are nice." A major milestone in ToM development is the understanding of false beliefs, which involves recognising that others can hold beliefs that do not reflect reality (Baillargeon et al., 2010). The well-established *Sally-Anne task* assesses false belief understanding by presenting a scenario in which Sally places a marble in her basket and leaves. In her absence, Anne moves the marble to a new box. Children are then asked where Sally will look for the marble upon her return (Baron-Cohen et al., 1985). Most three-year-olds fail to predict Sally's actions based on false beliefs (i.e., to look for the marble where Anne had placed it and perceive Sally could not know that Anne had moved the marble), whereas most four-year-olds succeed, marking significant progress in their ToM development (Rubio-Fernández & Geurts, 2016).

Moreover, ToM also encompasses the ability to understand and interpret others' emotions. From around age 3, children begin to use desires and beliefs to reason about emotional states, and this emotional understanding continues to develop over time

(Harris et al., 1989). A three-year-old sees others looking sad after not getting the toy says, “He’s sad because he wanted to play.” This shows the child is beginning to understand that emotions can be explained by both desires and beliefs. Regarding more advanced multiple emotions, children generally grasp that a person can experience one emotion internally while displaying another outwardly (e.g., smiling while feeling sad) between the ages of three and five, though three-year-olds show a limited capacity (Banerjee, 1997).

In the present study, the component of ToM receiving the most attention is knowledge. Early research conducted in the United States and Australia has shown that children around the age of three typically grasp others’ desires and pass false belief tasks before they are able to infer whether someone possesses specific knowledge or not (Wellman, 2018; Wellman & Liu, 2004). This indicated that children understand knowledge access after false belief and basic desire recognition (Wellman & Liu, 2004). However, empirical evidence from China, Iran, and Turkey (Wellman, 2018) has reported the reversal pattern. In these cultures, children understood knowledge acquisition before grasping diverse or false beliefs. Consequently, the developmental trajectory of ToM, particularly the understanding of knowledge and belief, may be shaped by cultural environment (Wellman, 2017, 2018).

2.2.2 Theory of Artificial Mind (ToAM)

Explored in ToM research for over thirty years, Wellman (2018), a leading pioneer in this field, emphasised the importance of studying how children understand extraordinary minds, such as those of God, superheroes, and Santa Claus, as well as what he defined as “state of the art questions in need of state of the art research,” including AI agents. In response to the growing trend of children attributing mental states to technological entities, Spektor-Precel and Mioduser (2015) defined the concept of Theory of Artificial Mind (ToAM). They observed that children often mentalise AI agents by attributing mental states such as desires, beliefs, and knowledge to them. When children observe or construe robot behaviours, they are more likely to interpret and describe robots as intentional agents, which reflects the development of ToAM. For example, during interactions with robots, children may attribute agency and emotional experiences to robots, sometimes treating them as human-like companions (Brink et al., 2019).

Just as in the early stages of ToM research, scholars have placed particular emphasis on false belief in ToAM investigation. Children aged 3 to 6 have demonstrated promising performance in false belief tasks involving robots. Surprisingly, when a humanoid robot, rather than a human, asked the classic false belief question (e.g., "Where will Maxi look?" in an unexpected location task), children aged three performed significantly better with the humanoid robot than with the human (Baratgin et al., 2020). Equally, typically developing children aged five to seven have shown the ability to attribute false beliefs to robots, suggesting they recognise robots as intentional beings. However, children aged 5 to 8 with autism often face greater difficulty in recognising false beliefs with humans and robots (Zhang et al., 2019). Supporting these findings, Di Dio et al. (2020) applied both first and second-order false belief tasks. First-order task involves understanding that someone can hold a belief that differs from reality (e.g. "She thinks the toy is in the box"), while second-order task is more complex and involves understanding what one person thinks about another person's thoughts (e.g. "He thinks that she thinks the toy is in the box"). They reported that most five- and seven- year-olds succeeded in first-order tasks involving robots, while nine-year-olds performed better in more advanced second-order tasks (Di Dio et al., 2020).

Two studies have investigated adults' and children's ToAM using standardised scales. Banks (2020) was the first to adapt the well-established and widely used ToM scale developed by Wellman and Liu (2004) in human-robot interactions. The study found that adults mentalised robots in ways that mirrored their mentalisation of human agents, particularly when the robots displayed human-like social cues. More recently, Mao and colleagues (2025) extended this work to children, revealing a developmental sequence in children's understanding of ToAM. Their findings indicated particular challenges in understanding false beliefs and emotions, as well as a distinction between ToM and ToAM. Although both studies used similar materials and procedures, their findings differed by age group, highlighting the developmental trajectory of ToAM from early childhood to adulthood. Of particular interest is the finding that Mao et al. (2025) also discovered that Chinese children tended to understand knowledge access earlier than false belief, a pattern consistent with ToM development in children from Asian cultural backgrounds.

2.3 Selective trust in human informants

2.3.1 Introduction of selective trust

Within a Vygotskian framework (Vygotskii & Cole, 1978), human activities take place within cultural contexts and are mediated by language and other symbolic systems. Children develop cognitively through meaningful dialogue with more knowledgeable others (John-Steiner & Mahn, 2003). While Piaget emphasised that children construct knowledge independently through hands-on exploration, Vygotsky argued that learning from others is also a vital mechanism for internalising knowledge. In particular, cognitively demanding domains such as geography, microorganism, history, and astronomy are often acquired through social input (Harris et al., 2006) rather than through individual experience or discovery. Indeed, children do not blindly trust testimonies, even from knowledgeable individuals. They engage in selective trust, evaluating both the information and the informant based on specific cues and contextual factors (Koenig et al., 2004). For example, preschoolers aged three to four prefer to learn novel labels (i.e., how to call a novel object) from informants who had previously labelled familiar objects correctly (Koenig et al., 2004). Even eighteen-month-old infants have been shown to preferentially select and interact with objects labelled by a reliable speaker (Crivello et al., 2021).

2.3.2 Epistemic and social cues

When deciding whom and what to believe, children primarily rely on two types of cues. The first are epistemic cues, which include the informant's prior accuracy, confidence, and expertise. Prior accuracy has received the most attention since the emergence of selective trust research. Preschoolers tend to trust individuals who have previously labelled objects accurately (Koenig et al., 2004). Pasquini et al. (2007) tested children's sensitivity to varying degrees of relative accuracy across four conditions: 100% vs. 0%, 75% vs. 0%, 75% vs. 25%, and 100% vs. 25%. Their results showed that three-year-olds only demonstrated selective trust when one informant was entirely accurate, whereas four-year-olds could reliably select the relatively more accurate informant across all conditions. Confidence is another important epistemic cue. Children are more likely to trust an informant who expresses confidence in their statements, such as saying, "I looked and I saw an apple in the box" (Koenig, 2012). Finally, expertise also impacts

children's trust. Koenig and Jaswal (2011) found that when naming unfamiliar dog breeds, young children tended to accept the testimony of a dog expert.

Another important factor influencing trust is the use of social cues such as familiarity, moral behaviour, race, and group size. Evidence from the Strange Situation experiment shows that even nine-month-old infants can distinguish between their mothers and strangers, highlighting the importance of familiarity in early development (Ainsworth et al., 1978). In the context of selective trust, Corriveau and Harris (2009) found that children aged 3 to 5 tended to trust familiar teachers than unfamiliar ones when learning about novel objects. Similarly, Kinzler et al. (2011) reported that children preferred informants who spoke with a familiar local accent. Second, children also place greater trust in individuals who display positive personality traits such as honesty, kindness, and intelligence. In a study by Lane et al. (2013), when given a choice between an informant with positive traits who could not see inside a box and another with negative traits who could see inside, children aged 3 to 5 trusted the informant with positive traits. Further, children show a stronger tendency to trust individuals from their own ethnic group (Cameron et al., 2001) and prefer information endorsed by a group of people rather than by a single individual (Fusaro & Harris, 2008).

When epistemic and social cues conflict, children must navigate a complex decision about whom to trust. When preschoolers were asked to choose between a "kind but inaccurate" informant and a "mean but accurate" one, children aged 3 to 5 showed mixed responses. But six-year-olds prioritised accuracy, choosing the mean but reliable informant (Lane et al., 2013). This interesting shift reflects a growing emphasis on epistemic reliability over social traits. Similarly, preschoolers do not always view adults as the most reliable informants. For example, children aged 3 to 5 were more likely to direct toy-related questions to a peer and nutritional questions to an adult, suggesting they view peers as better sources for "children world" knowledge (VanderBorght & Jaswal, 2009). Generally speaking, a meta-analysis by Tong et al. (2020) examined the role of both cues. With a moderate effect size (Hedges' $g = 0.59$), results showed that children generally preferred accurate and knowledgeable informants. However, when cues conflicted, older (more than aged four) children increasingly favoured epistemic cues, while younger children often favoured socially familiar or in-group informants, even if they were less accurate ($Q(1) = 4.65$, $p = .031$).

2.3.3 The measurement of selective trust

Two main experimental paradigms have been widely adopted in selective trust research: the single informant paradigm and the conflicting informants paradigm (Cao et al., 2025). In the single informant paradigm, only one informant is presented, and researchers manipulate cues such as the informant's expertise or moral traits. Children are then asked to decide whether to accept or reject the informant's testimony. However, most selective trust studies employ the conflicting informants paradigm. This typically involves a familiarisation-test procedure. During the familiarisation phase (also referred to as the "history phase" in some studies, see Danovitch et al., 2023), children are introduced to two informants with contrasting characteristics, allowing them to form initial impressions. In the test phase, both informants provide conflicting information, and children are asked to choose which informant to trust. For example, in a classic study by Koenig and Harris (2005), children observed two speakers label familiar objects — one consistently accurate and the other inaccurate (familiarisation phase). Later, children received novel labels from both informants and were asked to endorse one of them (test phase).

2.4 Learning STEM knowledge selectively

2.4.1 STEM education in preschool

Although early research on selective trust focused primarily on object labelling involving familiar and novel items (e.g., "what is this?" "It is a ball."), children also acquire more complex knowledge from surrounding cultural environment, such as moral norms (Doebel & Koenig, 2013), historical facts (Skjæveland, 2017), and professional expertise (Landrum et al., 2013). One extraordinarily complicated yet underexplored domain is STEM knowledge. Due to animistic thinking (Harvey, 2005), young children often attribute mental states to animals, plants, or even non-living entities (e.g., believing the sun feels hot and asking the clouds for help), which conflict with scientific explanations. In fact, most countries have increasingly recognised the transformative potential of STEM education and allocated substantial funding to support high-quality STEM initiatives (Department of Education, 2024b; Zhao, 2018). As Johnson (2013) noted, STEM education fosters children's development of problem-solving, critical thinking,

and digital literacy by integrating practices such as scientific inquiry, technological and engineering design, and mathematical analysis.

Across both Western and non-Western contexts, early childhood STEM education increasingly emphasises hands-on exploration, problem-based learning, and child-led experimentation. In the United States, the Next Generation Science Standards (NGSS) advocate for engaging preschoolers in foundational scientific practices such as asking questions, making observations, and interpreting data from an early age (Council, 2013). In the United Kingdom, the Early Years Foundation Stage (EYFS) integrates science and numeracy into everyday learning through informal play and guided discovery (Department of Education, 2024a). Particularly, the Forest School model has also gained popularity across Europe, offering young learners opportunities to engage with STEM principles through embodied interaction with the natural environment (Cudworth & Lumber, 2021). Activities involving spatial awareness, sensory exploration, and material experimentation allow children to encounter scientific and mathematical concepts in meaningful and unstructured ways (Kraftl, 2014). In China, national guidelines emphasise the integration of STEM within play-based pedagogy (Ministry of Education, 2012). A common classroom programme is the "sinking and floating" activity, in which children aged 4 to 6 explore the buoyancy of everyday objects (such as leaves, stones, and plastic toys) through guided experimentation and observation.

2.4.2 STEM knowledge in selective trust

Despite exposure to STEM content in preschool curricula, children often find it difficult to grasp advanced scientific concepts. While they are capable of conducting simple experiments, making observations, and documenting results, — such as tracking the growth of an onion or comparing the falling speed of a feather and a metal object — they struggle to understand phenomena that are abstract or removed from everyday experience (e.g., the secret of Mars). They may find it difficult to generalise or transmit scientific knowledge to others. As a result, children frequently rely on external sources, including adults, animations, movies, and AI agents, to acquire STEM-related information.

Given the scientific nature of STEM learning, researchers have begun examining how children selectively trust STEM-related information. Studies show that children as young

as three prefer to learn biological facts, such as food health and taste, from familiar mothers and professional teachers, rather than from strangers or clowns (Nguyen, 2012). When learning about unfamiliar animals, children are more likely to trust accurate informants or group consensus (Sampaio et al., 2019). Considering the physics, children are more inclined to believe physically impossible events when experienced through immersive technologies like virtual reality compared to traditional picture books (Schmitz et al., 2020). Another evidence revealed that fourth and fifth graders trust mathematical knowledge from an accurate informant rather than an inaccurate one (Durkin & Shafto, 2016). Thus, children learn STEM information from a broad resources, but still rely on social and epistemic cues.

Several studies have explored children's trust in various subdomains in STEM information. A key finding is that children demonstrate an early understanding of the division of cognitive labour which refers to the cognitive effort involved in tasks. Precisely, children aged 3 to 5 were able to correctly attribute observable knowledge, such as treating illness or fixing cars, to familiar experts like doctors or mechanics. However, only four- and five-year-olds could accurately assign deeper scientific knowledge to appropriate experts, recognising that an eagle expert might also possess broader knowledge about birds, animals, and biology (Lutz & Keil, 2002). When evaluating biological and physical explanations, children aged 4, 5, and 7 reduced their trust in informants who provided low-quality explanations, even when those informants were introduced as domain experts (Clegg et al., 2019). Nonetheless, when experimental designs were held constant, children aged 4 to 6 showed no significant differences in their trust toward STEM and non-STEM testimony, but transmit more accurate physical information to the third party rather than mathematical ones (Danovitch et al., 2023). Taken together, while STEM content is often more complex and incomprehensible, preschoolers are still capable of evaluating the accuracy of information and the reliability of informants, much like they do with simpler knowledge.

2.5 Selective trust in the robot informants

2.5.1 Selectively trust AI agents

The integration of AI agents into various contexts has created new learning approaches for children. In response, researchers have applied the selective trust paradigm to

examine children's interactions with different types of technological agents. Digital voice assistants (DVAs) are intelligent and network devices that respond to voice commands such as "Hi Siri" or "Tell me the weather," processing spoken input and providing verbal output. When children ask questions and retrieve information from DVAs, this interaction can be considered a form of selective trust. Girouard-Hallam and Danovitch (2022) investigated children's trust in DVAs by comparing the responses of children aged 4 to 5 and 7 to 8. The results showed that older children were more likely to seek factual information from DVAs and ask personal information from humans. Similarly, Li et al. (2023) found that children aged 5 to 6 demonstrated greater trust in DVAs compared to those aged 3 to 4, with adults showing the highest level of trust in the technology.

Second, previous studies have shown that preschool-aged children are capable of selectively trusting robots in ways that align with how they evaluate human informants. To the best of my knowledge, one of the earliest studies applying the selective trust paradigm to robots was conducted by Breazeal et al. (2016). In their study, children aged 3 to 5 were inclined to ask for and endorse information from a robot that responded contingently with real-time feedback and eye contact, rather than a non-contingent robot. Geiskkovitch et al. (2019) further adapted the conflicting informants paradigm within the context of CRI and found that children aged 3 to 5 assessed a robot's trustworthiness based on its history of errors. Particularly, in another task in Geiskkovitch et al.'s (2019) study, when both robots provided the same label for different unfamiliar objects, children were more likely to choose the object labelled by the previously reliable robot. Brink and Wellman (2020) reported similar findings and further emphasised that children selectively trust humanoid robots, but not non-humanoid machines, even when those machines are accurate. What's more, children between 4 and 7 with autism also preferred accurate robots over inaccurate ones, similar to their typically developing peers. However, they exhibited generally lower levels of selective trust, regardless of whether the informant was a robot or a human (Chen et al., 2024).

Most selective trust research in CRIs has focused on non-STEM domains, primarily using object-labelling tasks, (Baumann et al., 2023; Brink & Wellman, 2020; Chen et al., 2024; Geiskkovitch et al., 2019; Stower et al., 2024) rather than exploring complicated social information or STEM knowledge. When learning social emotional testimonies

(e.g., “How to make yourself happy when you feel sad?”) from a humanoid robot and a non-humanoid robot, children aged 4 to 7 tended to trust humanoid one more. However, in the single informant paradigm, appearance plays a smaller role in shaping trust (Cao et al., 2025). Some studies have explored domains closer to STEM knowledge. Breazeal et al. (2016) presented children with information about animals, such as their characteristics and habits (e.g., “My favourite animal is the loma! I like how it’s white with such big antlers!”), which aligns with early biological science content. Kory Westlund et al. (2017) found that even three-year-old children were able to learn novel animal names from robots by following non-verbal cues such as gaze direction and gesture, highlighting the potential of robots to support both verbal and non-verbal learning.

2.5.2 Human informants VS. robot informants

Another important direction of research in selective trust within CRI focuses on comparing the reliability of human and robot informants. Using a single informant paradigm, Cao et al. (2025) found that children aged 4 to 7, particularly those between 4 and 5 years old, were more likely to direct their questions to humanoid robots than to human informants. Several studies also employed the conflicting informants paradigm to assess children’s preferences when faced with contradictory information. For instance, children aged 3 to 6 were more likely to endorse labels from a previously accurate robot, even when a human informant was also reliable. Interestingly, they tended to interpret robot errors as accidental but judged human errors as intentional (Stower et al., 2024). Similarly, children aged 5 preferred learning new labels from a competent robot over an incompetent human, while acknowledging the robot’s mechanical nature (Baumann et al., 2023). These findings suggest that in selective trust, preschool-aged children often display a trust preference toward robots rather than humans, especially when the robots are perceived as accurate.

In contrast, mixed findings have emerged regarding children’s trust in human versus robot informants. On the one hand, younger children often show a preference for human informants. Li and Yow (2024) found that while five-year-olds trusted accurate robots and humans equally and distrusted inaccurate ones to a similar extent, three-year-olds were more likely to trust inaccurate humans over inaccurate robots. Similarly, in a trust game study, seven-year-olds displayed greater trust in robots, whereas three-year-olds

demonstrated a stronger preference for humans (Di Dio et al., 2020). On the other hand, the domain of testimony may play a crucial role in shaping children's trust. Children aged 3 to 6 preferred to consult robots for mechanical knowledge but were least likely to trust them for biological or psychological information (Oranç & Küntay, 2020). However, Kory Westlund et al. (2017) found no significant differences in children's word recall when novel animals were labelled by a human or a robot using identical non-verbal cues such as gaze and orientation. Taken together, the current body of research presents inconsistent patterns, and due to the limited number of studies focusing specifically on STEM-related testimony, it remains unclear whether children are more inclined to trust human or robotic informants in STEM learning contexts.

2.6 The role of ToM and ToAM on selective trust

2.6.1 The association between ToM and selective trust

A growing body of research has demonstrated a strong association between children's ToM abilities and their capacity for selective trust, suggesting that mental state reasoning plays an important role in guiding trust-related decisions. Several scholars argued that both ToM and selective trust typically emerge around the age of three or four (Ding et al., 2017). DiYanni et al. (2012) found that children's understanding of false beliefs was positively correlated with their ability to selectively trust reliable informants. Similarly, children with stronger ToM skills were more likely to favour an unfamiliar but reliable source over a familiar one with a history of inaccuracy (Palmquist et al., 2022). In another study, Brosseau-Liard et al. (2015) showed that children with higher ToM performance were more inclined to base their trust decisions on epistemic cues rather than superficial traits, such as physical strength. Evidence from infancy further supported this link: even at 18 months, early knowledge inference (a precursor to ToM), has been positively associated with selective trust (Crivello et al., 2021). Collectively, these findings suggested that ToM is a foundational cognitive capacity that enables children to evaluate the credibility of informants and make more informed learning choices.

However, the influence of ToM on selective trust may be moderated by additional factors. First, children aged 3 to 4 from middle socioeconomic backgrounds exhibited higher ToM scores but unexpectedly performed worse on selective trust tasks compared to their peers from lower socioeconomic backgrounds (Souza et al., 2021). Second,

inhibitory control which is a cognitive process that enables individuals to suppress automatic responses in favour of goal-directed actions has been shown to support trust regulation. Crivello et al. (2021) found that children with stronger ToM and inhibitory control were more capable of distrusting misleading social robots, suggesting that managing selective trust requires both mental state reasoning and self-regulation. Third, longitudinal findings indicate no stable predictive relationship between ToM and selective trust across early childhood. While these two abilities may develop concurrently, they appear to be functionally distinct. Notably, selective trust assessed in the first year was a significant positive predictor of ToM in the second year (Ding et al., 2017). In summary, the relationship between ToM and selective trust appears to be shaped by socioeconomic background, inhibitory control, and developmentally independent over time.

2.6.2 The association between ToAM and selective trust

Direct assessments of ToAM remain limited, with much of the existing literature instead discussing the concept of mental state attribution (MSA). Similar to ToAM, MSA is also coined as the ability to attribute mental states to others or technical agents (Thellman et al., 2022). Instead, it was evaluated through more explicit and decontextualised questions (e.g., "would the robot feel pain?") instead of contextual stories, implicit mental state inference and behaviour prediction in ToAM (e.g., in *Sally-Anne* task, children are required to reason Anne's next action based on implicit false belief) (Mao et al., 2025). Compared to ToAM, however, MSA research has predominantly addressed children's ontological understanding of robots. Thus, given the conceptual proximity between ToAM and MSA, and the current scarcity of dedicated ToAM research, the following discussion on the relationship between ToAM and selective trust also draws on empirical findings from MSA studies.

There is no clear consensus in the literature regarding the association between ToAM and selective trust, as findings have varied inconsistently. Above all, several studies have reported positive correlations. Li et al. (2024) conducted an explanatory study with three- to four-year-olds using a five-item ToAM scale encompassing desire, belief, false belief, and emotion components, alongside a single informant paradigm and an inhibitory control assessment. Results indicated that ToAM was positively associated with selective

trust ($r = 0.26$, $p < 0.05$), as was inhibitory control ($r = 0.48$, $p < 0.01$). Oranç and Küntay (2020) found that the more perceptual properties (e.g., seeing and hearing) children aged 3 to 6 attributed to a humanoid robot, the greater their tendency to trust it. Equally, children aged 3 trusted accurate social robots over inaccurate ones, and this trust increased when they perceived the robot as having human-like psychological agency (Brink & Wellman, 2020). In another study, children consistently attributed traits such as liking (e.g., "wants to be your friend") and competence (e.g., "is smarter") to the robot, while attributing agency (e.g., "acts on purpose") and blame (e.g., "makes mistakes") more frequently to the human informant (Stower et al., 2024).

On the contrary, two indirect but relevant studies have reported patterns suggesting a negative or more complex relationship. With a trust game called "Guess where it is", Di Dio et al. (2020) assessed children's trust in both human and robotic agents alongside their ToM and MSA. Findings indicated that children with higher ToM scores were more sceptical during the trust game, displaying reduced willingness to follow either informant's suggestion across ages 3, 5, 7, and 9. Notably, however, the study did not specifically analyse the relationship between MSA and selective trust. In a separate meta-analysis, Stower et al. (2021) examined children's trust in robots, distinguishing between social trust — defined as the belief that an agent will keep promises — and competency trust, which is more closely aligned with selective trust. Interestingly, the results showed that while a humanoid appearance increased children's social trust in robots ($g = 0.53$, 95% CI [0.28, 0.78]), it significantly reduced their competency-based trust ($g = -0.49$, 95% CI [-0.81, -0.17]). These findings suggest that while children may develop stronger social bonds with human-like robots, they may simultaneously question their informational reliability.

