UNIVERSITÀ CATTOLICA DEL SACRO CUORE

Sede di Milano

Dottorato di ricerca in "Persona, Sviluppo, Apprendimento. Prospettive Epistemologiche, Teoriche e Applicative"

Ciclo XXXIV

S.S.D. M/PSI-04

When a child meets a robot: the psychological factors that make interaction possible

Coordinatore:

Ch.mo Prof.ssa Antonella Marchetti

Tesi di Dottorato di:

Giulia Peretti

N. Matricola: 4814646

Anno Accademico 2020/2021

A Davide
Ai miei genitori
Alla mente che scodinzola che mi sta accanto

Table of Contents

Chapter 1: General Introduction 1

Chapter 2: Shall I Trust You? From Child-Robot Interaction to Trusting Relationship 24

Chapter 3: Can a robot lie? The role of false belief and intentionality understanding in children aged 5 and 6 years 53

CHAPTER 1

General Introduction

"The other offers itself to me as a concrete and evident presence
which I can in no way draw from myself and which can in no way be questioned,
nor be made the object of a phenomenological reduction or other epoché."[1]

(Jean Paul Sartre, from *Vivere con i robot. Saggio sull'empatia artificiale*)

"When a child meets a robot"

The Other. Throughout life, human beings may come into contact with a multiplicity of other agents with whom they weave and develop relational and interactional processes, implying a mutual understanding of the intentions behind each other's behaviour. These "others" can have a different agency than the human one; for example, the other can be identified in a divinity, such as God, an animal or, particularly in recent years, an artificial agent such as humanoid social robots (HSRs). The latter can be considered as effective social partners that can be treated as intentional agents, probably due to some of the physical characteristics that make them similar to human beings (Dario et al., 2001; Manzi et al., 2020a; Okumura et al., 2013a). In particular, in recent years, psychological research has been giving great attention to the study of human-robot interaction (Belpaeme et al., 2018; Breazeal et al., 2016; Wellman, 2020; Westlund, 2017). The flourishing interest is motivated by a technologically evolving society, in which high-precision machinery, among which robots in particular, permeate different contexts of everyday living, from work and school to the family context. Let's consider robots in surgery, industry or included in educational and formative settings, offering thus different opportunities of encounter and interaction. The increasing introduction of artificial agents has allowed, and is still allowing, adults and children to interface increasingly with robotic agents, such as HSRs, or with educational robotic tools proposed in educational and school contexts, such as the so-called tangible interfaces like floor robots (e.g. Cubetto, Bee Bot). Bearing in mind that, regarding children's first approaches to the robotic tool, generally, educational robotics activities are structured according to a playful approach and appropriate to the child's developmental phase (Bers & Horn, 2010; Kazakoff & Bers, 2014). Therefore, children may come in contact with

---

[1] Personal translation of the text from Italian into English.

technological tools through play or playful and educational activities, thus developing an interaction with the technological artefact, to which they can attribute meanings deriving basically from the information they develop around it and, also, attribute a specific meaning and function according to the cultural context within which they are known and used. Technological artefacts, seen as a material object, become a vehicle of knowledge and objects activities lead to the creation of shared attention, and interaction with the tool whereby relationships with peers and adults can be woven (Manzi et al., 2020b; Savarese et al., 2017). Hence, the same artefact can take different meanings with regard, for example, to the characteristics of the subject - age as a diversifying factor in the understanding of meaning - and about the narratives that a culture develops around an object to attribute meaning to it (Bruner, 1991). The robot, as an object, is thus identified as a real technological artefact with which the child can develop an interaction, and around which the subject can develop beliefs and attribute mental states depending on the design characteristics it presents and on the associated cultural meaning. Consequently, the knowledge matured on the object, seen as a cultural artefact, be it a work tool, a game or, as previously mentioned, a robot, becomes relevant from the earliest stages of a child's cognitive development (Moro, 2011). How do subjects develop knowledge around the object? According to Piaget's vision (1932), to know objects, the subject must act on them, and therefore transform them: he/she must move them, connect them, combine them, disassemble them, reassemble them. The child becomes the individual builder of his/her cognitive development, managing to formulate knowledge through his/her actions on the material world around him/her (Cole & Wertsch, 1996), due to innate factors interacting with the physical and social world. In this way, knowledge about the object increases from the interactions that the child has with the object with which it is interfacing; directing the understanding that the subject experiences about the object through a universality of developmental stages that leads the child, step by step, to elaborate a specific mental representation of reality (Bruner, 1991). Hence, in the field of human-robot interaction, child-robot interaction (cHRI) holds great interest for psychological scientific research because it permits to observe how children, at different stages of development, get in contact with a technological artefact, the robot, which possesses structural characteristics that identify it as an object, which, however, can sometimes show human-like behaviours. Moreover, the robot, precisely because of these behaviours, may be subject to categorical overlap (e.g., dogs can talk), which causes ontological confusion (Lindeman et al, 2015) consisting of a bias in the knowledge base of psychological, biological and physical phenomena. This can be seen with greater intensity when children and robots interact. Furthermore, the tendency to anthropomorphize both living and non-living entities, especially in the attribution of beliefs to non-human entities, could result in confusion between some psychological characteristics that are properly human, attributing them to non-human entities (Di Dio et al., 2018).

For me, these elements appear of great interest to understand how the robotic agent is perceived and observed by developing minds, such as those of children. Moreover, in-depth investigation of the aforementioned elements permits the narration, with increasing sophistication on the part of science, of the potential and applicative uses of a robotic tool within educational, clinical and care contexts, thus revealing its potential implications to those directly involved in the process of implementing the robotic device, such as teachers, educators and caregivers. As Belpaeme and colleagues (2013) observed, child-robot interaction is shown to be different from human-robot interaction (HRI) since children have a different and immature cognitive maturation than adults. For example, it has been observed that children ascribe characteristics to the robot that are usually attributed to living systems, not viewing the robotic device, compared to adult subjects, as a mechanical device programmed to act in a certain previously determined sequence (Belpaeme et al., 2013). Furthermore, the anthropomorphic features of the HSRs can increase humans' perception in terms of humanness, such as mind attribution and personality, and affect other psychological mechanisms and processes (Bartneck et al., 2008; Broadbent et al., 2013; Kiesler & Goetz, 2002; MacDorman et al, 2005; Marchetti et al., 2018, 2020; Manzi et al., 2020c; Powers & Kiesler, 2006; Złotowski et al., 2015) (more information on anthropomorphism is included in Chapter 4). Through my research study, I had the opportunity to note that the interest in observing the effect of different physical characteristics of robots in terms of attribution of intentions, understanding, and emotions has also been investigated in children (Bumby & Dautenhahn, 1999; Woods et al., 2004; Woods, 2006). From the age of 3 (Belpaeme et al., 2013; Berry & Springer, 1993), children are inclined to anthropomorphize and are more tending to do so than adults, or as indicated by Turkle and colleagues (2004) have a desire to believe for more time that a robot has human-like characteristics. The dimension of make-believe play and anthropomorphism seem to be important elements for children to engage in interaction with robots and treat them as life-like agents (Belpaeme et al., 2013). For this reason, I believe it is important to observe how different physical characteristics of HSRs can significantly influence the quality of interaction, throughout lifetimes, between humans and robots (Marchetti et al., 2018). Bearing in mind these elements as important for building robots and for understanding which types of robotic agents turn out to be more functional in interactional dynamics according to the age group to which they are proposed (e.g. kindergarten, primary school or secondary school). In particular, two different directions in robotic development have emerged, which are based on different, albeit related, theoretical perspectives, namely: Developmental Cybernetics (DC; Di Dio et al, 2019; Manzi et al., 2020a; Itakura, 2008; Itakura et al., 2008; Moriguchi et al., 2011; Kannegiesser et al., 2015; Okanda et al., 2018;Wang et al., 2020) and Developmental Robotics (DR; Cangelosi & Schlesinger, 2015; De La Cruz et al., 2014; 2018 Di Dio et al., 2020a,b; Lyon et al., 2016; Morse & Cangelosi, 2017;

Vinanzi et al., 2019; Zhong et al., 2019) (for a detailed explanation see Manzi et al., 2020c). These theoretical perspectives that aim to design robots that behave and develop like human beings must necessarily pay attention to constructs related to the Theory of Mind and psychological development (Marchetti et al., 2018), particularly when the actors in interaction turn out to be children and robots. Through the above considerations, I was able to understand how child-robot interaction can be significantly different from human-robot interaction based on the characteristics of the interacting subjects, i.e. children, whose neurophysiological, physical and mental development is still in progress (Belpaeme et al., 2013), can produce different results compared to the interaction between robots and adult subjects. When a child meets a robot, there are multiple psychological elements that make interaction possible. Therefore, I believe it is crucial to explore some of the cognitive and affective elements that characterize the development of children and that can influence the interaction with a different agents. In this psychological framework, two constructs, surely crucial to understanding the relational dynamics, are Theory of Mind and trust. Attachment is another important affective correlate of the relationship to be considered. In the following paragraphs, these elements will be treated more deeply in the following sections.

Theory of Mind

The attribution of a mental state to robotic agents is an important element for children to initiate a relational process with them (Di Dio et al., 2020a). What is being referred to when talking about the attribution of a mental state? I think it is important to start with a definition of Theory of Mind (ToM) to understand what it means to attribute a mental state. Theory of Mind (ToM) refers to one of the components that characterizes the development of the mind. It is understood as the capacity to attribute mental states to oneself and others (beliefs, emotions, desires, intentions, thoughts) and to explain and predict, based on such inferences, one's own and others' behavior (Premack & Woodruff, 1978). Theory as "*the child acquires a theory, i.e. a coherent conceptual system capable of explaining and predicting human actions in terms of constructs such as desires, intentions, thoughts and beliefs*" (Camaioni, 1995/2003, p.13). Premack & Woodruff (1978) developed this definition from their pioneering study of the observation of intentional understanding of chimpanzee behaviour. ToM is an essential prerequisite for social competence (Perner & Wimmer, 1985; Premack & Woodruff, 1978; Wellman et al., 2001). This social competence appears as a commonly used ability, as it refers to the reasoning processes that each individual uses daily in the encounter with the other, thus being defined as folk psychology (Lecciso, 2005). Over the years, as studies and observations on the subject have progressed, various terms have been used to define or encompass the same ability, for example, "mindreading" (Baron-Cohen et al., 1997); "reflective function" (Fonagy & Target, 1997);

"perspective-taking" (Carpendale & Lewis, 2006); "mentalization" (Fonagy & Allison, 2012). Since the 1980s, a great deal of research in the field of developmental psychology has been interested in the study of ToM. In particular regarding developmental studies, the works of Wimmer and Perner (1983), Perner and Wimmer (1985), Baron-Cohen and colleagues (1985) and Perner and colleagues (1987) began to study the ToM with deep interest. These works, starting from the experimental paradigm proposed by Wimmer & Perner (1983), have developed different versions of the so-called false belief task, aimed at measuring the presence of a first-order recursive thinking, i.e. a thought that implies a meta-representation; in other words, a mental representation that is included in another mental representation, i.e.: *"I think you think X*" (Battistelli, 1995; Valle et al., 2015). The experimental subject, for the false belief task, is asked to predict how a protagonist of a story will act, bearing in mind the protagonist's false belief and not the fact of reality (Castelli & Marchetti, 2018). In this regard, two specific tasks have been developed, namely: the Unexpected Transfer task (Wimmer & Perner, 1983), in which the subject must predict where the protagonist of the story will go to look for an object that he/she had previously placed in a container, then moved by another protagonist, without his/her knowledge, in another container; and the Unexpected Content task (Perner & Wimmer, 1985) in which the experimental subject is shown a closed box on which are represented some sweets (smarties), but in which are some pencils, unknowingly placed by the experimental subject, so the content does not correspond to what is represented on the box. Initially, the experimenter will ask the experimental subject what is in the box, before showing the real content, then, after the subject has seen the real content of the box, the experimenter will ask the child to predict what another person will think is in the box, before seeing the content. To be able to answer the questions of the two tasks described, the child must understand that the protagonist of the story possesses *"a representation of reality that is different from the actual state of reality (which in this case corresponds to the child's representation)"* (Camaioni, 1995/2003, p.5). The child must therefore take the perspective of the other and manage to represent the content of his/her mind, that is a false belief about reality. To do so, he/she must momentarily suspend his/her knowledge of reality, thus managing to predict how the other will act based on the false belief (Castelli & Marchetti, 2018).

Based on the tasks developed, it was found that 4 years-old children can solve the first-order false belief task, while 2-3 years-old children are not yet able to correctly solve the task, showing more of a difficulty in the linguistic comprehension of the test rather than mentalization that is still immature (Onishi & Baillargeon, 2005). Subsequently, approximately at the age of 7/8 years, children acquire second-order recursive thinking, i.e.: *"I think that you think that he/she thinks..."* (Perner & Wimmer, 1985), showing a more sophisticated recursive thinking (a depiction of a classical task of second-order false belief is shown in Figure 1). The perspective, presented up to this point, is based on the cognitive-meta-representational approach used in the Theory-Theory approach (Wellman, 1991), in which the 4 years-old is seen as a real watershed in the acquisition of the meta-representation capacity, risking to convey the message of "*all-nothing*"; instead it is considered to be an evolutionary ability, which presupposes a *continuum* (Lecciso, 2005), and not an "*all-nothing*". Therefore, several theoretical perspectives are in agreement in arguing that ToM ability is not possessed immediately from birth, but develops in an evolutionary perspective. Therefore, some theories support the link between some childhood cognitive skills and the subsequent development of ToM (Lecce et al., 2010). These include: 1) constructivist theories (Bosacki & Astington, 1999), which highlight the role of growth contexts to develop mental abilities, and a vision of the social construction of the mind; 2) modular theories (Baron-Cohen, 1995; Leslie, 1994), according to which mental abilities can be found in the action of specific modules, genetically determined and not modifiable by experience (Lecce et al., 2010). The modules are already present from birth, and during life, they are activated in a stereotyped and rigid manner, independently of experience. Environment, in this instance, plays only a marginal role. 3) the imitation theories (Meltzoff, 2002; Meltzoff & Moore, 1983), in which the role of imitation in being able to represent the mental states of others is highlighted for the development of intersubjectivity. Finally, the simulation approach (Harris, 1991), in which attention is focused on the role of first-person knowledge in the attribution of mental states, in which the child comes to understand the mental states of others through a mechanism of mental simulation.

Although these theories deal with the understanding of mental states from different perspectives, they consider it very interesting to observe what happens in the course of cognitive development before the age of 2 (Lecce et al., 2010) paying particular attention to ToM precursors, i.e. cognitive structures and schemes that prepare for the acquisition of this skill, such as: social referencing, joint-attention, declarative pointing, agency comprehension, visual perception comprehension, make-believe play. For my research on the child-robot relationship, it seems very interesting to dwell on agency comprehension. The term agency refers to *"the understanding that animate beings act autonomously, in turn causing effects on other objects/agents"* (Castelli & Marchetti, 2018). The comprehension of agency (Gergely et al., 1995) appears to be important to understand the distinction between animate and inanimate beings, will and intentions to see the other, different from oneself, endowed with mental states that guide their actions.

It is useful to underline that the comprehension of self and others does not finish at the age of 7-8, but evolves throughout life (Apperly, 2013). For this reason, several researchers (Happé, 1994; Stone et al., 1998); have begun to study ToM from a life-span perspective, requiring more complex assessment tools to be able to evaluate third-order false beliefs (Baglio et al., 2012; Baron-Cohen et al, 2001; Castelli et al., 2010; Kinderman et al., 1998; Valle et al., 2015), apt to assess more ecological than recursive elements (Castelli & Marchetti, 2018). The ToM, therefore, proves to be a construct that embraces the entire life of the individual, evolving according to the specific stage of development and, especially thanks to the increasing technological development that allows the interaction with numerous artificial agents, with new and innovative future perspectives of study. Studying human-robot interaction within the theoretical frame of ToM enables one to understand, in a more refined way, how the attribution of mental states to a robotic agent can make the intersubjective nature of this type of interaction more comprehensible (Marchetti et al., 2018).

Figure 1. Say Prediction task: four colour vignettes to illustrate the story about the second-order false belief task (drawing by C. Bignotti). The characters in the story are Gianna and her mother and, as an object, Gianna's birthday present: a dog. The story is told to the child, which finds its graphic correspondence in the vignettes, in which the mother wants to surprise Gianna by giving her a small dog, which she hides in the cantina, and which she knows the child wants very much. When the girl asks for a puppy, Mum replies that she has given her another present. While Mum is on the phone, Gianna says she is going out to play, but first, she goes down to the cantina and sees the little dog. The whole story is told to the child and then a series of questions are asked about the belief held by the mother regarding Gianna's discovery of the little dog, to observe whether or not the child has acquired the second-order false belief.

Trust

In the area of developmental psychology, I have had the opportunity to observe how the construct of trust proves to be a remarkable subject of study as it involves both the dimension of the self and the dimension of social relations. Trust, as shown by Earl (1987), can be differentiated between self-confidence and social and interpersonal trust. Self-confidence refers to the belief that one can cope with a task thanks to one's abilities, while interpersonal confidence refers to the attribution of trust to subjects considered reliable. Children, from the earliest stages of life, are inclined to attribute trust to information conveyed to them by adults through actions such as showing or telling (Tomasello, 2019). Therefore, in this perspective, the trust acquisition process during the child's developmental stages, especially during the early stages of attachment with primary caregivers, proves to be significantly important. Trust relationships, shaped by past relational experiences with caregivers, also extend to subsequent significant affective relationships with other individuals (Ainsworth, 1978; Bowlby, 1969, 1973, 1980). The generalization of a child's personal attachment story, e.g. secure attachment pattern vs. insecure attachment pattern, might influence how he or she is led to confer trust on an unknown informant (Bo et al., 2017; Fonagy, 1998; Fonagy & Allison, 2014; Fonagy et al., 2015). Harris and colleagues have focused particular attention, in their research, on how young children are led to trust an adult who conveys information to them (Harris, 2012), and the role of trust during the developmental period, particularly when the child is dealing with an informant who can be considered more or less trustworthy. They note that 3-year-olds place more trust in an informant who is consistently accurate (Clément et al, 2004; Koenig et al., 2004; Pasquini et al., 2007), whereas 5-year-olds consider what is told to them to be trustworthy, as they consider the information being conveyed as if they had seen it with their own eyes (Haux et al., 2017). With development, children gain more flexibility in attributing trust to an informant, bringing together multiple assessment elements (e.g. past accuracy and information of the moment) before conferring trust (Di Dio et al., 2020b). The attribution of trust to another subject, and the receptivity of the information provided, has played a key role throughout human evolution (Breazel et al., 2016; Nielsen, 2012; Richerson & Boyd, 2005), with impact on relational processes, especially human-human relationships. Currently, thanks to the constant and progressive inclusion of technological artefacts within multiple cultural contexts, children can receive information not only from people but also from different technological devices such as, for example, social robots (Brink & Wellman, 2020). It has been observed that children can interact with, and treat as social companions, robots that exhibit anthropomorphic features (Breazel et al., 2016), finding them interesting as informants and interacting in a friendly manner, communicating with robots more deeply than with a traditional device, and conferring mentalistic skills on them (Kahn et al., 2012; Kanda et al., 2004; Marchetti et al., 2018; Okumura et al., 2013b; Shiomi et al., 2006). However, although anthropomorphic robots are seen as interesting informants, it has been observed that children aged 3-5 years are more likely to receive information, and even learn, from subjects that are familiar to them or that possess physical or language characteristics, such as accent, similar to their own (Breazel et al., 2016; Corriveau & Harris, 2009; Kinzler et al., 2010). Hence, as trust turns out to be a psychological component of considerable importance for the establishment of long-term relationships with robots (Di Dio et al., 2020b), it seems important to understand what elements make social informant robots trustworthy,

worthy of receiving the trust of children. Developmental Robotics (Cangelosi & Schlesinger, 2015) is moving its research interests precisely in this direction, raising the challenge of creating social robots that can engage in a long-term relationship, within which there are components such as trust, with human subjects, including children.

Attachment relation

The relational dimension appears to be a fundamental element for the human being, relationships that he/she will weave throughout his life. In particular, parental relationships, which develop between caregiver and child, are shaped through attachment bonds. In the 1970s and 1980s, an initial formulation of attachment theory was by John Bowlby's Attachment and loss trilogy (1969; 1973; 1980; Cassidy & Shaver, 2002; Caravita et al., 2018), and the work of Mary Ainsworth about the mother-child relationship (Ainsworth 1967, 1982; Ainsworth et al, 1971). These works have created the conditions for the growth of numerous observations and theoretical elaborations concerning the attachment relationship and the consequences that it has on the development of the individuals personality throughout their lives. As noted by Riva Crugnola and Ierardi (2018), attachment theory (AT, Bowlby, 1978) can be considered a theory of personality development based on the analysis of developmental paths and individual differences (Riva Crugnola & Ierardi, 2018). Attachment (Bowlby, 1969, 1973,1980; Ainsworth et al., 1978; Cassidy & Shaver, 2002), thus describes the quality of the relationship between infants and young children with their primary caregiver. The work of Bowlby and Ainsworth consists, therefore, is a real change of perspective concerning previous psychoanalytic and behaviorist studies, drawing instead important reflections from ethological studies, such as those of Harlow (1958), with whom Bowlby was in contact from 1957 until the mid-1970s (Van der Horst et al., 2008), and the studies of Lorenz (1971; Lorenz & Martin, 1971). Bowlby and Ainsworth’s hypothesis concerned an innate biological predisposition of children to build attachment bonds through a search for proximity and physical contact with their reference figures, the caregivers, to have comfort and protection on two main levels: physical and emotional (Riva Crugnola & Ierardi, 2018). According to Bowlby, therefore, the child is driven to seek the closeness of caregivers, activating their attention through communicative modalities such as crying, smiling and vocalizations, and form attachment bonds, initially to search for protection and, gradually with motor development, to seek proximity with their caregiver (Riva Crugnola & Ierardi, 2018). Therefore, the relationship with caregivers plays an important role in the development of attachment and is also important for establishing relationships with other reference figures (e.g. teachers, friends, partners) and for the development of emotional and relational skills, as it takes on a role of significant importance in the dynamics of regulation of the child’s emotions, especially those related to fear and stress (Riva Crugnola & Ierardi, 2018). Bowlby and Ainsworth focused a lot on the study of attachment bonds, coming to distinguish different stages that characterize the development of attachment relationships from the first moments of a child's life up to about two years of age. In particular, they identify four phases (Riva Crugnola & Ierardi, 2018), which are distinguished by the attachment needs expressed by the child, which vary according to the child's stage of development. Continuing their studies on attachment, Bowlby and Ainsworth realized that attachment patterns varied depending on the children observed. Therefore, Ainsworth developed the experimental Strange Situation Procedure (SSP, 1969) paradigm to measure mother-child attachment, through the exposure of the child to mildly stressful situations, and identified three attachment patterns (secure, insecure-avoidant, insecure-ambivalent). Later, a fourth model, called disorganized/disoriented, was added by Main and Solomon (1986). Depending on the quality of the relationship with the caregiver, the child will consider him/her, on an emotional level, as reliable, unpredictable or unavailable (Di Dio et al., 2020a, 2020b). These relational patterns are mentally represented by internal operating models (Main et al., 1985), cognitive-affective structures based on the generalization of attachment experiences, which have a prototypical function for future relationships. Future relationships could, therefore, be based on the generalization of one's attachment bonds (Fonagy, 1998; Fonagy & Allison, 2014; Fonagy et al., 2015). For this reason, I think it is important for my research to consider whether attachment dynamics might have an influence on the mental representations of non-human agents, such as robots, and how this might affect the development of interaction with non-human agents (Di Dio et al., 2020a).

This thesis

As may be evident from the theoretical overview in the previous paragraphs, there is a multiplicity of elements that come into play when referring to the interaction between children and robots, especially for the particular phase of development in which children interact with another completely different from themselves. Thus, a multiplicity of constructs and affective correlates of the relationship come into play, which makes studying the interaction involving children and robots incredibly fascinating. Through this thesis I wished to explore all the elements that define and characterize this interaction, particularly from age 3 to 9 years, rich in cognitive and affective developmental changes. In other words, the child-robot interaction represents the common matrix around which multiple considerations related to the wide range of constructs and affective correlates that characterize this interaction can mature. The following chapters will present the research I carried out during my PhD and that focus on the constructs introduced here, specifically within the child-robot relationship, for example trust (Chapter 2), ToM (Chapter 3 and 4), and educational robotic related to cognitive skills (Chapter 5). Through the following chapters, I wished to present the many facets that can characterize child-robot interaction. Specifically, showing how trust and Theory of Mind play a key role in the interactional processes that develop with a robotic agent. Moreover, since nowadays children, since early childhood, can interact with a robotic tool, it seems interesting to observe what happens, what cognitive skills can mature through the first exposure to an educational robot. Through mainly playful modalities, children can begin to interact with a robotic tool, which will allow them to start the path of knowledge around the robotic agent. Below, a brief presentation of each chapter.

Chapter 2 deals with the study of trust acquisition, loss and restoration process when pre-school and school children play with a human and with the robot Nao in vivo. The study of trust, within human-robot interaction, proves to be very important, especially for the increasing presence, as mentioned in the previous paragraphs, of robotic agents within multiple areas of everyday life. In this research, the relationship between trust and the representation of the quality of attachment relationships was investigated, also exploring ToM and executive functions (EF). The results showed the impact of the cognitive development and attachment history of children involved in the study. Thus, I was able to observe that 3-year-olds children are more sensitive to the affective component of trust to attribute trust, especially in interactions with humans. In addition, I noted the role of understanding false belief in order to establish trusting relationships. I observed that, for children, the robot proved to be a more trusting partner than the human, as it was less susceptible to the dynamics associated with the quality of the attachment relationship. The results of this study are important since they may provide details to different disciplines, such as Developmental Robotics, on how to make a robot more trusting for a human, by understanding how to model its cognitive architectures to make it more trusting, leading the human-robot interaction to be based more on trust.

Chapter 3 reports an investigation about the ability of 5- and 6-year-olds children to differentiate between an intentionally false and an unintentionally false statement. The relationship with another agent, such as a robot, implies a mutual understanding of the intentions that guide the other's behaviour. Therefore, this also requires an understanding of whether another person is intentionally or unintentionally producing a false statement. This kind of understanding, which is fundamental in human-human relations, appears more significant when children and robots interact. Therefore, understanding whether or not children believe that a robot can deceive intentionally or unintentionally, knowing if this understanding is there or not, seems important to define the quality of the child-robot relationship. The ability to lie and also to recognize intentionality the other's to lie represents an important developmental step in children, which is also characterized by the acquisition of the ToM. Indeed, from the age of 4, children begin to distinguish between an intentionally false statement (a lie) and an unintentionally false one (a mistake). Furthermore, 4- and 5-year-olds, but not 3-year-olds (Gilli et al., 2001), evaluate intentionally false statements negatively, in contrast to unintentionally wrong statements. Through the present study, I was enabled to investigate how children were able to differentiate between intentionally or unintentionally false statements made by a human or the humanoid NAO robot after watching videos of both agents making a false statement that was either false or simply wrong. Children's first-order ToM skills, verbal skills, understanding of intentionality, and executive function skills were also assessed. I observed that the findings support previous data regarding ToM skills (Di Dio et al., 2020a, 2020b), showing that understanding intentionality and verbal skills are important for differentiating lies from mistakes. Furthermore, the results further showed that children who acquired a first-order false belief are shown to be less judgmental towards intentionally false statements produced by the robot than the same ones provided by the human. Implicitly, younger children tend to view a robot's behaviour as fundamentally unintentional.

Chapter 4 deals with the influence of a robot's design characteristics, e.g. human-like or less, that lead to anthropomorphizing a robotic agent more. Anthropomorphism, in the psychological language, refers to the tendency to attribute human characteristics, physical and/or psychological, to non-human agents (Duffy, 2003; Epley et al., 2007). To explore this, in the study presented in this chapter, I compared the attribution of mental states, using the Attribution of Mental Scale questionnaire (AMS, Di Dio et al, 2020a), it is a measure of mental states that participants attribute to when they look at images depicting specific characters, for this study at to two different humanoid robots, NAO robot and Robovie, which differ in their design characteristics and degree of anthropomorphism. NAO robot has more human-like structural features, while Robovie has features that make it look more mechanical. Children aged 5, 7 and 9 were then asked to attribute mental states to the robots. Through the study, I was able to highlight how different types of robots can elicit different attributions of mental states in children. In particular, the age of the children appears to be an important element to bear in mind while designing robots for children. I observed that, in 5-year-old children, anthropomorphism appears to be widespread independently of the type of robot, while the design characteristics of the robot influence older children aged 7 and 9 more, leading them to prefer the NAO robot, attributing more mental states to it compared to Robovie, as it is seen as more human-like. The results, therefore, showed that the evaluation of HSRs, in terms of attribution of mental states, may be a useful measure to study the effect of different robot designs for children.

Chapter 5 presents the study I carried out on a coding intervention, based on Computational Thinking (CT), using the Cubetto robot, to evaluate its effects on some cognitive skills, such as the sequential ability to reconstruct stories and visual-spatial skills, in 4-year-olds children. Computational Thinking refers to a thought process that uses algorithmic and analytical approaches to formulate, analyze and solve problems (Wing, 2006; 2008). Many educational robotics and coding interventions pursue this logic. Therefore, it seems crucial to understand the potential effects of activities related to Computational Thinking on cognitive abilities to understand, which developmental stage, coding activities and educational robotics could be more effective. In this research, the effects of the proposed coding intervention - in which the children were asked to programme Cubetto's pathways, on his specially designed map, this activity required children to manipulate and reason about the physical object presented to them - on some cognitive abilities (sequential, visual-spatial) were investigated. The sample was divided into two groups: the Experimental Group took part in the coding training, in which the coding rules were taught, and was tested both before and after the specific coding training, while the Control Group, waiting list, was tested at the same time as the Experimental sample but without the training, receiving the training only after the measurements had been completed. The results of the Cubetto task show that the Experimental Group was more able to program the Cubetto at the end of the task than the Control Group. An explanation for this result could be related to the effect of having received specific training on the task. With this last work, that can appear different in the modality of children's interaction with the robotic tool compared to the previous chapters, I wished to show also another modality of interaction with a robotic tool, in this case with specific educational characteristics and, moreover, to highlight which cognitive abilities can improve through the present use. In other words, a robotic tool that children can actually act on, manipulate, and program to make it perform programs. The first form of interaction that can make children more confident with robotic tools that they will encounter with increasing frequency during their life.

References

Ainsworth, M. D. S. (1967). *Infancy in Uganda: Infant care and the growth of love*. Johns Hopkins Press.

Ainsworth, M. D. S. (1969). Object relations, dependency and attachment: A theoretical review of the infant-mother relation- ship. *Child Development*, 40, 969-1025. https://doi.org/10.2307/1127008

Ainsworth, M. D. S. (1971). *Individual Differences in Strange-Situational Behaviour of One-Year-Olds*, in H.R. Schaffer (by), *The origin of human social relations*, London, The Academic Press, pp.17-57.

Ainsworth, M. D. S. (1978). The Bowlby-Ainsworth attachment theory. *Behavioral and Brain Sciences*, 1(3), 436-438. doi:10.1017/S0140525X00075828

Ainsworth, M. D. S. (1982). *The Development of Infant-Mother Attachment*. In *The beginning* (pp. 133-143). Columbia University Press. https://doi.org/10.7312/bels91552-016

Apperly, I. (2013). *Can theory of mind grow up? Mindreading in adults, and its implications for the development and neuroscience of mindreading*. In *Understanding other minds: Perspectives from developmental social neuroscience*, 3rd ed (pp. 72– 92). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199692972.003.0005

Baglio, F., Castelli, I., Alberoni, M., Blasi, V., Griffanti, L., Falini, A., Nemni, R., & Marchetti, A. (2012). Theory of mind in amnestic mild cognitive impairment: an FMRI study. *Journal of Alzheimer's Disease*, 29(1), 25-37.

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. MIT press.

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

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry*, 38(7), 813-822. https://doi.org/10.1111/j.1469-7610.1997.tb01599.x

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the Mind in the Eyes” Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines,* 42(2), 241-251. doi:10.1017/S0021963001006643

Bartneck, C., Croft, E., & Kulic, D. (2008). “Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots,” in Proceedings of the Metrics for Human-Robot Interaction Workshop in affiliation with the 3rd ACM/IEEE International

Conference on Human- Robot Interaction (HRI 2008), Technical Report 471 (Piscataway, NJ: IEEE), 37–44. https://doi.org/10.6084/m9.figshare.5154805

Battistelli, P. (1995). *Io penso che tu pensi... Le origini della comprensione della mente.* Milano: Franco Angeli.

Belpaeme, T., Baxter, P., De Greeff, J., Kennedy, J., Read, R., Looije, R., ... & Zelati, M. C. (2013, October). Child-robot interaction: Perspectives and challenges. *In International conference on social robotics* (pp. 452-459). Springer, Cham. https://doi.org/10.1007/978-3-319-02675-6_45

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social robots for education: a review. *Sci. Robot*. 3:eaat5954. doi: 10.1126/ scirobotics.aat5954 doi: 10.1016/j.ijhcs.2011.04.003.

Berry, D.S., Springer, K.: Structure, motion, and preschoolers' perceptions of social causality. *Ecological Psychology* 5(4) (1993) 273–283. https://doi.org/10.1207/s15326969eco0504_1

Bers M., & Horn M.(2010) *Tangible programming in early childhood: Revisiting developmental assumptions through new technologies*. In. Berson IR, Berson MJ (Eds), *High-tech tots: Childhood in a digital world,* Greenwich, CT: Information Age Publishing; 49-70.