In addition to studies showing positive or negative patterns, some research has found no significant association between ToAM and selective trust. Baumann et al. (2023) used a parent-report ToM questionnaire assessing children's understanding of emotion, intention, desire, perception, knowledge, and belief in social interactions. Their findings indicated that individual differences in ToM did not significantly predict selective trust behaviour. Similarly, while children aged 4 to 8 relied on prior accuracy when deciding whether to trust DVAs or humans, their MSA scores did not significantly influence their trust or learning outcomes, even though the MSA items only included epistemic

properties such as “know the answer” or “unsure of the answer” (Girouard-Hallam & Danovitch, 2022). These findings signified that the association between ToAM and selective trust may not be consistently observable across different measurement tools and task designs.

2.7 Conclusion of the literature review

This literature review systematically examined key research areas relevant to children’s selective trust in human and robot informants, specifically within STEM and non-STEM contexts, and explored the role of children’s ToM and ToAM. Initially, the review conceptualised ToM and its developmental trajectory, emphasising the centrality of understanding diverse beliefs, desires, emotions, and particularly knowledge states. Previous work highlighted cultural differences regarding knowledge access and false belief understanding. Expanding upon this, the review further introduced ToAM, exploring how children mentalise artificial agents, such as DVAs and robots, which they increasingly encounter in educational and social environments. Notably, literature reflected significant parallels between ToAM and ToM. However, limited studies to date have rigorously investigated the full spectrum of ToAM toward robots across different ages and cultural contexts.

Subsequently, the literature on selective trust was critically reviewed, revealing that children from as early as three years old show nuanced trust behaviours based on informants’ epistemic and social cues, including accuracy, confidence, familiarity, and group consensus. However, while extensive research has investigated selective trust in human informants, fewer studies have systematically explored trust in robots, especially within cognitively demanding domains such as STEM. Prior studies suggested preschool-aged children demonstrate considerable trust in robots when these agents are accurate, though findings remain mixed regarding comparative trust levels in human versus robotic informants. Crucially, research specifically addressing children’s selective trust in STEM-related testimonies delivered by robots remains sparse.

While studies on children’s selective trust in robots have flourished in recent years across China, the United States, and Singapore, this area of research remains relatively limited and is still in its infancy. Firstly, limited research has directly examined the relationship

between children's ToAM and selective trust, even though selective trust behaviours can be interpreted as an explicit performance of knowledge access. While some researchers have acknowledged this connection and incorporated relevant assessments, very few have employed both standardised ToAM and ToM scales to systematically explore their association with trust-related outcomes. Certain studies have relied on implicit MSA (Brink & Wellman, 2020; Girouard-Hallam & Danovitch, 2022), while others excluded ToM measures altogether, even when human informants were involved (Stower et al., 2024; Li et al., 2024). More importantly, the relationship between ToAM and selective trust remains inconsistent across various experimental paradigms, underscoring the need for further empirical validation.

Furthermore, existing selective trust studies in the context of CRIs have primarily focused on object-labelling tasks, often neglecting more complex domains of social learning. Unlike simple object labels, STEM content is generally more cognitively demanding and less accessible for children to learn independently, yet it remains a central component of early childhood education (Danovitch et al., 2023). Reversely, selective trust in CRIs which concerned STEM knowledge mainly discussed the learning of animal names, neglecting more essential animal growth or basic math. Danovitch and colleagues (2023) found that children aged 4 to 6 showed no significant difference in their willingness to trust and transmit non-STEM object labels and mathematical information. However, unlike human informants, robots originate from STEM disciplines and are often perceived as products of technological advancement and the AI era. Indeed, children aged 5 to 6, particularly younger children, frequently view robots as mechanical devices rather than as agents with mental states (Katayama et al., 2010). Therefore, it remains unclear to what extent children are willing to trust a robot in transmitting STEM knowledge, given that the robot itself may be conceptualised more as a STEM object than as a knowledgeable informant.

To sum up, limited systematic examination of the relationship between children's ToAM, ToM, and selective trust in robots, particularly on their potential relationships. Moreover, there is an insufficient exploration of selective trust in robots beyond simple object-labelling tasks. While children may perceive robots as machines or the products of technological advancement, the current body of selective trust research has not compared human and robot informants when learning STEM knowledge, such as physics, biology,

and mathematics. To my knowledge, this is the first study to directly compare children's trust in robot versus human informants within the context of STEM-related testimony, offering novel insights into how cognitive reasoning supports trust in artificial agents about important knowledge.

2.8 Theoretical framework in this study

Overall, the conceptual framework illustrates how the advent of AI not only introduces new conceptual pathways (ToAM) but also reconfigures existing psychological processes (ToM and selective trust), with implications for how children evaluate and learn from both human and robotic informants. The diagram summarises three central constructs derived from the literature review: ToM, ToAM, and selective trust, represented respectively by the blue, green, and pink rectangles. Each construct is further delineated by its key subcomponents. ToM pertains to children's understanding of human agents in terms of desire, belief, knowledge, and emotion. With the increasing presence of artificial intelligence in children's environments, ToAM has emerged to describe how children attribute similar mental states to AI agents such as robots. The transition from ToM to ToAM is indicated by a solid single-headed arrow, reflecting how children begin to mentalise AI entities. As children generally learn from robots, ToAM than linked to selective trust, especially when the informant is an AI agent. The selective trust component is structured around three nested layers: the informant (e.g., human or robot), trust cues (e.g., accuracy for epistemic cues and familiarity for social cues), and the type of information (e.g., non-STEM and STEM).

As shown in Figure 1, a solid double-headed arrow connects ToM and selective trust, signalling a well-established reciprocal relationship in which mental state attribution informs trust (see section 2.6.1) and, conversely, engagement in trust scenarios may impact ToM. In contrast, a dashed double-headed arrow links ToAM and selective trust, representing a less conclusive body of evidence with mixed findings across studies (see section 2.6.2).

The framework is framed by two tiers of contextual factors. The white boxes, placed adjacent to each core construct, indicate proximal influences — variables that more directly shape the development and functioning of ToM, ToAM, and selective trust.

These include, for example, age and developmental stage, socioeconomic background, and cognitive ability such as inhibitory control. Surrounding it is a larger grey frame, which signals distal factors: broader socio-technical changes, such as the increasing integration of AI in education and childhood settings, which indirectly prompt the evolution of these constructs.

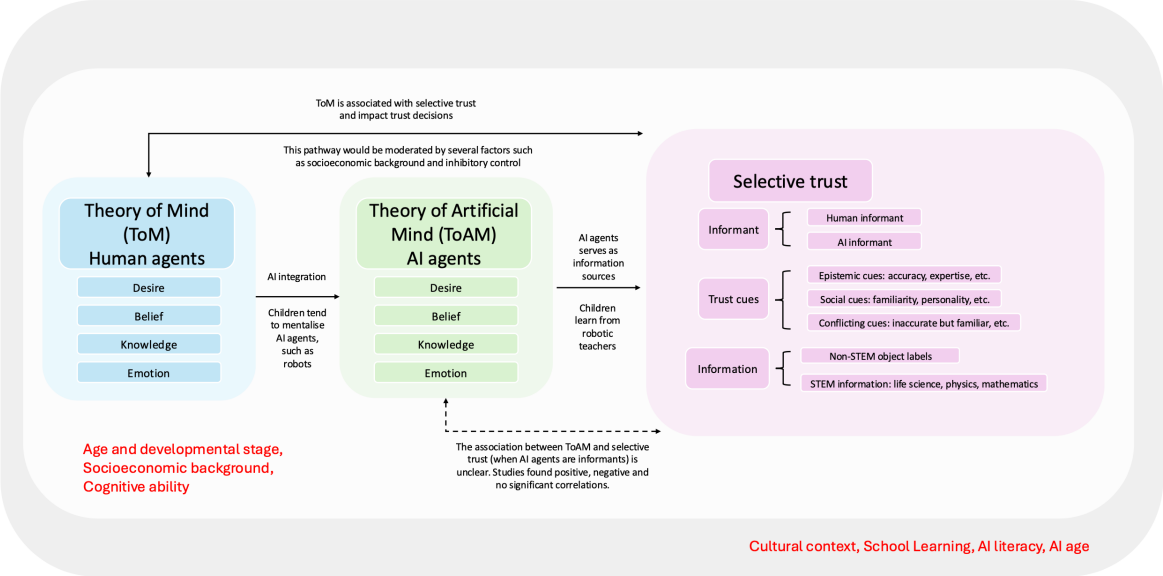


Figure 1 The theoretical framework which maps key concepts in the current study and point out a socio-cognitive scaffold.

Chapter 3 The present study

The present study aims to address research gaps by investigating how children aged 4 to 6 selectively trust a humanoid robot (Nao) versus a human informant across four domains of knowledge: object labelling (non-STEM), physical science, life science, and mathematics (STEM). Employing a conflicting informants paradigm, in which a robot and a human provide contradictory testimony, the study captures children's nomination preference, endorsement and accuracy judgement in a controlled setting.

The research adopts a mixed design, with accuracy condition (*Nao-accurate* vs. *Human-accurate*) as a between-subjects variable, and information domain (non-STEM, physical science, life science, and mathematics) as a within-subjects variable. This structure allows for the examination of both cross-condition differences in trust and within-child variation across knowledge domains. The primary outcome measures include children's responses to nomination, endorsement, and accuracy questions, which together reflect their selective trust patterns.

In addition, children complete validated scales of ToM and ToAM. These measures are used to explore the potential relationship between children's ToM/ToAM and selective trust. By analysing both implicit cognitive profiles and explicit behavioural choices, the study aims to clarify whether children apply similar trust strategies when evaluating artificial agents compared to human informants.

The findings may have significant educational implications. As robots and other AI agents become more integrated into early learning environments, it is essential to understand how children perceive and evaluate these non-human sources of knowledge. Insights from this research could inform the design of developmentally appropriate educational technologies and guide educators in supporting children's critical thinking about information and informants in both human and artificial forms. Moreover, findings in this study can be a theoretical basis for using robots in early STEM education.

Overall, this study proposes the following research questions and corresponding hypotheses:

RQ1 (research question 1): To what extent do children's selective trust in robots and humans differ?

RQ2: How does the domain of testimony (non-STEM versus STEM) influence children's selective trust?

RQ3: Do ToAM and ToM relate to children's selective trust in robot and human informants?

Regarding children aged 4 to 6 who have demonstrated sensitivity to epistemic cues such as accuracy and can distinguish accurately from inaccurately informed robots (Koenig et al., 2004; Brink & Wellman, 2020; Geiskkovitch et al., 2019), it is logical to expect this sensitivity to consistently apply irrespective of the informant type. It is hypothesised that children will show greater selective trust in the robot (Nao) when they think the robot is accurate, and greater trust in the human when they think the human is accurate, reflecting sensitivity to informant reliability regardless of agent type (H1).

RQ2 addresses domain-specific variations in selective trust. Research indicates that young children conceptualise robots as products of technology (Katayama et al., 2010) and are inclined to trust them for mechanical knowledge (Oranç & Küntay, 2020) and learning about habits of animals (Breazeal et al., 2016). Using expertise as an epistemic cue, children may view robots as experts in STEM and humans as experts in everyday knowledge, reflecting an early understanding of the division of cognitive labour. Thus, it is hypothesised that children will be more likely to trust the robot in STEM domains and more likely to trust the human in the non-STEM domain, indicating domain-specific preferences in informant trust (H2).

Finally, though the association between ToAM and selective trust in robots remains unclear in previous research, ToM is positively correlated with selective trust in human informants and varies by several factors. Mao et al. (2025) noted age-related differences in both ToM and ToAM development, suggesting age as a covariate. Consequently, it is hypothesised that children with higher ToAM scores will be more likely to trust an accurate robot, while those with higher ToM scores will be more likely to trust the

accurate human informant (H3). However, these associations are expected to be statistically controlled for children's age.

Chapter 4 Methods

4.1 Participants

Although most selective trust studies have focused on children aged 3 to 6, Danovitch et al. (2023) limited their sample to children aged 4 and above, noting that 3-year-olds may have difficulty understanding certain STEM-related testimonies. Considering the syllabus and learning aim of Chinese early childhood education, children aged 4 are generally confident in engaging with STEM knowledge. Based on simulations with a sample size of 100 and assuming a log-odds effect size of 0.80 for the condition × domain (2 × 4) interaction, the estimated statistical power was 48% (95% CI [37.9%, 58.2%]). Each knowledge domain included three trials, resulting in a total of 12 trials per participant. Although a sample of 100 participants is sufficient to detect moderate to large effects, the target sample size was increased to 120 children aged 4 to 6 to account for potential attrition due to participant withdrawal or absence. A detailed power analysis is provided in Appendix A.

Participant recruitment and data collection took place between April and May 2025. Initially, 120 children aged between 4 and 6 years were recruited from a kindergarten in southeastern China. Two participants were absent during the ToAM and ToM assessment session, while an additional two quitted due to difficulties comprehending the experimental procedures. A total of 116 children successfully completed the first session and proceeded to the subsequent selective trust session. However, six children were absent from the selective trust session due to illness, and three others withdrew as they were unable to understand the tasks. Ultimately, the final sample comprised 107 children aged between 4.62 and 6.61 years (M = 5.57, SD = 0.58, 48 (44.86%) girls). All of them completed the ToAM and ToM session. Further, participants were randomly allocated to one of two conditions in the selective trust session: the *Nao-accurate* condition (n = 51; M = 5.62, SD = 0.59; age range = 4.62 – 6.61; 22 (43.13%) girls), and the *Human-accurate* condition (n = 56; M = 5.52, SD = 0.58; age range = 4.65 – 6.61; 26 (52.34%) girls). Anecdotally, the city's residents consider the residential area where the kindergarten is located to be middle-class and the main residents are Han Chinese.

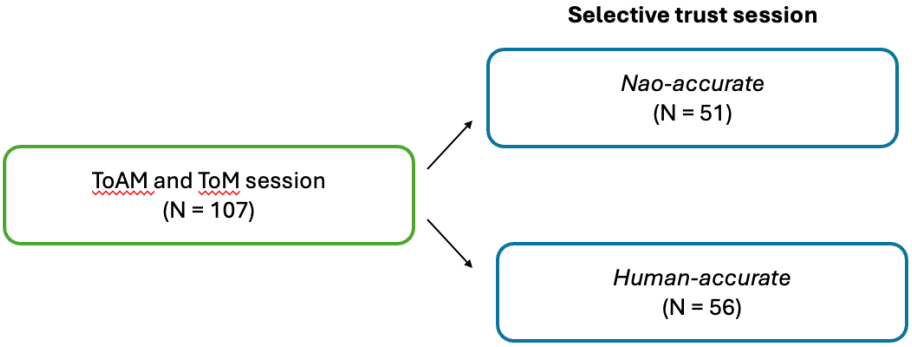


Figure 2 The assignment of participants.

4.2 Materials

4.2.1 Robot Nao

The present study selected a humanoid robot Nao (see Figure 3) as the main stimulus in both ToAM and ToM and selective trust sessions. Developed by Softbank robotics, Nao is a fully programmed humanoid robot widely applied in education, psychology and health care. Designed by SoftBank Robotics, Nao is 58cm tall and has a humanoid appearance with a head, torso, arms, and legs, allowing it to mimic human movements and gestures. Standing about 58 cm tall, Nao is equipped with a wide range of sensors, including cameras, microphones, and tactile sensors, allowing it to perceive and interact with surrounding environment. It has 25 degrees of freedom, enabling realistic and dynamic movement. Nao is the robot's official name and does not have an inherent gender. Throughout the study, researchers used the pronoun "it" to refer to Nao, as the Chinese pronouns for it (它), he (他), and she (她) are all pronounced "ta."

Instead of deploying the authentic robot, a static image of Nao accompanied by synthesised speech was used in each session for several reasons. First, this approach ensured consistency across participants by eliminating variability that might result from real-time behavioural fluctuations or potential malfunctions of the robot. All children were exposed to identical stimuli, allowing the assessment to proceed smoothly and reliably. Second, due to practical constraints, including the robot's weight (11.9 pounds) and the substantial costs associated with purchasing, transporting, maintaining, and programming a robot, the digital representation offered a more feasible and accessible

alternative. Third, prior research on children's perceptions of and trust in artificial agents has successfully employed screen-based stimuli (e.g., Nao in Banks, 2020 and Geiskkovitch et al., 2019; DVAs in Zhou et al, 2025 ; non-humanoid robot in Katayama et al, 2010). Such approaches were grounded in the existing literature and have produced robust and meaningful findings. Therefore, this research utilised the image due to the benefits of enhanced consistency and stability, limited resources and funding, and the methodological validity demonstrated by prior studies

Figure 3 The official image of robot Nao.

Note: The photo is shot by Softbank robot [https://robotsguide.com/robots/nao].

4.2.2 ToAM and ToM assessment

The materials and procedure for the ToAM and ToM assessments were adapted from a previous work (Mao et al., 2025). Similar to the settings in Mao et al. (2025), children in the current study completed both the ToM and ToAM scales. The ToM scale, originally developed by Wellman and Liu (2004), has been widely applied across diverse cultural contexts (Wellman, 2010). The ToAM version was adapted by Mao et al. (2025), maintaining the core structure and descriptions. Both scales comprised seven items: Diverse Desires, Diverse Beliefs, Knowledge Access, Contents False Belief, Explicit False Belief, Belief-Emotion, and Real-Apparent Emotion. The distinction between the two scales lies in the character involved — human for ToM and robot for ToAM. As shown in Table 1, items increased in difficulty, and children received one point for each completely correct response, yielding a total score ranging from 0 to 7. Task order and stimulus presentation were randomised to minimise bias. Each story was narrated aloud by the researcher and accompanied by contextually appropriate images.

Table 1 The description of ToAM and ToM items

Task	Description
Diverse Desires	Judging whether two persons or two robots have different desires regarding the same objects.
Diverse Beliefs	Judging whether two persons or two robots have different beliefs regarding the same objects, when the participant does not know which belief is true or false.
Knowledge Access	After seeing inside a box, the participant judges (yes-no) the knowledge of another person or robot who does not see what is in the box.
Contents False Belief	After seeing inside a box, the participant judges another person's or robot's false belief about what is in a distinctive container.
Explicit False Belief	Judging how a person or a robot will search with mistaken belief.
Belief-Emotion	Judging how a person or a robot will feel with mistaken belief.
Real-Apparent Emotion	Judging whether a person or a robot can feel one thing but display a different emotion.

Note: The description is quoted from Mao et al. (2025).

4.2.3 Informants in the selective trust

To investigate children's selective trust in robots, the study used the Nao robot as one of the informants across four knowledge domains. Robots in pink, blue, grey, and orange represented the non-STEM, physical science, life science, and mathematics domains, respectively. Given the influence of ToAM and ToM, and the central role of human informants in real-life learning, four Chinese female volunteers (M = 25.09, SD = 2.45, all were Han Chinese) were recruited to serve as human informants. To minimise colour-based preference bias, each human informant wore clothing that matched the corresponding robot's colour (see Figure 4). While Robots were presented as static images with synthesised speech, human informants appeared in videos, speaking in a calm tone without moving their bodies or limbs. As informants' prior accuracy and knowledge states may influence children's judgements, each robot – human dyad was introduced at the start of learning in each new domain. To support domain-specific character recognition, each informant was assigned a unique name, with robots named after their colour (e.g., "pink") and humans possessed nicknames (e.g., "Yueyue"). This helped children identify each dyad as a distinct character within its respective domain.

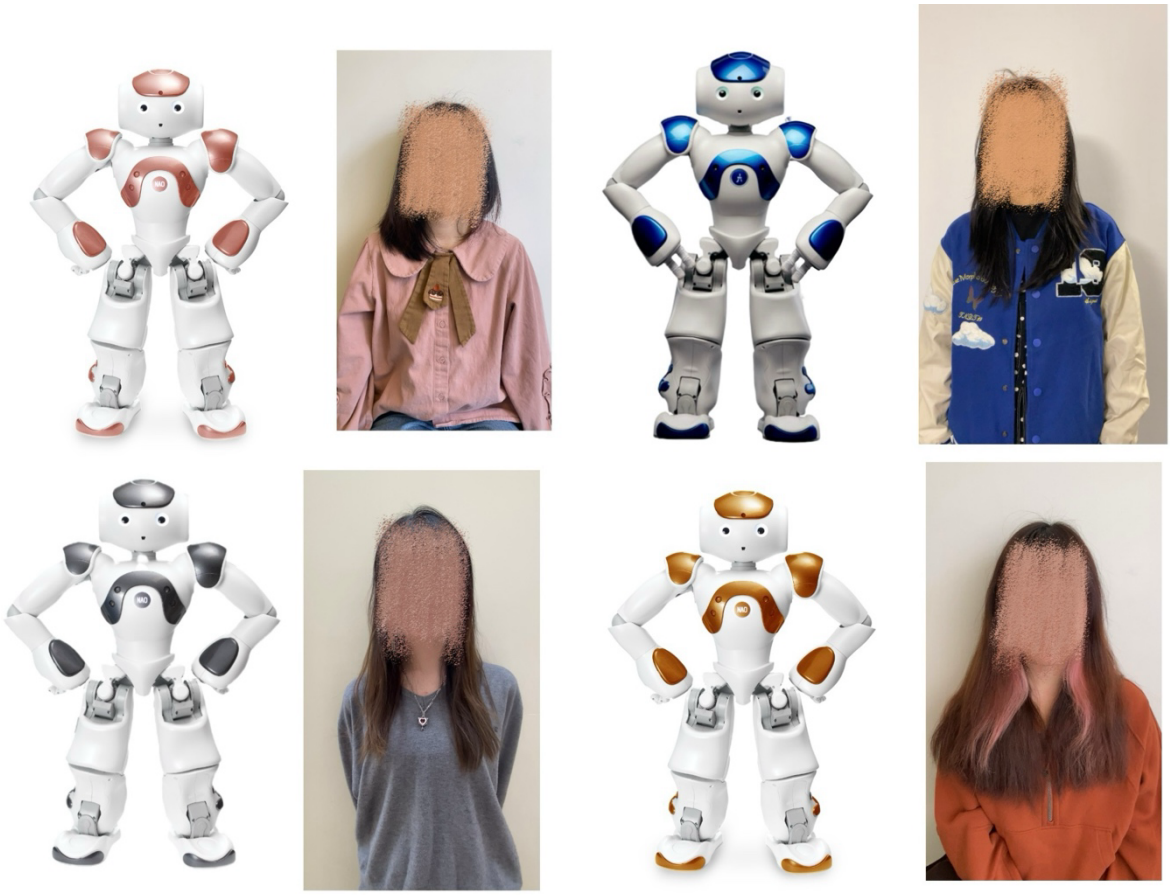


Figure 4 The four dyads of informants

Note: The image shows four dyads of informants (robot – human), each representing a different domain. Top left: non-STEM (pink – Yueyue); top right: physical science (blue – Shanshan); bottom left: life science (grey – Xixi); bottom right: mathematics (brown – Junjun). Each dyad includes a Nao robot (left) and a female human agent (right), presented as distinct characters with matching colours.

4.2.4 Information in the selective trust task

The non-STEM materials consisted of familiar everyday objects, aligning with the non-STEM materials in Danovitch et al. (2023). In the test phase, novel objects created by Li et al. (2023) were used, each labelled with a nonsensical, reduplicative pseudoword. Prior to the experiment, a separate group of 30 children (not involved in the main study) assessed the materials. All reported no preference for the artificial labels and were unfamiliar with the novel objects (Li et al., 2023). On the other hand, the STEM materials were adapted from Danovitch et al. (2023). During the familiarisation phase, children were presented with common STEM knowledge. In the test phase, they encountered novel and nonsensical items representing foundational concepts in physical sciences, life sciences, and mathematics, in line with U.S. and China early education

guidelines. Table 2 and Table 3 exhibit the materials and corresponding statements used during the familiarisation and test phases. All materials were displayed onscreen via PowerPoint during formal data collection, with both robot and human voices used to convey the information.

Table 2 Materials and statements in the familiarisation phase

Subdomain	Image	Statement one	Statement two
Non-STEM	Cup	This is called a cup.	This is called a shoe.
	Key	This is called a key.	This is called a spoon.
	Ball	This is called a ball.	This is called a book.
Physical Science	Ice cube melting	Ice melts when it gets hot.	Ice melts when it gets cold.
	Pile of dirt	Dirt gets wet in the rain.	Dirt gets dry in the rain.
	Broken glass	Glass breaks when hit by a rock.	Glass breaks when hit by a feather.
Math	3 dogs, 1 cat	There are more dogs than cats.	There are more cats than dogs.
	Tall girl, short boy	The girl is taller than the boy.	The boy is taller than the girl.
	Circle and triangle	Circles are round.	Triangles are round.
Life Science	Turtle	Turtles have a shell on their back.	Turtles have wings on their back.
	Bird	Birds fly in the air.	Birds fly in the water.
	Plant	Plants need water to grow.	Plants need candy to grow.

Table 3 Materials and statements in the test phase

Subdomain	Image	Statement one	Statement two
Non-STEM		Kaka	Hoho
		Biubiu	Kuku

		Soso	Nunu
Physical Science	Red object	This floats in water.	This sinks in water.
	White object	This turns red in sunlight.	This turns blue in sunlight.
	Black object	This is soft on the bottom.	This is hard on the bottom.
Math	Brown object	This is bigger than a shoe.	This is smaller than a shoe.
	Orange object	This is heavier than an apple.	This is less heavy than an apple.
	Bucket	There are two oranges in the bucket.	There are four oranges in the bucket.
Life Science	Animal that looks like a chubby frog	This lives in trees.	This lives in the ground.
	Animals that looks like a possum	This sleeps only at night.	This sleeps only during the day.
	Animal that looks like an armadillo	This has many sharp teeth.	This has no teeth.

4.3 Procedure

At the outset, the researcher explained the study procedure to the children and distributed information sheets and consent forms to their parents. Only children with authorised parental consent and oral assent were included in the study. Testing took place in a quiet room within the kindergarten, where each child completed a one-on-one session with the researcher, free from disruptions or distractions. During the first two weeks, children completed both the ToM and ToAM scales. In the following two weeks, children participated in the selective trust tasks. In the selective trust session, children were randomly assigned to one of two conditions: the *Nao-accurate* or *Human-accurate* condition. In accordance with ethical guidelines and setting rule, participating children did not receive any gifts or rewards after tests.

4.3.1 ToAM and ToM assessment

This session followed the procedure outlined in Mao et al. (2025). It began with a warm-up activity in which children were introduced to two main characters — a female human and the Nao robot — through a series of slides accompanied by audio recordings. These recordings, using either authentic or AI-simulated voices, provided each character's name and distinctive traits. To ensure comprehension, children were asked to recall and name each character, and recordings were replayed as needed until the correct responses were given. Following this introduction, participants completed the ToM tasks and their adapted ToAM counterparts, based on the validated scale developed by Wellman and Liu (2004). The child received one point for each task completed with a fully correct response. Using ToAM Explicit False Belief task as an example, the full assessment procedure is illustrated in Figure 5.

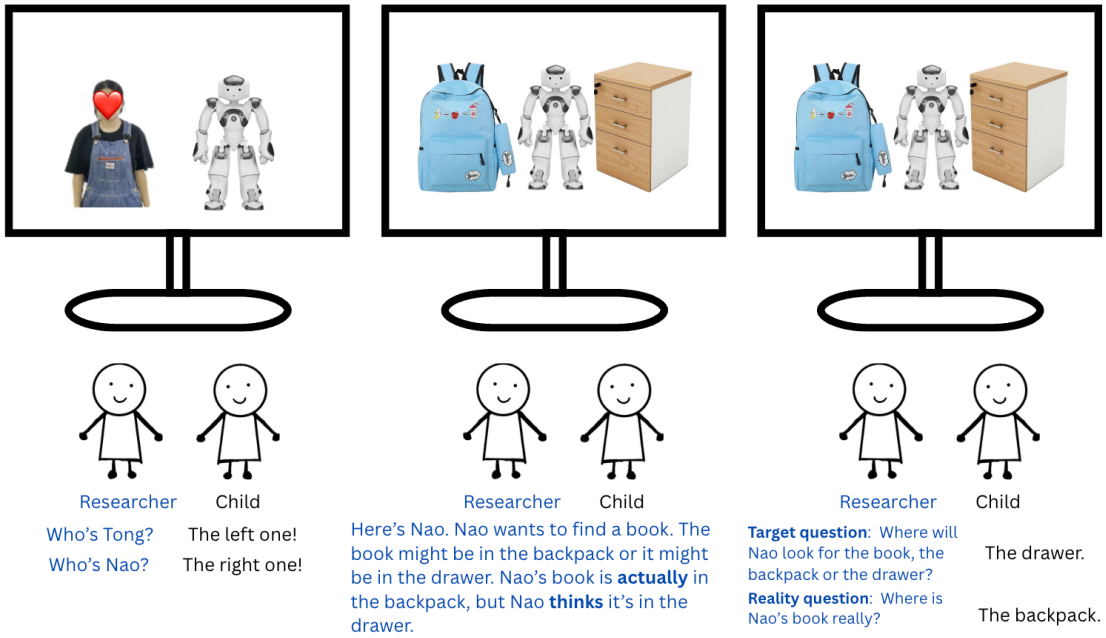


Figure 5 The procedure of Explicit False Belief task in ToAM

Note: This flowchart illustrates the warm-up and ToAM Explicit False Belief task. In the warm-up (first screen), children are introduced to the two characters — Tong (human) and Nao (robot) — and asked to identify them. In the Explicit False Belief task (second and third screens, only displays ToAM assessment), Nao is looking for his book, which is actually in the backpack, though he believes it is in the drawer. Children answer a **target question** assessing Nao's belief and a **reality question** to confirm factual understanding. The ToM assessment is similar.

4.3.2 Selective trust

Selective trust tasks adapted the well-established conflicting informants paradigm from Brink and Wellman (2020) and Koening et al. (2004). Children were randomly assigned to one of two between-subjects conditions: *Nao-accurate* or *Human-accurate*. In the first condition, Nao 100% correctly label the objects and answer the questions, while human is 0% accurate across four domains. In the second condition, vice versa. In each condition, children were presented with pairs of informants — one Nao robot and one human — who provided conflicting information. Each dyad represented a specific domain (non-STEM, physical science, life science, or mathematics), and informants were distinguished visually by colour. Overall, the position of each informant (left or right) and their speaking order (first or second) were counterbalanced across participants to minimise order and position bias. Figure 6 outlines the selective trust task using a non-STEM object labelling example in a *Nao-accurate* condition.

Familiarisation phase In this phase, each robot – human dyad provided information across the four domains. For the non-STEM labelling tasks, each informant labelled four familiar items. One informant consistently gave accurate labels, while the other provided incorrect ones. After each trial, children answered a name-check question to assess their judgement. For example, in the *Nao-accurate* condition: “Pink said it’s a cup, and Yueyue said it’s a shoe. What do you think it’s called?” In the three STEM subdomains, the procedure followed the same structure. For instance, in the *Human-accurate* condition, a question might be: “Grey said turtles have wings on their back, and Xixi said turtles have a shell on their back. What do you think is true?” In addition, only children who correctly named each informant advanced to the test phase.

Test phase In the subsequent test phase, each robot – human dyad labelled novel items using pseudowords, starting with the non-STEM domain. Children were first asked to select which informant they wished to consult regarding the novel object (**nomination question**), for example: “Which one do you want to ask, Nao or Yueyue?” Both informants then labelled the same object using different pseudowords (information provided). In each condition, one informant was consistently accurate (100%) based on the familiarisation phase, while the other was consistently inaccurate (0%).

Following the information presentation, children were asked to endorse one of the labels (**endorsement question**), such as: "Nao said it's a Kaka, and Yueyue said it's a Hoho. What do you think it's called? A Kaka or a Hoho?" Finally, they were asked to judge which informant was more trustworthy (**accuracy question**), for instance: "Which one is good at answering this question?" To sum up, in each trial, children responded to one nomination question, one endorsement question, and one accuracy question. This procedure was identical across all four domains, with content tailored to the respective subject area.

Familiarisation-test loop As noted earlier, children were presented with knowledge from four domains. Each domain followed a familiarisation-test loop, meaning that participants first completed both the familiarisation and test phases for one domain before proceeding to the next. The order of domains was randomised across participants to minimise possible order effects. Before each new domain, the researcher emphasised that the robot – human dyad was different from the previous one, introducing them with distinct names to reinforce their uniqueness.

Coding For analysis purposes, children's choices were coded as follows: selecting Nao was recorded as "1" and selecting the human informant as "0" for nomination, endorsement and accuracy questions. This coding was used across both *Nao-accurate* and *Human-accurate* conditions to ensure consistency between conditions and information domains. Although traditional selective trust paradigms typically code accuracy-based choices as "1", this study adopted a consistent agent-based coding approach to support cross-domain and cross-condition comparisons. In contrast, responses to the name-check questions during the familiarisation phase served solely to help children become familiar with the informants and recognise their relative accuracy.

These responses were not recorded for analysis.

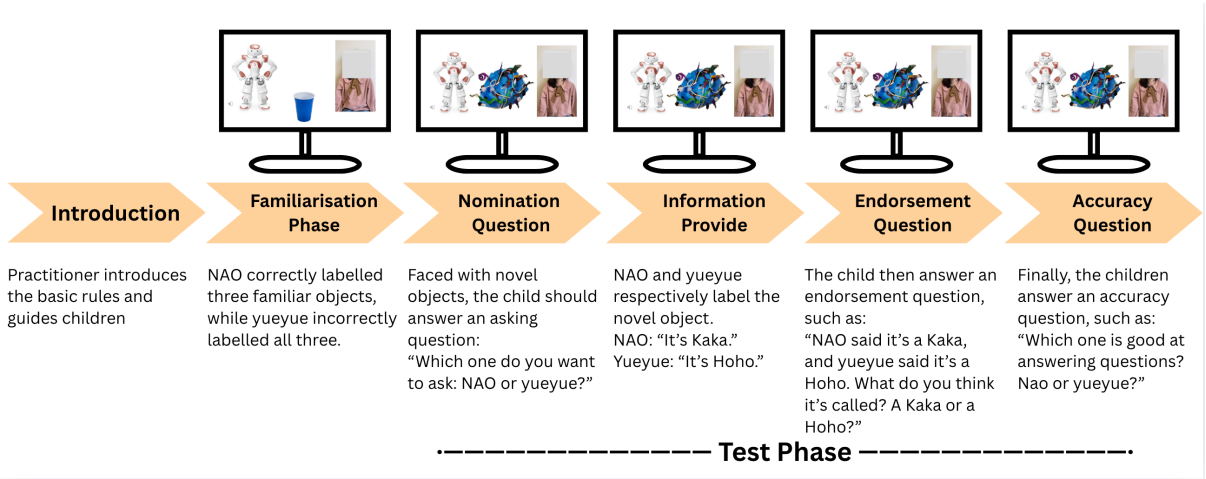


Figure 6 The procedure of selective trust in non-STEM domain

Note: This sequence assesses children's trust and decision-making. After the introduction of rules, children complete a familiarisation phase where Nao correctly labels three familiar objects, while Yueyue labels them incorrectly (*Nao-accurate* condition). In the test phase, children are shown a novel object and asked whom they want to ask (**nomination question**). Each informant provides a different pseudoword. Children then answer an **endorsement question** and an **accuracy question**. Once children completed the familiarisation-test loop in one domain, they proceeded to the next domain.

4.4 Ethical Consideration

4.4.1 Basic ethical considerations

To protect participants' privacy and rights, I referred closely to the Ethical Guidelines for Educational Research as set out by the British Educational Research Association (BERA, 2024). BERA emphasises that children should be treated with dignity, fairness, and respect, and that their participation should be developmentally appropriate and free from coercion or harm. The participated children are vulnerable due to their age and limited capacity to give informed consent. Therefore, all procedures were designed to be developmentally appropriate and ethically sound. Warm-up sessions were used to build rapport, while all questions were simplified and supported with visuals to reduce cognitive load. Furthermore, the study complies with relevant Chinese laws on the protection of minors and strictly avoids exclusion based on protected characteristics such as race, religion, sex, or sexual orientation. Only age and disability were considered as justified criteria in relation to study eligibility.

The current study received ethical approval from the Education Departmental Research Ethics Committee (DREC) at the University of Oxford (Ethics reference: Education (Educ) DREC – 1433726). The project is categorised under Approved Procedure 25 as outlined by the Central University Research Ethics Committee (CUREC). This procedure governs research involving children aged 3 to 16 years, provided that non-invasive methods are used and participants are recruited via an organisation, such as a school or early years setting. The study fully adhered to the University of Oxford's regulations and policies relating to research involving human participants, personal data, and vulnerable groups, ensuring that appropriate safeguards were in place throughout. Overall, the ethical approval and relevant materials are shown in Appendix I to L.

4.4.2 Consent and assent

To secure access and support, I submitted a cover letter to the headteacher of the participating kindergarten, outlining the research objectives, methodology, and ethical safeguards. The letter clearly described how data would be collected, stored, and anonymised, and emphasised the voluntary nature of participation for both children and their parents. As a result, the headteacher and teaching staff gave full approval and support for the research. Recruitment and data collection were conducted in close collaboration with the kindergarten, ensuring that all activities were aligned with the school's safeguarding practices and the children's everyday routines.

In line with CUREC guidelines, which stress the importance of informed consent and age-appropriate assent, the study followed a two-stage consent process. Children were introduced to the study using simple, developmentally appropriate language, while parents received a physical information sheet and consent form. The information sheet clearly outlined the study's purpose, procedures, data protection measures, and participants' rights. Written informed consent was obtained from parents or legal guardians, and oral assent was secured from each child before participation. The voluntary nature of participation was emphasised, and children were reminded that they could withdraw at any time without consequence. Although no objections were raised during the sessions, several children chose to discontinue due to the cognitive demands of the task, and their decisions were fully respected.

4.4.3 Data Storage

In line with BERA and CUREC guidelines, I took care to ensure the secure handling and storage of personal data. All participant information was anonymised at the point of analysis to protect privacy. During the dissertation project period, a fully anonymised version of the dataset will be stored in Nexus 365 OneDrive for Business, accessible only to the researcher and supervisor. Upon project completion, the data will be transferred to the Oxford Research Archive (ORA) for secure long-term storage and potential future secondary analysis, in accordance with institutional data management policies. No information related to participants' health, physiological data, financial details, or organisational records was collected.

4.5 Travel and fieldwork Risk assessment

Following the procedures of the Department of Education at the University of Oxford, I completed and submitted a comprehensive Travel and Fieldwork Risk Assessment, which has been formally approved. Although conducting research outside of Oxford, I was working in my home country of China, where I am a native speaker and deeply familiar with the local culture, language, and law. This familiarity gives me great confidence in navigating the environment safely and respectfully. I remained in regular contact with my supervisor and have local emergency support from family. I hold valid university travel insurance, and all COVID-19 and public health measures are strictly followed. No high-risk activities or politically sensitive topics are involved in this child-friendly, ethically sound study.

4.6 Statistical analyses

4.6.1 ToAM and ToM assessment

The dataset of the final sample (N = 107) proceeded in two stages: descriptive and inferential analyses. First, descriptive statistics were calculated to summarise participants' demographic information and overall performance across the ToM, ToAM, and selective trust tasks. For the ToM and ToAM assessments, binary scores (1 = correct, 0 = incorrect) were computed for each of the seven tasks, with total scores ranging from 0 to 7. These scale scores were treated as continuous variables and described using measures of central tendency and dispersion, including the mean and standard deviation.

Distribution characteristics were assessed using skewness and kurtosis values, and the Shapiro-Wilk test was applied to evaluate the normality of ToAM and ToM score distributions.

Further, to examine whether children's performance on ToAM and ToM tasks was influenced by stimulus type (robot versus human), a binary logistic regression analysis was conducted (Peng et al., 2002). Logistic regression predicts the probability of a binary outcome based on explanatory variables and estimates how a one-unit change in a predictor affects the odds of that outcome occurring. In this study, logistic regression allowed for the estimation of how the type of stimulus affected the log-odds of producing a correct answer. This approach is appropriate for binary outcome variables and estimates the probability of a correct response (coded as 1) based on the explanatory variable "stimulus." Specifically, robot Nao was set as the reference category, enabling direct comparison with the human condition.

4.6.2 Selective trust

For the selective trust data, measures including the nomination, endorsement, and accuracy questions (coded as 1 = Nao chosen, 0 = human chosen) were summarised using proportions across four domains (non-STEM, physical science, life science and mathematics) and two conditions (*Nao-accurate* and *Human-accurate*). Accordingly, in the following analyses, a positive coefficient indicates a greater likelihood of choosing the Nao, whereas a negative coefficient indicates a greater likelihood of choosing the human. These proportions were further compared against chance (50%) using chi-square goodness-of-fit tests. Given the categorical and discrete nature of the data, the chi-square test was more appropriate and robust than a one-sample t-test, which requires continuous and normally distributed variables. The test was applied within each combination of condition and domain, enabling a basic evaluation of children's trust behaviour in relation to the informant's accuracy and the type of knowledge presented.

Different from the ToAM and ToM assessment, although the response of selective trust (the dependent variable) is binary, there exists a random effect of repeated measure since we asked the same person three times (three trials) in one domain. Linear mixed model (LMM) allows us to calculate the random effect but fails to explore the binary variable.

Generalised Linear Model (GLM), such as logistic regression, handles non-normal response variables. Thus, combining the effects of LMM and GLM, generalised mixed model (GLMM) is applied. GLMMs are an extension of GLMs and are suitable for the analysis of non-normal data with a clustered structure (Tuerlinckx et al., 2006). It is acceptable to evaluate how condition (*Nao-accurate* or *Human-accurate*), information domain (non-STEM, physics, life science and mathematics), the scale score of ToAM/ToM contribute to children's selective trust responses.

Prior to analysis, key assumptions of binomial GLMMs were reviewed. First of all, each outcome variable (nomination question, endorsement question, and accuracy question) was binary and appropriately modelled using a binomial distribution with a logit link. In the model developing, a part of the R command "family = binomial" not only specified the binomial nature of the outcome, but also set the default with a logic link. Secondly, as responses were repeated measures, the data were structured across two levels: multiple questions nested within each participant and multiple trials within each domain. Since observations from the same participant may be correlated, random effects for participant ID and trial were included to account for this dependency. Thus, the models comprised both fixed and random effects. Last but not least, checks for multicollinearity and overdispersion were conducted for each model, confirming adequate model fit. Detailed results are reported in the Appendix E.

Furthermore, pairwise post hoc comparisons were conducted using estimated marginal means with Tukey's adjustment to control the familywise error rate (the probability of making a Type I error among a specified group) across multiple domain comparisons (Nicholson et al., 2022). From another angle, Tukey test can compare multiple within-subject groups and needs equal or near equal sample size in each group (Ryan, 1959). Under each condition, participants are informed all four domains of knowledge, resulting in an equal group. Tukey test, in this study, controlled Type I errors and was suitable for pairwise comparison.

Effect size quantifies the magnitude of a relationship or difference in a standardised way, enabling comparisons across studies or experiments. Unlike p-values, which indicate whether an effect exists, effect size provides practical significance and insight into the strength or importance of the effect (Fritz et al., 2012). R^2 is a statistical measure that

represents the proportion of variance in the dependent variable which is explained by the independent variables (Ozili, 2023). Therefore, the study applied R^2 to assess the effect size of logistic regression models and GLMMs. For the post-hoc test, one widely used effect size is Cohen's d. It standardises the difference between two means relative to their pooled standard deviation (Cohen, 1992). Cohen's *d* is a measure of statistical performance, with outstanding advantages such as single measure, easy-computered and independent of the cutpoint of the test (Hasselblad & Hedges, 1995). Here, outcome is binary and using GLMM with binomial family, the model returns log-odds instead of automatically returning Cohen's d. A commonly accepted approximate Cohen's *d* is log-odds ratio divided by 1.81 (Chinn, 2000). Since this approximate calculation is validated in several studies (e.g., Salgado, 2018), the current study also applied.

4.6.3 The association between ToAM, ToM and selective trust

The phi coefficient was used to assess the association between two binary variables (Akoglu, 2018). When both variables are coded as 0 and 1, the phi coefficient is equivalent to the product-moment correlation coefficient. In fact, a Pearson correlation calculated between two binary variables yields the same value as the phi coefficient (Guilford, 1936), making phi a straightforward and streamlined alternative. Accordingly, this study employed the Pearson correlation test to examine the correlations between the ToAM and ToM scales, as well as the individual binary-scored items within each scale.

To address the RQ3, a correlation analysis was conducted to examine how children's ToAM and ToM scores were related to their selective trust in both robots and humans. Unlike the phi coefficient matrix, which focused on individual binary-scored items, this analysis used total ToAM and ToM scores, which were continuous but not normally distributed, along with binary responses from the nomination, endorsement, and accuracy questions. A point-biserial correlation was selected as the most appropriate method. This statistical approach is a special case of Pearson's product-moment correlation and is suitable when one variable is continuous and the other is dichotomous (Kornbrot). It does not require both variables to be normally distributed or to have a linear relationship, making it well suited to the characteristics of the current data.

Notably, given the developmental nature of ToM and ToAM from ages 3 to 6, as demonstrated by Mao et al. (2025), the current study employed partial correlation analyses using for both the phi coefficient and point-biserial correlation methods. Partial correlation estimates the relationship between two variables while statistically controlling for the influence of a third variable (Epskamp & Fried, 2018) — in this case, age. This approach was selected to account for age-related variance, which, while developmentally important, was not the primary focus of the current study. This adjustment enhances the interpretability of the cognitive overlap between ToM, ToAM, and selective trust, independent of developmental variability.

Chapter 5 Results

5.1 Descriptive analysis

5.1.1 ToAM and ToM assessment

Table 4 displays proportions of correct responses and standard deviations for each of the seven tasks on the ToAM and ToM scales. Children's mean score on the ToAM scale was 4.48 (SD = 1.91), while their mean score on the ToM scale was 3.91 (SD = 2.17), suggesting an ability to attribute mental states to robots and humans in this sample. Shapiro-Wilk tests were performed and revealed that both ToAM ($W = 0.91$, $p < 0.001$) and ToM scale scores ($W = 0.92$, $p < 0.001$) significantly deviated from a normal distribution. ToAM displayed slight negative skewness (-0.17) and low kurtosis (-1.28), while ToM was symmetric (skewness = 0.00) but also platykurtic (kurtosis = -1.32). Although both distributions appear approximately symmetric, the combination of statistical significance and low kurtosis values suggests that the variables deviate from normality, particularly in terms of their tail distributions. For more details in descriptive analysis, see Appendix F and Appendix G.

Specifically, for both scales, largest proportion correctly answered early-developing items: Diverse Desires, Diverse Beliefs, and Knowledge Access. Prior evidence suggested that most children aged 3 successfully completed the first two items on both the ToM and ToAM scales (Wellman & Liu, 2004). By around age 5, children began to understand that both humans and robots can hold knowledge states that differ from their own (Mao et al., 2025). As items increased in complexity, correct response rates decreased. This trend was especially evident in the final three tasks — Explicit False Belief, Belief-Emotion, and Real-Apparent Emotion — which require more advanced reasoning. Notably, Real-Apparent Emotion had the lowest accuracy in both conditions, with a correct response rate of only 0.24 for both ToAM and ToM, indicating that children had particular difficulty understanding that an agent (human or robot) might feel one emotion while displaying another on face.