Bosacki, S., & Wilde Astington, J. (1999). Theory of mind in preadolescence: Relations between social understanding and social competence. *Social Development*, 8(2), 237-255. https://doi.org/10.1111/1467-9507.00093

Bo, S., Sharp, C., Fonagy, P., & Kongerslev, M. (2017). Hypermentalizing, attachment, and epistemic trust in adolescent BPD: Clinical illustrations. *Personality Disorders: Theory, Research, and Treatment*, 8(2), 172-182. doi:10.1037/per0000161

Bowlby, J. (1969). *Attachment. Attachment and Loss* (1). New York, NY: Basic Books.

Bowlby, J. (1973). *Separation: Anxiety & Anger. Attachment and Loss* (2). London: Hogarth Press.

Bowlby, J. (1978). Attachment theory and its therapeutic implications. *Adolescent Psychiatry*, 6, 5–33.

Bowlby, J. (1980). *Loss: Sadness & Depression. Attachment and Loss* (3). London: Hogarth Press.

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 Scie*nce, 8(2), 481-491. doi: 10.1111/tops.12192

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

Broadbent, E., Kumar, V., Li, X., Sollers, J., Stafford, R. Q., MacDonald, B. A., et al. (2013). Robots with display screens: A robot with a more humanlike face display is perceived to have more mind and a better personality. *PLoS One* 8:e72589. doi: 10.1371/journal.pone.0072589

Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry*, 18:1, 1–21. https://doi.org/doi.org/10.1086/448619

Camaioni, L. (2003). *La teoria della mente: origini, sviluppo e patologia*. Laterza. (Reprint of Camaioni, L. (1998). *La teoria della mente: origini, sviluppo e patologia*.

Cangelosi, A., & Schlesinger, M. (2015). *Developmental robotics: From babies to robots*. MIT Press.

Caravita, S. C. S., Milani, L., & Traficante, D. (2018). *Psicologia dello Sviluppo e dell'Educazione. Traiettorie evolutive e contesti educativi: una prospettiva integrata*. Il Mulino, Bologna

Carpendale, J., & Lewis, C. (2006). *How children develop social understanding*. Blackwell Publishing.

Cassidy, J., Shaver, P.R., a cura di, *Manuale dell'attaccamento. Teoria, ricerca e applicazioni cliniche*, Giovanni Fioriti Editore, Roma, 2002.

Castelli, I., Baglio, F., Blasi, V., Alberoni, M., Falini, A., Liverta-Sempio, O., Nemni, R., & Marchetti, A. (2010). Effects of aging on mindreading ability through the eyes: an fMRI study. *Neuropsychologia,*48(9),2586-2594. https://doi.org/10.1016/j.neuropsychologia.2010.05.005

Castelli, I., & Marchetti, A. (2018). *Capirsi: La Teoria della Mente*, (pp.127-144). In Caravita, S. C. S., Milani, L., & Traficante, D. (A c. Di), *Psicologia dello Sviluppo e dell'Educazione. Traiettorie evolutive e contesti educativi: una prospettiva integrata*. Il Mulino, Bologna, pp. 127-144.

Clément, F., Koenig, M., & Harris, P. (2004). *The ontogenesis of trust. Mind & Language*, 19(4), 360-379. Retrieved from https://www.fabriceclement.net/doc/13.pdf

Cole, M., & Wertsch, J. V. (1996). Beyond the individual-social antinomy in discussions of Piaget and Vygotsky. *Human development*, 39(5), 250-256. https://doi.org/10.1159/000278475

Corriveau, K. H., & Harris, P. L. (2009). Choosing your informant: Weighing familiarity and recent accuracy. *Developmental Science*, 12, 426–437. doi: 10.1111/j.1467-7687.2008.00792.x

Dario, P., Guglielmelli, E., and Laschi, C. (2001). Humanoids and personal robots: design and experiments. *J. Robot. Syst.* 18, 673–690. doi: 10.1002/rob. 8106

De La Cruz, V. M., Di Nuovo, A., Di Nuovo, S., & Cangelosi, A. (2014). Making fingers and words count in a cognitive robot. *Front. Behav. Neurosci.* 8:13. doi: 10.3389/fnbeh.2014.00013

Di Dio, C., Manzi, F., Itakura, S., Kanda, T., Hishiguro, H., Massaro, D., et al. (2019). It does not matter who you are: fairness in pre-schoolers interacting with human and robotic partners. Int. *J. Soc. Robot*. 1–15. doi: 10.1007/s12369- 019- 00528- 9

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020a). Come i bambini pensano alla mente del robot: il ruolo dell'attaccamento e della Teoria della Mente nell'attribuzione di stati mentali ad un agente robotico. *Sistemi Intelligenti*, 32(1), 41-56. doi:10.1422/96279

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020b). Shall I trust you? From child human-robot interaction to trusting relationships. *Frontiers in Psychology*, 11, 469. https://doi.org/10.3389/fpsyg.2020.00469

Duffy, B. R. (2003). Anthropomorphism and the social robot. Robot. *Auton. Syst.*42, 177–190. doi: 10.1016/s0921-8890(02)00374-3

Earl, W. L. (1987). Creativity and self-trust: A field study. *Adolescence*, 22(86), 419.

Epley, N., Waytz, A., and Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychol. Rev*. 114, 864–886. doi: 10.1037/0033-295x.114.4.864

Fonagy, P. (1998). Attachment theory approach to treatment of the difficult patient. *Bulletin of the Menninger Clinic*, 62(2), 147-169.

Fonagy, P., & Target, M. (1997). Attachment and reflective function: Their role in self-organization. *Development and Psychopathology*, 9(4), 679-700. doi:10.1017/S0954579497001399

Fonagy, P., & Allison, E. (2012). *What is mentalization? The concept and its foundations in developmental research.* In N. Midgley & I. Vrouva (A c. Di), *Minding the child: Mentalization-based interventions with children, young people and their families.* Routledge.

Fonagy, P., & Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy*, 372-380. doi:10.1037/a0036505

Fonagy, P., Luyten, P., & Allison, E. (2015). Epistemic petrification and the restoration of epistemic trust: A new conceptualization of borderline personality disorder and its psychosocial treatment. *Journal of Personality Disorders*, 29(5), 575-609. doi:10.1521/pedi.2015.29.5.575

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, 56(2), 165-193. https://doi.org/10.1016/0010-0277(95)00661-H

Gilli, G., Siegal, M., Marchetti, A., & Peterson, C. C. (2001). Children's incipient ability to distinguish mistakes from lies: An Italian investigation. *International Journal of Behavioral Development*, 25(1), 88-92. https://doi.org/10.1080%2F01650250042000195

Happé, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal* *of Autism and Developmental Disorders*, 24(2), 129-154. https://doi.org/10.1007/BF02172093

Harlow, H. F. (1958). *The nature of love*. American psychologist, 13(12), 673.

Harris, P.L. (1991). *The work of the imagination,* in A. Whiten (by), *Natural theories of mind: Evolution, development and simulation of everyday mindreading*, Oxford, Blackwell, pp. 283-304.

Harris, P. L. (2012). *Trusting what you're told*. Harvard University Press.

Haux, L., Engelmann, J. M., Herrmann, E., & Tomasello, M. (2017). Do young children preferentially trust gossip or firsthand observation in choosing a collaborative partner? *Social Development,* 26(3), 466-474. doi: 10.1111/sode.12225

Itakura, S. (2008). Development of mentalizing and communication: from viewpoint of developmental cybernetics and developmental cognitive neuroscience. *IEICE Trans. Commun.* 2109–2117. doi: 10.1093/ietcom/ e91- b.7.2109

Itakura, S., Ishida, H., Kanda, T., Shimada, Y., Ishiguro, H., & Lee, K. (2008). How to build an intentional android: infants' imitation of a robot's goal-directed actions. *Infancy* 13, 519–532. doi: 10.1080/15250000802329503

Kahn Jr., P. H., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., Ruckert, J. H., & Shen, S. (2012). "Robovie, you'll have to go inside the closet now": Children's social and moral relationships with a anthropomorphic robot. *Developmental Psychology*, 48, 303–314.

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human–Computer Interaction*, 19(1-2), 61-84. doi:10.1207/s15327051hci1901&2_4

Kannegiesser, P., Itakura, S., Zhou, Y., Kanda, T., Ishiguro, H., and Hood, B. (2015). The role of social eye-gaze in children's and adult's ownership attributions to robotic agents in three cultures. *Interact. Stud.* 16, 1–28. doi: 10.1075/is.16.1. 01kan

Kazakoff E.R., & Bers M.U. (2014). Put your robot in, put your robot out: Sequencing through programming robots in early childhood. *Journal of Educational Computing Research*, *50*(4), 553-573. https://doi.org/10.2190%2FEC.50.4.f

Kiesler, S., and Goetz, J. (2002). "Mental models of robotic assistants," *in Proceedings of the CHI'02 Extended Abstracts on Human Factors in Computing Systems* (New York, NY: ACM), 576–577.

Kinderman, P., Dunbar, R., & Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, 89(2), 191-204. https://doi.org/10.1111/j.2044-8295.1998.tb02680.x

Kinzler, K. D., Corriveau, K. H., & Harris, P. L. (2010). Children's selective trust in native-accented speakers. *Developmental Science*, 14, 106–111. doi: 10.1111/j.1467-7687.2010.00965.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. Retrieved from https://journals.sagepub.com/doi/pdf/10.1111/j.0956-7976.2004.00742.x

Lecce, S., Cavallini, E., & Pagnin, A. (2010). *La teoria della mente nell'arco di vita*. Il Mulino.

Lecciso, F. (2005). *Prospettive teoriche. Terre conosciute e nuovi orizzonti.* In Sempio Liverta, O., Marchetti, A., & Lecciso, F. (A c. Di), *Teoria della mente tra normalità e patolog*ia. Cortina, pp.1-33.

Leslie, A. M. (1994). *ToMM, ToBy, and Agency: Core architecture and domain specificity*. In Hirschfeld, L.A, & Gelman, S.A., (by). *Mapping the mind: Domain specificity in cognition and culture*, New York, Cambridge University Press 29, 119-48.

Lindeman, M., Svedholm-Häkkinen, A. M., & Lipsanen, J. (2015). Ontological confusions but not mentalizing abili- ties predict religious belief, paranormal belief, and belief in supernatural purpose. *Cognition*, 134, 63-76. doi:10.1016/j. cognition.2014.09.008

Lorenz, K., & Martin, R. (1971). Studies in animal and human behaviour. *British Journal for the Philosophy of Science*, 22(1).

Lyon, C., Nehaniv, L. C., Saunders, J., Belpaeme, T., Bisio, A., Fischer, K., & Cangelosi, A. (2016). Embodied language learning and cognitive bootstrapping: methods and design principles. *Int. J. Adv. Robotic Syst*. 13:105. doi: 10.5772/ 63462

MacDorman, K. F., Minato, T., Shimada, M., Itakura, S., Cowley, S., and Ishiguro, H. (2005). "Assessing human likeness by eye contact in an android testbed," *in Proceedings of the 27th annual meeting of the cognitive science society* (Austin, TX: CSS) 21–23.

Main, M., & Solomon, J. (1986). *Discovery of an insecure-disorganized/disoriented attachment pattern*, in T.B. Brazelton and M. Yogman (by), *Affective development in infancy*, Norwood, NJ, Ablex Publishing.

Main, M., Kaplan, N., & Cassidy, J. (1985). Security in infancy, childhood, and adulthood: A move to the level of representation. *Monographs Of The Society For Research in Child* Development. https://doi.org/10.2307/3333827

Manzi, F., Ishikawa, M., Di Dio, C., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020a). The understanding of congruent and incongruent referential gaze in 17-month-old infants: An eye-tracking study comparing human and robot. *Scientific Reports*, *10*(1), 11918. https://doi.org/10.1038/s41598-020-69140-6

Manzi, F., Savarese, G., Mollo, M., and Iannaccone, A. (2020b). Objects as communicative mediators in children with autism spectrum disorder. *Front. Psychol.* 11:1269. doi: 10.3389/fpsyg.2020.01269

Manzi, F., Peretti, G., Di Dio, C., Cangelosi, A., Itakura, S., Kanda, T., ... & Marchetti, A. (2020c). A robot is not worth another: exploring children's mental state attribution to different humanoid robots. *Frontiers in Psychology*, 11, 2011. https://doi.org/10.3389/fpsyg.2020.02011

Marchetti, A., Di Dio, C., Manzi, F., and Massaro, D. (2020). The psychosocial fuzziness of fear in the COVID-19 era and the role of robots. *Frontiers in Psychology*. 10:2245. doi: 10.3389/fpsyg.2020.02245

Marchetti, A., Manzi, F., Itakura, S., & Massaro, D. (2018). Theory of Mind and humanoid robots from a lifespan perspective. *Z. Psychol.* 226, 98–109. doi: 10.1027/2151-2604/a000326

Meltzoff, A. N. (2002). Imitation as a mechanism of social cognition: Origins of empathy, theory of mind, and the representation of action. *Blackwell handbook of childhood cognitive development,* 6-25.

Meltzoff, A. N., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, 54(3), 702-709. doi:10.2307/1130058

Moriguchi, Y., Kanda, T., Ishiguro, H., Shimada, Y., and Itakura, S. (2011). Can young children learn words from a robot? *Interact. Stud.* 12, 107–119. doi:10.1075/is.12.1.04mor

Moro, C. (2011). Material Culture, Semiotics and Early Childhood Development. In: Kontopodis M., Wulf C., Fichtner B. (eds) Children, Development and Education. International perspectives on early childhood education and development, vol 3. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0243-1_4

Morse, A. F., & Cangelosi, A. (2017). Why are there developmental stages in language learning? A developmental robotics model of language development. *Cognitive Sci.* 41, 32–51. doi: 10.1111/cogs.12390

Nielsen, M. (2012). Imitation, pretend play and childhood: Essential elements in the evolution of human culture. *Journal of Comparative Psychology*, 126, 170–181. https://psycnet.apa.org/doi/10.1037/a0025168

Okanda, M., Zhou, Y., Kanda, T., Ishiguro, H., and Itakura, S. (2018). I hear your yes−no questions: Children's response tendencies to a humanoid robot. *Infant Child Dev.* 27:e2079. doi: 10.1002/icd.2079

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., Itakura, S. (2013a). Can infants use robot gaze for object learning?: The effect of verbalization. *Interaction studies*, *14*(3), 351-365. https://doi.org/10.1075/is.14.3.03oku

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., & Itakura, S. (2013b). The power of human gaze on infant learning. *Cognition*, 128(2), 127-133. doi:10.1016/j.cognition.2013.03.011

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308(5719), 255-258. https://doi.org/10.1126/science.1107621

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. doi:10.1037/0012-1649.43.5.1216

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that..." attribution of second-order beliefs by 5-to 10-year-old children. *Journal of Experimental Child Psychology*, 39(3), 437-471. doi:10.1016/0022-0965(85)90051-7

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, 5(2), 125-137.

Piaget, J. (1932). *The moral judgment of the child*. San Diego, CA: Harcourt Brace.

Powers, A., & Kiesler, S. (2006). "The advisor robot: tracing people's mental model from a robot's physical attributes," *in Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction* (New York, NY: ACM), 218–225.

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

Richerson, P. J., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago, IL: University of Chicago Press.

Riva Crugnola, C., & Ierardi, E. (2018). *I legami di attaccamento nel ciclo di vita*. In Caravita, S. C. S., Milani, L., & Traficante, D. (A c. Di),. *Psicologia dello Sviluppo e dell'Educazione. Traiettorie evolutive e contesti educativi: una prospettiva integrata*. Il Mulino, Bologna, pp. 53-69.

Savarese, G., Manzi, F., & Iannaccone, A. (2017). Social Functions in Autistic Children and Interaction with/through Object: A Brief Report. *Psychology*, 8(08), 1129. https://doi.org/10.4236/psych.2017.88073

Shiomi, M., Kanda, T., Ishiguro, H., & Hagita, N. (2006). *Interactive anthropomorphic robots for a science museum*. In M. Goodrich, A. C. Schultz, & D. J. Bruemmer (Eds.), *Proceedings of the 1st ACM/IEEE international conference on human-robot interaction* (pp. 305–312). New York: Association for Computing Machinery.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10(5), 640-656.* https://doi.org/10.1162/089892998562942

Tomasello, M. (2019). *Diventare umani*. Raffaello Cortina Editore, Milano.

Turkle, S., Breazeal, C.L., Dast ́e, O., Scassellati, B.: Encounters with kismet and cog: children respond to relational artifacts. In: *IEEE-RAS/RSJ International Conference on Humanoid Robots*, Los Angeles, CA (2004)

Valle, A., Massaro, D., Castelli, I., & Marchetti, A. (2015). Theory of mind development in adolescence and early adulthood: the growing complexity of recursive thinking ability. *Europe's Journal of Psychology*, 11(1), 112. https://dx.doi.org/10.5964%2Fejop.v11i1.829

Van der Horst, F. C., LeRoy, H. A., & Van der Veer, R. (2008). "When strangers meet": John Bowlby and Harry Harlow on attachment behavior. *Integrative Psychological and Behavioral Science*, 42(4), 370-388. https://doi.org/10.1007/s12124-008-9079-2

Vinanzi, S., Patacchiola, M., Chella, A., & Cangelosi, A. (2019). Would a robot trust you? Developmental robotics model of trust and theory of mind. *Philos. Tr. R. Soc. B* 374:20180032. doi: 10.1098/rstb.2018.0032

Wellman, H. M. (1991). From desires to beliefs: Acquisition of a theory of mind. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 19–38). Basil Blackwell.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72(3), 655-684. doi:10.1111/1467-8624.00304

Wellman, H.M. (2020). *Reading Minds: How Childhood Teaches Us To Understand People*. Oxford, UK: Oxford University Press.

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

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33-35. https://doi.org/10.1145/1118178.1118215

Wing, J. M. (2008). Computational thinking and thinking about computing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 366*(1881), 3717-3725. https://doi.org/10.1098/rsta.2008.0118

Złotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: opportunities and challenges in human–robot interaction. *Int. J Soc. Robot*. 7, 347–360. doi: 10.1007/s12369-014- 0267-6

Zhong, J., Ogata, T., Cangelosi, A., & Yang, C. (2019). Disentanglement in conceptual space during sensorimotor interaction. *Cogn. Comput. Syst*. 1,103–112. doi: 10.1049/ccs.2019.0007

CHAPTER 2

Shall I Trust You? From Child-Robot Interaction to Trusting Relationships

Cinzia Di Dio1*, Federico Manzi[1], **Giulia Peretti[1]**, Angelo Cangelosi[2], Paul L. Harris[3], Davide Massaro[1] and Antonella Marchetti[1]

[1]Research Unit on Theory of Mind, Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy, [2]School of Computer Science, The University of Manchester, Manchester, United Kingdom, [3]Graduate School of Education, Harvard University, Cambridge, MA, United States

Published on Frontiers in Psychology, April 2020.

Keywords

Developmental robotics, HRI, Theory of Mind, attachment, social interaction

Abstract

Studying trust in the context of human–robot interaction is of great importance given the increasing relevance and presence of robotic agents in the social sphere, including educational and clinical. The acquisition, loss and restoration of trust when pre-school and school children played with a human or humanoid robot *in vivo* was investigated. The relationship between trust and the representation of the quality of attachment relationships, Theory of Mind, and executive function skills was also investigated. Additionally, to outline children's beliefs about the mental competencies of the robot, the attribution of mental states to the interactive agent was further evaluated. In general, no substantial differences were found in children's trust in the play partner as a function of agency (human or robot). Nevertheless, 3-year-olds showed a trend toward trusting the human more than the robot, as opposed to 7-year-olds, who displayed the reverse pattern. These findings align with results showing that, for 3- and 7-year-olds, the cognitive ability to switch was significantly associated with trust restoration in the human and the robot, respectively. Additionally, supporting previous findings, a dichotomy was found between attributions of mental states to the human and robot and children's behavior: while attributing to the robot significantly lower mental states than the human, in the Trusting Game, children behaved in a similar way when they related to the human and the robot. Altogether, the results of this study highlight that similar psychological mechanisms are at play when children are to establish a novel trustful relationship with a human and robot partner. Furthermore, the findings shed light on the interplay – during development – between children's quality of attachment relationships and the development of a Theory of Mind, which act differently on trust dynamics as a function of the children's age as well as the interactive partner's nature (human vs. robot).

Introduction

One of the challenges of contemporary robotics is *long-term interaction*, which assumes that competent robot partners will have many human-like characteristics, enabling the complexity and multidimensionality of human interactions. This objective has been strengthened by a new interdisciplinary approach to robotics, i.e. Developmental Robotics (Cangelosi and Schlesinger, 2015). For example, Vinanzi et al. (2019) have proposed an artificial cognitive architecture to simulate human decision making in the robot by using concepts from developmental theories, such as Theory of Mind (ToM). From this perspective, the implementation of an artificial architecture, together with an understanding of the human's response to the behavior of a robot within a relational context, aims to shed light on the processes involved in establishing a relationship with robotic agents (e.g. Wykowska et al., 2016; Wiese et al., 2017). Within this framework, trust comes into play as a key psychological component underpinning successful interpersonal relationships, particularly when these include at least one robotic agent. In the present study, children between the ages of 3 and 9 were observed establishing trusting relationships with a human or the humanoid robot NAO in a simple "guessing" game in which the child and the human or robot played together. Furthermore, not only did we assess trust acquisition, but also a key feature of real-life relational dynamics: trust restoration after trust loss. As a matter of fact, trust is a dynamic process based on past relational experiences and, as such, it is subject to fluctuations operationalized in this study via three phases of trust: acquisition, loss, and restoration. The latter phase is of particular interest. While human forgiveness has been studied in different conditions (see, for example, Grover et al., 2019), the investigation of how relational failures may affect trust restoration in a relationship with a robot is still unexplored.

In psychology, trust can be described as "a multidimensional psychological attitude involving beliefs and expectations about the reliability of the trustee resulting from social experiences involving uncertainty and risk" (Jones and George, 1998; in Lewis et al., 2018, p. 137). Trust in the choices of unknown people can be envisaged also in situations where we passively witness their behavior, with consequences on our own decisions (e.g. Rizzato et al., 2016). The multidimensional nature of trust encompasses the idea that trust can be built based on either (or both) objective factors or (and) an emotional, quite irrational, attitude toward the partner (Lahno, 2001). In this light, emotional trust can be conceived as somewhat independent of objective information. In this study, a situation of total uncertainty was recreated in which the choices of a partner, who should be trusted, are not based on the evaluation of objective elements, and also the decision of the child to trust in the partner are devoid of rational elements. Rather, the decision to trust or not to trust the partner's choices is consequentialist in nature considering that, until proven otherwise, the partner is always accurate in her/his/its choices. That is, trust is progressively built through constant endorsement of the play partners' reliability in providing correct responses (see, Rotenberg, 2010). From this perspective, conformation to the other's choices reflected levels of trust acquisition as well as acceptance of the other as a potential partner (Nass et al., 1995; Nass and Moon, 2000).

The establishment of trusting relationships is critical for effective interpersonal dynamics. This is particularly relevant where children are called to build new relationships with peers, educators, and other adults. An example of the importance of the construction of interpersonal trust is highlighted in a study with children under protection services (Petrocchi et al., 2018). In these critical circumstances, not only does the success of the social interventions rely on building trusting relationships between the child's parents and the social workers, but also between the latter and the child in need of psychosocial adjustment (Hafford- Letchfield and Spatcher, 2007). Developmental research on the construction of trusting relationships shows that trust dynamis change significantly as a function of age. For example, children aged 3 years tend to display trust if the informant is consistently accurate (Clément et al., 2004; Koenig et al., 2004; Pasquini et al., 2007) but are relatively unforgiving in case of mistakes (Harris, 2007), effectively showing a certain behavioral rigidity. With development, particularly from 4 years of age, children become more flexible: they do not rely on another's testimony in an indiscriminate fashion (Harris, 2007) and show selective trust in others' testimony (Clément et al., 2004; Chan and Tardif, 2013). They attend both to the information available at that moment, and to the reliability that a person has shown in the past.

Human trusting relationships are also shaped by past relational histories, originating with primary caregivers (e.g. Camisasca et al., 2017; Giovanelli et al., 2020; Marchetti et al., 2020) and extending to subsequent, significant affective relationships (Bowlby, 1969, 1973, 1980). It has been suggested that children's decision to place trust in an unknown informant, especially in a context of uncertainty, may also depend on generalizing from their personal attachment history (Fonagy, 1998; Allison and Fonagy, 2016; Fonagy et al., 2019; see also, Bo et al., 2017; Luo et al., 2018). For example, securely attached children are more flexible in establishing trustful relationships with epistemically reliable strangers than children with a fragile relational past (see, for example, Corriveau et al., 2009). In this view, it is possible to ask about interactions that involve partners with whom there is no affective history and with whom a relationship needs to be built on the basis of novel interactional dynamics that develop hic et nunc.

Likewise, the development of the individual's cognitive competencies is important, particularly for the definition of the informant's epistemic reliability. Cognitive skills allow individuals to reason about the other's perspective and to objectively evaluate informational access. In this respect, the development of a ToM enabling individuals to conceptualize the mental states that guide behavior (Wimmer and Perner, 1983) and social competence (Premack and Woodruff, 1978; Perner and Wimmer, 1985; see also, Lombardi et al., 2018; for a review, see Wellman et al., 2001) is a necessary prerequisite for the establishment of trusting relationships (Fusaro and Harris, 2008; Lecciso et al., 2011; Sharp et al., 2011; Lucas et al., 2013; Brosseau-Liard et al., 2015; Rotenberg et al., 2015; Van Reet, 2015). The association between the establishment of trust and the development of ToM competencies was first hypothesized by Koenig and Harris (2005) who found that only 4-year-olds, and not 3-year-olds, showed selective trust toward a previously accurate informant. More recently, associating trust beliefs with ToM abilities in children aged 9 years, Rotenberg et al. (2015) further showed that children's trust beliefs in others are associated with both second-order false belief ToM ability as well as with advanced ToM abilities (see also Van Reet et al., 2015). As well-documented (e.g. Carlson and Moses, 2001; Frye et al., 1995), there is a strict relationship between false belief understanding and more general executive function skills. One may then question about the overlap between these competencies in building trust. Still, socio-cognitive skills mediated by one's ability to understand the others' knowledge, like false belief, appear to be more influential in building selective trust rather than more general executive function skills, at least in some cultures (Lucas et al., 2013).

In relation to human–robot interaction, studies that have specifically investigated trust in a robot agent or system have typically involved adult participants. These studies have either used explicit measures of trust assessment, mostly involving self- reports (e.g. Yagoda and Gillan, 2012), or implicit measures of trust assessment. Explicit measures of trust are strongly subject to the idiosyncratic attitude and the impression that one has of the robot, which are often based on beliefs and not on actual interactional experiences with the robot; on the other hand, implicit measures of trust generally involve the postulation of hypotheses framed by specific environmental and theoretical conditions that are then tested during actual interaction with a robotic system. Gaudiello et al. (2016), for example, investigated the role played by functional acceptance (perceived ease of use, usefulness) and social acceptance (generally linked to social competencies) of the robot iCub for effective human–robot interaction. These two aspects appear to be most consistently associated with an enduring perception of the robot's skills, i.e. its usefulness and sociality (Shaw- Garlock, 2009; Heerink, 2010). As a most comprehensive measure of functional and social acceptance of the robot, the users' trust in the robot was assessed as a function of the robot's social and functional knowledge. The users' trust in the robot prevalently relied on its functional rather than social knowledge, although data generally highlighted adults' poor acceptance of, and a predominant distrust in robots. With children, the factors underpinning child human–robot interaction have not been systematically explored. There are several studies that inform about ways in which children interact, play, and learn from a robotic agent in school and educational contexts (Kanda et al., 2004; Okumura et al., 2013a,b; Breazeal et al., 2016; Baxter et al., 2017; Belpaeme et al., 2018; Cangelosi and Schlesinger, 2018; Di Dio et al., 2019). These studies have shown that children tend to interact with robot partners in a human- like manner, proving to be sensitive to verbal and non-verbal signals, such as eye gaze (Okumura et al., 2013a,b), and often attributing mentalistic competencies to the robot (for a review, see Marchetti et al., 2018). In this respect, the work by Short et al. (2010) shows that unfair/cheating robots in a common "rock-paper-scissors" child-game are able to elicit interest in the child as well as a greater tendency to attribute intentions to the robot. This study brings further support to the idea that human-like behavior (either trustful or even deceptive) is associated to a greater interactional potential toward a robot partner.

In the present study, trust was explored through a novel Trusting Game (TG) named "Guess where it is" requiring the interactive partner (either the human or the robot) and, subsequently, the child to guess the position of a doll hidden under a box. Through the structure of the game, the conditions are created for the child to consequentially make the same decisions as the play partner, thus ultimately establishing a trusting relationship (e.g. Nass and Moon, 2000): the other becomes trustworthy because it demonstrates that her/his/its choices, even if random, always lead to a correct answer. This procedure benefitted from having the child gradually build trust in the partner during a social interaction. It was chosen not to establish epistemic trust before the game following best known procedures (see, for example, Koenig and Harris, 2005; Corriveau and Harris, 2009) because it was also wanted to appreciate the dynamics of trust construction when interacting with different relational agents, i.e. the human and the robot. Once trust had been acquired, as indexed by a consistent agreement between the play partner's and the child's responses, the phases of loss of trust and trust restoration put the child's trust to test. These latter phases were most critical for the child because s/he had to reconsider the newly established trust in the robot or the human. To better understand what psychological factors are in place when building a trusting relationship with the robot, as compared to the human, specific different chronological ages were referenced (e.g. Lombardi et al., 2017) where the development of affective and cognitive processes may be distinctively influential on trust. Also, to better appreciate how trust is configured within robot–human and human–human interaction, we avoided creating competitive or collaborative conditions that could have polarized the dynamics of trust-building. As a matter of fact, the type of interaction can significantly influence trust (Hancock et al., 2011) by negatively or positively skewing trust in case of competition or collaboration, respectively (Kidd, 2003; Kuchenbrandt and Eyssel, 2012). Therefore, children were made to play for the mere fun of playing with a little thank-you gift delivered at the end of the game (the structure of the TG is detailed in section "Materials and Methods"). Finally, the distinctive contribution of ToM and executive function skills in building trust at different developmental ages was further assessed, thus extending current literature by also exploring these cognitive components when children interacted with a robot or a human agent. To make the child perceive the robotic agent NAO as a real interactional partner, it was introduced to children in a preliminary session when they were familiarized with some of the robot's physical and social skills (walking, moving its arms, talking, greeting, etc.) (see Vogt et al., 2017). To make its behavior human-like, when playing its turn during the TG, NAO used simple and clear verbal indications, accompanied by gestures indicating the possible target position of the doll. Additionally, the robot was programmed to alternate between looking at the play setup and the child, reproducing a realistic attentional shift (Zanatto et al., 2019). The children's perception of the robot's mental qualities as compared to the human was evaluated through the Attribution of Mental States (AMS) questionnaire (Di Dio et al., 2018, 2019). This measure has consistently shown that school-age children do discriminate between the human and the robot in mental terms, although, during interaction, children also typically display similar behaviors toward both. Accordingly, it was hypothesized to find substantial differences in the children's attribution of mental states to the human and the robot, whereas a similar trust- building dynamics when interacting with either partner during the TG. Additionally, it was hypothesized to find a greater tendency to trust, especially in the human, among younger children whose trust is possibly mainly driven by affect rather than cognition. On the other hand, it was hypothesized to find the establishment of more reflective trusting relationships among children given the development of ToM competencies. No specific predictions were advanced with respect to the role of executive functions in trust dynamics given the fair lack of specific evidence in this respect.

Materials and Methods

Participants

Ninety-four (94) Italian kindergarten and school-age children participated in the experiment. The children were divided into four age groups as follows: 3-year-olds (N = 22, 9 females), 5- year-olds (N = 24, 13 females), 7-year-olds (N = 25, 13 females), and 9-year-olds (N = 23, 12 females). The children were recruited from a preschool and a primary school of Milan. The children's parents received a written explanation of the procedure of the study, the measurement items, and the materials used, and they gave written consent. Children were not identified by parents or teachers for learning and/or socio-relational difficulties. The study was approved by the Local Ethics Committee (Università Cattolica del Sacro Cuore, Milan).

Tasks

The children were assessed in two experimental sessions on different days within a 2-week time frame. In the first session, the children were administered the following tests: AMS scale (inspired by the work of Martini et al., 2016), TG task [inspired by the work of Yang et al. (2017)], and a first-order and (for 5- to 9-year-olds) a second-order False-Belief task (Wimmer and Perner, 1983; Perner and Wimmer, 1985). In the second session, the children were administered a further version of the first-order and second-order False-Belief task, the quality of attachment relationships (SAT) test (Liverta Sempio et al., 2001), an executive function task (Dimensional Change Card Sort, DCCS; Zelazo, 2006) for the 3- and 5-year-olds, and the Developmental NEuroPSYchological Assessment (NEPSY II; Korkman et al., 2007) subtest for the 7- and 9-year-olds. Both tests assess the ability to switch between responses.