Table 4 Mean of correct responses and standard deviations for ToAM and ToM scale and items

	correct proportion	M	SD	skewness	kurtosis
ToAM					
Diverse Desires	97.20%				

Diverse Beliefs	85.05%				
Knowledge Access	77.57%				
Contents False Belief	57.01%				
Explicit False Belief	53.27%				
Belief-Emotion	53.27%				
Real-Apparent Emotion	24.30%				
total		4.48	1.91	-0.17	-1.28
ToM					
Diverse Desires	76.64%				
Diverse Beliefs	73.83%				
Knowledge Access	64.49%				
Contents False Belief	53.27%				
Explicit False Belief	50.47%				
Belief-Emotion	47.66%				
Real-Apparent Emotion	24.30%				
total		3.91	2.17	0.00	-1.32

5.1.2 Selective trust

Table 5, Table 6, and Table 7 presented the proportion of choices in three questions. To assess whether children's selective trust significantly deviated from chance (50%) across conditions and domains, chi-square goodness-of-fit tests were conducted. This non-parametric test was used to evaluate whether the observed frequencies of selecting either Nao or the human informant differed from an equal 50/50 distribution, which would suggest random responding. Appendix H plots the proportion of choosing Nao in each measure across domain and condition.

Children's responses to the **nomination question** were distinct across domains and conditions (see Table 5). In the *Nao-accurate* condition, children selected Nao significantly more often than would be expected by chance across all domains, indicating strong preference for the previously accurate informant. In the *Human-accurate* condition, children preferred to ask information from human more than Nao in most domains, except for the physical science, where choices did not differ significantly from chance ($\chi^2 = 0.02$, $p = 0.877$).

Table 5 Response to nomination questions in Nao vs. human informants across domains and conditions

	Nao-accurate				Human-accurate			
Domain	Nao	human	χ^2	p-value	Nao	human	χ^2	p-value
Non-STEM	71.90%	28.10%	29.3	<0.001	35.71%	64.29%	13.7	<0.001
STEM								
physics	62.75%	37.25%	9.94	0.002	50.60%	49.40%	0.02	0.877
life science	74.51%	25.49%	36.8	<0.001	38.10%	61.90%	9.52	0.002
mathematics	75.82%	25.18%	40.8	<0.001	59.52%	40.48%	6.10	0.014

Note: All tests were performed with 1 degree of freedom (*df* = 1).

Chi-square goodness-of-fit tests were conducted to determine whether children's choices on **endorsement question** differ from chance (see Table 6). In both *Nao-accurate* and *Human-accurate* conditions, children significantly endorse the accurate informants across all four domains.

Table 6 Response to endorsement questions in Nao vs. human informants across domains and conditions

	Nao-accurate				Human-accurate			
Domain	Nao	human	χ^2	p-value	Nao	human	χ^2	p-value
Non-STEM	86.93%	13.07%	83.5	<0.001	22.62%	77.38%	50.4	<0.001
STEM								
physics	74.51%	25.49%	36.8	<0.001	36.31%	63.69%	12.6	<0.001
life science	71.24%	28.76%	27.6	<0.001	32.74%	67.26%	20.0	<0.001
mathematics	72.55%	27.45%	31.1	<0.001	27.98%	72.02%	32.6	<0.001

Table 7 displayed whether children's judgements on **accuracy question** deviated from random responding. Briefly, in both conditions, children consistently trusted the accurate informant, whether Nao or a human, across all information domains.

Table 7 Response to accuracy questions in Nao vs. human informants across domains and conditions

	Nao-accurate				Human-accurate			
Domain	Nao	human	χ^2	p-value	Nao	human	χ^2	p-value
Non-STEM	88.89%	11.11%	92.6	<0.001	27.38%	72.62%	34.4	<0.001
STEM								
physics	77.12%	22.88%	45.0	<0.001	39.88%	60.12%	6.88	0.009
life science	75.16%	24.84%	38.8	<0.001	34.52%	65.48%	16.1	<0.001
mathematics	75.16%	24.84%	38.8	<0.001	27.98%	72.02%	32.6	<0.001

5.2 Inferential analysis in ToAM and ToM assessment

Key assumptions of logistic regression were also addressed (Peng et al., 2002). First, although logistic regression does not assume a linear relationship between predictors and the outcome, it assumes that any continuous predictors have a linear relationship with the logit-transformed outcome. This is a key assumption introduced by the logistic link function, which maps the linear combination of predictors onto a probability bounded between 0 and 1 (MacKenzie et al., 2018). As the present model did not include continuous predictors, this assumption was not applicable. Second, the assumption of independent errors was satisfied because each child contributed only one independent response and there was no clustering in the data. Finally, multicollinearity was not a concern, as the model included only one predictor. Overall, logistic regression was both statistically appropriate and theoretically aligned with the study design. Here, Nao is served as the reference category. The null model included no predictors, while the full model included stimulus type as the sole predictor.

Diverse Desires A logistic regression was conducted to examine the effect of stimulus type on children's performance in the Diverse Desires task. The model including stimulus as a predictor provided a significantly better fit to the data than the null model ($\chi^2(1) = 22.36$, $p < .001$). With a moderate effect, the regression coefficient was B = -2.36, with an odds ratio of 0.09, indicating that more children tend to respect Nao's diverse desire, but project their own desire onto the human.

Diverse Beliefs The full model also significantly outperformed the null model ($\chi^2(1) = 4.16$, $p = 0.041$). The regression coefficient was B = -0.70, with an odds ratio of 0.50. This suggests children were half as likely to respond correctly when the informant was a human compared to when it was the robot Nao.

Knowledge Access The model including stimulus as a predictor provided a significantly better fit than the null model ($\chi^2(1) = 4.48$, $p = 0.034$). The regression coefficient was B = -0.64, and the odds ratio was 0.52. This indicates that children are inclined to attribute knowledge to Nao rather than the human.

Content False Belief The full model did not differ from the null ($\chi^2(1) = 0.30$, $p = .0582$) and there was no significant difference between two stimuli on this item.

Explicit False Belief The full model did not significantly improve model fit compared to the null model ($\chi^2(1) = 0.17$, $p = 0.681$). Stimulus type did not significantly predict performance on this item.

Belief-Emotion The full model did not significantly improve model fit compared to the null model ($\chi^2(1) = 0.67$, $p = 0.412$), revealing no significant effect of stimulus.

Real-Apparent Emotion There was no variation in the outcome, and the full model did not improve fit compared to the null model ($\chi^2(1) = 0.00$, $p = 1.00$).

Table 8 Logistic regression results predicting children's performance on ToAM and ToM items

Model	B	SE	Nagelkerke R^2	p-value	odds ratio
Diverse desires	-2.36	0.62	0.18	<0.001	0.09
Diverse beliefs	-0.70	0.35	0.03	0.045	0.50
Knowledge access	-0.64	0.30	0.03	0.036	0.52
Contents false belief	-0.15	0.28	<0.01	0.583	0.86
Explicit false belief	-0.11	0.27	<0.01	0.682	0.89
Belief-emotion	-0.22	0.27	<0.01	0.412	0.80
Real-apparent emotion	<0.001	<0.001	0.00	1.00	1.00

Note: The regression models include stimulus (Nao versus human) as the predictor, enabling analysis of its effect on task performance and the distinction between ToAM and ToM. Odds ratios and Nagelkerke R^2 statistics were used to interpret the strength and quality of the model.

5.3 RQ1 and 2: Selective Trust Differences by Agent and Testimony Domain

5.3.1 Nomination question

GLMMs were employed to assess how various factors influenced children's responses to the **nomination question**. Candidate predictors were added stepwise, and models were compared using the Bayesian Information Criterion (BIC) and conditional R^2 as indicators of model fit and effect size. All models included random intercepts for participant ID and trial to account for repeated measures. The null model (BIC = 1633.11, $R^2 = 0.00$) included no predictors and served as a baseline. Model 1 added the

main effect of condition, resulting in a substantial BIC reduction (ΔBIC = -33.44, $p <$ 0.001; $R^2 = 0.32$). However, its explanatory scope was limited due to the inclusion of only a single predictor. Model 2 incorporated condition, domain, and their interaction, providing a stronger theoretical basis. Although its BIC (ΔBIC = 25.58, $p = 0.004$) was slightly higher than that of Model 1, the improvement in conditional R^2 ($R^2 = 0.34$) indicated better overall explanatory power. Model 3 added ToAM scores but showed no substantial gain in explanatory value ($R^2 = 0.34$), while BIC not significantly increased (ΔBIC = 49.72, $p = 0.479$), reducing model parsimony. The full model, which included all main effects and interactions, had the highest R^2 ($R^2 = 0.35$) but the largest BIC (ΔBIC = 104.09, $p = 0.843$), suggesting overfitting. In conclusion, Model 2 was selected as the final model as it balanced statistical performance, interpretability, and theoretical relevance. The results of each model are presented in Appendix B.

In the selected model 2, condition, domain and their 2-way interaction effect were entered with the **nomination question** response treated as the dependent variable. A Type II Wald chi-square test revealed a significant main effect of condition ($\chi^2(1) = 41.64$, $p < 0.001$), and a significant interaction between condition and domain ($\chi^2(3) = 17.87$, $p < 0.001$). The main effect of domain was not significant ($\chi^2(3) = 1.64$, $p = 0.651$). Multicollinearity was assessed using generalised variance inflation factors (GVIFs). All adjusted GVIF values were below 3.5, indicating acceptable levels of collinearity: condition (1.40), domain (3.32), and the interaction term (3.36). Overdispersion was tested using Pearson residuals and revealed no evidence of overdispersion ($\chi^2(1274) = 1044.78$, $p \approx 1.00$) (for the visualisation, see Appendix E). These diagnostics indicate that the model fits the data well and meets key assumptions for GLMMs.

From the Tukey adjusted post-hoc test, a main effect of condition confirmed that children were significantly more likely to trust Nao in the *Nao-accurate* condition compared to the *Human-accurate* condition ($z = 6.65$, $p < 0.001$, $d = 0.87$). In the *Nao-accurate* condition, children were more likely to choose Nao in the mathematics compared to the physical science with a medium-sized effect ($z = 2.72$, $p = 0.033$, $d = 0.41$). In the *Human-accurate* condition, a significant difference was found between the non-STEM and physics domains, with children showing more trust in Nao for physics information ($z = 2.97$, $p = 0.016$, $d = 0.39$).

5.3.2 Endorsement question

For the **endorsement question**, a similar model comparison procedure was conducted using stepwise GLMMs. The null model (BIC = 1417.00, $R^2 = 0.00$) and Model 1, which included only condition as a predictor ($\Delta BIC = -55.02$, $p < 0.001$; $R^2 = 0.56$), provided limited explanatory power. Model 2, which included condition, domain, and their interaction, offered a strong balance between model fit and parsimony ($\Delta BIC = 13.58$, $p < 0.001$; $R^2 = 0.58$). Although more complex models slightly increased R^2, they also resulted in higher BIC values, indicating overfitting (model 3: $\Delta BIC = 42.33$, $p = 0.060$; $R^2 = 0.59$; full model: $\Delta BIC = 99.13$, $p = 0.496$; $R^2 = 0.61$). Therefore, Model 2 was selected as the final model for analysing endorsement behaviours. The results of each model are presented in Appendix C.

In the model 2, the main effect of condition ($\chi^2(1) = 64.54$, $p < 0.001$) and the 2-way interaction effect ($\chi^2(3) = 24.71$, $p < 0.001$) were significant while there is no significance in the main effect of domain ($\chi^2(3) = 3.30$, $p = 0.347$). Multicollinearity diagnostics indicated acceptable GVIF values for all predictors (condition = 1.20; domain = 3.27; condition × domain = 3.29). An overdispersion test revealed no evidence of overdispersion ($\chi^2(1274) = 880.93$, $p = 1.00$), confirming adequate model fit (for the visualisation, see Appendix E).

Tukey test found that children were more likely to choose Nao in the *Nao-accurate* condition than in the *Human-accurate* condition ($z = 8.46$, $p < .0001$, $d = 1.63$), indicating a very large effect. Within the *Nao-accurate* condition, children were significantly more likely to trust Nao in the non-STEM domain compared to the life science ($z = 3.76$, $p = 0.001$, $d = 0.67$), mathematics ($z = 3.46$, $p = 0.003$, $d = 0.62$), and physical science domains ($z = 3.04$, $p = 0.013$, $d = 0.55$). In the *Human-accurate* condition, only an increased likelihood of selecting Nao in the physical science compared to the non-STEM domain ($z = 3.23$, $p = 0.007$, $d = 0.50$) detected. Children's selective trust varies by domain, particularly when robots are accurate informants, with reduced trust observed in object labelling compared to STEM-related domains.

5.3.3 Accuracy question

GLMMs were employed to assess how various factors influenced children's responses to the **accuracy question**. The null model contained no predictors (BIC = 1332.36, R^2 = 0.00). Model 1 added the main effect of condition and substantially improved fit, though it explained no additional variance (ΔBIC = -46.54, $p < 0.001$; $R^2 = 0.00$). Model 2 introduced condition, domain, and their interaction. Although its BIC increased slightly, it accounted for substantially more variance (ΔBIC = 13.80, $p < 0.001$; $R^2 = 0.65$). Model 3 and the full model did not improve R^2 beyond 0.65 and exhibited non-significant higher BIC values (model 3: ΔBIC = 42.97, $p = 0.074$; full model: ΔBIC =95.78, $p = 0.282$), suggesting overfitting. Therefore, Model 2 was selected as the final model for analysing accuracy judgements. The results of each model are presented in Appendix D.

A Type II Wald chi-square test revealed significant main effect of condition ($\chi^2(1) = 51.77$, $p < 0.001$) and interaction effect ($\chi^2(3) = 20.17$, $p < 0.001$). The main effect of domain showed no significance ($\chi^2(3) = 6.45$, $p = 0.092$). All adjusted GVIF values showed acceptable levels of multicollinearity: condition (1.16), domain (3.30), and condition × domain (3.32). Overdispersion was also assessed, with a non-significant result ($\chi^2(1274) = 831.96$, $p = 1.00$), suggesting no evidence of overdispersion (for the visualisation, see Appendix E).

Tukey test revealed that children are significantly more likely to identify Nao as the more accurate informant in the *Nao-accurate* condition with a very large effect size ($z = 7.57$, $p < .0001$, $d = 1.79$). Precisely, within the *Nao-accurate* condition, children were significantly inclined to trust Nao in the non-STEM domain compared to the life science ($z = 3.41$, $p = 0.004$, $d = 0.67$), mathematics ($z = 3.42$, $p = 0.004$, $d = 0.67$), and physical science domains ($z = 2.99$, $p = 0.015$, $d = 0.59$). Within the *Human-accurate* condition, the significant contrasts were between the physical science and both the mathematics ($z = 2.85$, $p = .023$, $d = 0.46$) and the non-STEM domain ($z = 2.99$, $p = 0.015$, $d = 0.49$), suggesting children were less likely to judge the human as accurate in the physical science. These findings support the view that children's accuracy judgements are influenced not only by who is accurate but also by the domain of information presented, with reduced trust in the informant (robot or human) within the physical science domain.

5.3.4 The pattern of selective trust

GLMMs and Post hoc analyses, alongside predicted probability plots shown in Figure 7, found consistent and interpretable patterns across the nomination, endorsement and accuracy questions. Across all three measures and four domains, children were significantly more likely to trust Nao in the *Nao-accurate* condition compared to the *Human-accurate* condition (*p*s < 0.001), with predicted probabilities (see Figure 7) showing a strong downward shift from *Nao-accurate* condition to *Human-accurate* condition. This highlighted that children reliably tracked the prior accuracy of informants when making a decision about trustworthiness.

The plots further illustrated domain-specific differences, particularly between non-STEM and physical science domains. For the endorsement and accuracy questions, children's trust in Nao was especially strong in the object labelling under the *Nao-accurate* condition. They kept trusting Nao in three questions. Conversely, under the *Human-accurate* condition, children generally preferred the human informant, which was most notable in the non-STEM domain. Trust in the life and physical science remained more equally balanced. Though Tukey comparisons revealed fewer domain contrasts, compared to the physics, human were recognised as more trustworthy in the non-STEM object labelling across three questions. This might demonstrate heightened sensitivity to informant reliability when the task involves simpler, more familiar content that requires lower cognitive demands. Notably, even in the *Human-accurate* condition, while most children tended to endorse the human and view them as the accurate informant, some still chose to ask Nao for information, particularly regarding physical science content. Based on the descriptive analysis, however, children's responses to the nomination question in the physical domain did not significantly differ from chance, indicating difficulty in making a clear decision.

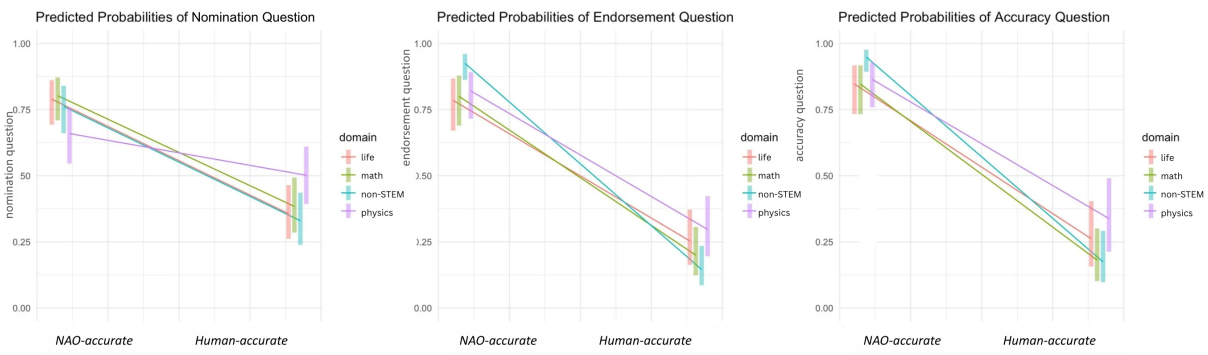


Figure 7 Predicted probabilities of choosing Nao for the nomination, endorsement, and accuracy questions. *Note*: From the left to right, panels show predicted probabilities of choosing Nao across conditions for the nomination, endorsement, and accuracy questions. Lines represent three domains, and shaded areas indicate 95% confidence intervals. Downward slopes across panels reflect decreased trust in Nao when the human informant is accurate, highlighting condition and domain effects in children's selective trust.

5.4 RQ3: The association between ToAM, ToAM and selective trust

This section focuses on RQ3 and corresponding hypothesis. First of all, to examine the potential inter-item relationships within the ToM and ToAM assessments, a partial correlation matrix was computed and visualised as a heat map (see Figure 8). This analysis accounts for the potential confounding effect of age on performance, providing a more precise assessment of the cognitive overlap between ToM and ToAM. Correlations among early-emerging tasks such as Diverse Desires and Diverse Beliefs were weak and mostly non-significant, suggesting minimal correspondence between ToM and ToAM reasoning at foundational developmental levels. The tasks in the assessment were structured in a developmental progression, beginning with simpler concepts such as desires and beliefs, and advancing to more complex tasks involving Knowledge Access, False Belief, Belief-Emotion, and ultimately Real-Apparent Emotion. However, significant associations began to emerge at intermediate levels of difficulty. Knowledge Access (ToM) was not only significantly correlated with the corresponding ToAM item but also associated with the other three items. Moreover, the strongest correlations appeared among items probing more a more advanced developmental level. Explicit False Belief, Belief-Emotion, and Real-Apparent Emotion in ToAM were each significantly associated with their ToM counterparts. Besides, the total scale score of ToAM and ToM are strongly associated ($r = 0.80$, $p < 0.001$). These findings demonstrated an increasing convergence between ToM and ToAM performance as cognitive demands grow, supporting the view that more advanced reasoning about robot-focused ToAM relies on similar meta-representational capacities as human-oriented ToM.

Secondly, the Point-Biserial correlation with age as a covariate was conducted to explore the association between ToAM/ToM and selective trust due to its interpretability and widespread use for binary × continuous comparisons. Correlation tests were computed separately for the *Nao-accurate* and *Human-accurate* conditions. The correlation was

exhibited in Figure 9. In the *Nao-accurate* condition, several small but statistically significant negative correlations were detected. ToM scores were negatively associated with the nomination (r = -0.10, $p < 0.05$), endorsement (r = -0.16, $p < 0.001$) and accuracy questions (r = -0.18, $p < 0.001$). Similarly, ToAM scores were also negatively correlated with all three selective trust measures, including nomination (r = -0.10, $p < 0.01$), endorsement (r = -0.13, $p < 0.001$), and accuracy questions (r = -0.14, $p < 0.001$). Although the correlations were relatively weak, these results still revealed that children with higher ToAM or ToM scores were somewhat less likely to trust the robot, even when the robot had been perfectly accurate. On the contrary, when the human informant was much more accurate, all correlations between ToAM and ToM scale scores and trust measures were near zero and statistically non-significant. In brief, neither scale showed a clear association with selective trust under the *Human-accurate* condition.

Overall, most items in the ToM scale showed significant correlations with their ToAM counterparts, with the exception of the early-developing tasks of Diverse Desires and Diverse Beliefs. The correlations were generally medium to strong and statistically significant. While ToAM and ToM were closely related, their associations with selective trust were more complex. In the *Nao-accurate* condition, small but statistically significant negative correlations emerged between children's ToM and ToAM scores and their selective trust measures. This indicated that children with stronger mental state attributing were slightly less likely to trust the robot, despite its prior accuracy. In contrast, in the *Human-accurate* condition, few meaningful associations were observed. No medium or large effect sizes were found in either condition.

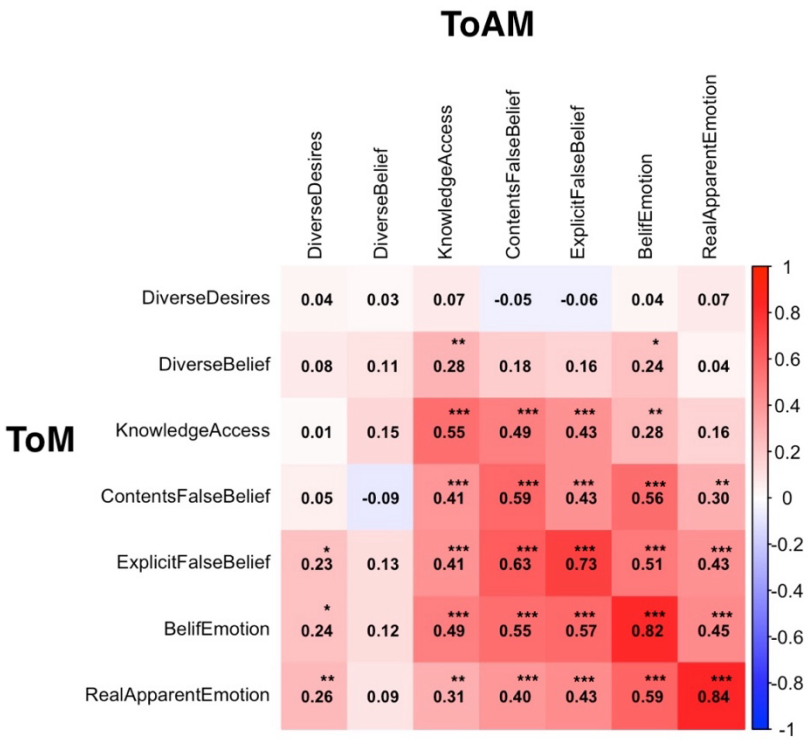


Figure 8 The associations between ToM and ToAM

Note: Associations between ToM and ToAM were examined across tasks ordered from developmentally simpler (e.g., Diverse Desires) to more complex levels of reasoning (e.g., Real-Apparent Emotion). Heat map of the phi (ϕ) correlations. The phi coefficient ranges from -1 to +1, where red indicates a positive correlation (with darker red reflecting stronger associations) and blue indicates a negative correlation. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. This significance notation is applied consistently in all subsequent figures and tables.