Attribution of Mental States

The AMS scale is a measure of the mental states that participants attribute when looking at pictures depicting specific characters, in this case a human and the robot NAO. The scale is an ad hoc questionnaire that was based on Martini et al. (2016). AMS has been used in previous works (Di Dio et al., 2018, 2019; see also, Di Dio et al., 2020; Manzi et al., 2020) and has proven fairly consistent in outlining age-specific response patterns with respect to attribution of mental states to both robots and humans. Children were asked 25 questions grouped in five different state categories: Perceptive, Emotional, Desires and Intentions, Imaginative, and Epistemic. The child had to respond "Yes" or "No" to each question. If the answer was Yes, then the experimenter asked a follow-up question: "How much? A little bit or very much?", yielding a 3-point scale. For example, in answer to the question: "Do you think that he/she/it can understand?", the range of answers could be: No (0), Yes, a little bit (1), or Yes, very much (2). The total score was the sum of all answers (range = 0–50); the five partial scores were the sum of the answers within each category (score range = 0–10).

Trusting Game

The TG was inspired by the work of Yang et al. (2017). The game requires the play partner (either the human or the robot) and the child guess the position of a doll hidden under a box. By its nature, the game is neither explicitly collaborative nor competitive since both players have to independently guess the position of the doll and correct guesses do not lead to any tangible reward. The TG involves two players (i.e. a child – participant – playing with either the experimenter or the robot) and a game-master (i.e. a second experimenter). The game consists of presenting to the players two boxes and a little doll that are positioned on a table that looks very much like a coffee table and at which both players are seated. The game-master, who sits on the opposite side of the table, hides the doll under one of the two boxes without being seen by the two players (see Figure 1 for a depiction of the experimental setup). The game consists of guessing where the game-master has hidden the doll. The game starts with the experimenter explaining verbally to the players the rules, showing them an example of a sequence: "Now you and (the other partner's name) will play a game together called "guess where it is." I'll show you how it is played. Here are two boxes and a little doll. I will hide the doll under one of the two boxes, but you won't see where I hide it because I will put this paper board in front of you, like this." After positioning the board, the experimenter moves the boxes around and then removes the board, placing it on the side of the table. The experimenter then informs the partners that they have to guess where the doll has been hidden by pointing at one of the boxes. Next, without revealing the doll's location, the experimenter asks: "It is all clear?". If both partners answer positively, then the play started. All children understood the instructions the first time. The children were also informed that they would receive a packet of stickers at the end of the game to thank them for their participation. Once the game began, the experimenter told the child that the partner (referred to by her/his/its name) would always make the first guess. The position of the doll was established a priori to correctly instruct (or program, if NAO) the play partner's choice during each phase of the game.

The TG involves three independent phases. The first phase [Trusting Acquisition (TA)] aims to assess the participant's acquisition of trust in the other player by calculating how many trials elapse before the child follows the other player's guess. Trust is assumed when the child follows the other player's guess on three consecutive trials. After trust acquisition, the game switches into the second phase [Mistrust Acquisition (MA)], which assesses the participant's acquisition of mistrust in the other player by calculating how many trials it takes for the child not to follow the other player's guess. Mistrust is assumed when the child does not follow the other player's guess on three consecutive trials. The last phase [Trusting Restoration (TR)] shares the same play structure as the initial phase. The game lasted, on average, between 10 and 20 min.

Each phase consisted of a maximum of 10 trials and ended after trust acquisition (phase 1), mistrust acquisition (phase 2), and trust restoration (phase 3). The switch to the following phase also occurred if the participant completed 10 trials within a given phase without completing the three-trial sequence.

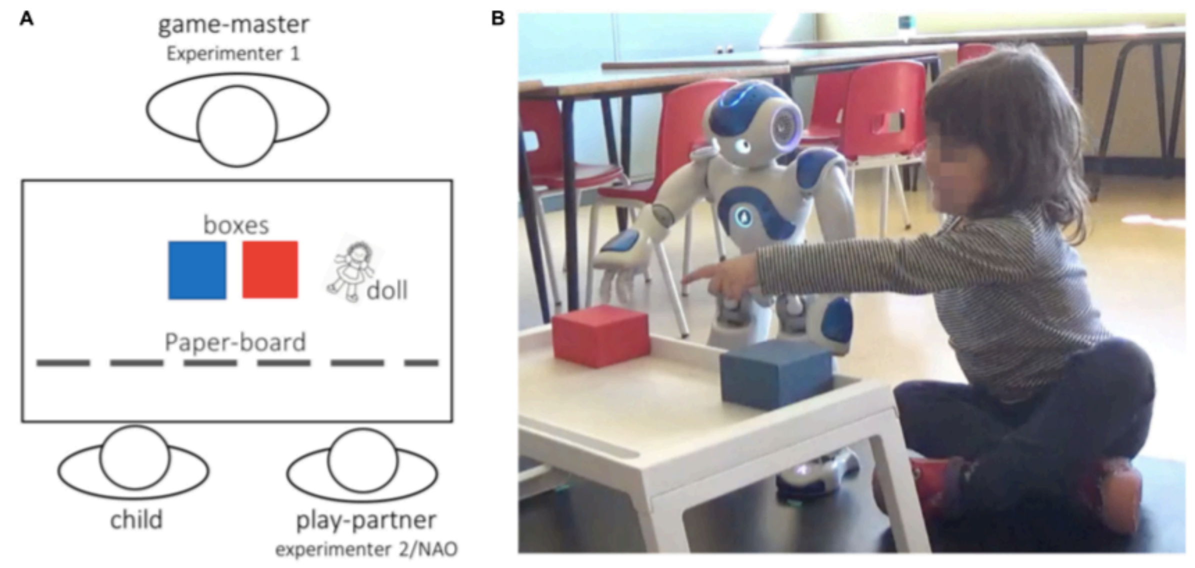


FIGURE 1 | Overview of the experimental setup in the Trust Game. **(A)** Overview of the participants' seating arrangement and stimuli. **(B)** Photograph representing the SoftBank Robotics NAO humanoid robot while playing the Trust Game with a child (subject).

The dependent variable (DV) was the number of trials the child required before acquiring trust or mistrust. For example, in the initial phase, if the child started to follow the other player for three consecutive trials after the second trial (i.e. 0 0 1 1 1), the participant scored 2. If the child displayed trust immediately (i.e. 1 1 1), s/he scored 0. If the child completed the 10 trials within each phase without acquiring trust or mistrust, she/he scored 8, which is the maximum value that could possibly be attributed before ending the phase with a three-trial sequence. To compare data in the analyses, trust and trust restoration indexes were reversed to indicate, alongside trust loss, a comparable measure of the tendency to trust. Thus, a child could score between 0 (low trust) and 8 (high trust).

For the treatment of missing cases, it was considered mean, median, and mode values, as well as children's most common response patterns. The median was ultimately chosen as the most representative index for replacing missing values. Accordingly, two children were recovered for age groups 3, 5, and 9 years; one child was recovered for the age group 7 years. When an entire session was missing, the values were not replaced and the child was removed from the analyses. Accordingly, one child was removed from age group 3 years and three children were removed from age group 7 years.

Theory of Mind

The Unexpected Transfer task (Wimmer and Perner, 1983) and the Unexpected Content task (Perner and Wimmer, 1985) were used to evaluate first-order ToM by assessing the acquisition of false beliefs understanding. First-order ToM entails a recursive thinking, which implies the meta-representation or the representation of a mental representation of a low complexity level, of the kind “I think that you think...”. Children exhibit this competence at around 4 years of age with the emergence of false beliefs. The child is told a story involving two doll characters. One of the characters is deceived with respect to either the location or contents of an object and the child is tested for his/her ability to understand the character’s false belief. For example, the unexpected transfer story is about two siblings playing with a ball in a room. One of the children puts the ball in a box and leaves the room. Meanwhile, the other child takes the ball out of the box, puts it in the basket and goes away. Finally, the first character comes back in the room and wants to play with the ball. At the end of the story, the experimenter asks the child the following questions: “What is the first place where she will look for the ball?”—referring to the first character (first-order false belief question); “Where did the child put the ball before going away?” (control memory question); “Where really is the ball?” (reality control question). The answers to the two control questions (memory and reality) were used to filter the children’s performance. Having passed control questions, the test question about false belief is scored 1 if correct and 0 if incorrect.

The development of a second-order false belief competence was assessed through the Ice-Cream Van task (Perner and Wimmer, 1985) and the Look-Prediction task (Liverta Sempio et al., 2001; Astington et al., 2002). Second-order ToM implies a meta-representation of a greater complexity with respect to first-order ToM, of the kind “I think that you think that s/he thinks...”. Children aged from 7 years have typically matured this competence, although it can also emerge at an earlier age. The second-order ToM stories involve three characters presented on a storyboard. For example, the ice-cream van story is about Maria and Giovanni, who – while playing in the park – see an ice-cream van. Maria wants to buy an ice cream, but she has no money. She therefore decides to go home to take the money, sure that the ice-cream van will stay in the park. However, while Maria is away, Giovanni sees the ice-cream van moving away. Giovanni asks the ice-cream man where he is going, and the ice-cream man replies that he is going in front of the school to sell more ice creams. While Maria is leaving home, she sees the ice-cream man and she asks him where he is going. After knowing that he is moving to school, she says that now that she has the money, she can follow him to school. At the end of the story Giovanni goes to Maria’s house, and asks her mother where her friend is. Maria’s mum answers that Maria has just gone out to buy an ice cream. The child (participant) is then asked the following questions: “Where does Giovanni think Maria went to buy the ice cream? (second-order false belief); "Why does Giovanni think so?" (justification); "Does Maria know that the ice-cream van is in front of the school?" (first-order false belief); "Does Giovanni know that the ice-cream man spoke with Maria while she was leaving her house?" (reality control question); "Where did Maria go to buy the ice cream?" (memory control question). For both second-order false belief tasks, having passed the control questions, children scored 1 for correct statements and 0 for incorrect statements on both test and justification questions. A second-order false belief task total score was then calculated ranging from 0 (no response) to 2 (completely correct response) (Perner and Wimmer, 1985).

Separation Anxiety Test–Family Version (F-SAT)

The Separation Anxiety Test is a semi-projective task that evaluates the child's mental representation of his/her attachment to the caregiver. The original version developed by Hansburg (1972) for adolescents was adapted by Klagsbrun and Bowlby (1976) for children aged 4 to 7 years. In the latter version, six pictures are presented to the child, each depicting a situation of separation from a familiar caregiver. The child is asked to describe the protagonist's feelings, to justify them, and to predict what the protagonist will do, thereby probing the coping strategy of the protagonist. The Italian version used in this study (Liverta Sempio et al., 2001) is based on a modification of other versions of the same task (Fonagy et al., 1997).

The coding reflects three dimensions: (1) attachment, i.e. the ability to express vulnerability and need; (2) self-confidence, i.e. the ability to autonomously face separation; and (3) avoidance, i.e. the propensity to speak about the separation. Participants score 1 for each dimension. The final score is the result of the sum of the scores in the attachment scale and in the self-confidence scale, and of the sum of the inverse of the avoidance scale, calculated by subtracting this score from the total amount potentially obtainable on this scale. Scores range from 6 to 36, with higher scores reflecting greater quality of attachment relationship.

Executive Function Skills

Children aged 3 and 5 years were administered the DCCS assessing the capacity to switch responses [for a full description of the test, please refer to Zelazo (2006)]. Seven and 9-year- olds' executive functions were assessed using "A Developmental NEuroPSYchological Assessment" subtest (NEPSY II; Korkman et al., 2007), testing the ability to inhibit automatic responses and to switch between response types. The child looks at a series of black and white shapes or arrows and names either the shape or direction or makes an alternate response, depending on the color of the shape or arrow. In the present study, the combined scores of the Inhibition NEPSY-II subtest, which combines accuracy and speed of response, were used. For a detailed description of the scoring criteria, please refer to the manual (Korkman et al., 2007).

Experimental Procedure

Introducing the Play Partners

On a day that preceded the main experimental session, children were introduced to three play partners (two humans – a boy and a girl – and the robot) through video clips displayed in class on a large projector. In the videos, each of the potential partners said the same sentence: “Hello, my name is. I will be playing with you in the next days. See you soon. Bye.” The videos represented the actors while exiting a room and waving their hand to say goodbye. In this way, the children saw that the robot NAO could walk, talk, and move its head and arms.

Experimental Sessions

The children were tested individually in a quiet room in their kindergarten or school. Tests were carried out by two researchers both in the morning and in the afternoon during normal activity. In the first session, the administration of the battery lasted approximately 20–30 min, depending on the child’s age. The administration of the task in the second session took about 35– 45 min.

The first session started with the administration of AMS. The five AMS state categories (Perceptive, Emotional, Desires and Intentions, Imaginative, and Epistemic) were randomized across children. Afterward, children participated in the TG. At this point, the partner (i.e. human or robot) entered the experimental room and was introduced by the experimenter by his/her name: “Do you remember, this is ...”. Then, both the child and the partner were invited to sit on the ground on a plastic carpet in front of an ad hoc table. The plastic carpet was used to correctly position NAO when children interacted with the robot. A paper black board was positioned next to the table, and was used to cover the setting when playing. A female experimenter played with girls and a male experimenter played with boys. Half of the children played with the robot in the first session and with the human in the second session. The other half underwent the reversed play order.

After the game, the child was administered one of the two first-order ToM tasks and, starting from 5 years of age upward, one of the two second-order ToM tasks. The order of the ToM tasks was randomized across children, so that those who performed, for example the unexpected transfer task in the first session, completed the unexpected content task in the second session. The same was true for the second-order ToM tasks. Finally, children were given two further assessments: SAT and executive function.

Statistical Analysis

Statistical analysis was carried out using the IBM Statistical Software Platform SPSS (v. 19.0). To evaluate possible differences in children's tendency to trust the human and robot partner as a function of the child's age and trust phase (acquisition, loss, and restoration), a repeated measures General Linear Model (GLM) analysis was carried out. The DV was the number of trials until children followed their partner (trust acquisition), stopped following their partner (trust loss), and again followed their partner (trust restoration) during the TG. To compare data from the three phases, trust and trust restoration indexes were reversed to indicate, together with trust loss, a comparable measure of the tendency to trust. That is, for all three phases of the TG, greater numbers correspond to a greater tendency to trust.

Additionally, correlation analyses (Pearson's r) were carried out to evaluate the relationship between the tendency to trust and (1) the quality of attachment relationships (SAT), (2) ToM (first- and second-order false beliefs tasks), and (3) executive function skills.

Finally, to assess possible differences in children's mental states attribution to the robot with respect to the human partner, a repeated measures GLM analysis comparing AMS scores between human and robot was carried out as a function of the children's age. For all the GLM analyses, the Greenhouse–Geisser correction was used for violations of Mauchly's Test of Sphericity, $P < 0.05$. All post hoc comparisons were Bonferroni corrected.

Results

Trusting Game

The GLM analysis, with three levels of phase (acquisition, loss, and restoration) and two levels of agency (HB and RB) as within-subjects factors, and age group (four levels) as the between-subjects factor (3, 5, 7, and 9 years), was carried out to assess children's tendency to trust in the human and in the robot. An inspection of the box plots displaying the performance of each age group showed no extreme cases.

The results revealed a main effect of phase (Figure 2A), $F(2, 172) = 10.51$, $P < 0.001$, partial-$\eta2 = 0.11$, $\delta = 0.99$, indicating that, independent of agency and age group, children exhibited a lower tendency to trust in phase 3 (trust restoration), compared to both phase 1 (trust acquisition), Mdiff = 1.19; SE, 0.26; $P < 0.001$, and phase 2 (trust loss), Mdiff = 0.94; SE, 0.27; $P < 0.01$. Additionally, age-related differences were found, $F(3, 86) = 8.76$, $P < 0.001$, partial-$\eta2 = 0.23$, $\delta = 0.99$. More specifically, 3-year-olds showed a greater tendency to trust than the other age groups including the

5-year-olds, Mdiff = 1.83; SE, 0.52; $P < 0.01$; 7-year-olds, Mdiff = 1.1; SE, 0.53; $P < 0.05$; and 9-year-olds, Mdiff = 2.64; SE, 0.53; $P < 0.001$. No interactions were found between phase and age group, $P > 0.05$. Additionally, agency did not have any impact as a main effect and in the interaction with the other variables, $P > 0.05$.

Having found a consistent correlation across ages between first-order ToM and performance in the TG as described below, a further GLM was carried out using first-order ToM as a covariate. This analysis revealed a main effect of agency (Robot – RB > Human Being – HB), $F(1,85) = 4.99$, $P < 0.05$, partial- $\eta 2 = 0.06$, $\delta = 0.60$, and a significant interaction of agency × age group, $F(3, 172) = 2.81$, $P < 0.05$, partial-$\eta 2 = 0.09$, $\delta = 0.66$. The post hoc analyses showed that while 3-year-olds tended to generally trust in the human more than in the robot, Mdiff = 1.04; SE, 0.67; $P < 0.05$, children aged 7 years tended to trust in the robot more than in the human, Mdiff = 1.33; SE, 0.56; $P < 0.05$. This interaction is plotted in Figure 2B.

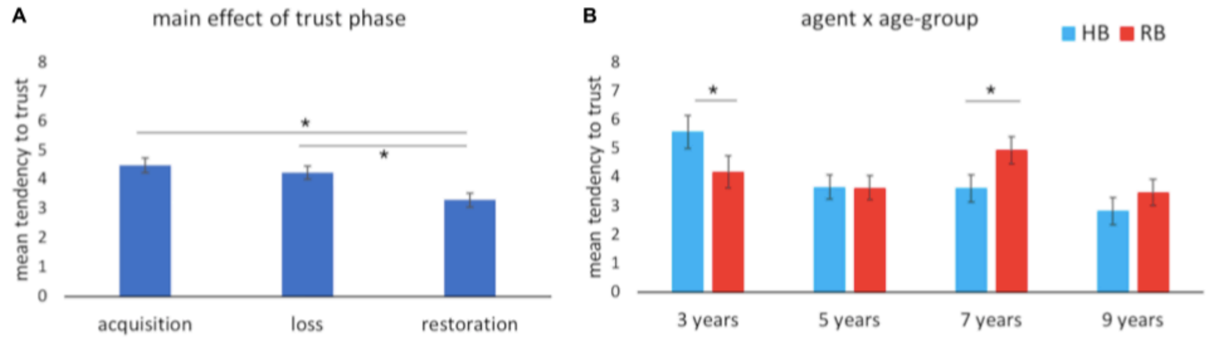


FIGURE 2 | Trust scores for the Trusting Game. Children's average tendency to trust during the Trusting Game **(A)** across the three phases of trust (acquisition, loss, and restoration) and **(B)** across age groups for the human (HB) and the robot (RB) when controlling for the effect of Theory of Mind. The bars represent the standard error of the mean. *indicates significant differences.

Correlations

Trusting and SAT

As shown in Table 1, most of the significant correlations between the quality of attachment relationships (SAT) and the tendency to trust were positive, i.e. more securely attached children showed a greater tendency to trust in the play partner's choice. These relationships were found mainly when children played with the human, and especially in the youngest age group. For 3-year-olds, all SAT dimensions (except for avoidance) correlated positively with a greater tendency to trust in the human, including during the trust loss phase. For 7-year-olds, quality of attachment positively correlated with the tendency to trust during the trust loss phase.

For 9-year-olds, the results also showed a positive relationship between trust acquisition and the SAT sub-dimension of attachment, indicating that more securely attached children tended to acquire trust quicker. Additionally, among 9- year-olds, there was a positive correlation between the tendency to trust during the trust loss phase and the SAT sub-dimension of avoidance. This correlation was also significant across ages.

A positive correlation was finally found between the SAT sub-dimension of attachment and the tendency to trust in the robot for 3-year-olds during the restoration of trust phase. No significant correlations were found for 5-year-olds.

Trusting and ToM

The scores on the two ToM tasks were merged into one single score for each level of complexity (first and second order).

A low level of ToM performance (coded 0) included children who scored 0 (failed) on both tasks, whereas a high level of performance (coded 1) included children who passed at least one ToM task at each complexity level. Table 2 reports descriptive data for the ToM tasks.

All correlations found between the tendency to trust and ToM scores were negative. Thus, greater ToM abilities were associated with a lower tendency to trust, i.e. with a more reflective tendency to trust. This relationship was independent of the partner's agency (human or robot) or the child's age. The tendency to trust was often significantly correlated with first-order ToM, which was therefore included as a covariate in the GLM model described above. Finally, a substantial negative correlation between the tendency to trust and second-order ToM was observed during the acquisition of trust for children aged 7 years when playing with the human. These statistics are reported in Table 3.

Trusting and Executive Function Skills

Children aged 3 and 5 years were administered the DCCS, which assesses the capacity to switch between responses (Zelazo, 2006). The same skill was assessed in 7- and 9- year-olds using the "Developmental NEuroPSYchological Assessment" subtest (NEPSY II; Korkman et al., 2007). To compare data across age groups, scores were standardized.

Significant age-related positive relationships were found between the ability to switch and the tendency to trust during the restoration phase among 3-year-olds when playing with the human, $r(19) = 0.49$, $P < 0.05$, and among 7-year-olds when playing with the robot, $r(22) = 0.43$, $P < 0.05$.

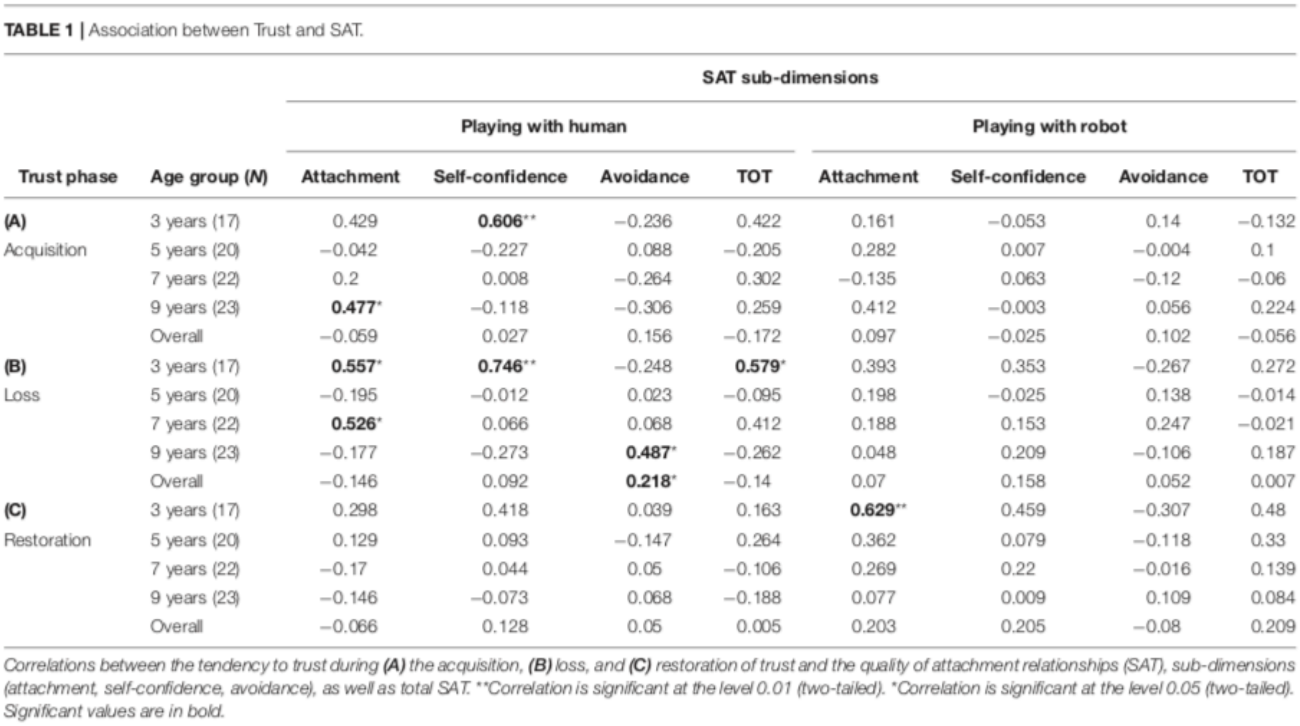


TABLE 1 | Association between Trust and SAT.

Trust phase	Age group (*N*)	SAT sub-dimensions Playing with human Attachment	Self-confidence	Avoidance	TOT	Playing with robot Attachment	Self-confidence	Avoidance	TOT
(A)	3 years (17)	0.429	**0.606****	−0.236	0.422	0.161	−0.053	0.14	−0.132
Acquisition	5 years (20)	−0.042	−0.227	0.088	−0.205	0.282	0.007	−0.004	0.1
	7 years (22)	0.2	0.008	−0.264	0.302	−0.135	0.063	−0.12	−0.06
	9 years (23)	**0.477***	−0.118	−0.306	0.259	0.412	−0.003	0.056	0.224
	Overall	−0.059	0.027	0.156	−0.172	0.097	−0.025	0.102	−0.056
(B)	3 years (17)	**0.557***	**0.746****	−0.248	**0.579***	0.393	0.353	−0.267	0.272
Loss	5 years (20)	−0.195	−0.012	0.023	−0.095	0.198	−0.025	0.138	−0.014
	7 years (22)	**0.526***	0.066	0.068	0.412	0.188	0.153	0.247	−0.021
	9 years (23)	−0.177	−0.273	**0.487***	−0.262	0.048	0.209	−0.106	0.187
	Overall	−0.146	0.092	**0.218***	−0.14	0.07	0.158	0.052	0.007
(C)	3 years (17)	0.298	0.418	0.039	0.163	**0.629****	0.459	−0.307	0.48
Restoration	5 years (20)	0.129	0.093	−0.147	0.264	0.362	0.079	−0.118	0.33
	7 years (22)	−0.17	0.044	0.05	−0.106	0.269	0.22	−0.016	0.139
	9 years (23)	−0.146	−0.073	0.068	−0.188	0.077	0.009	0.109	0.084
	Overall	−0.066	0.128	0.05	0.005	0.203	0.205	−0.08	0.209

*Correlations between the tendency to trust during **(A)** the acquisition, **(B)** loss, and **(C)** restoration of trust and the quality of attachment relationships (SAT), sub-dimensions (attachment, self-confidence, avoidance), as well as total SAT. **Correlation is significant at the level 0.01 (two-tailed). *Correlation is significant at the level 0.05 (two-tailed). Significant values are in bold.*

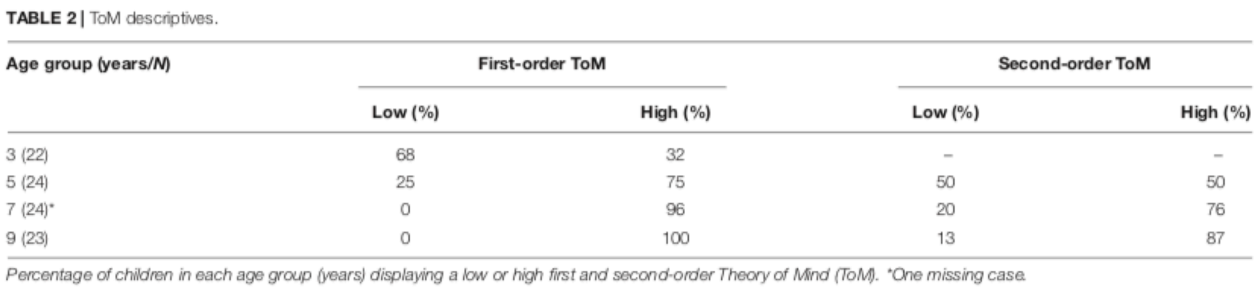


TABLE 2 | ToM descriptives.

Age group (years/*N*)	First-order ToM Low (%)	High (%)	Second-order ToM Low (%)	High (%)
3 (22)	68	32	–	–
5 (24)	25	75	50	50
7 (24)*	0	96	20	76
9 (23)	0	100	13	87

*Percentage of children in each age group (years) displaying a low or high first and second-order Theory of Mind (ToM). *One missing case.*

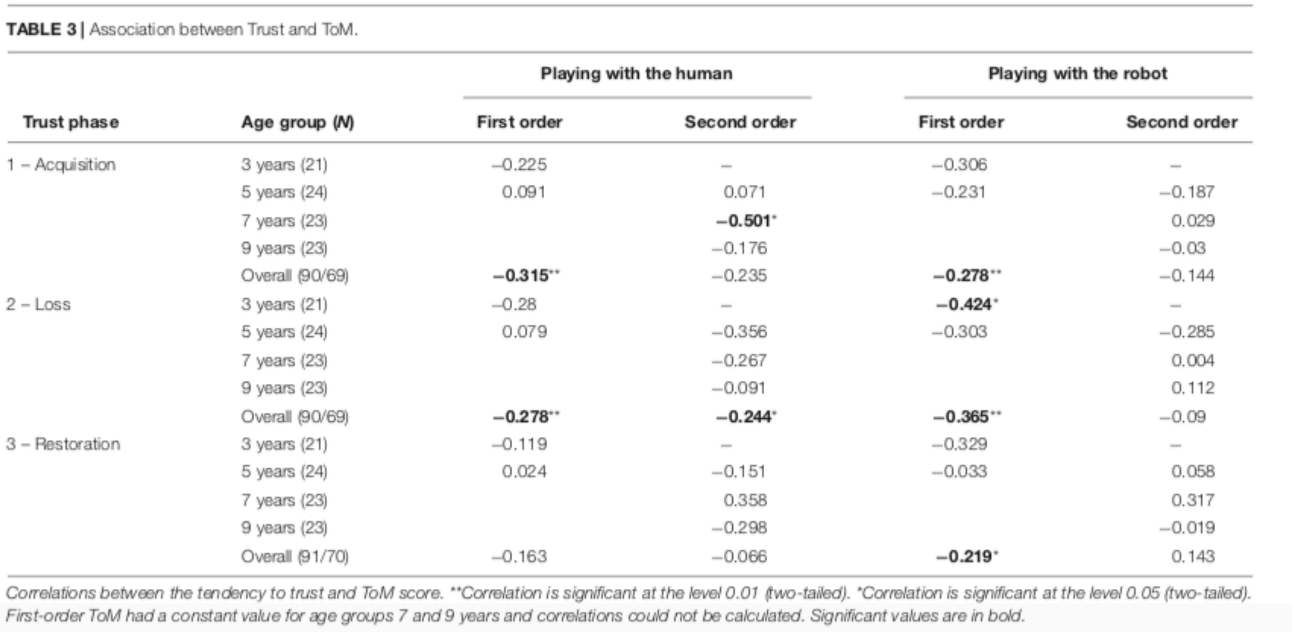


TABLE 3 | Association between Trust and ToM.

Trust phase	Age group (*N*)	Playing with the human First order	Second order	Playing with the robot First order	Second order
1 – Acquisition	3 years (21)	−0.225	–	−0.306	–
	5 years (24)	0.091	0.071	−0.231	−0.187
	7 years (23)		**−0.501***		0.029
	9 years (23)		−0.176		−0.03
	Overall (90/69)	**−0.315****	−0.235	**−0.278****	−0.144
2 – Loss	3 years (21)	−0.28	–	**−0.424***	–
	5 years (24)	0.079	−0.356	−0.303	−0.285
	7 years (23)		−0.267		0.004
	9 years (23)		−0.091		0.112
	Overall (90/69)	**−0.278****	**−0.244***	**−0.365****	−0.09
3 – Restoration	3 years (21)	−0.119	–	−0.329	–
	5 years (24)	0.024	−0.151	−0.033	0.058
	7 years (23)		0.358		0.317
	9 years (23)		−0.298		−0.019
	Overall (91/70)	−0.163	−0.066	**−0.219***	0.143

*Correlations between the tendency to trust and ToM score. **Correlation is significant at the level 0.01 (two-tailed). *Correlation is significant at the level 0.05 (two-tailed). First-order ToM had a constant value for age groups 7 and 9 years and correlations could not be calculated. Significant values are in bold.*

Attribution of Mental States

A repeated measures GLM analysis comparing AMS scores between human and robot, with five levels of state (perceptual, emotions, intentions and desires, imagination, and epistemic) and two levels of agency (HB and RB) as within-subjects factors, and age group (four levels) as the between-subjects factor, showed a main effect of state, F(4, 332) = 71.72, P < 0.001, partial- η2 =0.46,δ=1,amaineffectofagency(HB>RB),F(1, 83) = 82.10, P < 0.001, partial-η2 = 0.50, δ = 1, an interaction of state × age group, F(12, 332) = 5.18, P < 0.001, partial- η2 = 0.16, δ = 1, an interaction of agency × age group, F(3, 83) = 8.66, P < 0.001, partial-η2 = 0.24, δ = 0.99, an interaction of state × agency, F(12, 332) = 19.99, P < 0.001, partial-η2 = 0.19, δ = 1, and a three-way interaction between state, agency, and age group, F(12, 332) = 2.31, P < 0.01, partial-η2 = 0.85, δ = 0.96. This interaction is represented in Figure 3.

Exploring the three-way interaction, the most consistent difference was for the attribution of perception (HB > RB), which was significant for all four age groups, P < 0.01. Attribution of epistemic states was also greater for HB than RB for all age groups, P < 0.01, except 5-year-olds, for whom there was a trend toward significance, P = 0.07. Attributions of emotion and imagination were similar for HB and RB among 3- and 5-year-olds, but greater for HB among 7- and 9-year-olds, P < 0.05. Finally, only 9-year-olds ascribed greater intentions and desires to HB than RB, Mdiff = 4.00;

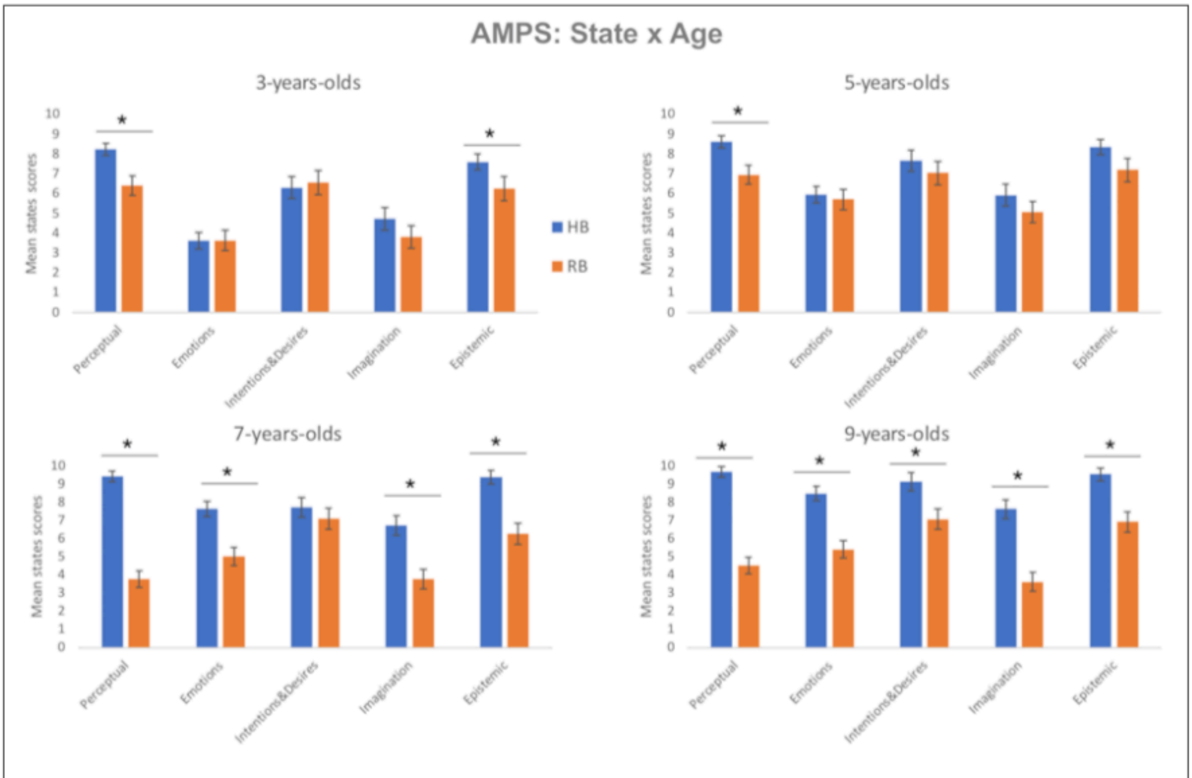


FIGURE 3 | Children's scores on the Attribution of Mental States (AMS) scale. AMS mean scores for the human (HB = blue bar) and the robot (RB = orange bar) for each age group (3-, 5-, 7-, and 9-year-olds) as a function of state (Perceptual, Emotions, Intentions and Desires, Imagination, Epistemic). The bars represent the standard error of the mean. * indicates significant differences.

SE, 0.63; P < 0.001. These post hoc analyses are summarized in Figure 3. Overall, these analyses confirm that humans and robots are differentiated even by 3-year-olds with respect to perception and epistemic states with that differentiation spreading to all five states among 9-year-olds.

Discussion

The present study investigated trust dynamics when children aged 3, 5, 7, and 9 years played a TG in vivo with either a human or a robot partner. Children's tendency to trust decreased across the three phases of the game, from acquisition to restoration of trust. Also, 3-year-olds displayed a greater tendency to trust in both play partners compared to the other age groups, although initially placing their trust more easily in the human than in the robot. The opposite was observed for the 7-year-olds, who generally placed more trust in the robot than the human.

To better understand age changes in trust, the results for quality of attachment relationships, false belief understanding, and executive function skills were examined. It has been previously shown that children aged 3 and 4 years are likely to endorse information provided by someone who proved accurate in the past (see also Koenig and Harris, 2005; Pasquini et al., 2007; Harris, 2007). The results for SAT deepen this observation. Among 3-year-olds, it was found that the SAT sub-dimension of self-confidence was positively associated with selective trust in the human, probably because these children's past relationships were secure, thus increasing the perception of the unfamiliar experimenter as trustworthy. On the contrary, the robot was an entity with which children had never had any relational experience, further skewing the youngest children's trust preference toward the human. Additionally, it was found that the youngest children – and particularly securely attached children – showed a tendency to retain trust during the loss of trust phase, confirming a certain behavioral rigidity as introduced above. However, when realizing that the other was no longer trustworthy, they switched to trusting the robot more. This result supports the observation that when very young children's expectations are betrayed (loss of trust), they are less forgiving than older children (Harris, 2007); additionally, our findings enrich previous results (Corriveau and Harris, 2009) by further showing that children who are securely attached in infancy are more flexible when investing their trust.

The development of a fundamental cognitive ability makes a substantial contribution to trust dynamics in child–robot interaction across all age groups. According to our findings, the development of ToM appears to temper the relation between quality of attachment relationships and trust, by introducing into the trust matrix a mentalistic evaluation of the other's judgment based on an awareness of her/his/its beliefs. More specifically, children who had developed at least first-order ToM also knew that the other player did not know the position of the doll, and was therefore an unreliable informant. Not by chance, the effect of ToM on trusting behavior was most evident at 3 and 7 years of age, typically marked by the development of increasingly complex levels of ToM. When children start developing the concept of the other's mind, they are also able to evaluate whether the other (either a human or a robot) is trustworthy on the basis of informational access. Preferential trust in either agent then moderates. The dichotomy found between the younger and older age groups in the AMS to the robot and the human further helps to delineate the children's perception of the robot as a mentalistic agent: For younger children, the robot is perceived as more mentalistically comparable to the human than for older children. Nevertheless, when younger children decided to trust a play partner, the affective component prevailed over the more "cool" mentalistic component, defining the preferred relational target accordingly (i.e. the human). On the other hand, the trust attributed to the robot by older children may stem from the dominance of a cognitive over an affective engagement.

In support of agent-specific differences in trusting behavior between 3- and 7-year-olds, the results also revealed a positive association between the ability to switch and trusting behavior during the trust restoration phase among children aged 3 and 7 years. Strikingly, and consistent with the data discussed above, these relations were specific to playing with the human partner for the 3-year-olds, and to playing with the robot for the 7- year-olds. In general, these correlations indicate that a greater tendency to recover trust in the other is associated with the development of the ability to switch. The specificity related to the play partner's agency further underlines the relevance of the interactive partner for the child and reflects children's engagement with one or the other player: 3-year-olds' selective trust in the human was plausibly influenced by the quality of attachment relationships – as also evidenced by data on attachment described above; 7-year-olds preferential trust in the robot was possibly due to an emerging familiarity with artificial devices typical of this age. These results shed light on previous findings (Lucas et al., 2013) that showed that, compared to false belief understanding, executive functions skills do not play an essential role in building selective trust in an informant, at least within specific cultural frames. In this study, executive function skills were found to play a role, although not during the confidence acquisition phase, but rather during the restoration of confidence. Here, the child's ability to switch was possibly required to re-organize information and re-establish trust in an informant that, during the loss of trust phase, became unreliable. Executive function skills may then be specifically involved in building trust only under specific conditions, which had not been empirically assessed so far.

Conclusion, limitations, and future directions

The present study provided some insight into the dynamics of trust both when relating to a human and a robot partner. These results highlighted the impact of cognitive development, as well as children's attachment history. It was found that cognition and attachment operated separately (given the absence of a direct correlation between these two dimensions) on the establishment of trust. Particularly for children aged 3 years, trust appears to be significantly influenced by the affective dimension of trust, especially when interacting with a human. Interestingly, although securely attached children exhibited a greater tendency to trust the human, they also shifted their trust more rapidly in the trust restoration phase with the robot. This may be due to the lack of any affective bond with the robot and to the child's cool relational attitude toward it. Effectively, this would render the robot a more "forgivable" partner.

Also, the development of false belief understanding proved to play a significant role in the establishment of trusting relationships. In particular, the development of mentalizing abilities enabled children to reflect rationally on the fact that the other player had exactly the same guessing opportunities as they did, and was therefore as susceptible to making mistakes as they were. This moderated the effect of the affective component of trust.

In the present study, the robot proved to be less susceptible to the dynamics associated with the quality of attachment relationships, and thus became a more stable trusted partner. For this reason, and particularly for children with fragile affective relational histories who have difficulties with trust, the robot might fulfill a significant scaffolding role in human–human interaction. However, an evolution of the robot as a social partner is also to be expected. Therefore, different relational dynamics may be anticipated, according to which, perhaps, an affective relation history will be created with this new entity. In this respect, a longitudinal study would further delineate the development of trust in the robot increasing the robustness of the findings. Also, a larger sample size would eventually confirm the observed tendencies.

Last, but not least, the findings from this study may inform disciplines such as Developmental Robotics on how cognitive architectures can be modeled in the robot so as to make it trusting in the human partner in a "human-like" fashion, as discussed above. This circular behavior would make the human–robot relationship increasingly ecological and, ultimately, trustful. Starting, for example, from the architectural model designed by Vinanzi et al. (2019), in which the robots' trust in an informant varied as a function ToM, the present findings clearly indicate further psychological factors that may be integrated in the robot to design the robot's trust in the human at different developmental levels. Recent technical and theoretical achievements in the field of social robotics have encouraged researchers to develop social robots as tutors and learning companions for children (e.g. Movellan et al., 2009; Tanaka and Matsuzoe, 2012; Breazeal et al., 2016). Therefore, studying the mechanisms by which children learn from robots, and vice versa, is of vital importance.

Data Availability Statement

All data needed to evaluate the conclusions in the article are present in the article.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethic Committee, Università Cattolica del Sacro Cuore, Milano, Italy. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

AM, AC, PH, DM, CD, and FM conceived and designed the experiment. FM and GP conducted the experiments in schools. AM and FM secured ethical approval. CD and DM carried out the statistical analyses. AC granted the use of the SoftBank Robotics NAO humanoid robot. All authors contributed to the writing of the manuscript.

Funding

This work has been in part supported by the Honda grant "DecIFER" and the US EOARD grant "THRIVE++" to AC, and by the Universitá Cattolica del Sacro Cuore, Milan.

Acknowledgments

We wish to thank Dr. Gabriella Fumagalli, Principal of the "I.C. Leonardo da Vinci", and Mrs. Fulvia Maria Frigerio for their deep involvement in the organization of and support to the study. We wish to thank Dr. Mario Uboldi, Principal of the "I.C. Giovanni Pascoli", and Mrs. Maura Panzone for their important contribution to the organization of the project and support to the study. We also wish to thank all the teachers, who actively helped in conducting the study and, in particular, the schools "Scuola dell'Infanzia Premenugo," "Scuola dell'Infanzia Rodano," "Scuola dell'Infanzia Caleppio," "Scuola Primaria Rasori," and "Scuola PrimariaSettala."

References

Allison, E., and Fonagy, P. (2016). When is truth relevant? *Psychoanal.* Q. 85, 275–303. doi: 10.1002/psaq.12074

Astington, J. W., Pelletier, J., and Homer, B. (2002). Theory of mind and epistemological development: the relation between children's second-order false-belief understanding and their ability to reason about evidence. *New Ideas Psychol.* 20, 131–144. doi: 10.1016/s0732-118x(02)00005-3

Baxter, P., Ashurst, E., Read, R., Kennedy, J., and Belpaeme, T. (2017). Robot education peers in a situated primary school study: personalisation promotes child learning. *PLoS One* 12:e0178126. doi: 10.1371/journal.pone.017 8126

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social robots for education: a review. *Sci. Robot.* 3:eaat5954. doi: 10.1126/ scirobotics.aat5954

Bo, S., Sharp, C., Fonagy, P., and Kongerslev, M. (2017). Hypermentalizing, attachment, and epistemic trust in adolescent BPD: clinical illustrations. *Personal. Disord.* 8, 172–182. doi: 10.1037/per0000161

Bowlby, J. (1969). Attachment. Attachment and Loss (1). New York, NY: Basic Books.

Bowlby, J. (1973). Separation: Anxiety & Anger. Attachment and Loss (2). London: Hogarth Press.

Bowlby, J. (1980). Loss: Sadness & Depression. Attachment and Loss (3). London: Hogarth Press.

Breazeal, C., Dautenhahn, K., and Kanda, T. (2016). *Social Robotics*. In Springer Handbook of Robotics (1935-1972). Cham: Springer.

Brosseau-Liard, P., Penney, D., and Poulin-Dubois, D. (2015). Theory of mind selectively predicts preschoolers' knowledge-based selective word learning. *Br. J. Dev. Psychol.* 33, 464–475. doi: 10.1111/bjdp.12107

Camisasca, E., Procaccia, R., Miragoli, S., Valtolina, G. G., and Di Blasio, P. (2017). Maternal mind-mindedness as a linking mechanism between childbirth-related posttraumatic stress symptoms and parenting stress. *Health Care Women Int.* 38, 593–612. doi: 10.1080/07399332.2017.1296840

Cangelosi, A., and Schlesinger, M. (2015). *Developmental Robotics: From Babies to Robots*. Cambridge, MA: MIT Press.

Cangelosi, A., and Schlesinger, M. (2018). From babies to robots: the contribution of developmental robotics to developmental psychology. *Child Dev. Perspect.* 12, 183–188. doi: 10.1111/cdep.12282

Carlson, S. M., and Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Dev.* 72, 1032–1053. doi: 10.1111/1467- 8624.00333

Chan, C. C. Y., and Tardif, T. (2013). Knowing better: the role of prior knowledge and culture in trust in testimony. *Dev. Psychol.* 49, 591–601. doi: 10.1037/ a0031336

Clément, F., Koenig, M., and Harris, P. (2004). The ontogenesis of trust. *Mind Lang.* 19, 360–379. doi: 10.1111/j.0268-1064.2004.00263.x

Corriveau, K., and Harris, P. L. (2009). Choosing your informant: weighing familiarity and recent accuracy. *Dev. Sci.* 12, 426–437. doi: 10.1111/j.1467-7687. 2008.00792.x

Corriveau, K. H., Harris, P. L., Meins, E., Fernyhough, C., Arnott, B., Elliott, L., et al. (2009). Young children's trust in their mother's claims: longitudinal links with attachment security in infancy. *Child Dev.* 80, 750–761. doi: 10.1111/j. 1467- 8624.2009.01295.x

Di Dio, C., Isernia, S., Ceolaro, C., Marchetti, A., and Massaro, D. (2018). Growing up thinking of God's beliefs: theory of mind and ontological knowledge. *Sage Open* 8, 1–14.

Di Dio, C., Manzi, F., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., et al. (2019). It does not matter who you are: fairness in pre-schoolers interacting with human and robotic partners. *Int. J. Soc. Robot.* 1, 1–15.

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., et al. (2020). Come i bambini pensano alla mente del robot: il ruolo dell'attaccamento e della Teoria della Mente nell'attribuzione di stati mentali ad un agente robotico. *Sist. Intell.* 1, 41–56.

Fonagy, P. (1998). Attachment theory approach to treatment of the difficult patient. *Bull. Menninger Clin.* 62, 147–169.

Fonagy, P., Luyten, P., Allison, E., and Campbell, C. (2019). Mentalizing, epistemic trust and the phenomenology of psychotherapy. *Psychopathology* 52, 94–103. doi: 10.1159/000501526

Fonagy, P., Redfern, S., and Charman, T. (1997). The relationship between belief- desire reasoning and a projective measure of attachment security (SAT). *Br. J. Dev. Psychol.* 15, 51–61. doi: 10.1111/j.2044-835x.1997.tb00724.x

Frye, D., Zelazo, P. D., and Palfai, T. (1995). Theory of mind and rule based reasoning. *Cogn. Dev.* 10, 483–527. doi: 10.1016/0885-2014(95)90 024-1

Fusaro, M., and Harris, P. L. (2008). Children assess informant reliability using bystanders' non-verbal cues. *Dev. Sci.* 11, 771–777. doi: 10.1111/j.1467-7687. 2008.00728.x

Gaudiello, I., Zibetti, E., Lefort, S., Chetouani, M., and Ivaldi, S. (2016). Trust as indicator of robot functional and social acceptance. An experimental study on user conformation to iCub answers. *Comput. Human Behav.* 61, 633–655. doi: 10.1016/j.chb.2016.03.057

Giovanelli, C., Di Dio, C., Lombardi, E., Tagini, A., Meins, E., Marchetti, A., et al. (2020). Exploring the relation between maternal mind-mindedness and children's symbolic play: a longitudinal study from 6 to 18 months. *Infancy* 25, 67–83. doi: 10.1111/infa.12317

Grover, S. L., Abid-Dupont, M.-A., Manville, C., and Hasel, M. C. (2019). Repairing broken trust between leaders and followers: how violation characteristics temper apologies. *J. Bus. Ethics* 155, 853–870. doi: 10.1007/s10551-017-3509-3

Hafford-Letchfield, T., and Spatcher, P. (2007). Getting to know you: social work students experiences of direct work with children in education settings. *Soc. Work Educ*. 26, 311–317. doi: 10.1080/02615470601049891

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., de Visser, E. J., and Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human- robot interaction. *Hum. Factors* 53, 517–527. doi: 10.1177/0018720811417254

Hansburg, H. G. (1972). Separation Problems of Displaced Children. Oxford: Stanwix House.

Harris, P. L. (2007). *Trust. Dev. Sci.* 10, 135–138.

Heerink, M. (2010). Assessing Acceptance of Assistive Social Robots by Aging Adults. Ph.D Dissertation, Amsterdam: University of Amsterdam.

Jones, G. R., and George, J. M. (1998). The experience and evolution of trust: implications for cooperation and teamwork. *Acad. Manage. Rev*. 23, 531–546. doi: 10.5465/amr.1998.926625

Kanda, T., Hirano, T., Eaton, D., and Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: a field trial. *Hum. Comput. Interaction* 19, 61–84. doi: 10.1207/s15327051hci1901%262_4

Kidd, C. D. (2003). *Sociable Robots: The Role of Presence and Task in Human-Robot Interaction.* Unpublished Master Thesis, Boston, MA: Massachusetts Institute of Technology.

Klagsbrun, M., and Bowlby, J. (1976). Responses to separation from parents: a clinical test for young children. *Br. J. Proj. Psychol. Pers. Study* 21, 7–27.

Koenig, M. A., Clément, F., and Harris, P. L. (2004). Trust in testimony: children's use of true and false statements. *Psychol. Sci.* 15, 694–698. doi: 10.1111/j.0956-7976.2004.00742.x

Koenig, M. A., and Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Dev.* 76, 1261–1277. doi: 10.1111/j.1467-8624.2005.00849.x

Korkman, M., Kirk, U., and Kemp, S. (2007). *NEPSY-II: A Developmental Neuropsychological Assessment.* San Antonio, TX: The Psychological Corporation.

Kuchenbrandt, D., and Eyssel, F. (2012). "The mental simulation of a human robot interaction: positive effects on attitudes and anxiety toward robots," *in Proceedings of the 21st IEEE International Symposium in Robot and Human Interactive Communication (RO-MAN 2012),* Paris.

Lahno, B. (2001). On the emotional character of trust. *Cultivat. Emot*. 4, 171–189.

Lecciso, F., Petrocchi, S., Liverta Sempio, O., and Marchetti, A. (2011). A contribution to a new tool of the relationships between affects and mentalization: the trust story. *Psicol. Clin. Dello Sviluppo* 15, 63–94.

Lewis, M., Sycara, K., and Walker, P. (2018). "*The role of trust in human-robot interaction"* In *Foundations of Trusted Autonomy* eds H. A., Abbass, J., Scholz, D. J., Reid. Cham: Springer, 135–159.

Liverta Sempio, O., Marchetti, A., and Lecciso, F. (2001). Il *SAT Famiglia e il SAT Scuola: Strumenti di Misura dell'Ansia da Separazione da Genitori ed Insegnanti*. Milan: Università Cattolica del Sacro Cuore, ISU.

Lombardi, E., Di Dio, C., Castelli, I., Massaro, D., and Marchetti, A. (2017). Prospective thinking and decision making in primary school age children. *Heliyon* 3:e00323. doi: 10.1016/j.heliyon.2017.e00323

Lombardi, E., Greco, S., Massaro, D., Schär, R., Manzi, F., Iannaccone, A., et al. (2018). Does a good argument make a good answer? Argumentative reconstruction of children's justifications in a second order false belief task. *Learn. Cult. Soc. Interact.* 18, 13–27. doi: 10.1016/j.lcsi.2018.02.001

Lucas, A. J., Lewis, C., Pala, F. C., Wong, K., and Berridge, D. (2013). Social- cognitive processes in preschoolers' selective trust: three cultures compared. *Dev. Psychol*. 49, 579–590. doi: 10.1037/a0029864

Luo, Y., Hétu, S., Lohrenz, T., Hula, A., Dayan, P., Ramey, S. L., et al., (2018). Early childhood investment impacts social decision-making four decades later. *Nat. Commun*. 9:4705. doi: 10.1038/s41467-018-07138-5

Manzi, F., Di Dio, C., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., et al. (2020). "Moral evaluation of human and robot interactions in Japanese preschoolers," *in ACM IUI Workshop Proceedings*, Cagliari.

Marchetti, A., Manzi, F., Itakura, S., and Massaro, D. (2018). Theory of mind and humanoid robots from a lifespan perspective. *Zeitschrift für Psychologie* 226, 98–109. doi: 10.1027/2151-2604/a000326

Marchetti, A., Miraglia, L., and Di Dio, C. (2020). Towards a socio-material approach to cognitive empathy in Autistic Spectrum Disorder. *Front. Psychol.* 10:2965. doi: 10.3389/fpsyg.2019.02965

Martini, M. C., Gonzalez, C. A., and Wiese, E. (2016). Seeing minds in others– Can agents with robotic appearance have human-like preferences. *PLoS One* 11:e0146310.doi:10.1371/journal.pone.0146310

Movellan, J., Eckhardt, M., Virnes, M., and Rodriguez, A. (2009). "Sociable robot improves toddler vocabulary skills," *in Proceedings of the 2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI),* (La Jolla, CA: IEEE).

Nass, C., and Moon, Y. (2000). Machines and mindlessness: social responses to computers. J. *Soc. Issues* 56, 81–103. doi: 10.1111/0022-4537.00153

Nass, C., Moon, Y., Fogg, B. J., and Reeves, B. (1995). Can computer personalities be human personalities? *Int. J. Human Comput. Stud*. 43, 223–239. doi: 10. 1006/ijhc.1995.1042

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., and Itakura, S. (2013a). Infants understand the referential nature of human gaze but not robot gaze. *J. Exp. Child Psychol*. 116, 86–95. doi: 10.1016/j.jecp.2013.02.007

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., and Itakura, S. (2013b). The power of human gaze on infant learning. *Cognition* 128, 127–133. doi: 10.1016/ j.cognition.2013.03.011

Pasquini, E. S., Corriveau, K. H., Koenig, M., and Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Dev. Psychol*. 43, 1216–1226. doi: 10.1037/0012-1649.43.5.1216

Perner, J., and Wimmer, H. (1985). "John thinks that Mary thinks that. . . " attribution of second-order beliefs by 5-to 10-year-old children. J. *Exp. Child Psychol*. 39, 437–471. doi: 10.1016/0022-0965(85)90051-7

Petrocchi, S., Rotenberg, K. J., Levante, A., and Lecciso, F. (2018). Children's trust in social workers: scale development and relations to Children's engagement with social workers. *Child Fam. Soc. Work* 23, 239–247. doi: 10.1111/cfs. 12410

Premack, D., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind. *Behav. Brain Sci.* 1, 515–526. doi: 10.1017/s0140525x00076512

Rizzato, M., Di Dio, C., Fasano, F., Gilli, G., Marchetti, A., and Sensidoni, A. (2016). Is food desirability affected by social interaction? *Food Qual. Prefer*. 50, 109–116. doi: 10.1016/j.foodqual.2016.02.005

Rotenberg, K. J. (2010). "*The conceptualization of interpersonal trust: a basis, domain, and target framework*," in *Interpersonal Trust During Childhood and Adolescence*, ed. K. J. Rotenberg (New York, NY: Cambridge University Press), 2–27.

Rotenberg, K. J., Petrocchi, S., Lecciso, F., and Marchetti, A. (2015). The relation between children's trust beliefs and theory of mind abilities. *Infant Child Dev*. 24, 206–214. doi: 10.1002/icd.1891

Sharp, C., Pane, H., Ha, C., Venta, A., Patel, A. B., Sturek, J., et al. (2011). Theory of mind and emotion regulation difficulties in adolescents with borderline traits. *J. Am. Acad. Child Adolesc. Psychiatry*. 50, 563–573. doi: 10.1016/j.jaac.2011.01.017

Shaw-Garlock, G. (2009). Looking forward to sociable robots. *Int. J. Soc. Robot*. 1, 249–260. doi: 10.1007/s12369-009-0021-7

Short, E., Hart, J., Vu, M., and Scassellati, B. (2010). "No fair!! an interaction with a cheating robot," *in proceedings of the 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI),* (Osaka: IEEE).

Tanaka, F., and Matsuzoe, S. (2012). Children teach a care-receiving robot to promote their learning: field experiments in a classroom for vocabulary learning. *J. Hum. Robot Interact*. 1, 78–95. doi: 10.5898/jhri.1.1.tanaka

Van Reet, J. (2015). Conflict inhibitory control facilitates pretense quality in young preschoolers. *J. Cogn. Dev.* 16, 333–350. doi: 10.1080/15248372.2013.83 3924

Van Reet, J., Green, K. F., and Sobel, D. M. (2015). Preschoolers' theory-of-mind knowledge influences whom they trust about others' theories of mind. *J. Cogn. Dev*. 16, 471–491. doi: 10.1080/15248372.2014.892875

Vinanzi, S., Patacchiola, M., Chella, A., and Cangelosi, A. (2019). Would a robot trust you? Developmental robotics model of trust and theory of mind. Philos. *Trans. R. Soc. B* 374:20180032. doi: 10.1098/rstb.2018. 0032

Vogt, P., De Haas, M., De Jong, C., Baxter, P., and Krahmer, E. (2017). Child-robot interactions for second language tutoring to preschool children. *Front. Hum. Neurosci*. 11:73. doi: 10.3389/fnhum.2017.00073

Wellman, H. M., Cross, D., and Watson, J. (2001). Meta-analysis of theory-of- mind development: the truth about false belief. *Child Dev*. 72, 655–684. doi: 10.1111/1467- 8624.00304

Wiese, E., Metta, G., and Wykowska, A. (2017). Robots as intentional agents: using neuroscientific methods to make robots appear more social. *Front. Psychol*. 8:1663. doi: 10.3389/fpsyg.2017.01663

Wimmer, H., and Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128. doi: 10.1016/0010-0277(83)90004-5

Wykowska, A., Chaminade, T., and Cheng, G. (2016). Embodied artificial agents for understanding human social cognition. *Philos. Trans. R. Soc. B Biol. Sci*. 371:20150375. doi: 10.1098/rstb.2015.0375

Yagoda, R. E., and Gillan, D. J. (2012). You want me to trust a ROBOT? The development of a human robot interaction trust scale. *Int. J. Soc. Robot*. 4, 235–248. doi: 10.1007/s12369-012-0144-0

Yang, Y., Tian, Y., Fang, J., Lu, H., Wei, K., and Yi, L. (2017). Trust and deception in children with autism spectrum disorders: a social learning perspective. *J. Autism Dev. Disord*. 47, 615–625. doi: 10.1007/s10803-016- 2983-2

Zanatto, D., Patacchiola, M., Cangelosi, A., and Goslin, J. (2019). Generalisation of Anthropomorphic stereotype. *Int. J. Soc. Robot*. 2019, 1–10.

Zelazo, P. D. (2006). The dimensional change card sort (DCCS): a method of assessing executive function in children. *Nat. Protoc*. 1, 297–301. doi: 10.1038/ nprot.2006.46

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

CHAPTER 3

Can a robot lie? The role of false belief and intentionality understanding in children aged 5 and 6 years

G. Peretti,[1]* F. Manzi,[1] C. Di Dio,[1] A. Cangelosi,[2] P. L. Harris,[3] D. Massaro,[1] A. Marchetti[1]

[1] Research Unit on Theory of Mind, Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy

[2] School of Engineering, University of Manchester, UK

[3] Graduate School of Education, Harvard University, Cambridge, Massachusetts, USA

**Corresponding author*

Under review on Cognitive Development

Keywords

Theory of Mind, intentionality understanding, social robot, children

Abstract

Human relationships involve mutual understanding of behaviour and underlying intentions. In social communication, understanding intentionality implies knowing whether the other person is lying or mistaken, i.e., intentionally or unintentionally making false statements. This understanding is crucial in human relationships and even more important when relationships include non-human agents, such as robots. The ability to recognize an intention to lie or not represents a significant step in child development, which is also characterized by the acquisition of a Theory of Mind (ToM). The present study explored the ability of 5-and 6-year-old children to discriminate between intentionally and unintentionally false statements made by a human and NAO robot. In addition to supporting previous findings regarding ToM abilities, these results showed that children with relevant ToM comprehension are less judgmental of a robot's lies than of a human's. By implication, young children tend to view a robot's behaviour as fundamentally unintentional.

Keywords

Theory of Mind, intentionality understanding, social robot, children

Introduction

The ability to lie and to recognize behaviour driven by an intention to deceive represents a significant developmental stage in children, which is also characterized by the acquisition of the ability to distinguish between one's own beliefs and those of others, i.e., a Theory of Mind (Wellman, 2020). Classic research on the development of the ability to lie highlighted children's moral realism (Piaget, 1932): their tendency to judge facts and responsibilities exclusively according to their consequences. Thus, children under the age of 6 tend to attribute the quality of a lie to behaviours that have received a punishment, regardless of whether they were driven by an intention to harm (Bussey, 1992). Despite the helpful theoretical starting point proposed by Piaget, the years that followed saw a significant decrease in interest in this area of research. It is only from the 80s onwards that new research emerged aimed at investigating how, at different ages, children recognize and differentiate between intentionally false (lies) and unintentionally false (mistaken) statements. This type of research increased along with interest in understanding the child's ability to attribute intentionality to the actions of others (Peterson et al.,1983; Wimmer et al., 1984, 1985). For example, Wimmer et al. (1984) showed that recognition of a lie also depends on the child's ability to attribute the intention to deceive to the liar. This ability appears to emerge before the age of 6, close to the end of the period defined by Piaget as moral realism. Indeed, on studies of 4- to 10-year-old children's judgments of true but intentionally deceptive statements and unintentionally false statements, it was found that as early as age 4 children become more sensitive to the intentions of the informant, thus tending to reward according to the truth value of the statement (Wimmer et al., 1984). From this perspective, the acquisition of false belief understanding and a greater level of sophistication in Theory of Mind (ToM; Astington et al., 2002; Baron-Cohen et al., 1985; Castelli et al., 2014; Ding et al., 2015; Fu et al., 2014; Lee, 2013; Leslie et al., 2006; Ma et al., 2015; Massaro et al., 2013; Polak & Harris, 1999; Premack & Woodruff, 1978; Sullivan et al., 1995; Talwar & Lee, 2002, 2008; Wellman, 2020; Wimmer & Perner, 1983) is thought to refine children's ability to attribute intentionality.

There is also interesting evidence regarding the meaning and use that children make of the word "lie" that children acquire between the ages of 4 and 7 alongside their mastery of other mental verbs (Wimmer et al., 1984). In support, several authors (Bussey, 1999; Chandler et al.,1989; Evans & Lee, 2013; Lewis et al., 1989; Peskin, 1992; Polak & Harris, 1999; Talwar & Lee, 2002; Talwar et al., 2006; Wilson et al., 2003) showed substantial differences in the understanding of lying between 3- and 4-year-olds. From the age of 3, children understand that lying to conceal a transgression represents an inappropriate behaviour (Lee, 2013; Lyon & Dorado, 2008; Talwar & Lee, 2008), and can discriminate between intentionally false and unintentionally wrong actions (Siegal & Peterson, 1996, 1998). By age four, children can also distinguish between intentionally false and unintentionally mistaken statements (Taylor et al., 2003), but they are not yet able to differentiate between intentional and unintentional behaviour accurately and thus cannot clearly distinguish between a lie and a mistake. Supporting these observations, Gilli et al. (2001) found that from 3 to 5 years of age children were increasingly able to correctly distinguish between a lie and a mistake and that most 4- and 5-year-olds, but not 3-year-olds, evaluate intentional false statements negatively, as opposed to unintentional mistaken statements. On the other hand, Adenzato and Bucciarelli (2008) subsequently found less accuracy in the differentiation of lies and mistakes among 4- and 5-year-olds, possibly due to these children's limited attention and verbal skills and poor mindreading abilities. However, between the ages of 7 and 11 children can distinguish and correctly name an act based on an intention to make a false statement compared to an unintentionally mistaken statement (Adenzato & Bucciarelli, 2008; Burton & Strichartz, 1992; Gilli et al., 1998; see Perner, 1997 for a review; Strichartz & Burton, 1990; Surian et al., 2002). It has also been shown that the ability to lie in children between 3 and 8 years of age is related to the development of inhibitory control and working memory, which are classically assessed by means of the Stroop task (Carlson & Moses, 2001; Talwar & Lee, 2008).