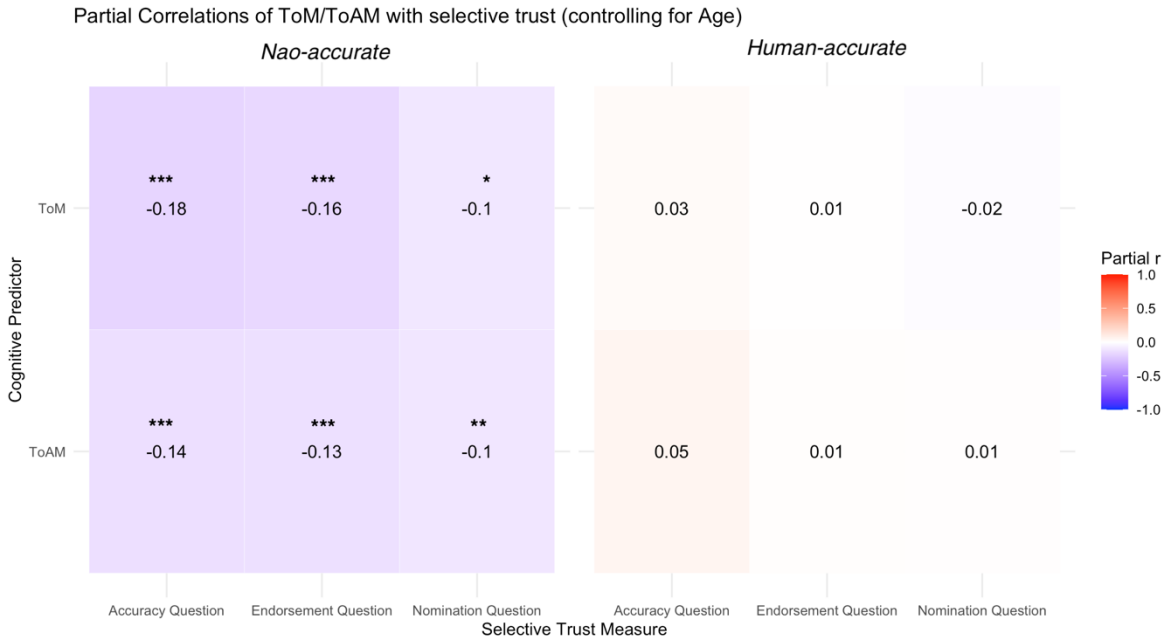


Figure 9 The associations between ToM/ToAM and selective trust

Note: Associations between ToM and ToAM beliefs and children's responses to accuracy, endorsement, and nomination questions, controlling for age, by condition. Correlation values are reported as Pearson's *r*, as the point-biserial correlation is mathematically equivalent to Pearson's product-moment correlation when one variable is continuous, and the other is discrete.

Chapter 6 Discussion

6.1 Findings and interpretations

This study aimed to bridge existing research gaps by investigating how 107 Chinese preschool children aged 4 to 6 (1) selectively trusted human and robotic informants (specifically, a humanoid robot called Nao) across four knowledge domains: non-STEM object labelling, physical science, life science, and mathematics (STEM), and (2) ToM, ToAM and selective trust were associated. Three research questions were posed: To what extent do children's selective trust in robots and humans differ (RQ1)? How does the domain of testimony (non-STEM versus STEM) influence children's selective trust (RQ2)? Do ToAM and ToM relate to children's selective trust in robot and human informants (RQ3)?

Using a conflicting informants paradigm, children were randomly assigned to a condition in which either the robot or the human consistently provided accurate information. Measures of selective trust included children's preferences for whom to ask for information, whose answers to endorse, and whom they judged as more accurate when learning both non-STEM (object labels) and STEM (physics, life science, and mathematics) information. Moreover, standardised assessments of ToM and ToAM were conducted using validated scales. These measures were used to examine the extent to which children's mentalisation on humans and robots influenced their selective trust. The findings provided new insights into children's interpretation about informant reliability across different knowledge domains and revealed important associations between mental state reasoning and trust. The main results are presented and discussed in the following sections.

6.1.1 RQ1: Children rely on prior accuracy to trust both robots and humans

As I hypothesised (H1), it was clear that children aged 4 to 6 detected and drew on the informants' prior accuracy when deciding whether to trust robot or human informants. When the robot Nao was entirely accurate during the familiarisation phase, children were significantly more likely to seek information from Nao, endorse its testimony, and evaluate it as a reliable source in the test phase. Likewise, when the human informant was accurate, children were more inclined to nominate, endorse, and trust the human

over the robot. This pattern was consistent across both non-STEM and STEM domains, indicating that children preferred to trust the informant with a history of accuracy regardless of the agent's identity. Children's selective trust behaviour closely resembles the trust patterns previously observed in studies involving only human informants (Corriveau & Harris, 2009; Koenig et al., 2004; Koenig & Harris, 2005). Moreover, the present findings align with prior research on selective trust in robots (Baumann et al., 2023; Brink & Wellman, 2020; Chen et al., 2024; Geiskkovitch et al., 2019) and other AI agents such as DVAs (Girouard-Hallam & Danovitch, 2022; Li et al., 2023). Whether informants are humans or novel AI agents, preschool children tended to rely on epistemic cues, prior accuracy in particular, to trust them and their information.

A noteworthy methodological similarity across these studies is the consistent use of the humanoid robot Nao as the informant (e.g., Brink & Wellman, 2020; Li & Yow, 2023). Across different contexts, children were more likely to trust the informant who had demonstrated prior accuracy, regardless of whether that informant was Nao or a human. In contrast, however, Brink and Wellman (2020) found that when non-humanoid machines (inanimate and blob-like) were used as informants, children's selective trust approached chance level (50%), suggesting that epistemic cues alone may be insufficient in the absence of human-like features. Moreover, although DVAs lack humanoid embodiment, their human-like voices and verbal interactivity continue to elicit considerable trust from young children (Li et al., 2023). In summary, preschool children are capable of selectively trusting both human and robotic informants with prior accuracy. However, this capacity appears to depend on the presence of human-like cues, such as a humanoid appearance or interactive audio features.

Overall, while the findings confirm H1 that children rely on prior accuracy to guide their trust in both human and robotic informants, from previous evidence, this trust criterion appears to be impacted by the human-like cues in robots. The consistent use of humanoid robots such as Nao across studies, including the present one, indicated that children are more willing to apply selective trust strategies when the robotic agent exhibits familiar social cues — such as a face, interactive audio, or bodily orientation. This principle was also reflected in ToAM abilities. Banks (2020) revealed that adults attributed mental states to robots (also using Nao) only when it displayed social cues such as engaging in social interactions. In contrast, when robots lack these social cues, children's selective

trust become less strategic and may revert to chance (Brink & Wellman, 2020). Consequently, selective trust in robots is not merely an assessment of epistemic cues but also shaped by the social cues that the robot conveys.

6.1.2 RQ2: Domain-Specific trust with uncertainty in the physical domain

Hypothesis 2 predicted that children would be more likely to trust the robot in STEM domains and the human in the non-STEM domain, reflecting domain-specific trust preferences. However, this hypothesis was not fully supported by the findings. Children consistently trust accurate informants in non-STEM domain across two conditions, but nominate and endorse Nao, and trust its information in the physical domain (e.g., ice cube melting, sinking and floating) when the human is accurate.

Preschool-aged children did not apply a generalised rule such as "robots are STEM experts" and "humans are everyday knowledge experts," which can be considered the expertise in epistemic cues. Specifically, in the non-STEM object-labelling tasks involving familiar and novel items (e.g., "cup" and "kaka"), children aged 4 to 6 were significantly more likely to endorse the robot's responses and judge it as trustworthy in the *Nao-accurate* condition than in the *Human-accurate* condition. Similarly, in the *Human-accurate* condition, they consistently nominated, endorsed, and evaluated the human as the more credible source for object labels. These findings contrast with prior research suggesting that children tend to prefer robots for mechanical or physical knowledge and humans for psychological or biological information (Oranç & Küntay, 2020). One possible explanation is that non-STEM tasks were more familiar and cognitively accessible, requiring less abstract reasoning. This may have enabled children to apply prior accuracy cues more confidently (Pasquini et al., 2007). In contrast, the abstract and counterintuitive nature of STEM content may make it more difficult for children to evaluate the plausibility of information, resulting in less consistent trust.

Interestingly, despite recognising the human informant's accuracy, children sometimes still preferred to consult Nao for physical knowledge. Specially, in the *Human-accurate* condition, while most children endorsed the human and judged them as the more accurate informant across the four domains, some favoured asking Nao for physical knowledge. Children showed significantly lower trust in the human informant when responding to

physical science (e.g., novel object floats or sinks in the water) tasks compared with non-STEM content (e.g., the label of novel object), across nomination, endorsement, and accuracy questions. However, their responses to the nomination question in the physical science domain did not significantly differ from chance, suggesting possible uncertainty in decision-making when faced with more cognitively demanding content and potentially less reliability of comparison between two informants. Similar findings were reported by Danovitch et al. (2023), who employed an identical materials for non-STEM information in the familiarisation phase and STEM information in both familiarisation and test phases. Their study showed that children aged 4 and 5 trusted and transmitted both non-STEM and STEM information from an accurate human informant but were significantly more likely to transmit physical than mathematical knowledge. Physical concepts may be particularly challenging for preschoolers to grasp, despite being part of national curricula (Valovičová et al., 2020). This difficulty arises because physical phenomena often involve abstract or invisible mechanisms, such as force or density, which are less intuitively understood than observable traits. While children can perform simple experiments such as sinking and floating, they often lack understanding of the underlying principles and struggle to generalise this knowledge to novel concepts, such as unfamiliar red objects used in the current study. Their chance-level responses in the nomination question suggest confusion or uncertainty when reasoning about abstract physical content.

The current study, conducted in China, reflected broader issues in early physics education. Compared to life science and mathematics, physics is less emphasised in Chinese kindergartens. Children often engage with animals and participate in counting tasks to support biological and mathematical learning, but exposure to physics is relatively rare. Although some kindergartens include physics-themed activities, these often lack conceptual explanation. For instance, while children enjoy building shadow patterns with multiple materials (e.g., wood, lamp), they are not taught why shadow sizes differ with the light changing. Furthermore, many Chinese preschool teachers report low confidence in delivering physics content, outlining inadequate training and teaching resources. Teachers always avoid having physics classes in public due to the lack of scientific instructions. Conversely, successful education depends not just on the resource but on how it is framed and communicated in dialogue with the child. Fridberg et al. (2019) highlighted that intersubjective communication, which is the teacher's ability to

align their explanations with children's perspectives, is key to effective physics teaching. Teachers' lack of content knowledge is an obstacle reported in the implementation of early education programmes in physical science (Roehrig et al., 2011).

6.1.3 RQ3: The association between ToM, ToAM and selective trust varies across conditions

First, controlling for the effect of age, partial phi correlation demonstrated that ToAM is not simply a mirror of ToM but rather a parallel construct with domain-specific variances. Most ToM items significantly correlated with their ToAM counterparts, except for early-emerging components like Diverse Desires and Diverse Beliefs. Logistic regression further supported these patterns, revealing that children were significantly more likely to respect Nao's diverse desires and beliefs and attribute knowledge to Nao rather than to a human. These results align with Mao et al (2025), who found that children aged 3 to 6 were more likely to project their own beliefs and knowledge onto peers than onto robots. Extending these findings to the selective trust research, children's greater attribution of knowledge to robots may support their willingness to accept testimony from them. In the current study, children consistently showed a preference for the robot informant in the physical science for nomination, even when the human informant was entirely accurate. This noted a robust trust in the robot's knowledge and belief states. This conceptual extension lends theoretical support to H3 — that children with higher ToAM scores would be more likely to trust the robot, while those with higher ToM scores would be more likely to trust the human informant — but this positive relationship requires statistical validation.

Point-Biserial correlation with age as a covariate was conducted to explore the association between ToAM/ToM and selective trust, yet revealing results against H3. The ToM and ToAM scores were weakly but significantly negatively associated with trust across all measures in the *Nao-accurate* condition, suggesting that children with stronger cognitive attributions were more cautious about trusting robots. Conversely, in the *Human-accurate* condition, correlations between ToM/ToAM and trust were non-significant, specifying that children's selective trust in the human informant was not shaped by their mentalising skills. All correlations were weak (below 0.2), no moderate or strong associations were identified in either condition.

On the one hand, there existed a negative relationship between ToAM/ToM and selective trust when robot Nao was consistently correct. Such negative patterns were also detected in one empirical study using the Trust Game (Di Dio et al., 2020) and one meta-analysis reviewing children's trust in social robots (Stower et al., 2021). Di Dio et al. (2020) found that children with higher ToM and ToAM were more sceptical and showed less willingness to rely on Nao during a Trust Game. Again, from the meta-analysis, it was clear that children may grow emotionally attached to robots without necessarily trusting their knowledge or testimony (Stower et al., 2021). This implied that mature cognitive functions may not lead to greater selective trust, but rather enable children to distinguish between social cues (e.g., friendliness, likeness) and epistemic cues (e.g., reliability, competency). Children with stronger mentalising abilities might adopt a more evaluative or critical stance, reflecting increased epistemic vigilance. This enhanced vigilance could lead them to scrutinise non-human agents more carefully, questioning their reliability even when they perform accurately.

On the other hand, both ToM and ToAM were not significantly related to selective trust when the human was accurate, showing a neutral perspective. Previous studies indicated that the relationship between ToM and selective trust in human may be swayed by socioeconomic status and inhibitory control, which are not accounted for in the present study. The participants in my study were exclusively from middle-class urban areas, and no data on their inhibitory control levels were collected. Souza et al. (2021), for example, found that although middle-class children demonstrated more advanced ToM and vocabulary skills, their selective trust performance was paradoxically lower than their lower-class peers, suggesting that cognitive development does not always translate into optimal epistemic reasoning. Therefore, it is possible that the homogeneous middle-class background of my sample reduce the potential variability in selective trust that ToM or ToAM might otherwise relate. Since adults represent a familiar and consistently reliable source of information in children's everyday lives, children may more easily access and apply prior accuracy about adult credibility, making trust in human informants less dependent on cognitive abilities such as ToM or ToAM.

In addition to this work, only one study to date has simultaneously examined how children mentalise both AI and human informants in relation to their selective trust.

Girouard-Hallam and Danovitch (2022) found that children's (from upper middle-class families) MSA to DVAs and humans (e.g., "Can [informant] think?", "Can [informant] learn things?") were generally not predictive of their selective trust in either type of agent, with the exception of pedagogical capacity (i.e., "Can [informant] teach you things?"). Combing with my findings, selective trust in AI agents versus humans appeared more cue-based (social and epistemic cues) than mentalisation-based (mental states). More broadly, ToM and ToAM (as well as inhibitory control) are all cognitive capacities that reflect the sophistication of children's cognitive development. Crucially, children with more advanced cognitive capacities may treat AI agents with increased scepticism and caution, indicating an emerging epistemic maturity. Rather than accepting technological agents uncritically, these children demonstrate a capacity to evaluate their reliability thoughtfully. This is a promising indication that, within rapid advancement and growing societal enthusiasm for AI, the next generation are equipped to engage with such agents critically and objectively, maintaining a discerning mindset.

6.2 Implications of the study

The findings of this study have important implications for multiple stakeholders engaged in child development, AI in education, and curriculum design.

For the practitioners in different fields, early childhood educators could be made aware of preschool-aged children's STEM learning, especially physics. Children possibly find difficulties in understanding abstract physics-related knowledge. In addition to the hands-on experiments and long-time observations, teachers could be made aware of elementary principles and master how to explain them to children with child-friendly language, helping children transmit knowledge to the close concepts. Moreover, it is essential to build AI literacy through guided interaction from early childhood as they are the digital natives. Teachers can scaffold children's understanding of robots not only as tools, but as informants whose knowledge is context-dependent. If it is possible, headteachers could introduce authentic robots or other AI agents in the daily activities, making it accessible for children to have closer emotional bond with robots and deepen their epistemic understanding.

For robotic designers, this study underscores the importance of creating child-friendly robots that align with children's developmental expectations and domain-specific trust. In STEM learning contexts, robots could be programmed to provide clear, age-appropriate explanations that support inquiry and critical thinking. Designers could also consider incorporating subtle social cues such as gaze, gesture, and voice tone.

For the policymakers, as digital technologies become more embedded in early childhood, policymakers play a key role in guiding ethical and developmentally appropriate practices. This study highlights the potential for robots to serve as instructional aids in STEM education. Policymakers could intend to support pilot programmes that responsibly introduce robots in early childhood classrooms. Further, given that ToAM and trust development may vary with age, culture, and socioeconomic status, efforts must ensure that AI-based learning tools are inclusive, culturally relevant, and accessible. Most importantly, there is a need for national or regional curricula to explicitly include early AI literacy, not only technical familiarity, but also critical thinking about artificial agents as knowledge sources.

6.3 Limitations and future directions

6.3.1 Limitations

While this study offers novel insights, some limitations are acknowledged. Firstly, the study included a sample with a relatively narrow age range (4 to 6 years) and set age as a covariate. While this age group was appropriate for examining early ToM and ToAM development, it limits the generalisability of findings to other developmental stages. A longitudinal or broader age-range design which focuses on the age difference could reveal how selective trust, ToM, and ToAM evolve over time and interact across developmental trajectories. Second, the range of covariates considered when evaluating the association between ToM, ToAM, and selective trust. Prior research has shown that both socioeconomic background and inhibitory control can significantly impact the relationship between ToM and selective trust (Crivello et al., 2021; Souza et al., 2021). This research only included children from middle-class families and did not evaluate their inhibitory control. By omitting these variables, the current study may not fully account for the complexity of the relationship between ToM/ToAM and selective trust. Third, the study's sample was drawn exclusively from China, which may limit the

generalisability of the findings across different socio-cultural contexts. Cross-cultural studies showed that while Western children typically understand false beliefs before knowledge access, Asian children often grasp knowledge access earlier (Wellman, 2017). Future research could include diverse cultural samples to investigate whether similar findings hold across contexts.

6.3.2 Future directions

Building on the current findings, several directions for future research are recommended. This study can be meaningfully extended to diverse cultural contexts. First, this study focused on a humanoid robot (Nao), but children interact with other AI agents, such as DVAs (e.g., Alexa), chatbots, or non-humanoid robots. Future research is recommended to examine whether and how children's selective trust generalises across different AI modalities and how it compares to their trust in humans. Second, future studies could broaden the range of information domains. While this research focused on STEM and non-STEM object labelling, children routinely encounter other kinds of testimony, such as personal, moral, or historical information. As children are more likely to forgive robots' errors while perceiving humans' errors as intentional (Stower et al., 2024), they may prefer to trust robots when it comes to moral norms.

6.4 Conclusion

In conclusion, this dissertation investigated how 107 Chinese children aged 4 to 6 selectively trust human and robotic informants (specifically the humanoid robot Nao) across STEM domains (physics, life science, and mathematics) and a non-STEM domain (object labelling), with particular attention to the roles of ToM and ToAM. Employing a conflicting informants paradigm, the study examined children's responses to nomination, endorsement, and accuracy questions, alongside standardised ToM and ToAM assessments. Findings showed that although children's trust patterns varied by domain and informant accuracy, they generally relied on prior accuracy when deciding whom to trust, especially in the non-STEM domain. Interestingly, in the physical science domain, children exhibited uncertainty: they tended to nominate inaccurate robots yet endorsed accurate human responses. Furthermore, in the *Nao-accurate* condition, both ToM and ToAM scores were negatively associated with selective trust, whereas these associations

were non-significant in the *Human-accurate* condition. These results advance our understanding of how young children navigate AI-integrated learning environments and assess accuracy of information and reliability of informants. Importantly, the findings underscore children's emerging sane awareness in evaluating AI agents. Based on these insights, practitioners and policymakers are encouraged to strengthen early physics education and support digital literacy through thoughtful integration of educational technologies, with robots being a possible candidate.

References

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation*. Lawrence Erlbaum.

Allen, M., Webb, A. W., & Matthews, C. E. (2016). Adaptive Teaching in STEM: Characteristics for Effectiveness. *Theory Into Practice*, *55*(3), 217–224. https://doi.org/10.1080/00405841.2016.1173994

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–118. https://doi.org/10.1016/j.tics.2009.12.006

Banerjee, M. (1997). Hidden Emotions: Preschoolers' Knowledge of Appearance-Reality and Emotion Display Rules. *Social Cognition*, *15*(2), 107–132. https://doi.org/10.1521/soco.1997.15.2.107

Banks, J. (2020). Theory of Mind in Social Robots: Replication of Five Established Human Tests. *International Journal of Social Robotics*, *12*(2), 403–414. https://doi.org/10.1007/s12369-019-00588-x

Baratgin, J., Dubois-Sage, M., Jacquet, B., Stilgenbauer, J.-L., & Jamet, F. (2020). Pragmatics in the False-Belief Task: Let the Robot Ask the Question! *Frontiers in Psychology*, *11*, 593807. https://doi.org/10.3389/fpsyg.2020.593807

Barlow, A., Qualter, P., & Stylianou, M. (2010). Relationships between Machiavellianism, emotional intelligence and theory of mind in children. *Personality and Individual Differences*, *48*(1), 78–82. https://doi.org/10.1016/j.paid.2009.08.021

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind" ? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. Oxford University Press.

Baumann, A.-E., Goldman, E. J., Meltzer, A., & Poulin-Dubois, D. (2023). People Do Not Always Know Best: Preschoolers' Trust in Social Robots. *Journal of Cognition and Development*, *24*(4), 535–562. https://doi.org/10.1080/15248372.2023.2178435

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, *3*(21), eaat5954. https://doi.org/doi:10.1126/scirobotics.aat5954

Breazeal, C., Harris, P. L., DeSteno, D., Kory Westlund, J. M., Dickens, L., & Jeong, S. (2016). Young Children Treat Robots as Informants. *Topics in Cognitive Science*, *8*(2), 481–491. https://doi.org/https://doi.org/10.1111/tops.12192

Brink, K. A., Gray, K., & Wellman, H. M. (2019). Creepiness Creeps In: Uncanny Valley Feelings Are Acquired in Childhood. *Child Development*, *90*(4), 1202–1214. https://doi.org/https://doi.org/10.1111/cdev.12999

Brink, K. A., & Wellman, H. M. (2020). Robot teachers for children? Young children trust robots depending on their perceived accuracy and agency. *Developmental Psychology*, *56*(7), 1268–1277. https://doi.org/10.1037/dev0000884

British Educational Research Association. (2024). *Ethical Guidelines for Educational Research* (5th ed.). British Educational Research Association.