Deception in social robots

Interpersonal relationships are underpinned by fundamental dynamics such as mutual trust. To be effective, the parties involved in a relationship must be able to trust each other, especially if the other is an entity whose behavioural processes are opaque (Di Dio et al., 2020a, b; Marchetti et al., 2020a). As I have seen, understanding whether the other is lying or simply mistaken requires a grasp of the intention behind the behaviour and a recognition that behaviour can be instigated by mental states different from our own (Perner & Wimmer, 1985; Premack & Woodruff, 1978; for a review, see Wellman et al., 2001; Wimmer & Perner, 1983). Importantly, the attribution of a lie or a mistake to the other presupposes that he or she has intentions of his or her own. From about 1 year of age, very young children interact with humanoid social robots (HSRs) by treating them (in brief and very controlled interactions) like intentional agents (Manzi et al., 2020b; Okumura et al., 2013a). Anthropomorphic robots displaying human-like physical features, the most important of which is a face, as well as ecological settings promote this inclination (Manzi et al., 2020b; Okumura et al., 2013b). Nevertheless, as children acquire sociocognitive skills, they increasingly attribute limited mental states to humanoid robots such as NAO or Robovie. In particular, starting from about 7 years of age - that is, when children acquire ToM skills that are qualitatively more sophisticated - the gap between the attribution of mental states to humans and robots becomes more and more substantial

(Di Dio et al., 2018, 2019, 2020a, 2020b; Manzi et al., 2020c, 2020a; Marchetti et al., 2018). Based on these assumptions, how do children conceive of the behaviour of a robot making intentionally false claims? Do they view it in the same way as a lie by a human? No empirical data are currently available regarding how children discriminate between an "intentionally" false and an "unintentionally" mistaken statement made by a social robot.

From a conceptual point of view, the research that comes closest to the construct of lie and mistake is a study that investigated trust-building between children aged from 3 to 9 years and a NAO robot (Di Dio et al., 2020b). In this study, a situation that can be characterized as "deceptive" was created in that the child's partner (either a human or a robot) began to offer incorrect/false information after having demonstrated some epistemic trustworthiness. In this setting, it was possible to observe the dynamics of trust restoration after trust loss with respect to both partners (human and robot).

As expected, trust in the partner, independently of whether a human or robot, was substantially affected by the children's ToM development. When, after establishing a trusting relationship with the partner, trust was lost due to systematic mistakes by the partner (deceptive situation), the child - regardless of age - took the same amount of time to regain trust in the robot as it did in the human. Curiously, however, children who had positive human relationships as indexed by a secure attachment to their primary caregiver, and by being open to new relationships (Chan & Tardif, 2013; Clément et al., 2004; Harris, 2007; Koenig et al., 2004; Vinanzi et al., 2019), were also those who tended to regain trust in the robot more quickly than in the human. By implication, the robot's actions were perceived as unintentional, and its "misleading/deceptive" behaviours were more accepted, effectively making the robot a more forgivable partner. Accordingly, one may wonder whether the same "tolerance" characterizes children's evaluation of a robot that makes a false statement, whether more or less intentional.

Study aims and predictions

The present study investigates the ability of 5- and 6-year-old children to discriminate between intentionally and unintentionally false statements (i.e., between lies and mistakes) of a human as compared to a humanoid NAO robot. Children were presented with lie-mistake videos in which the character (a human or NAO robot) made false statements either intentionally or unintentionally. Children were also presented with the Attribution of Mental States questionnaire (AMS; Di Dio et al., 2018, 2019, 2020a, 2020b; Manzi et al., 2017; 2020c) assessing mental state attributions to the two characters, notably two first-order Theory-of-Mind tasks, and an intentionality task evaluating children's ability to understand intentional actions. Also, potential correlates were assessed via the administration of executive function tasks and two sub-tests of NEPSY II which evaluates language abilities. In line with previous findings (e.g., Gilli et al., 2001), Differences in the ability to discriminate between intentionally and unintentionally false statements were expected to be found as a function of children's ToM skills. More specifically, children with first-order ToM understanding were expected to differentiate more accurately between the two types of false statements, regardless of the observed agent. Furthermore, in line with previous findings on trust (Di Dio, 2020a, 2020b), a higher tolerance for intentionally false statements by the robot than by the human was hypothesized.

Method

Participants

One hundred and twelve (112) Italian children were recruited from three kindergartens and four primary schools in the local area of Milan. Data were analyzed on 87 children (see data analysis below) aged 5 years (N = 36, 22 female; M= 65.6 months, SD = 2.56), and 6 years (N = 51, 26 female; M = 74.9 months, SD= 3.21). Parents received a detailed explanation of the experimental procedure, the tasks, and the materials used during the experiment and gave their written consent. The study was approved by the Local Ethics Committee (Università Cattolica del Sacro Cuore, Milan).

Task and Procedure

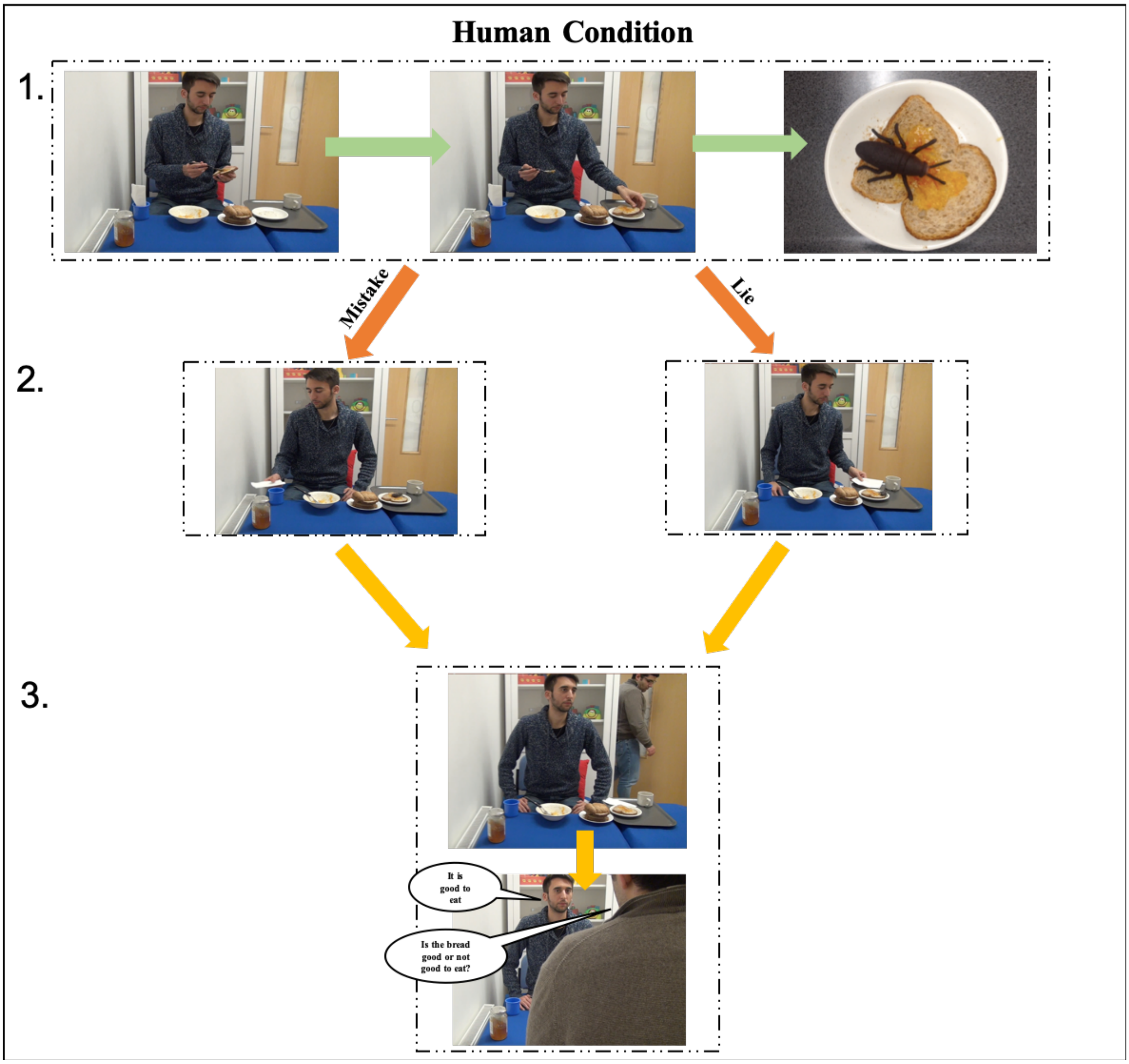
The children were given the following tests: Lie-Mistake videos; First-Order Theory-of-Mind tasks; Executive Function tasks (Dimensional Change Card Sort and Day-Night Stroop); Intentionality Task; Attribution of Mental States questionnaire (AMS); two sub-tests of NEPSY II (Comprehension of Instructions and Word Generation).

Lie-Mistake videos

Each child watched four videos in which the character (a girl or a boy, matched to the child's gender, or a NAO robot) intentionally or unintentionally made a false statement, i.e., lied or made a mistake, respectively (for an example, see Figure 1). The videos were an adapted version taken from a previous work by Siegal and Peterson (1996, 1998) and Gilli et al. (2001). In the lie situation, the character (H or R) spread jam on a slice of bread which was then put on a plate. The character then turned to pick up a napkin and, when turning back to the plate, saw a cockroach on the bread, which then left the scene. A different human character then entered in the room and asked: “Is the bread good or not good to eat?". The girl/boy/robot answered: “It is good to eat”, thereby intentionally making a false statement (lie). In the mistake situation, the character (H or R) who prepared the bread did not notice the cockroach on the bread because his/her back was turned. To the question. “Is the bread good to eat?” the girl/boy/robot answered as follows: “It is good to eat”, thereby unintentionally making a false statement (mistake). The main sequences are shown in Figure 1. At the end of the video presentations, the child was asked, with respect to both the lie and mistake situations, the following questions: 1) “Does the this boy/girl/robot know that the bread is not good to eat?” (control question); 2) “Did this boy/girl/robot tell a lie or make a mistake?”, randomizing the order of mention of the lie and mistake options; 3) “Is this boy/girl/robot naughty or not naughty?” 4) “When the boy/girl who had been told that the bread was good to eat found out that the boy/girl/robot told a lie or made a mistake, was he/she angry or not angry with him/her?”

These questions probed the following components (numbers refers to the above numbered questions): 2) Recognition of an intentionally or unintentionally false statement; a score of 1 was given for the correct answer and 0 for the incorrect answer; 3) Moral evaluation of the behaviour of the character who had intentionally or unintentionally made a false statement; a score of 1 was assigned to the correct answer (lie = naughty; mistake = not naughty) and 0 to the incorrect answer (lie = not naughty; mistake = naughty). 4) Attribution of feeling experienced by the character who found out that s/he had been (either intentionally or unintentionally) misled, a score of 1 was assigned to the correct answer (lie = angry; mistake = not angry) and 0 for the incorrect answer (lie = not angry; mistake = angry). Scores for questions 2 and 3 were subsequently combined to create the variable “Coherence” (range = 0-2), indicating the coherence between the recognition of the statement given by the subject (lie or mistake) and the behaviour attributed (naughty or not naughty). The Coherence variable is from now on referred to as Moral Reasoning since it reflects the appropriateness of the child's moral evaluation of the condition observed.

The Lie-Mistake videos were preceded by an informational video showing the characters in the stories (human – H, or robot – R) watching a video about food contamination by a cockroach, represented as a dirty insect landing on a slice of bread. This informational video was shown to ensure that children knew that the characters were aware of the negative effects of eating food contaminated by a dirty insect. About 87% of children responded correctly to the question on food contamination for all four lie-mistake situations. For those who did not respond correctly (note that children made errors with respect to no more than one video) the experimenter verbally gave the correct answer and explicitly justify it. Before the experimental videos, all children were equally aware of the negative effects of food contamination.


Human Condition
1.
2.
3.
Mistake
Lie
It is good to eat
Is the bread good or not good to eat?

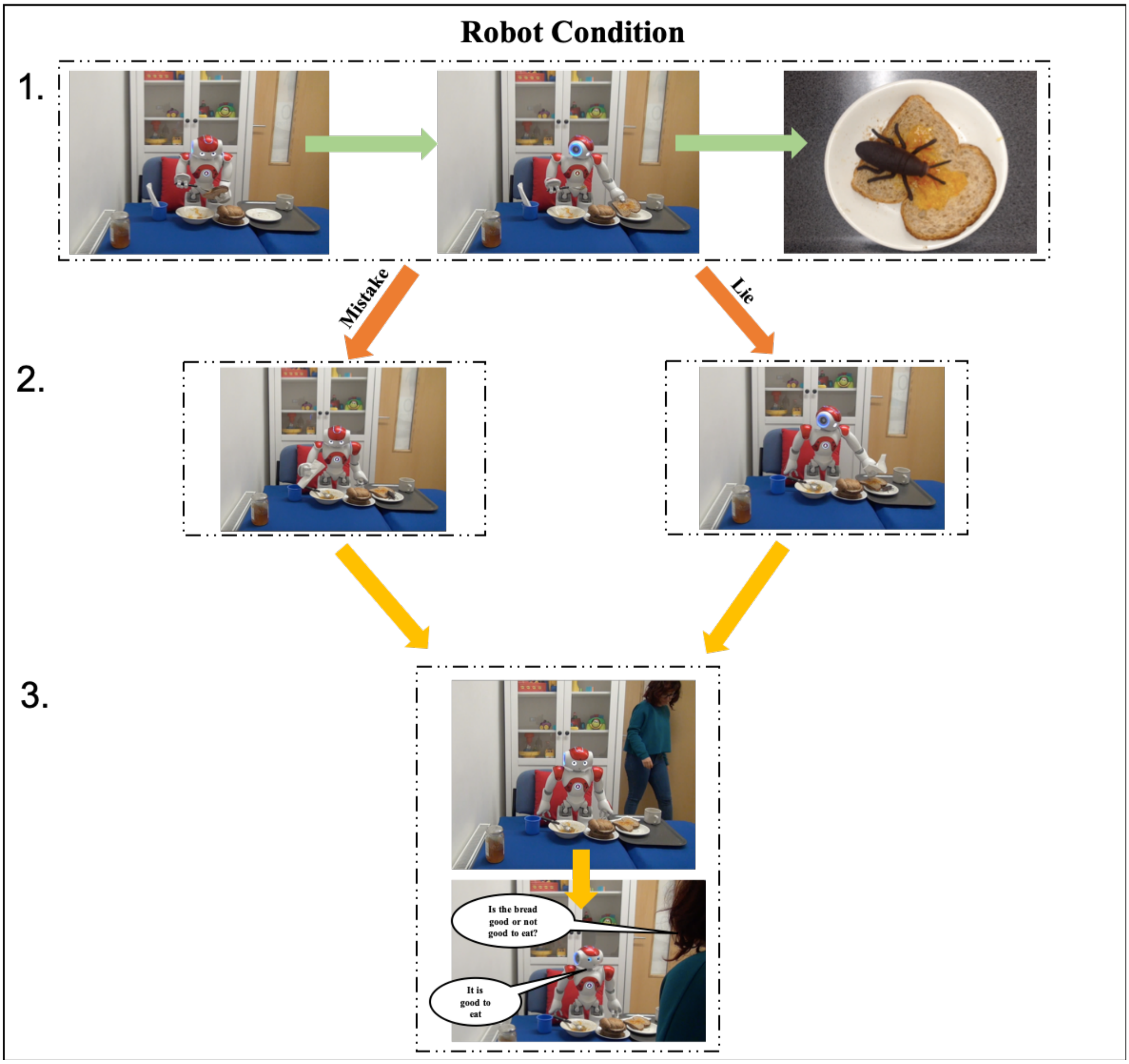


Figure 1. Main sequences of the Lie-Mistake videos for the human and robot conditions. 1) The human (male or female) and the robot are preparing the bread with jam; 2) Lie and Mistake situations, in which the character does not see (mistake) or sees (lie) the beetle on the bread; 3) concluding scene, in which a person enters the room and asks if the bread is good or not good to eat.

Theory of Mind (ToM)

To evaluate children's first-order ToM and their understanding of false beliefs, the following tasks were administered: the Unexpected Transfer task (Wimmer & Perner, 1983) and the Unexpected Content task (Perner & Wimmer, 1985). For both tasks, a story is read to the child. In the Unexpected Transfer task, the child is shown two characters (A and B), a ball, a box and a basket, all present in the same room. Character A puts the ball in a box and leaves the room. Character B, after character A leaves the room, takes the ball from the box and puts it in the basket. Thus, the experimenter asks the child a series of questions about what happened, as follows: (i) false belief question "Where is the first place where A will look for his ball?"; (ii) memory question "Where did character A put the ball before leaving the room?"; (iii) reality question "Where is the ball really located?".

In the Unexpected Content task, the child is shown a wax crayons box and asked (i) a reality question "Look at this, what's in here? Let's open the box and look inside". The experimenter then opens the box and shows that inside there is a puppet. Then puts the puppet back in the box and asks the child (ii) a question about the child's belief "What's in the box?". Finally, the child is shown a second character X, and asked (iii) the question about first-order false belief, i.e. "When X sees the box for the first time, what will he think is inside?", and (iv) a memory question "Before opening the box what had you told me was inside it?". If the child does not answer, the question can be repeated. Provided the control questions (reality and memory) were answered correctly, the test question relating to false beliefs was scored 1 if answered correctly and 0 if answered not correctly. For the correlational analysis (see below), a conservative approach was adopted: children's scores on the two ToM tasks were combined. More specifically, children who passed both tests scored 1, whereas those who failed either 1 or both tasks scored 0. Given this scoring system, 59% of children passed both first-order false belief tasks, and 41% of children failed.

Executive Function skills (EF)

To evaluate the development of Executive Function skills (FE), the following tasks were administered: Dimensional Change Card Sort (DCCS; Traverso & De Franchis, 2014; Zelazo, 2006) to assess the capacity to switch responses, and Day-Night Stroop (DNS; Gerstadt et al., 1994) to evaluate the ability to inhibit distractor stimuli or irrelevant information (Carlson & Moses, 2001). In DCCS the child has to classify a series of cards, representing red rabbits and blue boats, in three different test phases with an increasing degree of complexity. The child is required to classify the cards shown to him/her by placing them in two boxes with target cards depicting a blue rabbit and a red boat. In the first phase, the child should order the cards by colour, in the second phase by shape and in the third phase by colour if the card has a black border and by shape if there is no border. Each test is given a score of 1 if passed and 0 if not passed. The total score is the sum of the three tests (range 0-3).

The DNS involves the administration of 16 cards, divided into two series. In the first, the cards depict a chessboard and an X, and in the second series, the sun and the moon In the first phase, the child is to answer "day" when s/he is shown the card with an X and "night" when s/he is shown the card with the chessboard. This phase represents the control test. In the second phase, the answer “day” corresponds to the card with the moon and the answer “night” to the card with the sun. The total score is calculated as the difference between the second and the first test (range –16-0).

Intentionality Understanding Task

To evaluate children’s comprehension of intentionality, the Intentionality Understanding (IU) task (Astington, 1998) was administered. The task measures children’s comprehension of intentionality from a linguistic perspective. Children are asked to closely observe two pictures, A and B presented at the same time, for a total of seven trials except for the information task, and then a specific question about the pictures. The questions refer to intentional states underpinning the action of the depicted subject. For example, “Which little girl wants to play with the sand?”. In this example, picture A depicted a little girl who is about to go and play with the sand (correct choice), while picture B depicted a little girl who is already playing. Correct answers were scored 1, while wrong answers were scored 0. The total score represents the sum of all answers (range 0-7).

Attribution of Mental States questionnaire (AMS)

The Attribution of Mental States (AMS) questionnaire measures the mental states attributed by the child to specific characters depicted by an image, here either a human or NAO robot (the human character was matched to the child’s gender). The scale is an ad-hoc questionnaire that was based on the questionnaire described in Martini and colleagues (2016) and aims to assess children’s perception of the partner (H or R) in mental terms. AMS has been used in previous works with children (Di Dio et al., 2019, 2020a, b; Manzi et al., 2017, 2020c). Children are asked 25 questions grouped into five different state-categories: Emotions, Desires and Intentions, Imaginative, Epistemic, and Sensory. For example, within the Epistemic category, a question would be “Do you think that s/he/it can decide?”. The child was invited to respond “Yes” or “No” to each question. If the answer was “Yes”, then the experimenter asked a follow up question: “How much? A little bit or very much?”, resulting in a 3-point scale. So, for example, to the question: “Do you think that s/he/it can understand?”, the range of answers could be “No” (0), “Yes, a little bit” (1), or “Yes, very much”

(2). The total score was the sum of all answers (range = 0-50); the five partial scores were the sum of the answers within each category (score range = 0-10).

Language abilities

To evaluate language abilities, two sub-tests of NEPSY II were used: 1) Comprehension of Instructions (age 3–16 years), assessing the ability to perceive, process and execute verbal instructions with an increasing level of semantic complexity. The total score is the sum of all correct answers (range = 0-33). 2) Word Generation (age 3-16 years) assessing the child's ability to generate words within specific semantic and phonological categories. In this study, for the Word Generation test, the child was given the semantic category (item 1 concerning animals; item 2 concerning foods that can be eaten or drunk) and asked to produce as many words as possible in 60 seconds. The final score is given by the sum of the number of words not repeated for items 1 and 2; the score obtained refers to the total semantic score (NEPSY II; Korkman et al., 2007).

Procedure

Children were examined individually in a room in their kindergarten or primary school; the environment was quiet and suitable for testing. Children were assessed in two different sessions on different days. The second assessment took place about 10 days after the first. The two assessment phases lasted about 30-35 minutes. Each session was conducted by the researchers both in the morning and in the afternoon, always during school time. The following tests were administered in the first session: Lie-Mistake videos, a first-order ToM test, two sub-tests of NEPSY II and the AMS questionnaire. During the second session, the following tests were administered: Lie-Mistake videos, a different first-order ToM test, two Executive Functions tests, and the AMS questionnaire, which was always administered at the end of the assessment phase, once asking for attribution of mental stated to the human and once to the robot. The administration order of the AMS was randomized between sessions and across participants. In both sessions, the two Lie-Mistake videos were presented as the first task. As for the AMS, the presentation order of the Lie-Mistake videos (one involving the lie situation and one involving the mistake situation) and the two first-order ToM tasks (Unexpected Transfer and Unexpected Content) were randomized between sessions and across participants. In the first session, before the trials were administered, the researcher spent a few minutes with the tested child to become familiar to her/him.

Data analysis

Twenty-five children were removed from the data analysis because they did not complete the protocol (e.g., were not available for the second assessment or struggled to complete a whole session). In total, data for 87 children were analyzed

To assess children's recognition of an intentionally or unintentionally false statement (lie, mistake), their associated moral evaluation (naughty, not naughty), and their understanding of emotional impact on the deceived character in the lie-mistake tasks (angry, not angry), non-parametric analyses (Sign-test; McNemars) were conducted. Additionally, a repeated measures GLM analysis was carried out to evaluate potential *coherence* between children's recognition of a lie or mistake and the related moral evaluation (i.e., lie = naughty; mistake = not naughty) as a function of agent (human, robot) and ToM group (failed, passed). For the GLM analysis, the Greenhouse-Geisser correction was used for violations of Mauchly's Test of Sphericity, $p<.05$. All post-hoc comparisons were Bonferroni corrected.

Finally, to investigate factors associated with children's ability to make an appropriate moral evaluation of the characters in the lie-mistake task, Pearson's correlation analyses were carried out between the variable *coherence* in the lie-mistake task and language abilities, executive function skills, intentionality understanding, ToM acquisition, and the attribution of mental states to the human and robot (AMS).

Results

Lie-mistake video task

As shown in Table 1, test distributions revealed that children who had not acquired a first-order ToM competence failed to: (i) reliably discriminate between lies and mistakes; (ii) make an appropriate moral evaluation; and (iii) attribute the appropriate emotion to the misled character (lie-angry, mistake-not angry). This was independent of whether the scenario presented a misleading human or robot. On the other hand, children who had acquired a first-order ToM competence were able to correctly discriminate at above chance levels between both a lie and mistake for the human, and only the mistake for the robot. Also, children with ToM skills were able to correctly reason about the naughtiness of the action carried out by both the human and the robot, but only for the mistake situation, where behaviour was correctly evaluated as not naughty above chance level. On the other hand, the lie was judged naughty at a random level also by the ToM group. The attribution of emotion to the misled character in response to the intentional or unintentional statement was random for both human and robot.

These data show that, while no-ToM children responded randomly in all conditions, ToM children, who were generally more accurate, struggled to attribute a lie to the robot. In support of this conclusion, direct human-robot comparisons (McNemar test) showed that children in the ToM group recognized the lie better in humans than in robots (65% vs 45%, $p<.01$) and made a more appropriate moral evaluation in the lie situation (lie-naughty) for humans than for robots (61% vs 43%, $p<.01$); that is, when judging the robot the lie was sometimes judged to be a mistake and therefore not judged so severely in moral terms.

Table 1. Percentage of children who correctly responded to the lie-mistake questions in the human (H) and robot (R) conditions. ToM_F: children who failed the first-level ToM tasks (N=36); ToM_P: children who passed the first-level ToM tasks (N=51). *p<.05; **p<.01, two-tailed binomial test, assuming a 50% chance level of success. In bold are the significant differences between ToM_F and ToM_P, chi-square asymptotic significance (two-tailed).

	Lie situation				**Mistake situation**			
	H		**R**		**H**		**R**	
Question Type	ToM_F	ToM_P	ToM_F	ToM_P	ToM_F	ToM_P	ToM_F	ToM_P
1. Labelling lie or mistake correctly (recognition)	58	65*	50	45	50	69**	**53**	**75****
2. Naughty lie/not naughty mistake (moral evaluation of behaviour)	**33**	**61**	42	43	56	71**	64	69**
3. Angry lie, not angry mistake (emotion)	56	63	56	53	47	57	47	55

The above conclusions should be reflected in greater *coherence* between lie-mistake recognition and the moral evaluation of behaviour for humans as compared to the robot, which was often judged as not naughty in the lie situation. Also, greater coherence should be found in the mistake scenario compared to the lie scenario, in which both agents were correctly judged as not naughty. This was assessed through a GLM analysis that included 2 levels of situation (lie, mistake), two levels of agent (human, robot) as within-subject factors, and ToM group (failed, passed) as a between-subject factor. The dependent variable was the level of consistency between lie-mistake recognition and moral evaluation of the deceiver's behaviour (score range = 0-2).

The results showed a main effect of ToM group, $F(1,85) = 242$, $p < .05$, *partial*-$\eta^2 = .15$, $\delta = .97$, indicating, as expected, that children who passed the ToM tasks were generally more coherent than those who failed. There was also a main effect of condition (mistake > lie), $F(1,85) = 4.23$, $p < .05$, *partial*-$\eta^2 = .05$, $\delta = .53$, and an interaction of agent*ToM group, $F(1,85) = 4.99$, $p < .05$, *partial*-

$\eta^2 = .06$, $\delta = .6$. As predicted, the main effect revealed that children were more coherent in the mistake as compared to the lie situation. The interaction further outlined that children with a ToM competence, who were generally more coherent in both lie-mistake situations than children who failed the ToM tasks, $M_{diff} = .26$, *SE* =. 07, $p<.001$, were particularly coherent with respect to the human rather than the robot condition, $M_{diff} = .1$, *SE* = .05, $p<.05$. This interaction is shown in **Figure 2**.

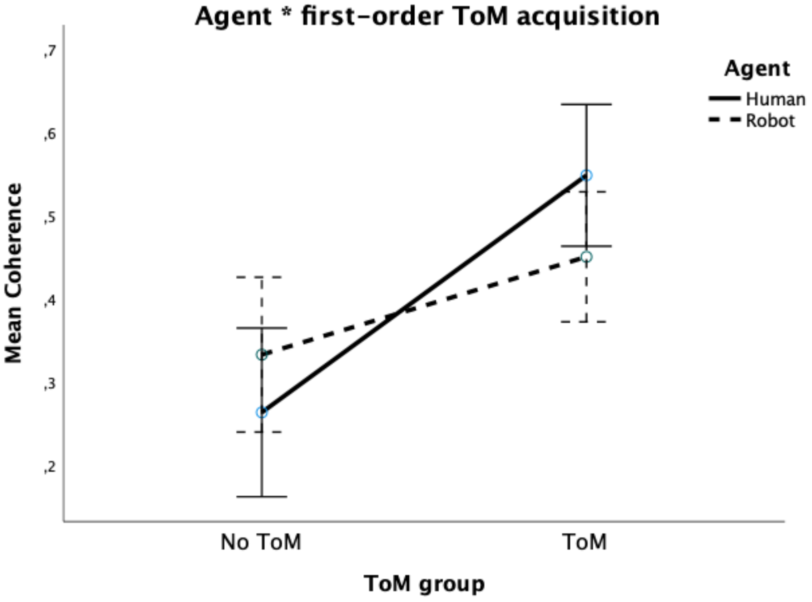


Figure 2. Line graph showing differences between children who had or had not acquired a first-level ToM competence in coherence between lie-mistake recognition and moral evaluation (averaged across the lie and mistake situations) with respect to the human and robot. The error bars represent the standard error of the mean. Significant differences were found between ToM and no ToM children, particularly in the human condition.

Correlations

To further investigate factors potentially associated with children's ability to make an appropriate moral evaluation of the characters' statements in the lie-mistake tasks, correlation analyses were carried out between the variable *coherence* in the lie-mistake task (i.e., lie-naughty; mistake-not naughty) and language abilities, executive function skills, intentionality understanding, and ToM acquisition. As shown in **Table 2**, the correlation analysis revealed that for children who failed the ToM tasks, coherence in the mistake situation for the human positively correlated with comprehension abilities as well as the ability to attribute intentionality. Also, for children with ToM competence, a positive correlation was found between intentionality understanding and coherence in the mistake situation in the robot condition. No associations were found with respect to executive function skills.

Table 2: Overall Correlations. Pearson's correlations for coherence between the recognition of an intentionally or unintentionally false statement and moral evaluation of behaviour, and language abilities (Verbal Comprehension, Word Generation), executive function skills (Day-Night Stroop - DNS, Dimensional Change Card Sort - DCCS), and Intentionality understanding, for the two ToM groups. *Correlation is significant at the 0.05 level (two-tailed); ** correlation is significant at the 0.01 level (two-tailed).

	H_Lie_ Coherence		**R_Lie_ Coherence**		**H_Mistake_ Coherence**		**R_Mistake_ Coherence**	
	No_ToM	ToM	No_ToM	ToM	No_ToM	ToM	No_ToM	ToM
Verbal Comprehension	-0.17	0.13	-0.11	-0.08	**.467****	0.12	0.13	0.16
Word Generation	-0.32	-0.23	-0.07	-0.07	0.28	-.22	-0.01	-0.08
DNS	-0.10	-0.12	0.18	0.04	-0.05	0.05	-0.13	0.09
DCCS	-0.02	-0.03	0.24	-0.03	0.21	0.13	0.14	0.00
Intentionality Understanding	-0.07	-0.09	0.03	-0.27	.**454****	0.25	0.01	**.303***
N	36	51	36	51	36	51	36	51

Attribution of Mental States

To evaluate differences in the attribution of mental states (AMS) between human and robots by children who had or had not acquired first-order ToM, a GLM analysis showed a main effect of agent (H>R), $F(1,85) = 51.025$, $p = .000$, $_{partial}\eta^2 = .375$, $\delta = 1$, and an interaction of agent*states, $F(4,85) = 14.821$, $p = .000$, $_{partial}\eta^2 = .148$, $\delta = 1$. No differences were found between ToM groups ($p>.05$). These data indicate that although the robot was generally given lower scores than the human for all state dimensions, a significant difference was only observed for sensory attribution $M_{diff} = .1$, $SE=.05$, $p<.05$, (**Figure 3**)

Finally, when correlating the lie-mistake *coherence* scores for the human and robot with the relative attribution of mental states (AMS), a positive correlation was found between children's coherence in the mistake situation (mistake-not naughty) for the human and the human's epistemic and sensory states. No other significant associations were found (**Table 3**).

Figure 3. AMS mean scores for the human (black bar) and the robot (line bar) as a function of state (Emotions, Desires and Intentions, Imaginative, Epistemic, Sensory).

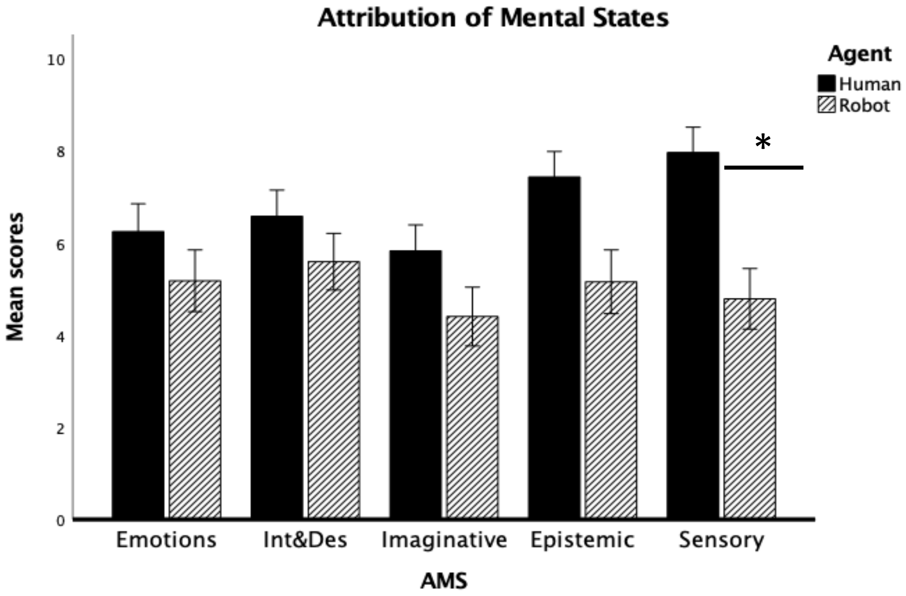


Table 3. Pearson's correlations between AMS dimensions and children's coherence scores (lie-naughty; mistake-not naughty) in the lie-mistake tasks for the human and robot conditions. ** correlation is significant at the 0.01 level (two-tailed).