Brosseau-Liard, P., Penney, D., & Poulin-Dubois, D. (2015). Theory of mind selectively predicts preschoolers' knowledge-based selective word learning. *British Journal of Developmental Psychology*, *33*(4), 464–475. https://doi.org/10.1111/bjdp.12107

Budiharto, W., Cahyani, A. D., Rumondor, P. C. B., & Suhartono, D. (2017). EduRobot: Intelligent Humanoid Robot with Natural Interaction for Education and Entertainment. *Procedia Computer Science*, *116*, 564–570. https://doi.org/10.1016/j.procs.2017.10.064

Cameron, J. A., Alvarez, J. M., Ruble, D. N., & Fuligni, A. J. (2001). Children's Lay Theories About Ingroups and Outgroups: Reconceptualizing Research on Prejudice. *Personality and Social Psychology Review*, *5*(2), 118–128. https://doi.org/10.1207/s15327957pspr0502_3

Cao, X., Wu, Y., Nielsen, M., & Wang, F. (2025). Does appearance affect children's selective trust in robots' social and emotional testimony? *Journal of Applied Developmental Psychology*, *96*, 101739. https://doi.org/10.1016/j.appdev.2024.101739

Chen, Z., Zheng, J., Gao, Y., Fang, J., Wang, Y., Chen, H., & Wang, T. (2024). Do Children with Autism Spectrum Disorders Show Selective Trust in Social Robots? *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-024-06474-4

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, *19*(22), 3127–3131. https://doi.org/10.1002/1097-0258(20001130)19:22<3127::AID-SIM784>3.0.CO;2-M

Clegg, J. M., Kurkul, K. E., & Corriveau, K. H. (2019). Trust me, I'm a competent expert: Developmental differences in children's use of an expert's explanation quality to infer trustworthiness. *Journal of Experimental Child Psychology*, *188*, 104670. https://doi.org/10.1016/j.jecp.2019.104670

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Corriveau, K., & Harris, P. L. (2009). Choosing your informant: weighing familiarity and recent accuracy. *Developmental Science*, *12*(3), 426–437. https://doi.org/j.1467-7687.2008.00792.x

Council, N. R. (2013). *Next Generation Science Standards: For States, By States*. The National Academies Press. https://doi.org/10.17226/18290

Crivello, C., Grossman, S., & Poulin-Dubois, D. (2021). Specifying links between infants' theory of mind, associative learning, and selective trust. *Infancy*, *26*(5), 664–685. https://doi.org/10.1111/infa.12407

Cudworth, D., & Lumber, R. (2021). The importance of Forest School and the pathways to nature connection. *Journal of Outdoor and Environmental Education*, *24*(1), 71–85. https://doi.org/10.1007/s42322-021-00074-x

Danovitch, J. H., Scofield, J., Williams, A. J., Davila, L., & Bui, C. (2023). Children's selective information transmission in STEM and non-STEM domains. *Cognitive Development*, *66*, 101332. https://doi.org/10.1016/j.cogdev.2023.101332

Department of Education. (2024a). *Early years foundation stage statutory framework: For group and school-based providers*. Retrieved from https://www.gov.uk/government/publications/early-years-foundation-stage-framework--2

Department of Education. (2024b). *YOU Belong in STEM initiative*. Retrieved from https://www.ed.gov/about/initiatives/you-belong-stem

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020). Shall I Trust You? From Child–Robot Interaction to Trusting Relationships. *Frontiers in Psychology*, *11*, 469. https://doi.org/10.3389/fpsyg.2020.00469

Ding, X.-C., Sang, B., & Pan, T.-T. (2017). Relations between Selective Trust, Theory of Mind, and Executive Function in Preschoolers: Evidence from Longitudinal Study. *Journal of Psychological Science*, *40*(5), 1129–1135.
DiYanni, C., Deniela, N., Whitney, R., & and Livelli, A. (2012). 'I Won't Trust You if I Think You're Trying to Deceive Me': Relations Between Selective Trust, Theory of Mind, and Imitation in Early Childhood. *Journal of Cognition and Development*, *13*(3), 354–371. https://doi.org/10.1080/15248372.2011.590462
Doebel, S., & Koenig, M. A. (2013). Children's Use of Moral Behavior in Selective Trust: Discrimination Versus Learning. *Developmental Psychology*, *49*(3), 462–469. https://doi.org/10.1037/a0031595
Durkin, K., & Shafto, P. (2016). Epistemic Trust and Education: Effects of Informant Reliability on Student Learning of Decimal Concepts. *Child Development*, *87*(1), 154–164. https://doi.org/10.1111/cdev.12459
Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological methods*, *23*(4), 617. https://doi.org/10.1037/met0000167
Fridberg, M., Jonsson, A., Redfors, A., & Thulin, S. (2019). Teaching chemistry and physics in preschool: a matter of establishing intersubjectivity. *International Journal of Science Education*, *41*(17), 2542–2556. https://doi.org/10.1080/09500693.2019.1689585
Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2–18. https://doi.org/10.1037/a0024338
Fusaro, M., & Harris, P. L. (2008). Children assess informant reliability using bystanders' non-verbal cues. *Developmental Science*, *11*(5), 771–777. https://doi.org/10.1111/j.1467-7687.2008.00728.x
Geiskkovitch, D. Y., Thiessen, R., Young, J. E., & Glenwright, M. R. (2019, 11–14 March 2019). What? That's Not a Chair!: How Robot Informational Errors Affect Children's Trust Towards Robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, https://doi.org/10.1109/HRI.2019.8673024
Girouard-Hallam, L. N., & Danovitch, J. H. (2022). Children's trust in and learning from voice assistants. *Developmental Psychology*, *58*(4), 646–661. https://doi.org/10.1037/dev0001318
Guilford, J. P. (1936). *Psychometric methods* (1st ed.). McGraw-Hill.
Harris, P. L., N., J. C., Deborah, H., Giles, A., & and Cooke, T. (1989). Young Children's Theory of Mind and Emotion. *Cognition and Emotion*, *3*(4), 379–400. https://doi.org/10.1080/02699938908412713
Harris, P. L., Pasquini, E. S., Duke, S., Asscher, J. J., & Pons, F. (2006). Germs and angels: the role of testimony in young children's ontology. *Developmental Science*, *9*(1), 76–96. https://doi.org/j.1467-7687.2005.00465.x
Harvey, G. (2005). *Animism: Respecting the living world*. Wakefield Press.
Hasselblad, V., & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological bulletin*, *117*(1), 167–178. https://doi.org/10.1037/0033-2909.117.1.167
John-Steiner, V., & Mahn, H. (2003). Sociocultural contexts for teaching and learning. *Handbook of psychology*, *7*, 125–148. https://doi.org/10.1002/0471264385.wei0707
Johnson, C. C. (2013). Conceptualizing Integrated STEM Education. *School science and mathematics*, *113*(8), 367–368. https://doi.org/10.1111/ssm.12043

Katayama, N., Katayama, J. I., Kitazaki, M., & Itakura, S. (2010). Young Children's Folk Knowledge of Robots. *Asian Culture and History*, *2*(2), 111. https://doi.org/10.5539/ach.v2n2p111

Keren, G., & Fridin, M. (2014). Kindergarten Social Assistive Robot (KindSAR) for children's geometric thinking and metacognitive development in preschool education: A pilot study. *Computers in Human Behavior*, *35*, 400–412. https://doi.org/10.1016/j.chb.2014.03.009

Koenig, M. A. (2012). Beyond Semantic Accuracy: Preschoolers Evaluate a Speaker's Reasons. *Child Development*, *83*(3), 1051–1063. https://doi.org/10.1111/j.1467-8624.2012.01742.x

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in Testimony: Children's Use of True and False Statements. *Psychological Science*, *15*(10), 694–698. https://doi.org/10.1111/j.0956-7976.2004.00742.x

Koenig, M. A., & Harris, P. L. (2005). Preschoolers Mistrust Ignorant and Inaccurate Speakers. *Child Development*, *76*(6), 1261–1277. https://doi.org/10.1111/j.1467-8624.2005.00849.x

Koenig, M. A., & Jaswal, V. K. (2011). Characterizing Children's Expectations About Expertise and Incompetence: Halo or Pitchfork Effects? *Child Development*, *82*(5), 1634–1647. https://doi.org/10.1111/j.1467-8624.2011.01618.x

Kornbrot, D. Point Biserial Correlation. In *Wiley StatsRef: Statistics Reference Online*. https://doi.org/10.1002/9781118445112.stat06227

Kory Westlund, J. M., Dickens, L., Jeong, S., Harris, P. L., DeSteno, D., & Breazeal, C. L. (2017). Children use non-verbal cues to learn new words from robots as well as people. *International Journal of Child-Computer Interaction*, *13*, 1–9. https://doi.org/10.1016/j.ijcci.2017.04.001

Kraftl, P. (2014). *Geographies of alternative education*. Policy Press Bristol.

Landrum, A. R., Mills, C. M., & Johnston, A. M. (2013). When do children trust the expert? Benevolence information influences children's trust more than expertise. *Developmental Science*, *16*(4), 622–638. https://doi.org/10.1111/desc.12059

Lane, J. D., Wellman, H. M., & Gelman, S. A. (2013). Informants' Traits Weigh Heavily in Young Children's Trust in Testimony and in Their Epistemic Inferences. *Child Development*, *84*(4), 1253–1268. https://doi.org/10.1111/cdev.12029

Lantolf, J. P. (2000). Introducing sociocultural theory. *Sociocultural theory and second language learning*, *1*, 1–26. https://doi.org/10.2307/328580

Lecce, S., Caputi, M., & Hughes, C. (2011). Does sensitivity to criticism mediate the relationship between theory of mind and academic achievement? *Journal of Experimental Child Psychology*, *110*(3), 313–331. https://doi.org/10.1016/j.jecp.2011.04.011

Li, J., Pang, Y., & Jia, Q. (2024). The Impact of Age on Children's Selective Trust in Misleading Social Robots. In *HCI International 2024 Posters. HCII 2024. Communications in Computer and Information Science*, Cham. https://doi.org/10.1007/978-3-031-61932-8_42

Li, X., & Yow, W. Q. (2024). Younger, not older, children trust an inaccurate human informant more than an inaccurate robot informant. *Child Development*, *95*(3), 988–1000. https://doi.org/10.1111/cdev.14048

Li, Z., Liu, Z., Mao, K., Li, W., Li, T., & Li, J. (2023). The epistemic trust of 3-to 6-year-olds in digital voice assistants in various domains. *Acta Psychologica Sinica*, *55*(9), 1411–1423. https://doi.org/10.3724/sp.J.1041.2023.01411

Lutz, D. J., & Keil, F. C. (2002). Early Understanding of the Division of Cognitive Labor. *Child Development*, *73*(4), 1073–1084. https://doi.org/10.1111/1467-8624.00458

Mao, K., Zisong, L., Zhongjie, L., Yuki, N., & and Li, T. (2025). NAO, I Can Read Your Mind: Preschool-Aged Children Construct Theory of Artificial Mind (ToAM). *International Journal of Human–Computer Interaction*, 1–14. https://doi.org/10.1080/10447318.2025.2474488

Mertzani, V., & Drigas, A. (2023). Exploring the Learning Gains of Implementing Teacher Humanoid Robots in STEM Education: A Systematic Review. *International Journal of Online and Biomedical Engineering (iJOE)*, *19*, 18–30. https://doi.org/10.3991/ijoe.v19i18.43893

Ministry of Education. (2012). *3–6 Years Old Learning and Development Guidelines*.

Nguyen, S. P. (2012). The Role of External Sources of Information in Children's Evaluative Food Categories. *Infant and Child Development*, *21*(2), 216–235. https://doi.org/10.1002/icd.745

Nicholson, K. J., Sherman, M., Divi, S. N., Bowles, D. R., & Vaccaro, A. R. (2022). The Role of Family-wise Error Rate in Determining Statistical Significance. *Clinical Spine Surgery*, *35*(5), 222–223. https://doi.org/10.1097/bsd.0000000000001287

Notopoulos, K. (2024, 2024, December 6). Moxie, the $1,500 AI robot toy for kids, is shutting down. *Business Insider*.

Oranç, C., & Küntay, A. C. (2020). Children's perception of social robots as a source of information across different domains of knowledge. *Cognitive Development*, *54*, 100875. https://doi.org/10.1016/j.cogdev.2020.100875

Ozili, P. K. (2023). The Acceptable R-Square in Empirical Modelling for Social Science Research. In C. A. Saliya (Ed.), *Social Research Methodology and Publishing Results: A Guide to Non-Native English Speakers* (pp. 134–143). IGI Global. https://doi.org/10.4018/978-1-6684-6859-3.ch009

Palmquist, C. M., Floersheimer, A., Crum, K., & Ruggiero, J. (2022). Social cognition and trust: Exploring the role of theory of mind and hostile attribution bias in children's skepticism of inaccurate informants. *Journal of Experimental Child Psychology*, *215*, 105341. https://doi.org/10.1016/j.jecp.2021.105341

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, *43*(5), 1216–1226. https://doi.org/10.1037/0012-1649.43.5.1216

Peng, C.-Y. J., Lida, L. K., & and Ingersoll, G. M. (2002). An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational Research*, *96*(1), 3–14. https://doi.org/10.1080/00220670209598786

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Roehrig, G. H., Dubosarsky, M., Mason, A., Carlson, S., & Murphy, B. (2011). We Look More, Listen More, Notice More: Impact of Sustained Professional Development on Head Start Teachers' Inquiry-Based and Culturally-Relevant Science Teaching Practices. *Journal of Science Education and Technology*, *20*(5), 566–578. https://doi.org/10.1007/s10956-011-9295-2

Rosanda, V., & Istenic Starcic, A. (2020). The Robot in the Classroom: A Review of a Robot Role. In *Emerging Technologies for Education. SETE 2019. Lecture Notes in Computer Science*, Cham. https://doi.org/10.1007/978-3-030-38778-5_38

Rubio-Fernández, P., & Geurts, B. (2016). Don't Mention the Marble! The Role of Attentional Processes in False-Belief Tasks. *Review of Philosophy and Psychology*, *7*(4), 835–850. https://doi.org/10.1007/s13164-015-0290-z

Ryan, T. A. (1959). Multiple comparison in psychological research. *Psychological bulletin*, *56*(1), 26–47. https://doi.org/10.1037/h0042478

Salgado, J. F. (2018). Transforming the area under the normal curve (AUC) into Cohen'sd, Pearson's rpb, odds-ratio, and natural log odds-ratio: Two conversion tables. *European Journal of Psychology Applied to Legal Context*, *10*(1), 35–47. https://doi.org/10.5093/ejpalc2018a5

Sampaio, L. R., Harris, P. L., & Barros, M. L. (2019). Children's selective trust: When a group majority is confronted with past accuracy. *British Journal of Developmental Psychology*, *37*(4), 571–584. https://doi.org/10.1111/bjdp.12297

Schmitz, A., Richard, J., & and Golds, P. (2020). Is seeing believing? The effects of virtual reality on young children's understanding of possibility and impossibility. *Journal of Children and Media*, *14*(2), 158–172. https://doi.org/10.1080/17482798.2019.1684964

Skjæveland, Y. (2017). Learning history in early childhood: Teaching methods and children's understanding. *Contemporary Issues in Early Childhood*, *18*(1), 8–22. https://doi.org/10.1177/1463949117692262

Souza, D. d. H., Sarah, S., & and Koenig, M. A. (2021). Selective Trust and Theory of Mind in Brazilian Children: Effects of Socioeconomic Background. *Journal of Cognition and Development*, *22*(2), 169–184. https://doi.org/10.1080/15248372.2020.1867553

Spektor-Precel, K., & Mioduser, D. (2015). *The influence of constructing robot's behavior on the development of theory of mind (ToM) and theory of artificial mind (ToAM) in young children* Proceedings of the 14th International Conference on Interaction Design and Children, Boston, Massachusetts. https://doi.org/10.1145/2771839.2771904

Stower, R., Calvo-Barajas, N., Castellano, G., & Kappas, A. (2021). A Meta-analysis on Children's Trust in Social Robots. *International Journal of Social Robotics*, *13*(8), 1979–2001. https://doi.org/10.1007/s12369-020-00736-8

Stower, R., Kappas, A., & Sommer, K. (2024). When is it right for a robot to be wrong? Children trust a robot over a human in a selective trust task. *Computers in Human Behavior*, *157*, 108229. https://doi.org/https://doi.org/10.1016/j.chb.2024.108229

Thellman, S., Graaf, M. d., & Ziemke, T. (2022). Mental State Attribution to Robots: A Systematic Review of Conceptions, Methods, and Findings. *ACM Transactions on Human-Robot Interaction*, *11*(4), 1–51. https://doi.org/10.1145/3526112

Tong, Y., Wang, F., & Danovitch, J. (2020). The role of epistemic and social characteristics in children's selective trust: Three meta-analyses. *Developmental Science*, *23*(2), e12895. https://doi.org/10.1111/desc.12895

Tuerlinckx, F., Rijmen, F., Verbeke, G., & De Boeck, P. (2006). Statistical inference in generalized linear mixed models: A review. *British Journal of Mathematical and Statistical Psychology*, *59*(2), 225–255. https://doi.org/10.1348/000711005X79857

Valovičová, Ľ., Trníková, J., Sollárová, E., & Katrušín, B. (2020). Stimulation and Development of Intellectual Abilities in Preschool-Age Children. *Education Sciences*, *10*(2), 43.

VanderBorght, M., & Jaswal, V. K. (2009). Who knows best? Preschoolers sometimes prefer child informants over adult informants. *Infant and Child Development*, *18*(1), 61–71. https://doi.org/10.1002/icd.591

Vygotskii, L. S., & Cole, M. (1978). *Mind in society : the development of higher psychological processes*. Harvard University Press.
Wellman, H. M. (2010). Developing a Theory of Mind. In *The Wiley-Blackwell Handbook of Childhood Cognitive Development* (pp. 258–284). https://doi.org/https://doi.org/10.1002/9781444325485.ch10
Wellman, H. M. (2017). The Development of Theory of Mind: Historical Reflections. *Child Development Perspectives*, *11*(3), 207–214. https://doi.org/10.1111/cdep.12236
Wellman, H. M. (2018). Theory of mind: The state of the art*. *European Journal of Developmental Psychology*, *15*(6), 728–755. https://doi.org/10.1080/17405629.2018.1435413
Wellman, H. M., & Liu, D. (2004). Scaling of Theory-of-Mind Tasks. *Child Development*, *75*(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x
Zhang, Y., Song, W., Tan, Z., Wang, Y., Lam, C. M., Hoi, S. P., Xiong, Q., Chen, J., & Yi, L. (2019). Theory of Robot Mind: False Belief Attribution to Social Robots in Children With and Without Autism. *Frontiers in Psychology*, *10*, 1732. https://doi.org/10.3389/fpsyg.2019.01732
Zhao, W. (2018). STEM education in English of early childhood in China. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(6), 2367–2378.
Zhou, K., Li, Z., Zhang, X., Li, T., & Li, J. (2025). Epistemic trust in digital voice assistants in conflict situations among 4- to 6-year-olds. *Journal of Applied Developmental Psychology*, *98*, 101804. https://doi.org/10.1016/j.appdev.2025.101804

Appendix A

The procedure and result of power analysis

To evaluate the appropriate sample size required to detect statistically meaningful findings, a simulation-based power analysis was conducted using the simr package in R. The target model was a generalised linear mixed-effects model (GLMM) predicting a binary outcome (i.e., choosing NAO or the human) with fixed effects for condition, domain, and their interaction, and random intercepts for participant ID and trial. For example, the model structure for the nomination question was specified as:

nomination_question ~ condition * domain + (1 | id) + (1 | trial).

The parameter of interest was the model-based interaction term condition × domain. Under each condition (*Nao-accurate* and *Human-accurate*), every child was presented with information from four domains, with each domain comprising three trials, resulting in a total of twelve observations per participant. A large effect size (log-odds = 0.80) was manually imposed on the interaction coefficient to simulate a conservative but meaningful difference. An initial sample size of 100 participants (each completing 12 trials, totalling 1200 observations) was tested. The simulation was run across 100 iterations using a z-test. The resulting estimated power was 48% (95% CI [37.90%, 58.22%]), indicating that the current sample size was insufficient to reliably detect a large interaction effect. Considering potential attrition, the study initially recruited 120 children aged 4 to 6 years. Ultimately, 107 children fully participated in both two sessions.

Appendix B

Table Fixed and interaction effect in GLMMs for the nomination question

	BIC	conditional R^2	χ^2	*df*	p-value
Null model	1633.11	0.00			
Model 1	1599.67	0.32			
condition			44.49	1	<0.001***
Model 2	1623.25	0.34			
condition			41.64	1	<0.001***
domain			1.64	3	0.651
condition × domain			17.87	3	<0.001***
Model 3	1672.97	0.34			
condition			38.02	1	<0.001***
domain			1.62	3	0.656
ToAM			1.60	1	0.206
condition × domain			18.62	3	<0.001***
condition × ToAM			1.19	1	0.275
domain × ToAM			1.77	3	0.620
condition × domain × ToAM			2.92	3	0.404
Full model	1777.06	0.35			
condition			37.97	1	<0.001***
domain			1.67	3	0.644
ToM			0.86	1	0.353
ToAM			<0.01	1	0.986
condition × domain			18.36	3	<0.001***
condition × ToM			0.03	1	0.863
domain × ToM			2.39	3	0.495
condition × ToAM			0.37	1	0.544
domain × ToAM			0.98	3	0.805
ToM × ToAM			<0.01	1	0.936
condition × domain × ToM			0.19	3	0.978
condition × domain × ToAM			1.37	3	0.712
condition × ToM × ToAM			0.47	1	0.491
domain × ToM × ToAM			2.65	3	0.448
condition × domain × ToM × ToAM			3.25	3	0.355

Appendix C

Table Fixed and interaction effect in GLMMs for the endorsement question

	BIC	conditional R^2	χ^2	*df*	p-value
Null model	1417.00	0.00			
Model 1	1361.98	0.56			
condition			71.21	1	<0.001***
Model 2	1375.56	0.58			
condition			65.54	1	<0.001***
domain			3.30	3	0.347
condition × domain			24.71	3	<0.001***
Model 3	1417.89	0.59			
condition			61.51	1	<0.001***
domain			3.50	3	0.321
ToAM			0.66	1	0.415
condition × domain			25.05	3	<0.001***
condition × ToAM			1.98	1	0.159
domain × ToAM			8.94	3	0.030*
condition × domain × ToAM			3.06	3	0.383
Full model	1517.02	0.61			
condition			54.48	1	<0.001***
domain			3.00	3	0.392
ToM			1.05	1	0.305
ToAM			0.30	1	0.582
condition × domain			22.97	3	<0.001***
condition × ToM			0.62	1	0.432
domain × ToM			1.59	3	0.662
condition × ToAM			0.10	1	0.749
domain × ToAM			0.19	3	0.978
ToM × ToAM			0.37	1	0.544
condition × domain × ToM			4.55	3	0.208
condition × domain × ToAM			0.66	3	0.883
condition × ToM × ToAM			1.75	1	0.186
domain × ToM × ToAM			1.26	3	0.738
condition × domain × ToM × ToAM			2.80	3	0.424

Appendix D

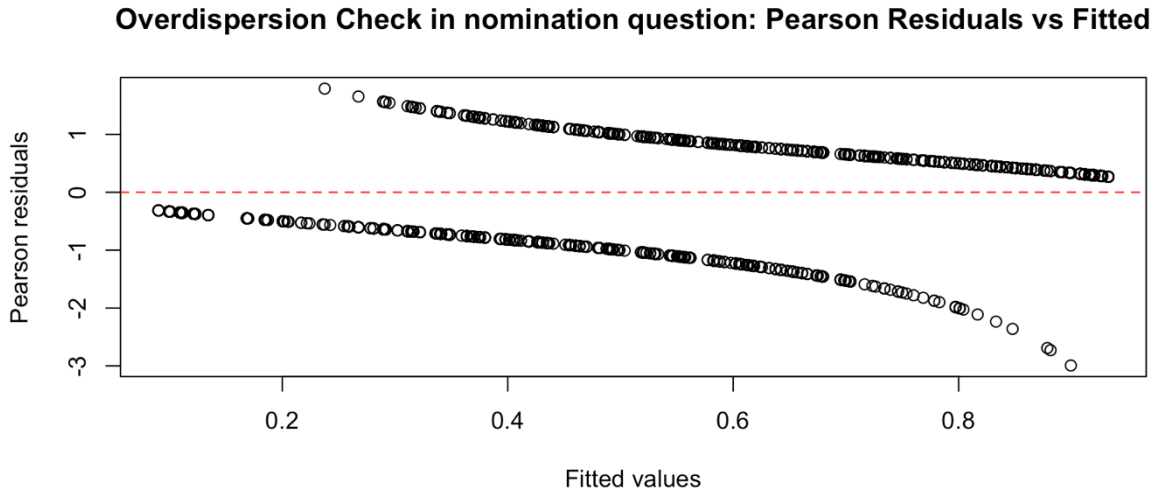
Table Fixed and interaction effect in GLMMs for the accuracy question.