	HUMAN		ROBOT	
AMS	Lie_Naughty	Mistake_NotNaughty	Lie_Naughty	Mistake_NotNaughty
Emotions	-0.063	0.21	-0.119	0.045
Intention&Desire	-0.192	0.061	-0.131	0.08
Imaginative	-0.125	0.119	0.026	-0.12
Epistemic	-0.17	**0.216***	-0.164	0.025
Sensory	-0.082	**0.221***	-0.112	-0.08
N	87	87	87	87

Discussion

In this study, it was investigated the comprehension of the distinction between an intentionally false statement (lie) and an unintentionally false statement (mistake) by 5- and 6-year-old children with respect to a human and NAO robot, as well as the moral evaluation of an intentionally or unintentionally false statement. In general, the results revealed a difference in the ability to appropriately discriminate a lie from a mistake between children who had acquired a first-order ToM competence and those who had not, i.e., those who made random judgments in all lie-mistake conditions, both for the human and the robot. On the other hand, children who had acquired a ToM competence were able to correctly discriminate between a lie and a mistake for the human, but struggled in attributing a lie to the robot. Consequently, the robot's behaviour was judged as not particularly bad in the lie situation. Additionally, regardless of their acquisition of a ToM competence, children failed in the attribution of an appropriate feeling to the deceived character (angry, not angry). These results were related to children's understanding of intentionality and to their verbal comprehension, but not to their executive function skills.

More specifically, the results of the present study showed that children who failed the first-order ToM tasks could not correctly distinguish between an intentionally or unintentionally false statement for both the human and the robot. This is in line with previous findings (e.g., Gilli et al., 2001; Wimmer et al., 1984) suggesting that a genuine conceptualization of lies and mistakes requires children to have acquired a ToM competence that includes an understanding not just of intentionality with respect to an *action* – an understanding that is acquired by age 3 (Lee, 2013; Lyon & Dorado, 2008; Siegal & Peterson, 1996, 1998; Talwar & Lee, 2008) – but also of intentionality with respect to a *statement,* an understanding that is acquired between 4 and 6 years (Adenzano & Bucciarelli, 2008; Gilli et al., 2001; Strichartz & Burton, 1990; Surian, 2008; Taylor et al., 2003). The acquisition of a ToM enabling the understanding of intentions beyond actions and statements presumably encompasses both levels of understanding. This was highlighted by correlational findings whereby greater coherence in correctly assessing the moral value of an unintentionally false statement (mistake-not naughty) made by a human was positively related, in no-ToM group, to both verbal comprehension and intentionality understanding. Specificity for the mistake situation may actually entail more complex reasoning abilities than a lie situation because understanding a *non*-intentional false statement requires not only grasping the concept of intentionality, but also some mental reversibility skills that, according to Piaget, develop from about age 6-7 (Piaget, 1932). Additionally, at this age, intentionality understanding and verbal competence appear to play a greater role than executive function skills in the recognition of a lie and a mistake. In fact, contrary to previous findings (Carlson & Moses, 2001; Talwar & Lee, 2008), these data showed no substantial correlation between children's executive function skills and performance in the lie-mistake tasks.

The picture changes with the emergence of first-order ToM skills. Children who had acquired a first-order ToM generally performed above chance in recognizing a mistake and attributing a moral value (not naughty) to the deceiver's behaviour. This was true for both the human and the robot. Interestingly, though, while these children correctly attributed a lie to the human in the lie situation, they were not so good at recognizing the lie in the robot condition. That is, ToM children failed to correctly attribute a lie to the robot when it made an intentional false statement. As a consequence, the moral evaluation of the robot's intentional deceiving behaviour was also inappropriate, because children often regarded the "lying" robot as not naughty. Supporting these observations, it was found greater coherence between lie-mistake recognition and the moral evaluation of behaviour for the humans as compared to the robot, which was at times judged as not naughty in the lie situation. Furthermore, as compared to non-ToM children, ToM children showed greater consistency between judgments of lies and mistakes and their moral evaluation - echoing the observations above - and less consistency for the robot than the human. It is noteworthy in this respect that there was a positive correlation between intentionality understanding skills and coherence in the mistake situation specifically for the robot. That is, among children who had acquired a first-order ToM, those who developed a higher level of intentionality understanding were also those who showed a better grasp of the concept of mistake, thus generalizing this concept also to the behaviour of the robot.

These results are largely in line with data on the attribution of mental states to the human and robot. Although children attributed greater mental competence to the human than the robot on all state-dimensions (emotions, desires and intentions, imaginative, epistemic, sensory), the robot significantly differed from the human only in the sensory dimension. This indicates that, in mental terms, the robot was perceived by the 5-6-year-olds in a rather anthropomorphic fashion, which may account for the fact that children responded to the lie-mistake situations in a very similar way for the human and the robot. A similar result was also found in a study by Di Dio et al. (2020b) where 5-year-old children interacting with the NAO robot in vivo essentially discriminated between the human and the robot on the sensory dimension. Placing the robot in a scenario where it exhibits complex behaviour can facilitate the perception of a robotic entity as a human-like entity, even though it is seen as different in nature, as demonstrated by significant differences in the perceptual/sensory domain. This is particularly so for young children because results in interactive situations with the robot show that with age children tend to discriminate between humans and robots in terms of mental states, with robots being significantly downgraded (Di Dio et al., 2020a,b; Manzi et al., 2020c).

Nevertheless, children treated the robot as an agent that is not really capable of lying. That is, in contrast to the human, the robot cannot “intentionally” harm. The resistance of ToM children to attributing a lie to the robot and, consequently considering it as less “naughty” than the human exhibiting exactly the same behaviour, parallels findings in the study by Di Dio et al. (2020b). In that study which investigated the development of trust between children aged 3 to 10 years and the robot NAO, children, particularly those with a secure attachment, regained trust in the robot more easily than in the human after trust had been betrayed. In other words, the robot was regarded as more “forgivable” than the human. In that study, as in the present one, children’s attitude toward robot misbehaviour seems to be more tolerant perhaps precisely because the robot is viewed as a different entity created by humans and, as such, cannot be programmed to be intentionally deceptive or cause harm. It is also interesting to note that in a study that compared the attribution of mental states to the

NAO and Robovie robots – the former being more physically anthropomorphic, and the latter more mechanical (though anthropomorphized) – young children tended to attribute only positive emotional qualities to the NAO robot, as if it could not experience negative emotional states, such as anger (Manzi et al., 2020c). Arguably, this positive stance toward the NAO robot may also have contributed to its characterization as an entity that cannot lie.

Finally, these data showed that children were not significantly able to attribute an appropriate emotion to the misled character (lie-angry; mistake-not angry). When they were asked whether the misled character was angry children responded randomly with respect to both the human and the robot. This result is not surprising because the question, "*When the boy/girl who has been told that bread was good to eat found out that the boy/girl/robot told a lie or made a mistake, was he/she angry or not angry with him/her"* requires the child to reason at a higher level of ToM competence, of the type *"I think that s/he thinks that.."* This second-level ToM competence develops around the age of 7 (Premack & Woodruff, 1978; Wimmer & Perner, 1983), thereby explaining why children in our study failed to correctly respond to this question.

Concluding remarks, some considerations and limitations

This study investigated 5- and 6-year-old children's understanding of the distinction between an intentionally false statement (a lie) and an unintentionally false statement (a mistake) by a human and the NAO robot, as well as the relevant moral evaluation. These data show that the ability to understand an agent intentionality and the acquisition of first-order ToM allow children to grasp the concepts of lie and mistake and, thus, to accurately distinguish between intentionally and unintentionally false statements. Moreover, with the acquisition of a more refined mentalistic ability, children can make a more cogent moral evaluation of a lie and mistake (lie = naughty, mistake = not naughty). However, although children who had acquired a first-order ToM competence were generally more accurate in their answers regarding *human* lie-mistake situations than children who had not, they nonetheless responded randomly both when attributing a lie to the robot and when making a relevant moral evaluation. Assuming that the appropriate attributions by these children to a human demonstrate an understanding of intentionality, their failure to attribute a lie to the robot suggests that, to some extent, these children do not attribute fully intentional behaviour to the robot. The robot's behaviour is more acceptable than that of a human because its actions are somehow outside of its will. For example, its actions could be attributed to someone who programmed it to act in a specific way, such as being "naughty" and intentionally making false statements. This observation is quite relevant when considering settings, such as educational or clinical ones, where robots are increasingly involved in children's daily activities (Marchetti, et al., 2020b). In line with previous findings, these findings show that, under specific conditions, robots can potentially be associated with the concept of human by young children, as also demonstrated by the AMS results; however, when children interact with robots adopting an educational or even a health-care role, one should keep in mind that an intentionally false statement made a robot may be misjudged.

The present study also produced some notable findings with respect to the concepts of lies and mistakes *per se*. ToM children were more accurate in their attribution of a mistake than a lie, and this may reflect a tendency to favor, from a psychological perspective, the making of mistakes. In fact, mistakes may not carry particularly serious consequences for the person who makes them. For example, a caregiver or an educator will not reprimand a child who has damaged something unintentionally. By contrast, lying is more significant in the child's mind because it carries with it negative consequences, such as punishment and/or weighty emotional and moral costs. In fact, in everyday life, intentionally bad actions are significantly more verbally remarked upon and acknowledged than mistakes, which are often overlooked or forgiven.

This study also has important implications for the design and use of applications on child-robot interaction, such as with robot tutors and education (Belpaeme et al. 2018). The literature on robot tutors has for example looked at different roles for the robot: from a robot "teacher" (i.e. one robot leading a class of children) to a robot "tutor" (one-to-one education interaction between a child learner and a robot knowledgeable of the subject) to robot "peer" (one-to-one where the child has to learn by teaching a novice robot) (Alnajjar et al., 2021). The transparent design of the robot's capability to make, or not, a mistake will have implications on the child's learning process and the efficacy of the learning process. This is also linked to the design of trustworthy behaviour in robots (Di Dio et al., 2019, 2020a, 2020b; Manzi et al., 2020b).

Two limitations of this study should be acknowledged: First, only one robot (the NAO robot) was used, and second the lie-mistake situations were presented in a video format. Future studies could therefore include different types of anthropomorphic robots in order to better assess the attribution of intentional or unintentional false statements (lies, mistakes) to agents with a range of different anthropomorphic features (e.g., NAO robot vs. Robovie robot) in real situations.

References

Adenzato, M., & Bucciarelli, M. (2008). Recognition of mistakes and deceits in communicative interactions. *Journal of Pragmatics, 40*(4), 608-629. https://doi.org/10.1016/j.pragma.2008.01.001

Alnajjar, F., Bartneck, C., Baxter, P., Belpaeme, T., Cappuccio, M., Di Dio, C., Eyssel, F., Handke, J., Mubin, O., Obaid, M., Reich-Stiebert, N. (2021). *Robots in Education: An Introduction to High-Tech Social Agents, Intelligent Tutors, and Curricular Tools*. London, UK:Routledge, Taylor & Francis group. ISBN 9780367655396

Astington, J. W. (1998). Theory of mind, Humpty Dumpty, and the icebox. *Human Development, 41(1)*, 30-39. from https://www.jstor.org/stable/26763339

Astington, J. W., Pelletier, J., & Homer, B. (2002). Theory of mind and epistemological development: The relation between children's second-order false-belief understanding and their ability to reason about evidence. *New ideas in Psychology, 20*(2-3), 131-144. https://doi.org/10.1016/S0732-118X(02)00005-3

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

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

Burton, R. V., & Strichartz, A. F. (1992). *Liar! Liar! Pants on fire*. In S. Ceci, M. Leichtman, & M. Putnick (Eds.), *Cognitive and social factors in early deception* (pp. 11-28). Hillside, NJ: Erlbaum.

Bussey, K. (1992). Lying and truthfulness: Children's definitions, standards, and evaluative reactions. *Child Development, 63*(1), 129-137. https://doi.org/10.1111/j.1467-8624.1992.tb03601.x

Bussey, K. (1999). Children's categorization and evaluation of different types of lies and truths. *Child Development, 70*(6), 1338-1347. https://doi.org/10.1111/1467-8624.00098

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*(4), 1032-1053. https://doi.org/10.1111/1467-8624.00333

Castelli, I., Massaro, D., Sanfey, A. G., & Marchetti, A. (2014). “What is fair for you?” Judgments and decisions about fairness and Theory of Mind. *European Journal of Developmental Psychology, 11*(1), 49-62. https://doi.org/10.1080/17405629.2013.806264

Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-scale deceit: Deception as a marker of two-, three-, and four-year-olds’ early theories of mind. *Child Development, 60*, 1263–1277. https://doi.org/10.2307/1130919

Clément, F., Koenig, M., & Harris, P. (2004). The ontogenesis of trust. *Mind & Language, 19*(4), 360-379. https://doi.org/10.1111/j.0268-1064.2004.00263.x

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020a). Come i bambini pensano alla mente del robot: il ruolo dell'attaccamento e della Teoria della Mente nell'attribuzione di stati mentali ad un agente robotico. *Sistemi Intelligenti*, *32*(1), 41-56. doi:10.1422/96279

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020b). Shall I trust you? From child human-robot interaction to trusting relationships. *Frontiers in Psychology, 11*, 469. https://doi.org/10.3389/fpsyg.2020.00469

Di Dio, C., Manzi, F., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2019). It Does Not Matter Who You Are: Fairness in Pre-schoolers Interacting with Human and Robotic Partners. *International Journal of Social Robotics, 12*(5)*,* 1045-1059. https://doi.org/10.1007/s12369-019-00528-9

Di Dio, C., Isernia, S., Ceolaro, C., Marchetti, A., & Massaro, D. (2018). Growing Up Thinking of God's Beliefs: Theory of Mind and Ontological Knowledge. *SAGE Open, 8*(4), 1-14. https://doi.org/10.1177%2F2158244018809874

Ding, X. P., Wellman, H. M., Wang, Y., Fu, G., & Lee, K. (2015). Theory-of-Mind Training Causes Honest Young Children to Lie. *Psychological Science*, *26*(11), 1812-1821. https://doi.org/10.1177%2F0956797615604628

Evans, A. D., & Lee, K. (2013). Emergence of lying in very young children. *Developmental Psychology, 49*(10), 1958-1963. https://psycnet.apa.org/doi/10.1037/a0031409

Fu, G., Xiao, W. S., Killen, M., & Lee, K. (2014). Moral judgment and its relation to second-order theory of mind. *Developmental Psychology, 50*(8), 2085-2092. https://psycnet.apa.org/doi/10.1037/a0037077

Gerstadt, C., Hong, Y., & Diamond, A. (1994). The relationship between cognition and action: Performance of 31⁄2-7-year-old children on a Stroop-like day-night test. *Cognition, 53*(2), 129-153. https://doi.org/10.1016/0010-0277(94)90068-X

Gilli, G., Marchetti, A., & Siegal, M. (1998). Teorie della mente e comprensione della bugia: effetto delle massime conversazionali e della rilevanza delle prove [Theories of mind and understanding of lies: The effect of conversational rules and experimental relevance]. *Archivio di Psicologia, Neurologia e Psichiatria, 59*(1), 77–91.

Gilli, G., Siegal, M., Marchetti, A., & Peterson, C. C. (2001). Children's incipient ability to distinguish mistakes from lies: An Italian investigation. *International Journal of Behavioral Development, 25*(1), 88-92. https://doi.org/10.1080%2F01650250042000195

Harris, P. L. (2007). Trust. *Developmental Science, 10*(1), 135-138. doi:10.1111/j.1467-7687.2007.00575.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%2Fj.0956-7976.2004.00742.x

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II: A developmental neuropsychological assessment*. San Antonio, TX: The Psychological Corporation.

Lee, K. (2013). Little liars: Development of verbal deception in children. *Child Development Perspectives, 7*(2), 91-96. https://doi.org/10.1111/cdep.12023

Leslie, A. M., Knobe, J., & Cohen, A. (2006). Acting Intentionally and the Side-Effect Effect: Theory of Mind and Moral Judgment. *Psychological science, 17*(5), 421-427. https://doi.org/10.1111%2Fj.1467-9280.2006.01722.x

Lewis, M., Stranger, C., & Sullivan, M. W. (1989). Deception in 3-year-olds. *Developmental Psychology, 25*(3), 439–443.

Lyon, T. D., & Dorado, J. S. (2008). Truth induction in young maltreated children: The effects of oath-taking and reassurance on true and false disclosures. *Child Abuse & Neglect, 32*(7), 738–748. https://doi.org/10.1016/j.chiabu.2007.08.008

Ma, F., Evans, A. D., Liu, Y., Luo, X., & Xu, F. (2015). To lie or not to lie? The influence of parenting and theory-of-mind understanding on three-year-old children's honesty. *Journal of Moral Education, 44*(2), 198-212. https://doi.org/10.1080/03057240.2015.1023182

Manzi, F., Di Dio, C., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020a). Moral evaluation of human and robot interactions in Japanese pre-schoolers. *CEUR Workshop Proceedings*, *2696*, 20–27.

Manzi, F., Ishikawa, M., Di Dio, C., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020b). The understanding of congruent and incongruent referential gaze in 17-month-old infants: An eye-tracking study comparing human and robot. *Scientific Reports*, *10*(1), 11918. https://doi.org/10.1038/s41598-020-69140-6

Manzi, F., Peretti, G., Di Dio, C., Cangelosi, A., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020c). A Robot Is Not Worth Another: Exploring Children's Mental State Attribution to Different Humanoid Robots. *Frontiers in Psychology, 11*, 2011. https://doi.org/10.3389/fpsyg.2020.

Manzi, F., Massaro, D., Kanda, T., Tomita, K., Itakura, S., & Marchetti, A., Teoria della Mente, bambini e robot: l'attribuzione di stati mentali, Abstract de <<XXX Congresso Nazionale, Associazione Italiana di Psicologia, Sezione di Psicologia dello Sviluppo e

dell’Educazione>>, (Messina, 14-16 September 2017), Alpes Italia srl, ROMA -- ITA 2017: 65-66 [http://hdl.handle.net/10807/106022]

Marchetti, A., Di Dio, C., Massaro, D., & Manzi, F. (2020a). The psychosocial fuzziness of fear in the coronavirus (COVID-19) era and the role of robots. *Frontiers in Psychology, 11,* 2245. https://doi.org/10.3389/fpsyg.2020.02245

Marchetti, A., Di Dio, C., Manzi, F., & Massaro, D. (2020b). Robotics in Clinical and Developmental Psychology. *Reference Module in Neuroscience and Biobehavioral Psychology*, B978-0-12-818697-8.00005-4. doi:10.1016/B978-0-12-818697-8.00005-4

Marchetti, A., Manzi, F., Itakura, S., & Massaro, D. (2018). Theory of Mind and humanoid robots from a lifespan perspective. Z. *Psychologie, 226*(2*)*, 98-109. https://doi.org/10.1027/2151-2604/a000326

Martini, M. C., Gonzalez, C. A., & Wiese, E. (2016). Correction: Seeing minds in others – Can agents with Robotic Appearance Have Human-Like Preferences? *PLOS ONE, 11*(2): e0146310. https://doi.org/10.1371/journal.pone.0149766

Massaro, D., Valle, A., & Marchetti, A. (2013). Irony and second-order false belief in children: What changes when mothers rather than siblings speak?. *European Journal of Developmental Psychology, 10*(3), 301-317. https://doi.org/10.1080/17405629.2012.672272

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., Itakura, S. (2013a). Can infants use robot gaze for object learning?: The effect of verbalization. *Interaction studies*, *14*(3), 351-365. https://doi.org/10.1075/is.14.3.03oku

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., Itakura, S. (2013b). Infants understand the referential nature of human gaze but not robot gaze. *Journal of Experimental Child Psychology*, *116*(1), 86–95. https://doi.org/10.1016/j.jecp.2013.02.007

Piaget, J. (1932). *The moral judgment of the child*. San Diego, CA: Harcourt Brace.

Perner, J. (1997). Children's competency in understanding the role of a witness: Truth, lies, and moral ties. *Applied Cognitive Psychology*, *11*(7), S21-S35. https://doi.org/10.1002/(SICI)1099-0720(199712)11:7%3CS21::AID-ACP513%3E3.0.CO;2-Z

Perner, J., & Wimmer, H. (1985). “John thinks that Mary thinks that...” attribution of second-order beliefs by 5-to 10-year-old children. *Journal of Experimental Child Psychology, 39*(3), 437-471. https://doi.org/10.1016/0022-0965(85)90051-7

Peskin, J. (1992). Ruse and representations: On children’s ability to conceal information. *Developmental Psychology, 28*(1), 84–89. https://psycnet.apa.org/doi/10.1037/0012-1649.28.1.84

Peterson, C. C., Peterson, J. L., & Seeto, D. (1983). Developmental changes in ideas about lying. *Child Development*, *54*(6), 1529-1535. https://doi.org/10.2307/1129816

Polak, A. & Harris, P. L. (1999). Deception by young children following noncompliance. *Developmental Psychology, 35*(2), 561–568. https://psycnet.apa.org/doi/10.1037/0012-1649.35.2.561

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

Siegal, M., & Peterson, C. C. (1996). Breaking the mold: A fresh look at children's understanding of questions about lies and mistakes. *Developmental Psychology, 32*(2), 322-334. https://psycnet.apa.org/doi/10.1037/0012-1649.32.2.322

Siegal, M., & Peterson, C. C. (1998). Preschoolers' understanding of lies and innocent and negligent mistakes. *Developmental Psychology, 34*(2), 332-341. https://psycnet.apa.org/doi/10.1037/0012-1649.34.2.332

Strichartz, A. F., & Burton, R. V. (1990). Lies and truth: a study of the development of the concept. *Child Development, 61*(1), 211-220. https://doi.org/10.1111/j.1467-8624.1990.tb02773.x

Sullivan, K., Winner, E., & Hopfield, N. (1995). How children tell a lie from a joke: The role of second-order mental state attributions. *British Journal of Developmental Psychology, 13*(2), 191-204. https://doi.org/10.1111/j.2044-835X.1995.tb00673.x

Surian, L., Peterson, C., Siegal, M., & Nemeroff, C. (2001). Lies, Mistakes, and Blessings: Defining and Characteristic Features in Conceptual Development. *Journal of Cognition and Culture, 1*(4), 323-339. https://doi.org/10.1163/156853701753678314

Talwar, V., & Lee, K. (2002). Development of lying to conceal a transgression: Children’s control of expressive behaviour during verbal deception. *International Journal of Behavioral Development*, *26*(5), 436–444. https://doi.org/10.1080%2F01650250143000373

Talwar, V., & Lee, K. (2008). Social and Cognitive Correlates of Children’s Lying Behavior. *Child Development, 79*(4), 866–881. https://doi.org/10.1111/j.1467-8624.2008.01164.x

Talwar, V., Murphy, S. M., & Lee, K. (2006). White lie-telling in children for politeness purposes. *International Journal of Behavioral Development, 31*(1), 1–11. https://doi.org/10.1177%2F0165025406073530

Taylor, M., Lussier, G. L., & Maring, B. L. (2003). The Distinction Between Lying and Pretending. *Journal of Cognition and Development, 4*(3), 299-323. https://doi.org/10.1207/S15327647JCD0403_04

Traverso, L., & De Franchis, V. (2014). Dimensional Change Card Sort. Validazione italiana di una prova per la valutazione delle Funzioni Esecutive. *Psicologia clinica dello sviluppo, 18*(2), 231-256. doi: 10.1449/77636

Vinanzi, S., Patacchiola, M., Chella, A., & Cangelosi, A. (2019). Would a robot trust you? Developmental robotics model of trust and theory of mind. *Philosophical Transactions of the Royal Society B, 374*(1771), 20180032. https://doi.org/10.1098/rstb.2018.0032

Wellman, H.M. (2020). *Reading Minds: How Childhood Teaches Us To Understand People*. Oxford, UK: Oxford University Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child Development, 72*(3), 655-684. https://doi.org/10.1111/1467-8624.00304

Wilson, A. E., Smith, M. D., & Ross, H. S. (2003). The Nature and Effects of Young Children's Lies. *Social Development, 12*(1), 21–45. https://doi.org/10.1111/1467-9507.00220

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128. https://doi.org/10.1016/0010-0277(83)90004-5

Wimmer, H., Gruber, S., & Perner, J. (1984). Young children's conception of lying: Lexical realism—moral subjectivism. *Journal of Experimental Child Psychology, 37*(1), 1-30. https://doi.org/10.1016/0022-0965(84)90055-9

Wimmer, H., Gruber, S., & Perner, J. (1985). Young children's conception of lying: Moral intuition and the denotation and connotation of" to lie.". *Developmental Psychology, 21*(6), 993-995. https://psycnet.apa.org/doi/10.1037/0012-1649.21.6.993

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*, 297-301. https://doi.org/10.1038/nprot.2006.46

CHAPTER 4

A Robot Is Not Worth Another: Exploring Children's Mental State Attribution to Different Humanoid Robots

Federico Manzi[1]*†, **Giulia Peretti[1]**†, Cinzia Di Dio[1], Angelo Cangelosi[2], Shoji Itakura[3], Takayuki Kanda[4,5], Hiroshi Ishiguro[5,6], Davide Massaro[1] and Antonella Marchetti[1]

[1] *Research Unit on Theory of Mind, Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy,*

[2] *School of Computer Science, University of Manchester, Manchester, United Kingdom,*

[3] *Centre for Baby Science, Doshisha University, Kyoto, Japan,*

[4] *Human-Robot Interaction Laboratory, Graduate School of Informatics, Kyoto University, Kyoto, Japan,*

[5] *Advanced Telecommunications Research Institute International, IRC/HIL, Keihanna Science City, Kyoto, Japan,*

[6] *Department of Systems Innovation, Osaka University, Toyonaka, Japan*

† These authors have contributed equally to this work

Published on Frontiers in Psychology, 30 September 2020.

Keywords

child–robot interaction (cHRI), social robots, humanoid and anthropomorphic robots, differences among robots, children, anthropomorphism, mental states attribution

Abstract

Recent technological developments in robotics has driven the design and production of different humanoid robots. Several studies have highlighted that the presence of human- like physical features could lead both adults and children to anthropomorphize the robots. The present study aimed to compare the attribution of mental states to two humanoid robots, NAO and Robovie, which differed in the degree of anthropomorphism. Children aged 5, 7, and 9 years were required to attribute mental states to the NAO robot, which presents more human-like characteristics compared to the Robovie robot, whose physical features look more mechanical. The results on mental state attribution as a function of children's age and robot type showed that 5-year-olds have a greater tendency to anthropomorphize robots than older children, regardless of the type of robot. Moreover, the findings revealed that, although children aged 7 and 9 years attributed a certain degree of human-like mental features to both robots, they attributed greater mental states to NAO than Robovie compared to younger children. These results generally show that children tend to anthropomorphize humanoid robots that also present some mechanical characteristics, such as Robovie. Nevertheless, age-related differences showed that they should be endowed with physical characteristics closely resembling human ones to increase older children's perception of human likeness. These findings have important implications for the design of robots, which also needs to consider the user's target age, as well as for the generalizability issue of research findings that are commonly associated with the use of specific types of robots.

Introduction

Currently, we are witnessing an increasing deployment of social robots (Bartneck and Forlizzi, 2004) in various contexts, from occupational to clinical to educational (Murashov et al., 2016; Belpaeme et al., 2018; Marchetti et al., in press). Humanoid social robots (HSRs), in particular, have proven to be effective social partners, possibly due to their physical human likeness (Dario et al., 2001). Humanoid social robots can vary in the degree of their anthropomorphic physical characteristics, often depending on the target user (children, adults, elderly, students, clinical populations, etc.) and the context (household, education, commercial, and rehabilitation). For example, the humanoid KASPAR robot that resembles a young child (with face, arms and hands, legs and feet), was specifically built for children with autism spectrum disorder (Dautenhahn et al., 2009; Wainer et al., 2014). In other instances, however, the same HSRs are used both for different purposes and different populations, like the NAO robot, which is largely used both with clinical and non-clinical populations (Shamsuddin et al., 2012; Mubin et al., 2013; Begum et al., 2016; Belpaeme et al., 2018), or the Robovie robot, that is employed both with adults and children (Shiomi et al., 2006; Kahn et al., 2012). A recent review of the literature by Marchetti et al. (2018) showed that different physical characteristics of HSRs may significantly affect the quality of interaction between humans and robots at different ages. The construction of robots that integrate and expand the specific biological abilities of our species led to two different directions in robotic development based on different, though related, theoretical perspectives: developmental cybernetics (DC; Itakura, 2008; Itakura et al., 2008; Moriguchi et al., 2011; Kannegiesser et al., 2015; Okanda et al., 2018; Di Dio et al., 2019; Wang et al., 2020; Manzi et al., 2020a) and developmental robotics (DR; De La Cruz et al., 2014; Cangelosi and Schlesinger, 2015, 2018; Lyon et al., 2016; Morse and Cangelosi, 2017; Vinanzi et al., 2019; Zhong et al., 2019; Di Dio et al., 2020a,b). The first perspective (DC) consists of creating a human-like system, by simulating human psychological processes and prosthetic functions in the robot (enhancing the function and lifestyle of persons) to observe people's behavioral response toward the robot. The second perspective (DR) is related to the development of cognitive neural networks in the robot that would allow it to autonomously gain sensorimotor and mental capabilities with growing complexity, starting from intricate evolutionary principles. From these premises, the next two paragraphs briefly outline current findings concerning the effect that physical features of the HSRs have on human perception, thus outlining the phenomenon of anthropomorphism, and a recent methodology devised to measure it.

Anthropomorphism

Anthropomorphism is a widely observed phenomenon in human – robot interaction (HRI; Fink et al., 2012; Airenti, 2015; Złotowski et al., 2015), and it is also greatly considered in the design of robots (Dario et al., 2001; Kiesler et al., 2008; Bartneck et al., 2009; Sharkey and Sharkey, 2011; Zanatto et al., 2016, 2020). In psychological terms, anthropomorphism is the tendency to attribute human characteristics, physical and/or psychological, to non-human agents (Duffy, 2003; Epley et al., 2007). Several studies have shown that humans may perceive non-anthropomorphic robots as anthropomorphic, such as Roomba (a vacuum cleaner with a semi-autonomous system; Fink et al., 2012). Although anthropomorphism seems to be a widespread phenomenon, the attribution of human traits to anthropomorphic robots is significantly greater compared to non-anthropomorphic robots. A study by Krach et al. (2008) compared four different agents (computer, functional robot, anthropomorphic robot, and human confederate) in a Prisoner's Dilemma Game, and showed that the more the interactive partner displayed human-like characteristics, the more the participants appreciated the interaction and ascribed intelligence to the game partner. What characteristics of anthropomorphic robots (i.e., the HSRs) increase the perception of anthropomorphism? The HSRs can elicit the perception of anthropomorphism mainly at two levels: physical and behavioral (Marchetti et al., 2018). Working on the physical level is clearly easier than on intrinsic psychological features, and – although anthropomorphic physical features of robots are not the only answer to enhance the quality of interactions with humans – the implementation of these characteristics can positively affect HRIs (Duffy, 2003; for a review see Marchetti et al., 2018). It should be stated, however, that extreme human-likeness can result in the known uncanny valley effect, according to which HRIs are negatively influenced by robots that are too similar to the human (Mori, 1970; MacDorman and Ishiguro, 2006; Mori et al., 2012). Thus, the HSRs' appearance represents an important social affordance for HRIs, as further demonstrated by the psychological research on racial and disability prejudice (Todd et al., 2011; Macdonald et al., 2017; Sarti et al., 2019; Manzi et al., 2020b). The anthropomorphic features of the HSRs can increase humans' perception of humanness, such as mind attribution and personality, and influence other psychological mechanisms and processes (Kiesler and Goetz, 2002; MacDorman et al., 2005; Powers and Kiesler, 2006; Bartneck et al., 2008; Broadbent et al., 2013; Złotowski et al., 2015; Marchetti et al., 2020).

The study of the design of physical characteristics of the HSRs and their classification has been already investigated in HRI, but not systematically. A pioneering study by DiSalvo et al. (2002) explored the perception of humanness using 48 images of different heads of HSRs, and showed that three features are particularly important for the robot's design: the nose, eyes, and mouth. Furthermore, a study by Duffy (2003) categorized different robots' head in a diagram composed of three extremities: “human head” (as-close-as-possible to a human head), “iconic head” (a very minimum set of features) and “abstract head” (a more mechanistic design with minimal human-like aesthetics). Also, in this instance, human likeness was associated with greater mental abilities. Furthermore, a study by MacDorman (2006) analyzed the categorization of 14 types of robots (mainly androids and humanoids) in adults. It was shown that humanoid robots displaying some mechanical characteristics – such as the Robovie robot – were classified average on a “humanness” scale and rated lower on the uncanny valley scale. Recent studies compared one of the most widely used HSRs, the NAO robot, with different types of robots. It was shown that the NAO robot is perceived less human-like than an android – which is a highly anthropomorphic robot in both appearance and behavior (Broadbent, 2017)-, but more anthropomorphic than a mechanical robot, i.e., the Baxter robot (Yogeeswaran et al., 2016; Zanatto et al., 2019). However, there were no differences in perceived ability to perform physical and mental tasks between NAO and the android (Yogeeswaran et al., 2016), indicating that human-likeness (and not “human-exactness”) is sufficient to trigger the attribution of psychological features to a robot. In addition, a database has recently been created that collects more than 200 HSRs classified according to their level of human likeness (Phillips et al., 2018). In this study the NAO robot was classified with a score of about 45/100, in particular thanks to the characteristics of its face and body. Robovie and other similar robots were classified with a score ranging between 27 and 31/100, deriving mainly from body characteristics. These findings corroborate the hypothesis that NAO and Robovie are two HSRs with different levels of human-likeness due to their physical anthropomorphic features.