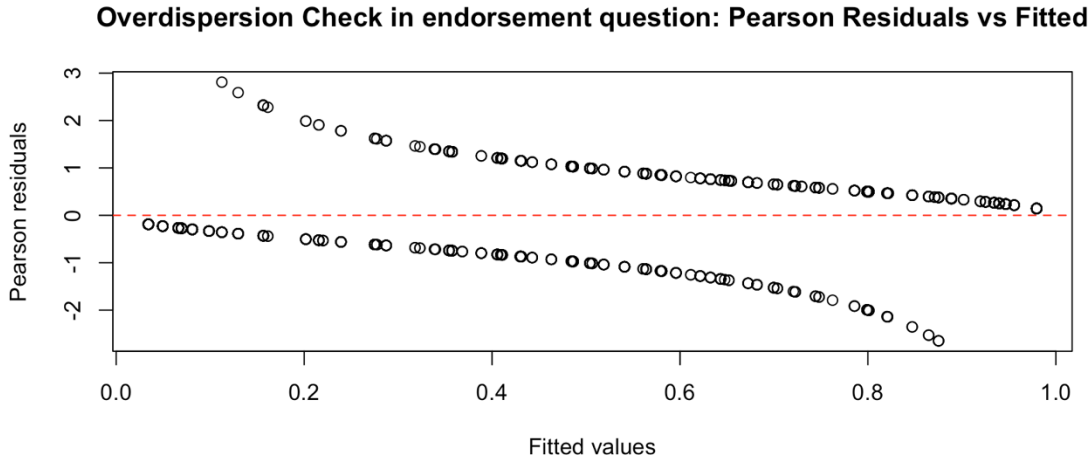
	BIC	conditional R^2	χ^2	*df*	p-value
Null model	1332.36	0.00			
Model 1	1285.82	0.00			
condition			56.86	1	<0.001***
Model 2	1299.62	0.65			
condition			51.77	1	<0.001***
domain			6.45	3	0.092
condition × domain			20.17	3	<0.001***
Model 3	1342.58	0.65			
condition			48.90	1	<0.001***
domain			6.61	3	0.085
ToAM			0.10	1	0.758
condition × domain			20.39	3	<0.001***
condition × ToAM			2.64	1	0.104
domain × ToAM			2.82	3	0.420
condition × domain × ToAM			6.52	3	0.089
Full model	1312.2	0.65			
condition			46.39	1	<0.001***
domain			5.68	3	0.128
ToM			1.81	1	0.179
ToAM			1.17	1	0.280
condition × domain			22.85	3	<0.001***
condition × ToM			0.85	1	0.355
domain × ToM			0.73	3	0.867
condition × ToAM			0.11	1	0.738
domain × ToAM			0.03	3	0.998
ToM × ToAM			0.37	1	0.542
condition × domain × ToM			0.15	3	0.985
condition × domain × ToAM			1.68	3	0.642
condition × ToM × ToAM			3.68	1	0.055
domain × ToM × ToAM			5.36	3	0.148
condition × domain × ToM × ToAM			3.60	3	0.308

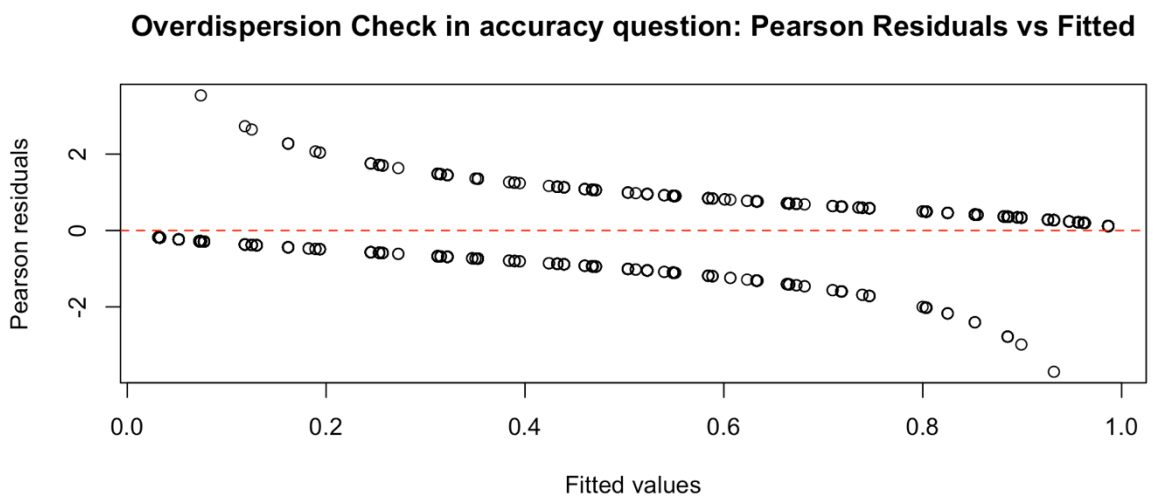
Appendix E

To evaluate the assumption of appropriate dispersion in the GLMM for the **nomination, endorsement and accuracy question**, a plot of Pearson residuals versus fitted values was generated. In the present plot, the residuals are clustered and symmetrically distributed around the upper and lower bounds, with no obvious fanning or curvature. The absence of a funnel shape or vertical banding indicates that overdispersion is not evident, which is consistent with the numeric overdispersion test

Figure E1 – E3 Overdispersion diagnostics for GLMMs






Overdispersion Check in accuracy question: Pearson Residuals vs Fitted
Pearson residuals
-2
0
2
0.0
0.2
0.4
0.6
0.8
1.0
Fitted values

Appendix F

Table F1 Thorough descriptive analyses of ToAM and ToM scales

Scale	M	SD	median	trimmed	mad	min	max	range	skewness	kurtosis	SE
ToAM	4.48	1.91	5	4.53	2.97	1	7	6	-0.17	-1.28	0.18
ToM	3.91	2.17	4	3.93	2.97	0	7	7	0	-1.32	0.21

Descriptive statistics were computed to summarise the distributions of children's ToAM and ToM scale scores. The mean ToAM score was 4.48 (SD = 1.91), indicating that, on average, children performed relatively well on the ToAM tasks. The median score was 5, with a trimmed mean of 4.53, suggesting that the distribution was not strongly affected by outliers. The median absolute deviation (MAD) was 2.97, reflecting moderate variability around the median. ToAM scores ranged from 1 to 7, with a skewness of -0.17 and kurtosis of -1.28, indicating a slightly left-skewed. The standard error (SE) was 0.18, suggesting a precise estimate of the mean. Similarly, ToM scores showed a mean of 3.91 (SD = 2.17), with a median of 4 and a trimmed mean of 3.93. The MAD was also 2.97. Scores ranged from 0 to 7, with a skewness of 0, indicating a symmetrical distribution, and kurtosis of -1.32, again pointing to a relatively flat shape. The SE was 0.21.

Overall, both ToAM and ToM showed moderate variability, with slightly a lower peak and wide tails compared with a normal distribution, and only minimal skew. These metrics collectively provide a comprehensive understanding of the distributional properties of children's performance on both cognitive constructs.

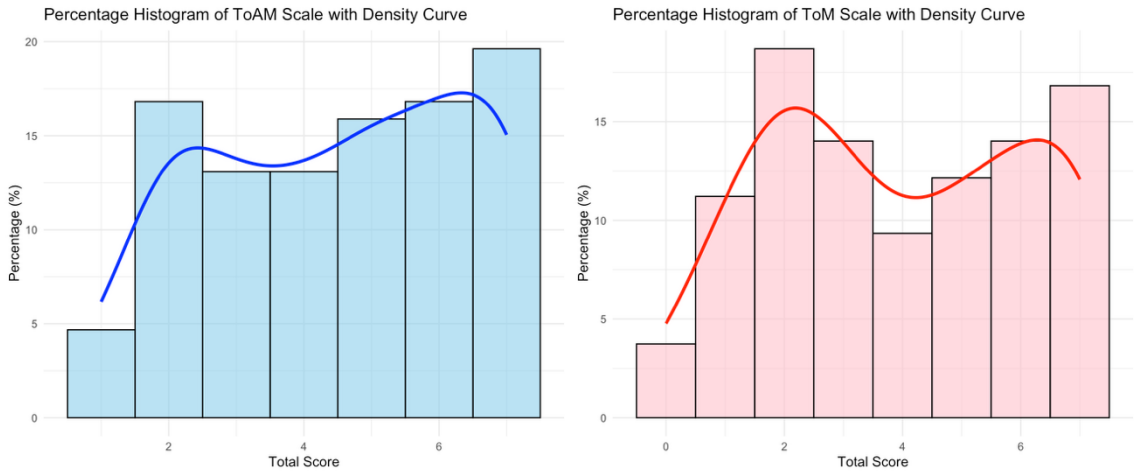


Figure F1 Percentage histogram of ToAM and ToM scale score with density curve

Appendix G

Correlation Matrix For Key Variables

Correlation analysis was conducted to examine the associations among the key variables in the study, including ToM, ToAM, children's responses to the selective trust tasks (nomination question, endorsement question, and accuracy question), age, condition (*Nao-accurate and Human-accurate*), and domain (non-STEM, physical science, life science and mathematics). All reported associations were calculated using Pearson's product-moment correlation coefficient, and significance levels are reported in the corresponding matrix.

Table G1 Correlation matrix for diverse key variables

	ToM	ToAM	nomination	endorsement	accuracy	age	condition	domain	gender
ToM	1								
ToAM	0.801***	1							
nomination	-0.099***	-0.105***	1						
endorsement	-0.121***	-0.13***	0.235***	1					
accuracy	-0.111***	-0.108***	0.312***	0.716***	1				
age	-0.117***	-0.214***	-0.013	0.008	0.012	1			
condition	0.14***	0.191***	-0.302***	-0.464***	-0.468***	-0.83**	1		
domain	0	0	-0.004	0.032	0.039	0	0	1	
gender	0.1***	0.08**	-0.034	-0.043	-0.078**	0.018	0.033	0	1

According to the matrix, ToM and ToAM total scores were strongly positively correlated (r = 0.83), indicating that children's understanding of artificial minds was closely aligned with their reasoning about human mental states. Weak negative correlations were found between ToM/ToAM and the nomination, endorsement and accuracy questions, suggesting that children with stronger mind reasoning were slightly less inclined to trust the robot Nao. The three selective trust questions were moderately interrelated, particularly the endorsement and accuracy questions, indicating consistency across trust-related judgments and strong internal reliability. However, age was modestly negatively associated with both ToAM (r = -0.21) and ToM (r = -0.12), implying that younger children tended to score higher on these measures, which was conflicted with the results in Mao et al. (2025). Based on this interesting result, the associations between age and ToM/ToAM were plotted.

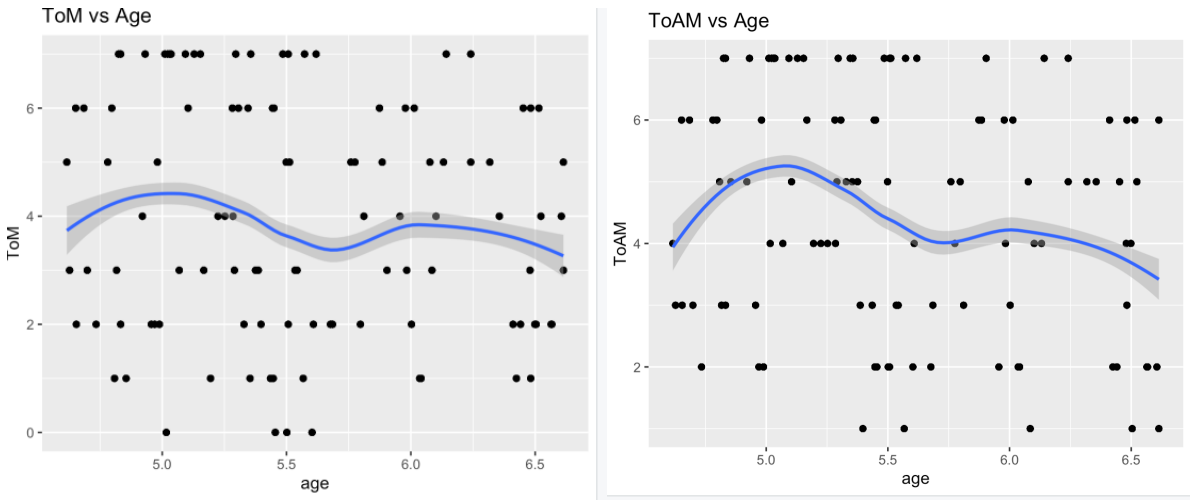


Figure G1 associations between age and children's ToM and ToAM scores

Contrary to the typical developmental expectation that older children perform better on ToM tasks, the plots revealed a non-linear trend. This unexpected pattern may be partially originated from the uneven age distribution in the sample. The largest group consisted of children aged 5-6 years (n = 56), with smaller samples from the 4-5 age group (n = 21) and the 6-7 group (n = 30). Given the limited number of younger and older participants, statistical noise or other latent factors could have contributed to the apparent decline in performance with age.

Correlation Matrix in each condition

Table G2 Correlation matrix for diverse key variables in *Nao-accurate* condition

	ToM	ToAM	nomination	endorsement	accuracy	age	domain	gender
ToM	1							
ToAM	0.791***	1						
nomination	-0.107**	-0.101*	1					
endorsement	-0.165***	-0.133**	0.104**	1				
accuracy	-0.176***	-0.136***	0.135***	0.800***	1			
age	0.047	-0.022	-0.120**	0.010	0.033	1		
domain	0	0	-0.097*	0.064	0.054	0	1	
gender	0.263***	0.235***	-0.022	-0.097*	-0.096*	-0.09*	0	1

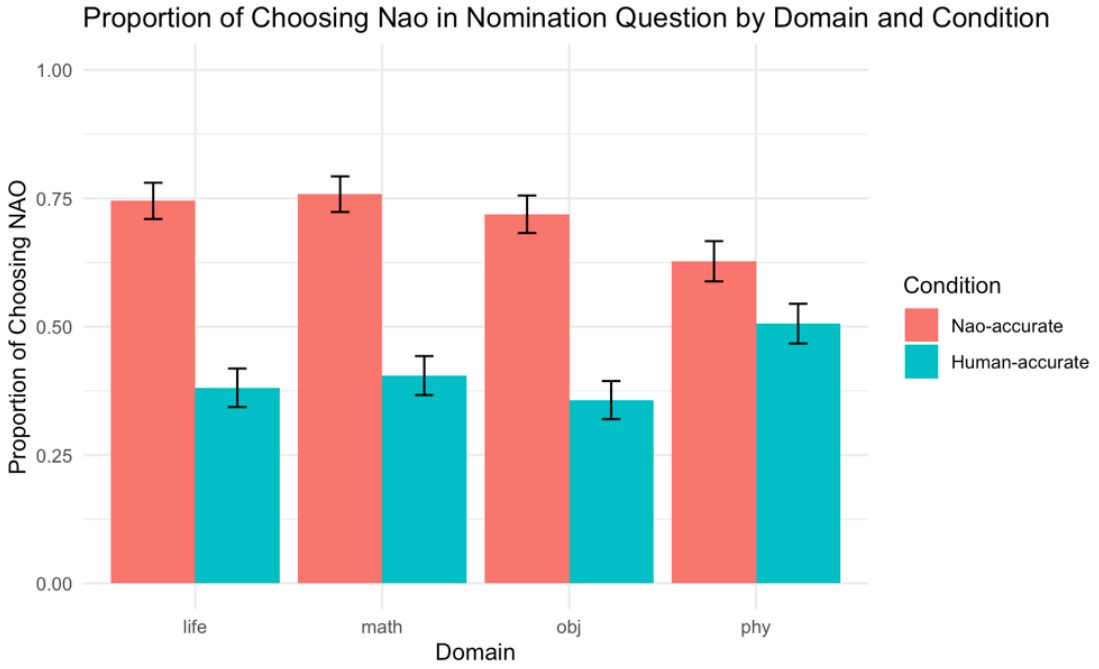
Table G2 Correlation matrix for diverse key variables in *Human-accurate* condition

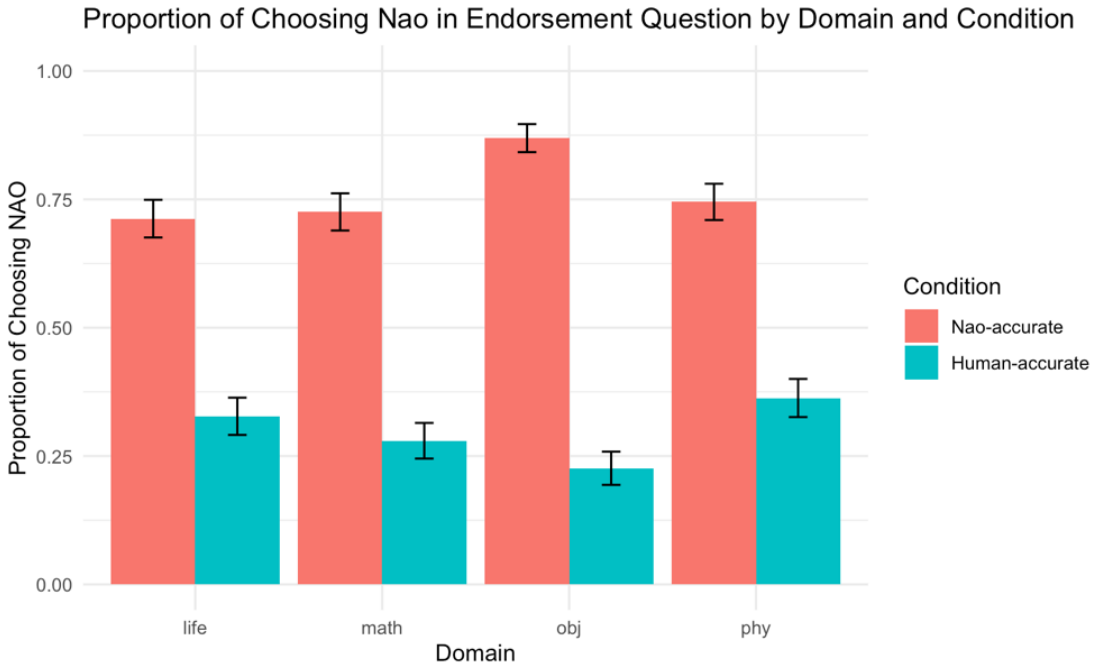
	ToM	ToAM	nomination	endorsement	accuracy	age	domain	gende

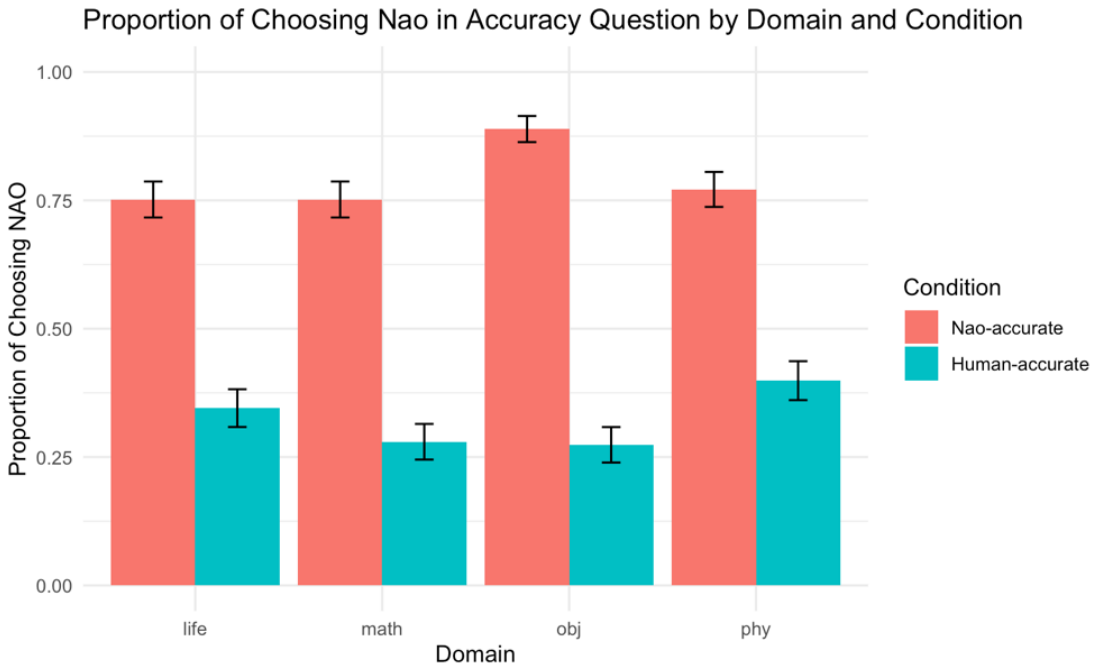
ToM	1						
ToAM	0.808***	1					
nomination	-0.023	-0.003	1				
endorsement	0.019	0.034	0.120**	1			
accuracy	0.044	0.082	0.253***	0.519***	1		
age	-0.244***	-0.394***	0.028	-0.056	-0.082*	1	
domain	0	0	0.074	0.013	0.037	0	1
			(p = 0.054)				
gender	-0.051	-0.090*	-0.028	0.024	-0.052	0.123**	0

Appendix H

Three plots illustrated the proportion of children who selected the robot (Nao) in the nomination, endorsement and accuracy questions across different knowledge domains and conditions, highlighting how trust varies depending on informant accuracy and information domain.






Proportion of Choosing Nao in Accuracy Question by Domain and Condition
Proportion of Choosing NAO
1.00
0.75
0.50
0.25
0.00
life
math
obj
phy
Domain
Condition
Nao-accurate
Human-accurate

Appendix I

DREC ethical approval letter.

Education (Educ) DREC
15 Norham Gardens, Oxford, OX2 6PY

Applicant: Keyu Mao
Principal Investigator: Lars-Erik Malmberg
Department: Education

Study title: Artificial Minds, Real Decisions: Children Selectively Trust the STEM and non-STEM Information from Robots
(Version: 1.0)

Ethics reference: Education (Educ) DREC - 1433726

Dear Lars-Erik Malmberg,

On behalf of the Committee, I confirm that the above research study described in the application and other supporting documentation submitted to the committee has been carefully considered by the Education (Educ) DREC in accordance with the University's regulations and policy for ethics approval of research involving human participants, human tissue and/or personal data. The opinion is as follows:

Opinion of Research Ethics Committee: Favourable Opinion

Subject to the following conditions:

Decision Date: 1 Apr 2025, 15:08

Opinion End Date: 1 Apr 2026

If favourable, insurance-provided indemnity arrangements will be in place between the decision date and opinion end date and you may now commence your study activities. Should you plan to continue the research beyond the end date above, it is your responsibility to ensure that you request, and receive, an extension (via amendment) from the committee for indemnity to remain in place. You may be required to provide a justification.

Please note the following:

Amendments: Should there be any subsequent changes to the reviewed study, applications for amendments can be made via the Oxford Ethics Application System (Worktribe Ethics).

Reports: Studies considered by OxTREC are expected to submit an *annual progress report* on each anniversary of study approval, until the study is completed. An end of study report is also required.

Audit: This study may be selected for audit at the discretion of the Research Governance, Ethics and Assurance Team.

Data safety: It is the responsibility of the PI to ensure that all data collected during the course of the study is stored and transferred safely and securely in accordance with University requirements. Further guidance and advice are available from the Research Data Team. Additional information is available at https://researchsupport.web.ox.ac.uk/governance/ethics

Appendix J

Parent/guardian consent form.