The interest in observing the effect of different physical characteristics of robots in terms of attribution of intentions, understanding, and emotions has also been investigated in children (Bumby and Dautenhahn, 1999; Woods et al., 2004; Woods, 2006). In particular, a study by Woods (2006), comparing 40 different robots, revealed that children experience greater discomfort with robots that look too similar to humans, favoring robots with mixed human-mechanical characteristics. These results were confirmed in a recent study by Tung (2016) showing that children preferred robots with not too many human-like features over robots with many human characteristics. Overall, these results suggested that an anthropomorphic design of HSRs may increase children’s preference toward them. Still, an excessive implementation of human features can negatively affect the attribution of positive qualities to the robot, again in line with the Uncanny Valley effect above.

Attribution of Mental States

Different scales were developed to measure psychological anthropomorphism toward robots in adults. These scales typically assess attribution of intelligence, personality and emotions, only to mention a few. In particular, the attribution of internal states to the robot, i.e., to have a mind, is widely used and very promising in HRI (Broadbent et al., 2013; Stafford et al., 2014).

In psychology, the ability to ascribe mental states to others is defined as the Theory of Mind (ToM). Theory of mind is the ability to understand one's own and others' mental states (intentions, emotions, desires, beliefs), and to predict and interpret one's own and others' behaviors on the basis of such meta-representation (Premack and Woodruff, 1978; Wimmer and Perner, 1983; Perner and Wimmer, 1985). Theory of mind abilities develop around four years of age, becoming more sophisticated with development (Wellman et al., 2001). Theory of mind is active not only during humans' relationships but also during interactions with robots (for a review, see Marchetti et al., 2018). Recent studies have shown that adults tend to ascribe greater mental abilities to robots that have a human appearance (Hackel et al., 2014; Martini et al., 2016). This tendency to attribute human mental states to robots was also observed in children. Generally, children are inclined to anthropomorphize robots by attributing psychological and biological characteristics to them (Katayama et al., 2010; Okanda et al., 2019). Still, they do differentiate between humans and robots' abilities. A pioneering study by Itakura (2008) investigating the attribution of mental verbs to a human and a robot showed that children did not attribute the epistemic verb "think" to the robot. More recent studies have further shown that already from three years of age, children fairly differentiate a human from a robot in terms of mental abilities (Di Dio et al., 2020a), although younger children appear to be more inclined to anthropomorphize robots compared to older children. This effect may be due to the phenomenon of animism, particularly active at three years of age (Di Dio et al., 2020a,b).

Aim of the Study

The present study aimed to investigate the attribution of mental states (AMS) in children aged 5–9 years to two humanoid robots, NAO and Robovie, varying in their anthropomorphic physical features (DiSalvo et al., 2002; Duffy, 2003). Differences in the attribution of mental qualities to the two robots were then explored using the robots' degree of physical anthropomorphism and the child's chronological age. The two humanoid robots, NAO and Robovie, have been selected for two main reasons: (1) in relation to their physical appearance, both robots belong to the category of HSRs, but differ for their degree of anthropomorphism (for a detailed description of the robots, see section "Materials"); (2) both robots are largely used in experiments with children (Kanda et al., 2002; Kose and Yorganci, 2011; Kahn et al., 2012; Shamsuddin et al., 2012; Okumura et al., 2013a,b; Tielman et al., 2014; Cangelosi and Schlesinger, 2015, 2018; Hood et al., 2015; Di Dio et al., 2020a,b).

In light of previous findings associated with the use of these specific robots described above, the following has been assumed: (1) independent of age, children would distinguish between humans and robots in terms of mental states by ascribing lower mental attributes to the robots; (2) children would tend to attribute greater mental qualities to NAO compared to Robovie because of its greater human-likeness; and (3) younger children would tend to attribute more human characteristics to robots (i.e., to anthropomorphize more) than older children.

Materials and Methods

Participants

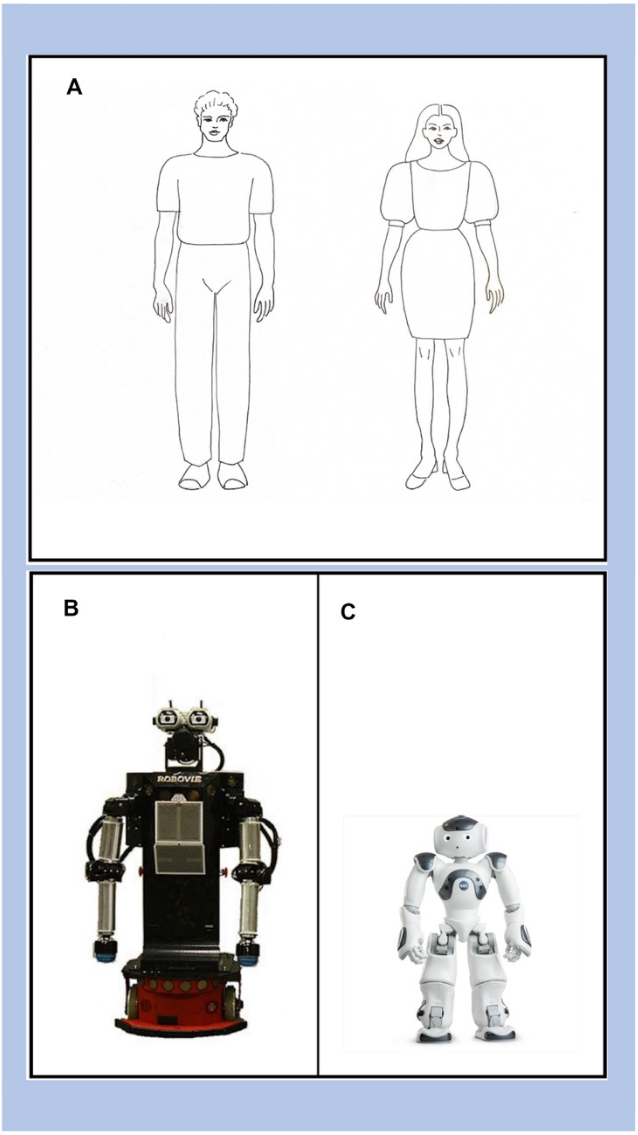
Data were acquired on 189 Italian children from kindergarten and primary school age. The children were divided into three age groups for each robot as follows: (1) for the NAO robot, 5 years (N = 24, 13 females; M = 68.14; SD = 3.67); 7 years (N=25, 13 females; M=91.9; SD=3.43);and 9 years (N=23, 12 females; M = 116.38, SD = 3.91); (2) for the Robovie robot, 5 years (N=33, 13females; M=70.9, SD=2.95);7 years (N=49, 26 females; M = 93.4, SD = 3.62); and 9 years (N = 35, 15 females; M = 117.42, SD = 4.44). The initial inhomogeneity between sample sizes in the NAO and Robovie conditions were corrected by the random selection of children in the Robovie condition, caring to balance by gender. Accordingly, the sample for the Robovie condition used for statistical analysis was composed as follows: 5 years (N = 24, 8 females; M = 70.87, SD = 3.1); 7 years (N=25, 14females; M=92.6, SD=3.73);and 9 years (N=23, 10 females; M = 117.43, SD = 4.62). The children's parents received a written explanation of the procedure of the study, the measurement items, and gave their written consent. The children were not reported by teachers or parents for learning and/or socio-relational difficulties. The study was approved by the Local Ethic Committee (Università Cattolica del Sacro Cuore, Milan).

Materials, Task, and Procedure

Materials

The two HSRs selected for this study were the Robovie robot (Hiroshi Ishiguro Laboratories, ATR; Figure 1B) and the NAO robot (Aldebaran Robotics, Figure 1C). Two robots were chosen because, although they both belong to the category of HSRs, they differ in their degree of anthropomorphic features (DiSalvo et al., 2002; Duffy, 2003; MacDorman, 2006; Zhang et al., 2008;

Phillips et al., 2018). Robovie is a HSR with more abstract anthropomorphic features: no legs but two driving wheels to move, two arms without hands. In particular, the head can be considered “abstract” (Duffy, 2003) because of two important human-like features: two eyes and a microphone that looks like a mouth (DiSalvo et al., 2002). Robovie is an HSR that can be rated as average in the continuum of mechanical-humanlike (Ishiguro et al., 2001; Kanda et al., 2002; MacDorman, 2006). NAO is a HSR with more pronounced anthropomorphic features compared to Robovie: two legs, two arms, and two hands with three moving fingers (Figure 1C). Besides, the face can be classified as “iconic” and consists of three cameras suggesting two eyes and a mouth. However, considering the whole body and the more detailed shape of the face, NAO is a HSR that can be rated as more human-like than Robovie (DiSalvo et al., 2002; MacDorman, 2006; Phillips et al., 2018).



Attribution of Mental States

The AMS questionnaire[2] is a measure of mental states that participants attribute to when they look at images depicting specific characters, in this case a human (female or male based on the participant's gender; Figure 1A), and, according to the group condition, the Robovie or the NAO robot (Figures 1B,C). The AMS questionnaire was inspired by the methodology described in Martini et al. (2016) and is already used in several experiments with children (Manzi et al., 2017; Di Dio et al., 2019, 2020a,b). The construction of the content of the questionnaire is based on the theoretical model of Slaughter et al. (2009) on the categorization of children's mental verbs resulting from communication exchanges between mother and child. This classification divides mental verbs into four categories: perceptive, volitional, cognitive, and dispositional. For the creation of the AMS questionnaire an additional category related to imaginative verbs has been added. It has been considered it necessary to distinguish between cognitive, epistemic, and imaginative states, since – especially for the robot – this specification enables the analysis of different psychological processes in terms of development. The AMS therefore consists of five dimensions: Perceptive, Emotive, Desires and Intentional, Imaginative, and Epistemic.

The human condition was used as a baseline measure to evaluate children's ability to attribute mental states. In fact, as described in the results below, children scored quite high when ascribing mental attributes to the human character, thus supporting children's competence in performing the mental states attribution task. Also, the human condition was used as a comparison measure against which the level of psychological anthropomorphism of NAO and Robovie was evaluated. The Cronbach's alfa for each category is as follows: Perceptive ($\alpha = 0.8$), Emotive ($\alpha = 0.8$), Desires and Intentional ($\alpha = 0.8$), Imaginative ($\alpha = 0.8$), and Epistemic ($\alpha = 0.7$).

Children answered 25 questions grouped into the five different state categories described above (see Appendix 1 for the specific items). The child had to answer "Yes" or "No" to each question, obtaining 1 when the response is "Yes" and 0 when the response is "No". The sum of all responses (range = 0–25) gave the total score ($\alpha = 0.9$); the five partial scores were the sum of the responses within each category (range = 0–5).

---

[2] [1]http://www.teoriadellamente.it, "Strumenti" section

Procedure

The children were tested individually in a quiet room inside their school. Data acquisition was carried out by a single researcher during the normal school activities.

The experimenter showed each child the image on a paper depicting a human - gender matched - and one of the two robots, NAO or Robovie. The presentation order of the image - human and robot- was randomized. Afterward, the experimenter asked children the questions on the five categories of the AMS (Perceptive, Emotive, Intentions and Desires, Imaginative, and Epistemic). The presentation order of the five categories was also randomized. The total time required to complete the test was approximately 10 min.

Results

Data Analysis

To evaluate the effect of age, gender, states, agent, and type of robots on children's mental state attribution to robots, a GLM analysis was carried out with five levels of states (Perceptive, Emotive, Intentions and Desires, Imaginative, and Epistemic) and two levels of agent (Human, Robot) as within-subjects factors, and age (5-, 7-, 9-year-olds), gender (Male, Female) and robot (Robovie, NAO) as the between-subjects factor. The Greenhouse-Geisser correction was used for violations of Mauchly's Test of Sphericity ($p < 0.05$). Post hoc comparisons were Bonferroni corrected.

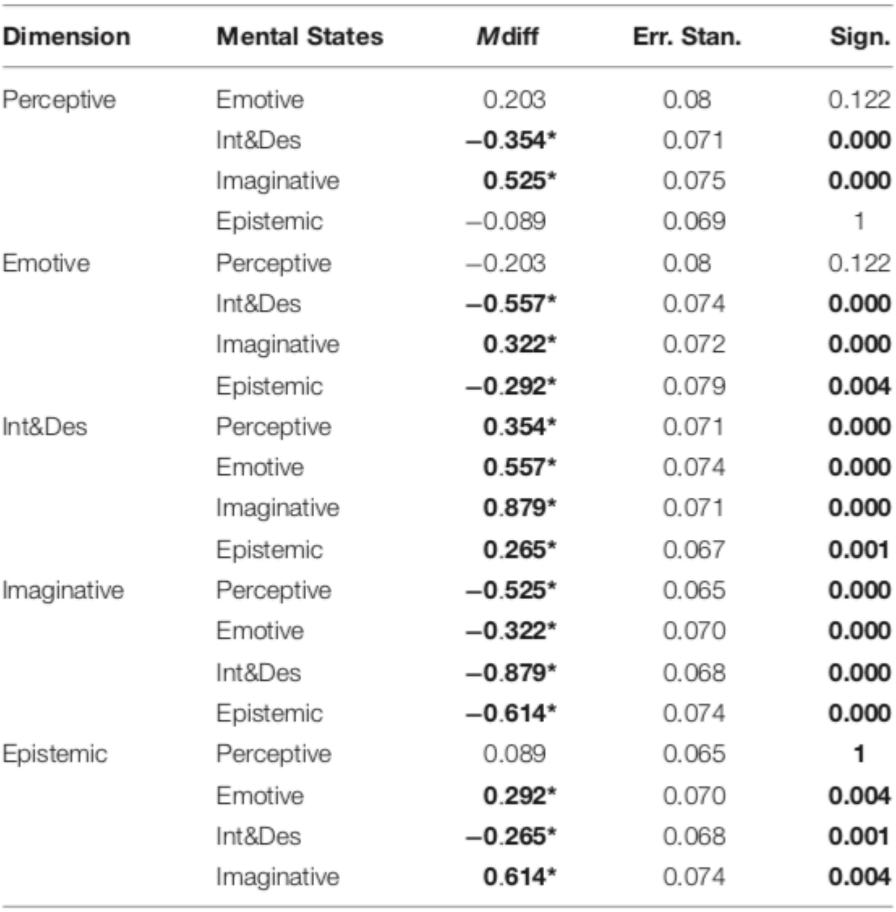
Results

The results showed (1) a main effect of agent, F(1, 126) = 570.9, $p < 0.001$, *partial*-$\eta^2 = 0.819$, $\delta = 1$, indicating that children attributed greater mental states to the human ($M = 4.6$, SD = 0.27) compared to the robot ($M = 2.7$, SD = 0.21; Mdiff = 1.75, SE = 0.087); (2) a main effect of states, $F(4,504) = 40.33$, $p < 0.001$, *partial*-$\eta^2 = 0.243$, $\delta = 1$, mainly indicating that children attributed greater intention and desires and lower imaginative states (for a full description of the statistics, see **Table 1**); (3) a main effect of robot, $F(1,126) = 39.4$, $p < 0.001$, *partial*-$\eta^2 = 0.238$, $\delta = 1$, showing that children attributed greater mental states to NAO ($M = 3.98$, SD = 0.17) compared to Robovie ($M = 3.4$, SD = 0.14; Mdiff = 0.568, SE = 0.099).

A two-way interaction was also found between (1) states and agent, $F(1,126) = 16.51$, $p < 0.001$, *partial*-$\eta^2 = 0.183$, $\delta = 1$ (for a detailed description of the differences see **Table 2**), and (2) agent and age, $F(2,126) = 25.17$, $p < 0.001$, $partial\text{-}\eta^2 = 0.285$, $\delta = 1$, showing that 5-year-old children attributed greater mental states to the robotic agents compared to older children (see **Table 2**).

Additionally, three-way interaction was found between states, age, and robot, $F(8,126) = 4.95$, $p < 0.001$, $partial\text{-}\eta^2 = 0.073$, $\delta = 1$. The planned comparisons on the three-way interaction revealed that children attributed greater mental states to NAO compared to Robovie, with the youngest children differentiating on the Perceptive and Epistemic dimensions, and with this difference spreading to all dimensions (but imaginative) in the older children (see Figure 2).

TABLE 1 | Statistics comparing the attribution of all AMS dimensions (Perceptive, Emotive, Intentions and Desires, Imaginative, Epistemic).



Dimension	Mental States	*M*diff	Err. Stan.	Sign.
Perceptive	Emotive	0.203	0.08	0.122
	Int&Des	**−0.354***	0.071	**0.000**
	Imaginative	**0.525***	0.075	**0.000**
	Epistemic	−0.089	0.069	1
Emotive	Perceptive	−0.203	0.08	0.122
	Int&Des	**−0.557***	0.074	**0.000**
	Imaginative	**0.322***	0.072	**0.000**
	Epistemic	**−0.292***	0.079	**0.004**
Int&Des	Perceptive	**0.354***	0.071	**0.000**
	Emotive	**0.557***	0.074	**0.000**
	Imaginative	**0.879***	0.071	**0.000**
	Epistemic	**0.265***	0.067	**0.001**
Imaginative	Perceptive	**−0.525***	0.065	**0.000**
	Emotive	**−0.322***	0.070	**0.000**
	Int&Des	**−0.879***	0.068	**0.000**
	Epistemic	**−0.614***	0.074	**0.000**
Epistemic	Perceptive	0.089	0.065	**1**
	Emotive	**0.292***	0.070	**0.004**
	Int&Des	**−0.265***	0.068	**0.001**
	Imaginative	**0.614***	0.074	**0.004**

*Based on estimated marginal averages *The average difference is significant at the level of b Adaptation for multiple comparisons: Bonferroni. Significant values are in bold.*

Discussion and Conclusion

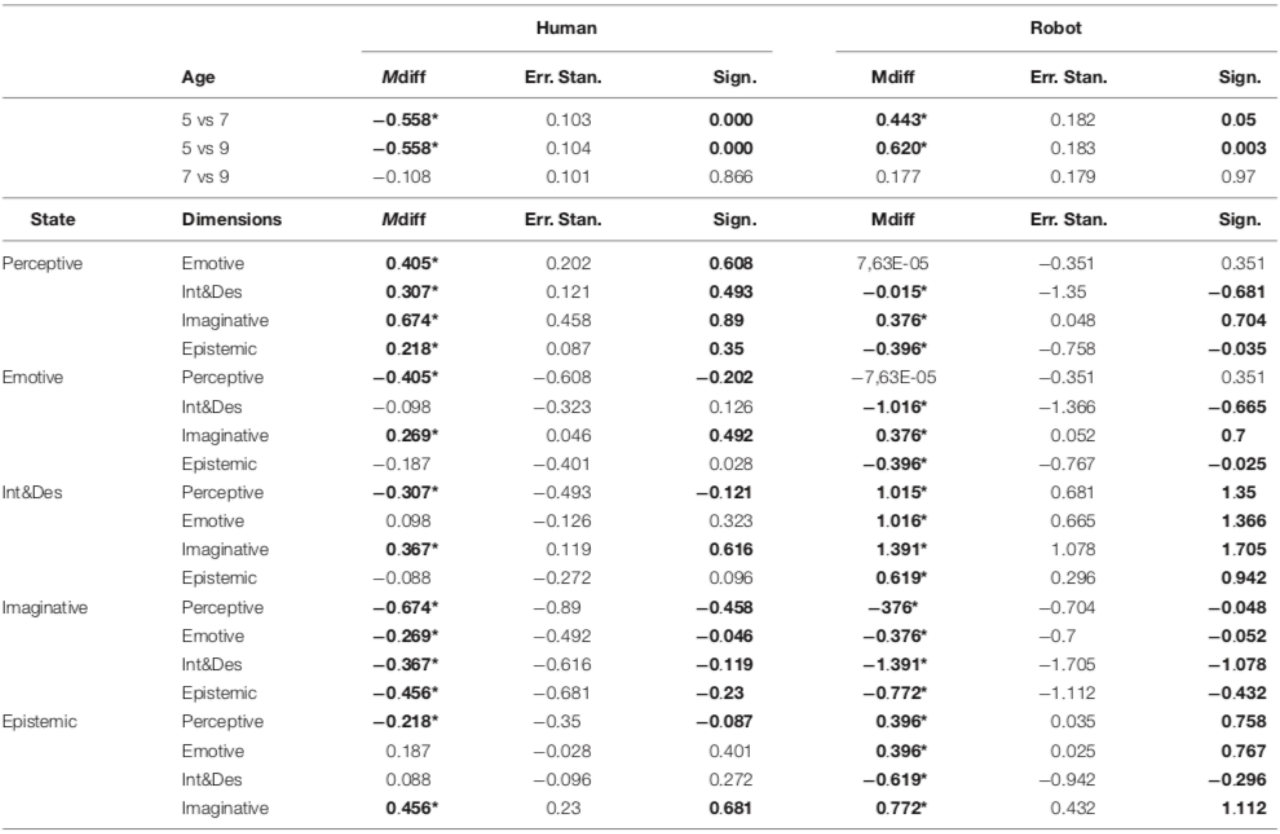
Discussion

In the present AMS was compared in children aged 5– 9 years between two HSRs, NAO and Robovie, also with respect to a human. The aim was to explore children's patterns of mental attribution to different types of HSRs, varying in their degree of physical anthropomorphism, from a developmental perspective.

These results on the AMS to the human and robot generally confirmed the tendency of children to ascribe lower human mental qualities to the robots, thus supporting previous findings (Manzi et al., 2017; Di Dio et al., 2018, 2019, 2020a,b). In addition, children generally attributed greater mental states to the NAO robot than to the Robovie robot, although differences were found in the quality of mental states attribution as a function of age, with older children discriminating more between the types of robots that the younger ones. As a matter of fact, the important role played by the type of robot in influencing children's AMS can be appreciated by evaluating differences in state attribution developmentally.

Firstly, 5-year-old children generally attributed greater human-like mental states to the robotic agents compared to older children. Additionally, while 5-year-old children discriminated between robots' mental attribution only on the perceptive and epistemic dimensions – with the NAO robot being regarded as more anthropomorphic than Robovie –, children aged 7 and 9 years were particularly sensitive to the type of robots, and attributed greater mental states to NAO than Robovie on most of the tested mental state dimensions.

TABLE 2 | Statistics comparing the attribution of all AMS dimensions (Perceptive, Emotive, Intentions and Desires, Imaginative, Epistemic) and the AMS for the two agents (Human, Robot) across ages (5-, 7- and 9-years).



		Human			Robot		
	Age	***Mdiff***	**Err. Stan.**	**Sign.**	**Mdiff**	**Err. Stan.**	**Sign.**
	5 vs 7	**−0.558***	0.103	**0.000**	**0.443***	0.182	**0.05**
	5 vs 9	**−0.558***	0.104	**0.000**	**0.620***	0.183	**0.003**
	7 vs 9	−0.108	0.101	0.866	0.177	0.179	0.97
State	**Dimensions**	***Mdiff***	**Err. Stan.**	**Sign.**	**Mdiff**	**Err. Stan.**	**Sign.**
Perceptive	Emotive	**0.405***	0.202	**0.608**	7,63E-05	−0.351	0.351
	Int&Des	**0.307***	0.121	**0.493**	**−0.015***	−1.35	**−0.681**
	Imaginative	**0.674***	0.458	**0.89**	**0.376***	0.048	**0.704**
	Epistemic	**0.218***	0.087	**0.35**	**−0.396***	−0.758	**−0.035**
Emotive	Perceptive	**−0.405***	−0.608	**−0.202**	−7,63E-05	−0.351	0.351
	Int&Des	−0.098	−0.323	0.126	**−1.016***	−1.366	**−0.665**
	Imaginative	**0.269***	0.046	**0.492**	**0.376***	0.052	**0.7**
	Epistemic	−0.187	−0.401	0.028	**−0.396***	−0.767	**−0.025**
Int&Des	Perceptive	**−0.307***	−0.493	**−0.121**	**1.015***	0.681	**1.35**
	Emotive	0.098	−0.126	0.323	**1.016***	0.665	**1.366**
	Imaginative	**0.367***	0.119	**0.616**	**1.391***	1.078	**1.705**
	Epistemic	−0.088	−0.272	0.096	**0.619***	0.296	**0.942**
Imaginative	Perceptive	**−0.674***	−0.89	**−0.458**	**−376***	−0.704	**−0.048**
	Emotive	**−0.269***	−0.492	**−0.046**	**−0.376***	−0.7	**−0.052**
	Int&Des	**−0.367***	−0.616	**−0.119**	**−1.391***	−1.705	**−1.078**
	Epistemic	**−0.456***	−0.681	**−0.23**	**−0.772***	−1.112	**−0.432**
Epistemic	Perceptive	**−0.218***	−0.35	**−0.087**	**0.396***	0.035	**0.758**
	Emotive	0.187	−0.028	0.401	**0.396***	0.025	**0.767**
	Int&Des	0.088	−0.096	0.272	**−0.619***	−0.942	**−0.296**
	Imaginative	**0.456***	0.23	**0.681**	**0.772***	0.432	**1.112**

*Based on estimated marginal averages *The average difference is significant at the level of b Adaptation for multiple comparisons: Bonferroni. Significant values are in bold.*

From a developmental perspective, the tendency of younger children to anthropomorphize HSRs could be reasonably explained by the phenomenon of animism (Piaget, 1929). Already Piaget in 1929 suggested that children younger than 6 years tend to attribute a consciousness to objects, i.e., the phenomenon of animism, and that this fades around 9 years of age. Recently, this phenomenon has been defined as a cognitive error in children (Okanda et al., 2019), i.e., animism error, characterized by a lack of differentiation between living and non-living things. In this respect, several studies showed that, although children are generally able to discriminate between humans and robots, children aged 5–6 years tend to overuse animistic interpretations for inanimate things, and to attribute biological and psychological properties to robots (Katayama et al., 2010; Di Dio et al., 2019, 2020a,b), in line with the results of this study. Interestingly, a difference in emotional attribution to NAO was also found between 5-year-olds and 7- and 9-year-olds children: younger children attributed lower emotions to NAO compared to the older ones. This result may seem counterintuitive in light of what has been discussed above; however, by finely looking at the scores obtained from the

5-year-olds for each single emotional question, it was found that younger children attributed significantly lower negative emotions to NAO compared to the other age groups, favoring positive emotions ($\chi2 < 0.01$). This resulted in an overall decrease of scores in the emotional dimension for the young children. Therefore, not only does this result not contradict the idea of a greater tendency to anthropomorphize robots in younger children compared to older ones, but also highlights that 5-years-olds perceive NAO as a positive entity that cannot express negative emotions such as anger, sadness, and fear: the "good" play-partner.

From the age of 7, children's belief of the robots' mind is significantly affected by a sensitivity to the type of the robot, as shown by differences between NAO and Robovie on most mental dimensions, except for Imaginative. The lack of differences between robots on the Imaginative dimension (for all age-groups), which encompasses psychological processes like pretending, and making jokes, appears to be regarded by children as a human prerogative. Interestingly, this result supports findings from a previous study (Di Dio et al., 2018) that compared 6-year-old children's mental state attribution to different entities (human, dog, robot, and God). Also, in that study, imagination was specific to the human entity.

Generally, the findings for older children indicate that the robot's appearance does affect mental state attribution to the robot, and this is increasingly evident with age. However, the judgment of older children could also be significantly influenced by the robot's behavioral characteristics, as demonstrated in

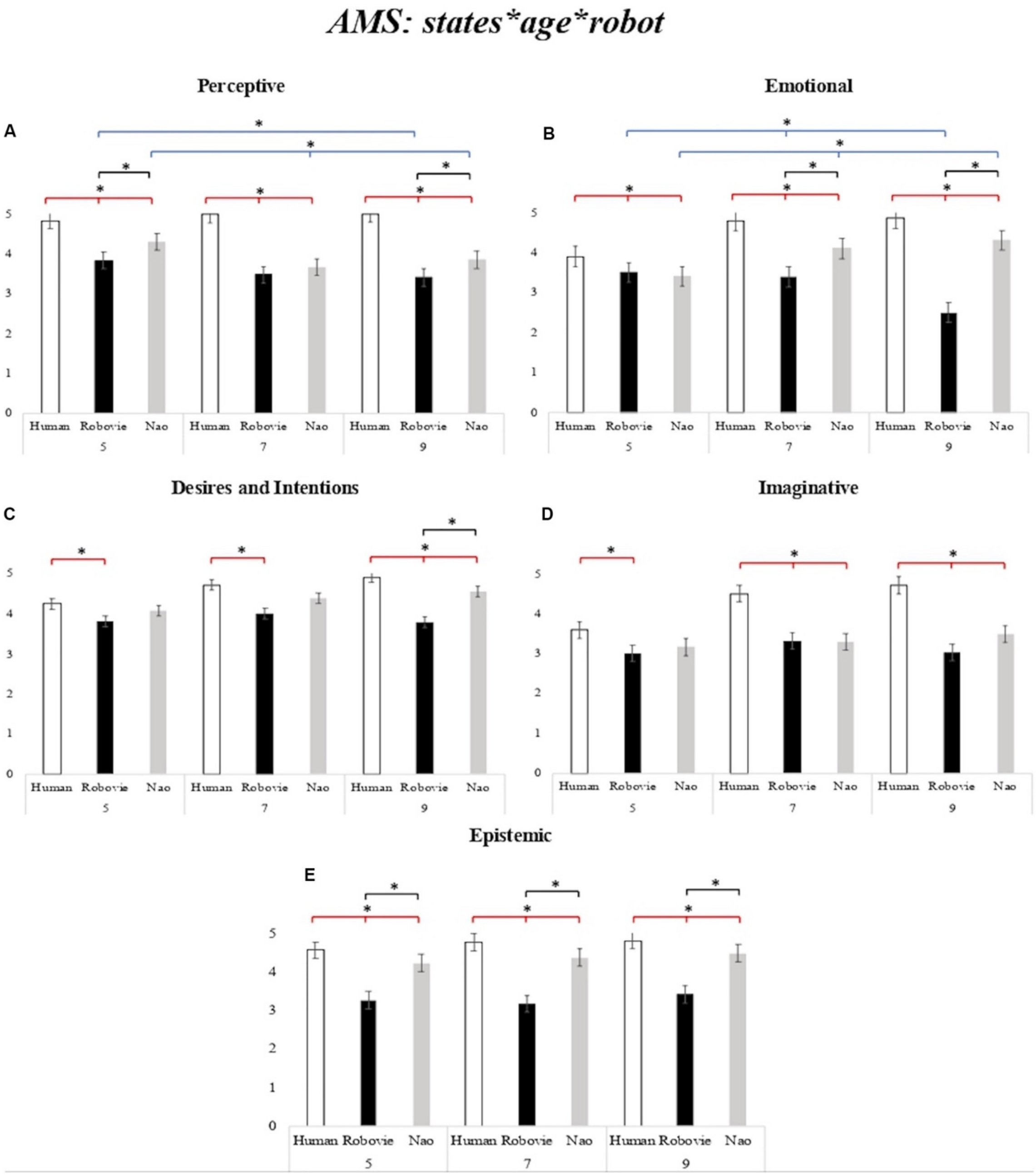


FIGURE 2 | (A–E) Children’s scores on the attribution of mental states (AMS) scale. AMS mean scores for the Human (white bar), for Robovie robot (black bar), and NAO robot (gray bar) for each state (Perceptive, Emotions, Intentions and Desires, Imagination, and Epistemic) as a function of age group (5-, 7-, and 9-year-olds). The bars represent the standard error of the mean. *Indicates significant differences. The red lines indicate the differences between agents (Human, Robot); the blue lines indicate the differences between ages (5-, 7-, and 9-year-olds); the black lines indicate the differences between robots (Robovie, NAO).

a long-term study conducted with children aged 10–12 years (Ahmad et al., 2016). In this study, children played a snakes and ladders game with a NAO robot three times across 10 days, whose behavior in terms of personality for a social robot in education was adapted to maintain and create long-term engagement and acceptance. It was found that children positively reacted to the use of the robot in education, stressing a need to implement robots that are able to adapt based on previous experiences in real time. Of course, this is very much in line with the great vision of disciplines such DR (Cangelosi and Schlesinger, 2015) and DC (Itakura, 2008). In this respect, it is also important to consider further aspects related to the effectiveness in human relations of constructs such as understanding the perspective of others (e.g., Marchetti et al., 2018) and empathy, on which several research groups are actively working. For example, in an exploratory study Serholt et al. (2014) highlighted the perceived need both for teachers and learners to deal with robots showing such a competence.

In the same vein, other studies that used Robovie as an interactive partner in educational contexts, have also shown that when the robot is programed to facilitate interactional dynamics with children, it can be considered by the children as a group member and even part of the friendship circle. In these studies, the robot is typically programed to act as an effective social communicative partner using strategies, like calling children by their name, or adapting the interactive behaviors for each child by means of behavioral patterns drawn from developmental psychology (Kanda et al., 2007; see also, Kahn et al., 2012). The study by Kahn et al. (2012) further showed that after interacting with Robovie, most children believed that Robovie had mental states (e.g., was intelligent and had feelings) and was a social being (e.g., could be a friend, offer comfort, and be trusted with secrets).