DEPARTMENT OF EDUCATION

[Lars-Erik Malmberg]
[lars-erik.malmberg@education.ox.ac.uk]

PARENT/GUARDIAN CONSENT FORM家长/监护人知情书

CUREC Approval Reference伦理审批参考编号: Education (Educ) DREC – 1433726

[Children Learning Knowledge from Robots儿童向机器人学习]

- Your child's school has agreed to take part in a study run by the University of Oxford looking at how children perceive and selectively trust robots. 您孩子所在的学校已同意参与一项由牛津大学开展的研究，该研究旨在探讨儿童如何感知机器人及信任机器人所提供的知识。
- If your child takes part, a researcher would come and visit them at school, do some activities and play some fun games with them. 如果您的孩子参与研究，研究人员将到学校与孩子们进行一些活动和有趣的游戏。
- To find out more about the study, please read the attached information sheet. You can e-mail me at keyu.mao@education.ox.ac.uk, if you have any questions. 如需了解更多研究详情，请阅读随附的研究说明材料。如果您有任何问题，欢迎通过电子邮件联系我：keyu.mao@education.ox.ac.uk。
- If you are happy for your child to take part, please fill in the form below and return it to your child's class teacher as soon as possible. **如果您同意您的孩子参与研究，请填写下方的表格，并将回执交于教师。**

Name of child 幼儿姓名: _____

I have read and understood the details of the above study, and have had the opportunity to ask questions and discuss the study with others. I have received satisfactory answers to my questions. I understand that the project has received ethics clearance through the University of Oxford's ethical approval process for research involving human participants, and I understand who will have access to the data, how it will be stored and what will happen to the data at the end of the study. I understand that participation is voluntary and that my child and I are free to withdraw at any time, without giving any reason and without my child's education being affected in any way. I understand how to raise a concern or make a complaint.

我已阅读并理解上述研究的详细信息，并知晓自己有机会提出问题或与他人讨论研究内容。我了解到，该项目已通过牛津大学针对人类参与者研究的伦理审查程序，并清楚谁将访问数据、数据如何存储以及研究结束后数据的处理方式。我明白参与研究是完全自愿的，**我和我的孩子可以随时退出**，无需提供任何理由，且不会对孩子的教育造成任何影响。我也了解如何提出疑虑或投诉。

I give permission for my child to take part in the above study. 我允许我的孩子参与该研究。

Name of parent/guardian 监护人姓名: _____

Artificial Minds, Real Decisions: Children Selectively Trust The STEM And Non-STEM Information From Robots
Consent Form, version 1.0, March 2025

Signature 监护人签名: _____ **Date日期:** _ _ _ _ _ _ _ _

Name of researcher 研究者姓名: _____

Signature 研究者签名: _____ **Date:** _ _ _ _ _ _ _ _

Appendix K

The letter to head teacher.

DEPARTMENT OF EDUCATION

[Lars-Erik Malmberg]
[lars-erik.malmberg@education.ox.ac.uk]

[Yanhong Huang]
[Yuedu Kindergarten, 10 Jinjishan Road, Zhuji, Zhejiang, China]
[15 March 2025]

Children Learning Knowledge from Robots
儿童向机器人学知识

Ethics Approval Reference伦理审批参考编号: Education (Educ) DREC - 1433726

Dear Mrs. Yanhong Huang, 尊敬的黄园长，

I am writing to enquire about conducting some research in your school between 1st April to 15th May 2025. I am a taught master student at the University of Oxford, supervised by Professor Lars-Erik Malmberg. In my research study, Children Learning The STEM And Non-STEM Information From Robots Depending On Theory Of Artificial Mind (ToAM), I will explore how children's Theory of Artificial Mind —their understanding and interpretation of robots' mental states—affects their selective trust across STEM (physics, mathematics and life science) and non-STEM (object labelling) domains.

我目前是牛津大学教育系的一名硕士研究生，导师是Lars-Erik Malmberg教授。我写信是想询问是否可以在贵园进行我的毕业论文研究。我的研究课题是《儿童基于人工心智理论（ToAM）从机器人身上学习STEM与非STEM信息》，旨在探讨儿童对机器人心理状态的理解（即"人工心智理论ToAM"）如何影响他们在STEM（物理、生物和数学）和非STEM（如物体标签）领域中对机器人的选择性信任。研究计划的时间为2025年4月1日至5月15日。我希望通过这项研究，能够更好地理解儿童与机器人互动时的学习模式，从而为未来的教育实践提供参考。

The research will take place with children aged 4 to 6 from preschool. I am not aiming to change what or how the teacher chooses to teach, and will not be making any judgements about teaching. The study is separate from regular classroom teaching and will not interfere with the school's curriculum. The research consists of simple and engaging games where children interact with a robot and a human, listen to information, and answer questions about who they trust more.

这项研究将面向幼儿园4至6岁的儿童开展。我的研究并不试图改变老师选择的教学内容或方式，也不会对教学进行任何评价。研究将与日常课堂教学分开进行，不会干扰学校的正常课程安排。

研究内容主要包括一些简单而有趣的小游戏。在游戏中，孩子们会与机器人和人类互动，听取信息，并回答一些问题，比如他们更信任谁（机器人还是人类）。整个过程轻松愉快，旨在通过游戏的形式了解孩子们的选择性信任模式。

By participating in the research, your school would be contributing to research that will help researchers and educators better understand how children interact with technology and how social robots might be used in education in the future.

通过参与这项研究，贵园将为一项重要的研究课题贡献力量。这项研究将帮助研究人员和教育工作者更好地理解儿童如何与技术互动，以及未来如何将社交机器人应用于教育领域。您的支持将为我们探索儿童学习的新方式提供宝贵的数据和见解。

The University of Oxford has strict ethics procedures on conducting research with teachers and children, consistent with current British Educational Research Association guidelines. Before beginning the research, I would inform parents/guardians about the research and offer the children and parents/guardians the opportunity to refuse to participate. Throughout the research, children and parents/guardians will be able to refuse to participate at any time. **No personal or identifiable information** (such as names, addresses, or school details) will be recorded, and **no photos, videos, or audio recordings** will be taken. I will collect **children's age, gender, and responses to tasks**

牛津大学对与教师和儿童进行的研究有严格的伦理规范，这些规范与英国教育研究协会（BERA）的最新指南一致。在研究开始之前，我会向监护人详细介绍研究内容，并为幼儿和监护人提供拒绝参与的机会。在整个研究过程中，幼儿和监护人也可以随时选择退出。

研究过程中不会记录任何个人或可识别的信息（如姓名、地址或学校信息），也不会拍摄照片、视频或录音。我只会收集儿童的年龄、性别以及他们在任务中的回答，所有数据将严格保密，仅用于研究目的。

All participants would be anonymised in all research reports. The data collected would be kept strictly confidential, available only to my supervisor and myself and not used other than specified without the further consent of all involved being obtained. I have enclosed copies of the information for parents/guardians and children with this letter.

所有参与者的信息在研究报告中都将被匿名处理。收集到的数据将严格保密，仅限我的导师和我本人访问。除非获得所有相关方的进一步同意，否则这些数据不会用于指定范围以外的其他用途。

随信附上了提供给监护人和幼儿的研究说明材料，供您参考

If your school would like to take part in the study, or you need more information about what is involved, please contact me. Whether or not you feel it would be appropriate for your school to participate, I would be grateful if you would complete the pro-forma below, and return it to me via email.

如果贵园有兴趣参与这项研究，或者需要了解更多相关信息，请随时与我联系。无论贵园是否决定参与，我都非常感激您能填写下方的回执表，并通过电子邮件回复我。

Thank you for your time and attention. I look forward to hearing from you.

感谢您抽出时间阅读此信。期待您的回复！

Yours Sincerely,

祝身体健康，工作顺利

[Keyu Mao毛珂妤]

[Artificial Minds, Real Decisions: Children Selectively Trust the STEM and non-STEM Information from Robots 人工心智，真实选择：儿童对机器人提供的STEM与非STEM信息的选择性信任**]**

[Keyu Mao毛珂妤]
[Department of Education教育系]

[Yuedu Kindergarten越都幼儿园]

[10 Jinjishan Road, Zhuji, Zhejiang, China 浙江省绍兴市诸暨市金鸡路10号]

[Yanhong Huang黄燕虹]

☐ We do not wish to participate in this project. 我不愿意参与本项目

☐ We would like to find out more about this project. 我需要更多相关信息

☐ We would like to take part in this project. 我愿意参与本项目

If you would like further information, or are **interested in taking part**, please give the name of a **contact person** for your school, and details of the best way to contact them.
如果您希望了解更多信息，或有意参与研究，敬请提供贵园的联系人姓名及其最佳联系方式。

Contact name联系人: _____

Contact email电子邮箱: _____

Contact telephone number手机号: _____

Please return this form via email [keyu.mao@education.ox.ac.uk].
请将回执发送至邮箱
Thank you for your help.
感谢您的帮助

Appendix L

Information sheet for parents/guardians.

[Lars-Erik Malmberg]
[lars-erik.malmberg@education.ox.ac.uk]
[Primary researcher contact details and status [MSc Education Student]
Oxford University telephone number: 01865274024
Oxford University email address: keyu.mao@education.ox.ac.uk

Children Learning Knowledge from Robots儿童向机器人学习知识

INFORMATION SHEET FOR PARENTS / GUARDIANS 研究信息说明 (家长版)

Central University Research Ethics Committee Approval Reference 伦理审批参考编号: Education (Educ)
DREC - 1433726

In partnership with researchers at the University of Oxford, Yuedu Kindergarten has agreed to take part in a research study investigating [how children perceive and selectively trust robots]. We would like to invite your child to be part of this research. We very much hope you would like your child to take part, but before you decide, it is important that you understand why the research is being done and what it will involve.

越都幼儿园已同意与英国牛津大学研究人员合作，参与研究儿童如何感知机器人并信任机器人所提供的信息。我们诚挚邀请您的孩子参与这项研究。但在您做出决定之前，请您详细了解研究的目的和具体内容。

Why is this research being conducted? 为什么进行该研究？

Children today interact with Artificial Intelligence, including robots, more than ever. However, we don't fully understand how children decide whether to trust information from robots compared to humans, especially in subjects like science and maths. Thus, this study explores how children perceive and learn from both human teachers and robots. We aim to find out whether children are more likely to trust robots for certain types of information, and how this changes with age and understanding of robots. Investigating this can help improve how robots are used in education.

如今，儿童与AI的互动比以往任何时候都更加频繁。心理学研究者尝尝探讨人与人之间的信任问题，尤其是认知发展不成熟的幼儿对专家、母亲、老师的信任。然而，我们尚不完全了解幼儿对机器人的信任程度。因此，这项研究旨在探讨儿童如何从人类教师和机器人学习信息。我们希望通过研究了解儿童是否更倾向于信任机器人提供的某些类型的信息（如STEM领域的物理、数学、生物），以及这种信任如何随着年龄和对机器人理解的变化而改变。因此，这项研究的结果将有助于改进机器人在教育中的应用方式。

Why has my child been invited to be involved in this research? 为什么我的孩子被邀请参与该研究？

We are inviting your child to take part because they are a young person, aged between 4 and 6 years, attending Yuedu Kindergarten

We are inviting 120 young people to take part.

我们计划在越都幼儿园招募120名4-6岁幼儿参与研究。

Does my child have to be involved? 我的孩子必须参加吗？

No. You can ask questions about the research before deciding whether to allow your child to participate. If you do agree to participation, you may **withdraw** your child and their data **at any time**, without giving a reason and without any effect on their education, by advising the school or researchers of this decision. If the data has been collected, you can still withdraw your child and their data freely before the thesis submission; the data will be excluded.
If your child is not involved in the research, they will receive alternative provision of equivalent educational value

不。在决定是否让孩子参与研究之前，您可以随时提出问题。如果您同意孩子参与，也可以随时退出，无需提供任何理由，且不会对孩子的教育造成任何影响。您只需通知学校或研究人员即可。如果数据已经收集，您仍可以在论文提交前自由退出，相关数据将被排除在研究之外。

What will happen if my child takes part? 我的孩子会在研究中做什么？

If you agree for your child to take part, you will be asked to sign a consent form before the study begins. On the day of the study, the researcher will explain the activity to your child in simple terms and ask if they are happy to take part **(child assent)**. If they do not wish to participate, they can choose not to, and they can **stop** at any time during the study. The research will take place over two sessions, during school hours, in a quiet and familiar space. In **Week 1**, your child will have a short session lasting about **10 minutes**. They will complete a few fun tasks to see how they understand thoughts and feelings, both in people and robots. In **Week 2**, they will take part in a learning activity where they will hear different information from both a robot and a human. After listening, they will answer simple questions about whose information they trust more. The session will be designed to be enjoyable, and there are no right or wrong answers—we are only interested in their thoughts! Week 2 process will take about **15 minutes**. No personal details will be collected, and no photos, videos, or recordings will be taken.

如果您同意孩子参与研究，您需要在研究开始前签署一份同意书。在研究当天，研究人员会用简单的语言向您的孩子解释活动内容，并询问他们是否愿意参与。如果孩子不愿意参与，他们可以选择不参加，也可以在研究过程中随时退出。

研究将分为两次，在幼儿园内安静的房间进行，。

第一周：您的孩子将参与一个约10分钟的简短活动。他们会完成一些有趣的小任务，帮助我们了解他们如何理解人类和机器人的想法与感受。

第二周：他们将参与一个学习活动，聆听机器人和人类提供的不同信息。听完后，他们会回答一些简单的问题，比如他们更信任谁提供的信息。活动设计得非常有趣，且没有对错之分——我们只关心他们的想法！第二周的活动大约需要15分钟。

研究过程中不会收集任何个人信息，也不会拍摄照片、视频或录音。

What are the possible disadvantages and risks in taking part? 研究有没有潜在危害或风险？

There are no significant risks associated with taking part in this study. The activities are designed to be simple, enjoyable, and similar to everyday social experiences. If your child does not want to continue at any point, they can stop without giving a reason.

参与这项研究没有显著风险。活动设计简单有趣，类似于日常的社交体验。如果您的孩子在任何时候不想继续参与，他们可以随时退出，无需提供任何理由。

Are there any benefits in taking part? 研究对于您的孩子有什么好处？

There will be no direct or personal benefit to your child from taking part in this research. However, it is hoped that the research study will help to better integrate robots in educational settings and design child-friendly robots.

您的孩子参与这项研究不会获得直接或个人化的收益。然而，我们希望这项研究能够帮助更好地将机器人融入教育环境，并设计出更适合儿童使用的机器人。

What information will be collected and why is the collection of this information relevant for achieving the research objectives? 将收集哪些数据？为什么？

During the study, we will collect **your child's age, gender, and responses to tasks** about how they learn from robots and humans. **No personal or identifiable information** (such as names, addresses, or school details) will be recorded, and **no photos, videos, or audio recordings** will be taken.

All data will be stored securely with a password and will only be accessible to the researcher and supervisor. Consent forms will be stored separately from the research data in a locked cabinet.

The researchers will retain Consent forms for 3 years after publication of the work of the research. Regular summaries of our findings will be given to the school and will be available to interested families. I will not identify the school, teacher or any students in any reports of the research.

在研究过程中，我们将收集您孩子的年龄、性别以及他们在任务中的回答，以了解他们如何从机器人和人类那里学习信息。我们不会记录任何个人或可识别的信息（如姓名、地址或学校信息），也不会拍摄照片、视频或录音。

所有数据将通过密码安全存储，仅限研究人员和导师访问。同意书将与研究数据分开存放，并保存在上锁的柜子中。

研究人员将在研究论文发表后保留同意书3年。在研究报告中，我们不会提及学校、老师或任何学生的身份信息。

Will the research be published? Could my child be identified from any publications or other research outputs? 研究会被发表吗？出版物会透露我孩子的信息吗？

The findings from the research will/may be written up in a master's dissertation. Since I will not include raw data in the dissertation, participants cannot identify themselves from it. A copy of my dissertation will be deposited both in print and online in the Oxford University Research Archive where it will be publicly available to facilitate its use in future research/ its access will be restricted.

研究结果可能会写入我的硕士论文中。由于论文中不会包含原始数据，参与者无法从中识别自己的信息。论文的副本将以印刷版和电子版形式存入牛津大学研究档案库，供未来研究使用（或访问权限将受到限制）。

Data Protection 数据保护

The University of Oxford is the data controller with respect to your personal data, and as such will determine how your child's personal data is used in the research. Any paper records and field notes need to be shredded upon data-entry. A completely anonymised version of the data will be kept indefinitely in the Oxford Research Archive to allow for future secondary analysis.

The University will process your child's personal data for the purpose of the research outlined above. Research is a task that we perform in the public interest.

Further information about your rights with respect to your personal data is available from https://compliance.web.ox.ac.uk/individual-rights.

牛津大学是您孩子个人数据的责任方，因此将决定这些数据在研究中如何使用。所有纸质记录和现场笔记将在数据录入后销毁。经过完全匿名化处理的数据将永久保存在牛津研究档案库中，以供未来开展二次分析使用。

大学将根据上述研究目的处理您孩子的个人数据。研究是一项我们为公共利益而开展的工作。

如需了解更多关于您对个人数据的权利，请访问：https://compliance.web.ox.ac.uk/individual-rights。

Who has reviewed this research? 谁来审查本研究？

This research has received ethics approval from a subcommittee of the University of Oxford Central University Research Ethics Committee. (Ethics reference: Education (Educ) DREC - 1433726).

这项研究已获得牛津大学中央大学研究伦理委员会下属委员会的伦理批准。（伦理批准编号：xxxxx）。

Who is organising and funding the research? 谁来管理和资助本研究？

The study organiser is Keyu Mao who is a master student in Education (Child Development and Education). I will conduct the research under the supervision by Professor Lars-Erik Malmberg and University of Oxford.

The Barbinder Watson Trust Fund from St Hugh's College, University of Oxford awared Keyu Mao £200 travel grant for this project.

本人毛珂妤，系英国牛津大学教育学（儿童发展与教育）专业硕士学生。我将在Lars-Erik Malmberg教授和牛津大学的指导下开展这项研究。

牛津大学圣休学院（St Hugh's College）的Barbinder Watson信托基金为本项目提供了200英镑的旅行补助。

Who do I contact if I have a concern about the research, or I wish to complain? 如果我对研究有疑虑或希望投诉，应该联系谁？

If you have a concern about any aspect of this research, please contact **Keyu Mao** through keyu.mao@education.ox.ac.uk and I will do my best to answer your query. I will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with. If you remain unhappy or wish to make a formal complaint, please contact the University of Oxford Research Governance, Ethics & Assurance (RGEA) team at rgea.complaints@admin.ox.ac.uk or on 01865 616480.

如果您对研究的任何方面有疑虑，请通过keyu.mao@education.ox.ac.uk联系毛珂妤，我会尽力解答您的疑问。我将在10个工作日内确认收到您的意见，并告知您处理方式。如果您仍然不满意或希望正式投诉，请联系牛津大学研究治理、伦理与保证（RGEA）团队，邮箱：rgea.complaints@admin.ox.ac.uk，电话：01865 616480。

What should I do next? 接下来我该做什么？

Please fill in the **enclosed consent form** and return it to your child's class teacher if you would like your child to take part in this research. Please remember that you may withdraw your child at any time, without affecting their education and without giving a reason, by notifying the researcher.

如果您同意孩子参与研究，请填写随附的**知情书**并交回给孩子的老师。请记住，您可以随时通知研究人员退出研究，无需提供理由，且不会对孩子的教育造成任何影响。

Further Information and Contact Details 进一步的信息与联系方式

If you would like to discuss the research with someone beforehand (or if you have questions afterwards), please contact:

如果您想进一步了解、探讨研究相关问题，请联系我：

[Keyu Mao毛珂妤]
[Department of Education教育系]
[15 Norham Gardens, Oxford OX2 6PY]
University email: [keyu.mao@education.ox.ac.uk]

Appendix M

ToAM and ToM scales

Diverse Desires

The child is introduced to Nao (robot) /Tong (human), who has just woken up from a nap and would like a snack. The researcher shows the child two options — a boiled egg and an ice cream — and asks, “Which of these do you like more? Do you prefer the egg or the ice cream?” After the child responds, the researcher continues, “Okay. But Nao/Tong likes [the other item]. Nao/Tong doesn’t like [child’s choice]. Its/Her favourite is [the other item].” The child is then asked, “Nao/Tong is going to choose one of them now — Nao/Tong can’t have both. Which one do you think Nao/Tong will choose, the egg or the ice cream?” The task is considered passed only if the child answers with the item that is not their own stated preference.

Diverse Beliefs

The second task introduces a new situation: Nao/Tong is searching for the ball, which might be either under the bed or inside the cupboard. The child is asked, “Where do you think the ball is, under the bed or in the cupboard?” Regardless of the child’s response, the researcher says, “Okay! But Nao/Tong thinks the ball is in [the opposite location].” The child is then asked, “Where will Nao/Tong look for the ball, under the bed or in the cupboard?” The child must give an answer opposite to their own belief to pass this task.

Knowledge Access

The child is shown a closed drawer and asked, “What do you think is inside the drawer?” After the child responds, the researcher opens the drawer dramatically to reveal a toy dog, saying, “Wow, it’s a dog!” After closing the drawer again, the researcher confirms understanding by asking, “So, what’s inside the drawer?” Then, the researcher introduces Nao/Tong, saying, “Here comes Nao/Tong. Nao/Tong hasn’t seen what’s inside the drawer.” The child is then asked, “Does Nao/Tong know what’s in the drawer? [knowledge question]” and “Did Nao/Tong see what was inside the drawer? [memory question]” The task is scored correct only if the child answers both “No” to knowledge and memory control questions.

Contents False Belief

The researcher shows the child a crisps container and asks, “What do you think is inside this container?” Then, the researcher opens the container with excitement and reveals, “Oh! It’s a black pen!” The researcher checks comprehension: “So, what is really inside the container?” Once this is confirmed, Nao/Tong is introduced with the explanation, “Nao/Tong hasn’t seen what’s inside this container.” The key question follows: “What does Nao/Tong think is inside the container, crisps or a black pen?” The researcher also asks, “Did Nao/Tong see what was inside the container?” A correct response requires the child to answer “crisps” (based on the label) and “no” (Nao did not see), showing false belief attribution.

Explicit False Belief

The child is told that Nao/Tong is looking for a book, which might be either in the backpack or in the drawer. The researcher explicitly states: “Nao’s/Tong’s book is actually in the backpack, but Nao/Tong thinks it’s in the drawer.” The child is then asked, “Where will Nao/Tong look for the book, the backpack or the drawer?” and, separately, “Where is Nao’s/Tong’s book really?” To pass this task, the child must correctly identify Nao’s false belief (he will look in the drawer) and also identify the true location (backpack).

Belief-Emotion

The researcher shows the child a chocolate box and asks, “What do you think is inside the box?” When the child says “chocolate,” the researcher acts out Nao/Tong saying, “Oh great! I love chocolate. It’s my favourite food. Now I’m off to play.” Nao/Tong is then placed out of sight. The researcher opens the box and reveals, “Actually, there’s no chocolate, only stones!” After closing the box again, the researcher asks, “What is Nao’s/Tong’s favourite treat?” Then, the key question is asked: “When Nao/Tong comes back and gets the box, how will he feel, happy or sad?” This is followed by, “And after Nao/Tong opens the box and sees what’s inside, how will he feel, happy or sad?” A correct response requires the child to say “happy” (based on the belief) and then “sad” (based on reality).

Real-Apparent Emotion

The child is told a story: Nao’s/Tong’s uncle has returned and promised to buy him a toy gun. However, instead, the uncle brings back a book. Nao/Tong doesn’t like books and really wanted the toy gun, but Nao/Tong needs to hide its/her true feelings because if uncle knows it/she’s upset, he might not get gifts in the future. The researcher checks understanding: “What did the uncle give to Nao/Tong?” and “What would happen if the uncle knew how

Nao/Tong really felt?" The child is then asked, "When Nao/Tong received the book, how did it/she really feel inside, happy, sad, or just okay?" followed by, "What expression did it/she show on the face, happy, sad, or just okay?" The task is passed if the child identifies a more negative internal emotion than the external facial expression (e.g., sad inside, but shows happy or neutral on the face).