The above studies highlight the prospective use of robots, particularly in the educational field. However, in reality, today’s robots are not yet able to sustain autonomous behavior in the long term, even though research is actively laying a good foundation for this. What can certainly be worked on with direct effects on children’s perception of the mental abilities of robots are their physical attributes. By outlining differences in mental states attribution to different types of humanoid robots across ages based on robots’ physical appearance, these findings could help map the design of humanoid robots for children: in early ages, robots can display more abstract and mechanical features (possibly also due to the phenomenon of animism as described above); conversely, in older ages, the tendency to anthropomorphize robots is at least partially affected by the design of the robot. However, it has to be kept in mind that excessive human-likeness may be felt as uncomfortable, as suggested by findings showing that children experience less discomfort with robots displaying both human and mechanical features compared to robots whose physical features markedly evoke human ones (Bumby and Dautenhahn, 1999; Woods et al., 2004; Woods, 2006). Excessive resemblance to the human triggers the Uncanny Valley effect (the more the appearance of robots is similar to humans, the higher the sense of eeriness). These data suggest that a well-designed HSR for children should combine both human and mechanical dimensions, which, in this study, seems to be better represented by the NAO robot.

Conclusion

This study enabled us to analyze the AMS to two types of HSRs, highlighting how different types of robot can evoke different attributions of mental states in children. More specifically, these findings suggest that children’s age is an important factor to consider when designing a robot, and provided us with at least two important insights associated with the phenomenon of anthropomorphism from a development perspective, and the design of HSRs for children. Anthropomorphism seems to be a widespread phenomenon in 5-year-olds, while it becomes more dependent on physical features of the robot in older children, with a preference ascribed to the NAO robot that is perceived as more human-like. This effect may then influence the design of robots, which can be more flexible in terms of physical features, as with Robovie, when targeted to young children.

Overall, these results suggest that the assessment of HSRs in terms of mental states attribution may represent a useful measure for studying the effect of different robots’ design for children. However, it has to be noted that the current results involved only two types of HSRs. Therefore, future studies will have to evaluate the mental attribution to a greater variety of robots by also comparing anthropomorphic and non-anthropomorphic robots, and across different cultures. In addition, in future studies it will be important to assess children’s socio-cognitive abilities such as language, executive functions, and ToM, to analyze the effect of these abilities on the AMS to robots developmentally. Finally, this study explored the mental attributions through images depicting robots. Future studies should include a condition where children interact with the robots in vivo to explore the intersectional effect between the robot’s physical appearance and its behavioral patterns. This would enable us to highlight the relative weight of each factor on children’s perception of the robots’ mental competences.

Data availability statement

The datasets generated for this study are available on request to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethic Committee, Università Cattolica del Sacro Cuore, Milano, Italy. Written informed consent to participate in this study was provided by the participants’ legal guardian/next of kin.

Author contributions

All the authors conceived and designed the experiment. FM and GP conducted the experiments in schools. AM, FM, and GP secured ethical approval. FM and CDD carried out the statistical analyses. All authors contributed to the writing of the manuscript.

Funding

This publication was granted by Università Cattolica del Sacro Cuore of Milan.

References

Ahmad, M. I., Mubin, O., and Orlando, J. (2016). “Children views’ on social robot’s adaptations in education,” i*n Proceedings of the 28th Australian Conference on Computer-Human Interaction* (Tasmania: University Of Tasmania), 145–149.

Airenti, G. (2015). The cognitive bases of anthropomorphism: from relatedness to empathy. *Int. J. Soc. Robot.* 7, 117–127. doi: 10.1007/s12369-014-0263-x

Bartneck, C., Croft, E., and Kulic, D. (2008). “Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots,” *in Proceedings of the Metrics for Human-Robot Interaction Workshop in affiliation with the 3rd ACM/IEEE International Conference on Human- Robot Interaction (HRI 2008)*, Technical Report 471 (Piscataway, NJ: IEEE), 37–44.

Bartneck, C., and Forlizzi, J. (2004). “A design-centered framework for social human-robot interaction,” *in Preceedings of the RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication* (Piscataway, NJ: IEEE), 591–594.

Bartneck, C., Kanda, T., Mubin, O., and Al Mahmud, A. (2009). Does the design of a robot influence its animacy and perceived intelligence? *Int. J. Soc. Robot*. 1, 195–204. doi: 10.1007/s12369-009-0013-7

Begum, M., Serna, R. W., and Yanco, H. A. (2016). Are robots ready to deliver autism interventions? A comprehensive review. *Int. J. Soc. Robot*. 8, 157–181. doi: 10.1007/s12369-016-0346-y

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social robots for education: a review. *Sci. Robot*. 3:eaat5954. doi: 10.1126/ scirobotics.aat5954 doi: 10.1016/j.ijhcs.2011.04.003.

Broadbent, E. (2017). Interactions with robots: the truths we reveal about ourselves. *Annu. Rev. Psychol*. 68, 627–652. doi: 10.1146/annurev-psych-010416-043958

Broadbent, E., Kumar, V., Li, X., Sollers, J., Stafford, R. Q., MacDonald, B. A., et al. (2013). Robots with display screens: A robot with a more humanlike face display is perceived to have more mind and a better personality. *PLoS One* 8:e72589. doi: 10.1371/journal.pone.0072589

Bumby, K., and Dautenhahn, K. (1999). “Investigating children’s attitudes towards robots: A case study,” *in Proceedings of the CT99, The 3rd International Cognitive Technology Conference*, San Francisco, 391–410.

Cangelosi, A., and Schlesinger, M. (2015). *Developmental Robotics: from Babies to Robots*. London: MIT Press.

Cangelosi, A., and Schlesinger, M. (2018). From babies to robots: the contribution of developmental robotics to developmental psychology. *Child Dev. Perspect*. 12, 183–188. doi: 10.1111/cdep.12282

Dario, P., Guglielmelli, E., and Laschi, C. (2001). Humanoids and personal robots: design and experiments. *J. Robot. Syst*. 18, 673–690. doi: 10.1002/rob. 8106

Dautenhahn, K., Nehaniv, C. L., Walters, M. L., Robins, B., Kose-Bagci, H., Mirza, N. A., et al. (2009). KASPAR–a minimally expressive humanoid robot for human–robot interaction research. *Appl. Bionics Biomech*. 6, 369–397. doi: 10.1155/2009/708594

De La Cruz, V. M., Di Nuovo, A., Di Nuovo, S., and Cangelosi, A. (2014). Making fingers and words count in a cognitive robot. *Front. Behav. Neurosci*. 8:13. doi: 10.3389/fnbeh.2014.00013

Di Dio, C., Isernia, S., Ceolaro, C., Marchetti, A., and Massaro, D. (2018). Growing up Thinking of God's Beliefs: Theory of Mind and Ontological Knowledge. *Thousand Oaks, CA: Sage Open*, 1–14. doi: 10.1177/2158244018809874

Di Dio, C., Manzi, F., Itakura, S., Kanda, T., Hishiguro, H., Massaro, D., et al. (2019). It does not matter who you are: fairness in pre-schoolers interacting with human and robotic partners. *Int. J. Soc. Robot.* 1–15. doi: 10.1007/s12369- 019- 00528- 9

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., et al. (2020a). Come i bambini pensano alla mente del robot. Il ruolo dell'attaccamento e della Teoria della Mente nell'attribuzione di stati mentali ad un agente robotico. *Sistemi Intell*. 32, 41–56. doi: 10.1422/96279

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., et al. (2020b). Shall I trust you? From child human-robot interaction to trusting relationships. *Front. Psychol*. 11:469. doi: 10.3389/fpsyg.2020.00469

DiSalvo, C. F., Gemperle, F., Forlizzi, J., and Kiesler, S. (2002). "All robots are not created equal: The design and perception of humanoid robot heads," *Proceedings of the 4th conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*. (New York, NY: ACM).

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robot. Auton. Syst.*42, 177–190. doi: 10.1016/s0921-8890(02)00374-3

Epley, N., Waytz, A., and Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychol. Rev*. 114, 864–886. doi: 10.1037/0033-295x.114.4.864

Fink, J., Mubin, O., Kaplan, F., and Dillenbourg, P. (2012). "Anthropomorphic language in online forums about Roomba, AIBO and the iPad," *in Proceedings of the 2012 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*, Munich: IEEE, 54–59.

Hackel, L. M., Looser, C. E., and Van Bavel, J. J. (2014). Group membership alters the threshold for mind perception: the role of social identity collective identification and intergroup threat. *J. Exp. Soc. Psychol.* 2014, 15–23. doi: 10.1016/j.jesp.2013.12.001

Hood, D., Lemaignan, S., and Dillenbourg, P. (2015). “When children teach a robot to write: An autonomous teachable humanoid which uses simulated handwriting,” *in Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Piscataway, NJ: IEEE) 83–90.

Ishiguro, H., Ono, T., Imai, M., Kanda, T., and Nakatsu, R. (2001). Robovie: an interactive humanoid robot. *Int. J. Industrial Robot* 28, 498–503. doi: 10.1108/ 01439910110410051

Itakura, S. (2008). Development of mentalizing and communication: from viewpoint of developmental cybernetics and developmental cognitive neuroscience. *IEICE Trans. Commun.* 2109–2117. doi: 10.1093/ietcom/ e91- b.7.2109

Itakura, S., Ishida, H., Kanda, T., Shimada, Y., Ishiguro, H., and Lee, K. (2008). How to build an intentional android: infants’ imitation of a robot’s goal-directed actions. *Infancy* 13, 519–532. doi: 10.1080/15250000802329503

Kahn, P. H. Jr., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., et al. (2012). “Robovie, you’ll have to go into the closet now”: children’s social and moral relationships with a humanoid robot. *Dev. Psychol.* 48:303. doi: 10.1037/a0027033

Kanda, T., Ishiguro, H., Ono, T., Imai, M., and Mase, K. (2002). “Development and Evaluation of an Interactive Robot “Robovie”,” *in Proceedings of the IEEE International Conference on Robotics and Automation*, Washington, DC, 1848– 1855. doi: 10.1109/ROBOT.2002.1014810

Kanda, T., Sato, R., Saiwaki, N., and Ishiguro, H. (2007). A two-month field trial in an elementary school for long-term human–robot interaction. *IEEE Trans. Robot.* 23, 962–971. doi: 10.1109/TRO.2007.904904

Kannegiesser, P., Itakura, S., Zhou, Y., Kanda, T., Ishiguro, H., and Hood, B. (2015). The role of social eye-gaze in children’s and adult’s ownership attributions to robotic agents in three cultures. *Interact. Stud.* 16, 1–28. doi: 10.1075/is.16.1. 01kan

Katayama, N., Katayama, J. I., Kitazaki, M., and Itakura, S. (2010). Young children’s folk knowledge of robots. *Asian Cult. Hist.* 2:111.

Kiesler, S., and Goetz, J. (2002). “Mental models of robotic assistants,” *in Proceedings of the CHI’02 Extended Abstracts on Human Factors in Computing Systems* (New York, NY: ACM), 576–577.

Kiesler, S., Powers, A., Fussell, S. R., and Torrey, C. (2008). Anthropomorphic interactions with a robot and robot–like agent. *Soc. Cogn.* 26, 169–181. doi: 10.1521/soco.2008.26.2.169

Kose, H., and Yorganci, R. (2011). "Tale of a robot: Humanoid robot assisted sign language tutoring," *in Proceedings of the IEEE International Conference on Robotics and Automation2011 11th IEEE-RAS International Conference on Humanoid Robots* (Piscataway, NJ: IEEE), 105–111.

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., and Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS One* 3:e2597. doi: 10.1371/journal.pone.0002597

Lyon, C., Nehaniv, L. C., Saunders, J., Belpaeme, T., Bisio, A., Fischer, K., and Cangelosi, A. (2016). Embodied language learning and cognitive bootstrapping: methods and design principles. *Int. J. Adv. Robotic Syst.* 13:105. doi: 10.5772/ 63462

Macdonald, S. J., Donovan, C., and Clayton, J. (2017). The disability bias: understanding the context of hate in comparison with other minority populations. *Disabil. Soc.* 32, 483–499. doi: 10.1080/09687599.2017.1304206

MacDorman, K. F. (2006). "Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley," *in Proceedings of the ICCS/CogSci-2006 Long Symposium: Toward social Mechanisms of Android Science*, Vancouver, 26–29.

MacDorman, K. F., and Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interact. Stud.* 7, 297–337. doi: 10.1075/is.7.3.03mac

MacDorman, K. F., Minato, T., Shimada, M., Itakura, S., Cowley, S., and Ishiguro, H. (2005). "Assessing human likeness by eye contact in an android testbed," *in Proceedings of the 27th annual meeting of the cognitive science society* (Austin, TX: CSS) 21–23.

Manzi, F., Ishikawa, M., Di Dio, C., Itakura, S., Kanda, T., Ishiguro, H., et al. (2020a). The understanding of congruent and incongruent referential gaze in 17-month-old infants: An eye-tracking study comparing human and robot. *Sci. Rep.* 10:11918. doi: 10.1038/s41598-020-69140-6

Manzi, F., Massaro, D., Kanda, T., Tomita, K., Itakura, S., and Marchetti, A. (2017). "Teoria della Mente, bambini e robot: l'attribuzione di stati mentali," *in Proceedings of the Abstract de XXX Congresso Nazionale, Associazione Italiana di Psicologia, Sezione di Psicologia dello Sviluppo e dell'Educazione* (Messina, 14-16 September 2017), Italy: Alpes Italia srl, 65–66. Available online at: http: //hdl.handle.net/10807/106022

Manzi, F., Savarese, G., Mollo, M., and Iannaccone, A. (2020b). Objects as communicative mediators in children with autism spectrum disorder. *Front. Psychol.* 11:1269. doi: 10.3389/fpsyg.2020.01269

Marchetti, A., Di Dio, C., Manzi, F., and Massaro, D. (in press). "*Robotics in clinical and developmental psychology*," in *Comprehensive Clinical Psychology*, 2nd Edn, Amsterdam: Elsevier.

Marchetti, A., Di Dio, C., Manzi, F., and Massaro, D. (2020). The psychosocial fuzziness of fear in the COVID-19 era and the role of robots. *Front. Psychol.* 10:2245. doi: 10.3389/fpsyg.2020.02245

Marchetti, A., Manzi, F., Itakura, S., and Massaro, D. (2018). Theory of Mind and humanoid robots from a lifespan perspective. *Z. Psychol.* 226, 98–109. doi: 10.1027/2151-2604/a000326

Martini, M. C., Gonzalez, C. A., and Wiese, E. (2016). Seeing minds in others – Can agents with robotic appearance have human-like preferences? *PLoS One* 11:e0146310. doi: 10.1371/journal.pone.0146310

Mori, M. (1970). The uncanny valley. *Energy* 7, 33–35.

Mori, M., MacDorman, K. F., and Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robot. Autom. Mag.* 19, 98–100. doi: 10.1109/mra.2012.2192811

Moriguchi, Y., Kanda, T., Ishiguro, H., Shimada, Y., and Itakura, S. (2011). Can young children learn words from a robot? *Interact. Stud.* 12, 107–119. doi:10.1075/is.12.1.04mor

Morse, A. F., and Cangelosi, A. (2017). Why are there developmental stages in language learning? A developmental robotics model of language development. *Cognitive Sci.* 41, 32–51. doi: 10.1111/cogs.12390

Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., and Dong, J. J. (2013). A review of the applicability of robots in education. *J. Technol. Educ. Learn.* 1:13.

Murashov, V., Hearl, F., and Howard, J. (2016). Working safely with robot workers: Recommendations for the new workplace. J. *Occup. Environ. Hyg.* 13, D61–D71.

Okanda, M., Taniguchi, K., and Itakura, S. (2019). "The role of animism tendencies and empathy in adult evaluations of robot," *in Proceedings of the 7th International Conference on Human-Agent Interaction, HAI '19* (New York, NY: Association for Computing Machinery), 51–58. doi: 10.1145/3349537.3351891

Okanda, M., Zhou, Y., Kanda, T., Ishiguro, H., and Itakura, S. (2018). I hear your yes−no questions: Children's response tendencies to a humanoid robot. *Infant Child Dev.* 27:e2079. doi: 10.1002/icd.2079

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., and Itakura, S. (2013a). Can infants use robot gaze for object learning? The effect of verbalization. *Int. Stud.* 14, 351–365. doi: 10.1075/is.14.3.03oku

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., and Itakura, S. (2013b). Infants understand the referential nature of human gaze but not robot gaze. *J. Exp. Child Psychol.* 116, 86–95. doi: 10.1016/j.jecp.2013.02.007

Perner, J., and Wimmer, H. (1985). "John thinks that Mary thinks that." attribution of second-order beliefs by 5- to 10-year- old children. *J. Exp. Child Psychol.* 39, 437–471. doi: 10.1016/0022-0965(85)90051-7

Phillips, E., Zhao, X., Ullman, D., and Malle, B. F. (2018). "What is human-like? Decomposing robots' human-like appearance using the anthropomorphic roBOT (ABOT) database," *in Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. (Piscataway, NJ: IEEE).

Piaget, J. (1929). *The Child's Conception of the World*. London: Routledge.

Powers, A., and Kiesler, S. (2006). "The advisor robot: tracing people's mental model from a robot's physical attributes," *in Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction* (New York, NY: ACM), 218–225.

Premack, D., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind. *Behav. Brain Sci.* 1, 515–526. doi: 10.1017/s0140525x000 76512

Sarti, D., Bettoni, R., Offredi, I., Tironi, M., Lombardi, E., Traficante, D., et al. (2019). Tell me a story: socio-emotional functioning, well-being and problematic smartphone use in adolescents with specific learning disabilities. *Front. Psychol.* 10:2369. doi: 10.3389/fpsyg.2019.02369

Serholt, S., Barendregt, W., Leite, I., Hastie, H., Jones, A., Paiva, A., et al. (2014). "Teachers' views on the use of empathic robotic tutors in the classroom," *in Preceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication* (Piscataway, NJ: IEEE), 955–960.

Shamsuddin, S., Yussof, H., Ismail, L. I., Mohamed, S., Hanapiah, F. A., and Zahari, N. I. (2012). Initial response in HRI-a case study on evaluation of child with autism spectrum disorders interacting with a humanoid robot Nao. *Procedia Eng.* 41, 1448–1455. doi: 10.1016/j.proeng.2012. 07.334

Sharkey, A., and Sharkey, N. (2011). Children, the elderly, and interactive robots. *IEEE Robot. Autom. Mag.* 18, 32–38. doi: 10.1109/mra.2010.94 0151

Shiomi, M., Kanda, T., Ishiguro, H., and Hagita, N. (2006). "Interactive humanoid robots for a science museum," *in Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction* (New York, NY: ACM) 305–312.

Slaughter, V., Peterson, C. C., and Carpenter, M. (2009). Maternal mental state talk and infants' earlygestural communication. *J. Child Lang.* 36, 1053–1074. doi: 10.1017/S0305000908009306

Stafford, R. Q., MacDonald, B. A., Jayawardena, C., Wegner, D. M., and Broadbent, E. (2014). Does the robot have a mind? Mind perception and attitudes towards robots predict use of an eldercare robot. *Int. J. Soc. Robot.* 6, 17–32. doi: 10.1007/s12369- 013- 0186- y

Tielman, M., Neerincx, M., Meyer, J. J., and Looije, R. (2014). "Adaptive emotional expression in robot-child interaction," *in Proceedings of the 2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Piscataway, NJ: IEEE), 407–414.

Todd, A. R., Bodenhausen, G. V., Richeson, J. A., and Galinsky, A. D. (2011). Perspective taking combats automatic expressions of racial bias. *J. Pers. Soc. Psychol.* 100, 1027–1042. doi: 10.1037/a0022308

Tung, F. W. (2016). Child perception of humanoid robot appearance and behavior. *Int. J. Hum. Comput. Interact.* 32, 493–502. doi: 10.1080/10447318.2016. 1172808

Vinanzi, S., Patacchiola, M., Chella, A., and Cangelosi, A. (2019). Would a robot trust you? Developmental robotics model of trust and theory of mind. *Philos. Tr. R. Soc. B* 374:20180032. doi: 10.1098/rstb.2018.0032

Wainer, J., Robins, B., Amirabdollahian, F., and Dautenhahn, K. (2014). Using the humanoid robot KASPAR to autonomously play triadic games and facilitate collaborative play among children with autism. *IEEE Trans. Auton. Ment. Dev.* 6, 183–199. doi: 10.1109/tamd.2014.2303116

Wang, Y., Park, Y. H., Itakura, S., Henderson, A. M. E., Kanda, T., Furuhata, N., et al. (2020). Infants' perceptions of cooperation between a human and robot. *Infant Child Dev.* 29:e2161. doi: 10.1002/icd.2161

Wellman, H. M., Cross, D., and Watson, J. (2001). Meta-analysis of theory-of- mind development: the truth about false belief. *Child Dev.* 72, 655–684. doi: 10.1111/1467- 8624.00304

Wimmer, H., and Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128. doi: 10.1016/0010-0277(83)90 004-5

Woods, S. (2006). Exploring the design space of robots: Children's perspectives. *Interact. Comput.* 18, 1390–1418. doi: 10.1016/j.intcom.2006. 05.001

Woods, S., Dautenhahn, K., and Schulz, J. (2004). "The design space of robots: Investigating children's views," *in Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication* (Piscataway, NJ: IEEE), 47–52.

Yogeeswaran, K., Złotowski, J., Livingstone, M., Bartneck, C., Sumioka, H., and Ishiguro, H. (2016). The interactive effects of robot anthropomorphism and robot ability on perceived threat and support for robotics research. *J. Hum. Robot Interact*. 5, 29–47. doi: 10.5898/jhri.5.2.yogeeswaran

Zanatto, D., Patacchiola, M., Cangelosi, A., and Goslin, J. (2020). Generalisation of Anthropomorphic stereotype. *Int. J. Soc. Robot*. 12, 163–172. doi: 10.1007/ s12369- 019- 00549- 4

Zanatto, D., Patacchiola, M., Goslin, J., and Cangelosi, A. (2016). "Priming anthropomorphism: can the credibility of humanlike robots be transferred to non-humanlike robots?" *in Proceedings of the 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI),* Christchurch: IEEE, 543–544. doi: 10.1109/HRI.2016.7451847

Zanatto, D., Patacchiola, M., Goslin, J., and Cangelosi, A. (2019). Investigating cooperation with robotic peers. *PLoS One* 14:e0225028. doi: 10.1371/journal. pone.0225028

Zhang, T., Zhu, B., Lee, L., and Kaber, D. (2008). "Service robot anthropomorphism and interface design for emotion in human-robot interaction," *in Proceedings of the 2008 IEEE International Conference on Automation Science and Engineering* (Piscataway, NJ: IEEE), 674–679.

Zhong, J., Ogata, T., Cangelosi, A., and Yang, C. (2019). Disentanglement in conceptual space during sensorimotor interaction. *Cogn. Comput. Syst*. 1,103–112. doi: 10.1049/ccs.2019.0007

Złotowski, J., Proudfoot, D., Yogeeswaran, K., and Bartneck, C. (2015). Anthropomorphism: opportunities and challenges in human–robot interaction. *Int. J Soc. Robot*. 7, 347–360. doi: 10.1007/s12369-014- 0267-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

CHAPTER 5

Coding with me: exploring the effect of coding intervention on preschoolers' cognitive skills

Giulia Peretti[a-b,1], Daniela Villani[b], Michele Marangi[c], Federica Pelizzari[c], Sara DI Bruno[d], Igor Guida[d], Antonella Marchetti[a-b], Giuseppe Riva[b]; Pier Cesare Rivoltella[c], Davide Massaro[a-b]

[a] *Research Unit on Theory of Mind, Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy*
[b] *Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy*
[c] *Cremit, Department of Pedagogy, Università Cattolica del Sacro Cuore, Milan, Italy*
[d] *Stripes Cooperativa Sociale Onlus, Rho, Italy*

[1] *Corresponding author*

Published on Annual Review of Cybertherapy and Telemedicine, summer 2020

Keywords

Coding, Computational Thinking, educational robotic, cognitive skills

Abstract

In the last ten years, the topic of Computational Thinking (CT) has been gaining increasing attention from researchers in the education field. Starting from kindergarten, increasingly programming activities such as coding and educational robotics are proposed to enhance CT and some cognitive skills, such as problem- solving, spatial and reasoning skills. The most commonly used tools are the so- called tangible interfaces, such as floor-robots (e.g. Cubetto, Bee and Blue-Bot and others), through which children can interact with the object and learn playfully. Investigating the effects of CT activities on children's cognitive abilities is important to understand the impact in kindergarten and to comprehend in which developmental periods these activities might be most successful. The aim of the present study is that of evaluating the effect of a coding intervention, based on CT, through the use of the Cubetto robot, on the cognitive skills of 4-years-old children. The coding intervention included three sessions and required the manipulation of physical objects to plan and conduct a Cubetto journey. Results showed that children of the experimental group performed better than those of the control group in programming the Cubetto path after the intervention.

Introduction

In the last ten years, the topic of Computational Thinking (CT), such as a thought process that uses analytic and algorithmic approaches to formulate, analyze and solve problems (Wing 2006; 2008; 2010), has been gaining increasing attention from researchers in the education field. In particular, several studies implementing coding and educational robotics activities in kindergarten found significant influence between CT and some cognitive skills, such as problem-solving, spatial and reasoning skills and short-term memory (Bers, 2008; 2010; Bers et al., 2002; Kazakoff et al., 2013; Nulli & Di Stasio, 2017; Rogers & Portsmore, 2004; Sung et al., 2017).

Thus, starting from kindergarten programming activities such as coding and educational robotics can allow children to become code-literate, that is to be able to read, write and think through the computer language and to be able to think in a computational way (Román-González et al., 2017). The most commonly used tools to sustain CT are the so-called tangible interfaces, such as floor robots (e.g. Cubetto, Bee and Blue-Bot and others), through which children can manipulate the object with which they are interacting and understand the activity they are performing step by step (Manches & Plowman, 2017; Nulli & Di Stasio, 2017). Specifically, children are required to program he correct sequence of actions to achieve a specific goal, thus enhancing sequential skills that are recognized as very important for cognitive development since the early stages of kindergarten (Kazakoff et al., 2013).

Moreover, in addition to the most current reflection on the use of digital media in education, literature has highlighted how the use of story-telling tasks promotes the development of narrative thinking and other relevant skills such as Theory of Mind and perspective-taking skills (Paris & Paris, 2003). Story-telling tasks, based on a sequential logic similar to the programming language, can be used to support and integrate robotic programming learning in education.

Generally, in the field of computer technology, educational robotics activities are structured according to a playful approach and appropriate to the child's developmental phase (Bers & Horn, 2010; Kazakoff & Bers, 2014), thus sustaining independent learning and discovery (Bers et al., 2019). Coding activities and educational robotics can be seen as real constructivist programming environments, in which children are encouraged to reflect on their thinking processes through activities in which abstract ideas are concretely and precisely conceived (Alimisis & Kynigos, 2009; Kazakoff et al., 2013; Papert, 1980).

Moreover, in addition to stimulating the cognitive abilities of the single child, such as problem-solving, spatial skills, reasoning skills and short-term memory, the use of robotics in kindergarten promotes different types of learning, such as new ways of social interaction with peers and opportunities for social and cognitive development (Kazakoff & Bers, 2014). Furthermore, the use of educational robots can stimulate the potential development level that the child can reach through social support from more expert peers, teachers and educators, that play the role of real scaffolders (Kazakoff & Bers, 2014).

Exploring the effects of CT activities on children's cognitive skills is important to evaluate the impact of their introduction at infancy school. This paper aims to contribute to this reflection, by exploring how instructional activities that teach the initial elements of CT by guided exposure to coding can boost the development of cognitive skills of 4- years-old children. The coding intervention included three sessions and required the manipulation of physical objects with symbolic meaning to plan and conduct a robot journey, thus stimulating the children visual spatial and story-telling ability.

Methods

Participants

The experimental study involved a sample of 40 children aged 4, attending three different kindergartens in the province of Milan, The children's parents received a detailed description of the study and expressed their written consent for their child participation to the research. All tasks selected were deemed appropriate for age of subjects participating in this experimental study and approved by the Ethics Committee of the Department of Psychology of Università Cattolica del Sacro Cuore of Milan.

Apparatus and Material

The children were divided into two groups: 1) an experimental group (N=20, 11 females) that followed the coding intervention and whose children were evaluated individually before and after the intervention; 2) a waiting list group (N=20, 10 females) whose children were evaluated at the same time but followed the intervention after the experimental group. The coding intervention lasted four weeks and consisted of three 60-minute laboratory sessions scheduled during the regular kindergarten day. It involved the use of Cubetto, an innovative tangible coding technology that facilitates young children's engagement with basics of coding, founded on the principle of visual programming (i.e. the child can design a route directly with his/her own hands without the use of a computer, yet it incorporates traditional play elements such as patterns, colour recognition and shape sorting). Specifically, children were assisted in defining Cubetto's orientation and the direction needed to reach a specific target throughout subsequent path episodes.

The assessment phase lasted about 20-25 minutes per child both pre (T0) and post (T1) intervention. Each child was invited to actively participate to the assessment sessions and was instructed to carefully listen to the instructions, ask questions when in doubt and perform the test according to his/her own skills. Specifically, the assessment included: an adapted version of Understanding of pictures stories test (Baron-Cohen et al., 1986), used to assess the child's ability to reorder images with a predetermined sequence; the Children's Mental Transformation Task (Hawes et al., 2019) used to evaluate the children visual-spatial abilities according to their cognitive development; an ad hoc task that involved the use of Cubetto, to observe what specific methods of achieving a target were used by the child during the task. Children involved in the test were asked to program Cubetto's path over the map using a keyboard in which they could insert coloured cards, where the different colours (green to go forward, red to turn right and yellow to turn left) represented the useful commands to move Cubetto on the map. Drawings were shown on the map that could be used to create educational stories or, as in this case, to construct coding paths to move Cubetto. The task involved the programming, by the children, three Cubetto paths for achieving a precise target on the Cubetto's map, previously indicated by the researcher through the description of a story divided into three parts. Following the preliminary instructions to perform the task, Cubetto was placed on a precise starting point on the map, equal for all, and the following three paths were narrated one by one to enable the child to focus on a single path. Each path started from the previously achieved end point . The final score of the Cubetto task ranged from 0 to 3: 1 point was allocated when the child was able to reach the target of the path. The total score is given by the sum of the three paths'points.

Results

To test the effect of coding intervention on children understanding of stories and visual-spatial abilities, A 2 (groups) x 2 (pre vs. post) repeated - measures ANOVA was performed. The primary purpose of two-way repeated measures ANOVA is to understand whether there is an interaction between these two factors on the dependent variable. In this case, results did not show significant interaction effects both related to children ability to reorder images with a predetermined sequence [F (1, 41) = 1.246, *p* = .27] and to children visual-spatial ability [F (1, 41) = .008, *p* = .93]. Thus, after the coding intervention children did not significantly improve these abilities.

Furthermore, to test the effect of coding intervention on children abilities to program Cubetto to achieve a target on the map, A 2x2 repeated measures ANOVA was performed with the Cubetto task score as the dependent variable. In this case, it was found both a significant time effect [F (1, 41) = 30.022, *p* = .001] in both groups and a significant time x group interaction effect [F (1, 41) = 9.825, *p* = .003]. This means that both the experimental and control groups increase their abilities in programming Cubetto to achieve the goal, but children of experimental group performed better than those of control group.

Discussion

In this study, the effects of a coding intervention, using the floor robot Cubetto, on some cognitive skills, such as sequential ability to reconstruct stories and visual-spatial skills, in 4-years-old children were investigated. No differences existed between the two groups at T0. The statistical analysis showed both a significant main effect of time (pre vs post) and interaction effect (time x group) on the sequential programming abilities measured by the ad hoc Cubetto task. Thus, the experimental group performed better than the control group in programming the Cubetto path after the intervention. Nevertheless, no significant interaction effects were found related to the children understanding of pictures stories and visuo-spatial skills.

A plausible explanation of the result obtained with the Cubetto task could be linked to the learning effect of a specific task. Various cognitive competences related to that specific type of learning are probably involved but no generalization effect was found. Unfortunately, the research focusing on the age considered by this study (4 years-old) is still too limited to draw conclusions on the beneficial effects of coding for these younger learners, whose cognitive skills are still largely immature. Furthermore, other studies (Arfé et al., 2020) involving 5-6 years children proposed a longer intervention (8 sessions over four weeks) and found significant improvements in two executive functions: planning and response inhibition. Thus, future studies are encouraged in testing the effects of a coding intervention whose duration and articulation is greater in sustaining the interconnections between specific programming skills and children cognitive abilities.

References

Alimisis, D., Kynigos, C. (2009). Constructionism and robotics in education. *Teacher education on robotic-enhanced constructivist pedagogical methods*; 11-26.

Arfé, B., Vardanega, T., & Ronconi, L. (2020). The effects of coding on children's planning and inhibition skills. *Computers & Education*, *148*, 103807, https://doi.org/10.1016/j.compedu.2020.103807

Baron-Cohen, S., Leslie, A.M., Frith, U. (1986). Mechanical, behavioural and Intentional understanding of picture stories in autistic children, *British Journal of Developmental Psychology*; 4, 113-125. https://doi.org/10.1111/j.2044-835X.1986.tb01003.x

Bers, M. U. (2008). *Blocks to robots: Learning with technology in the early childhood classroom*. New York, NY: Teacher's College Press.

Bers, M.U. (2010) The TangibleK robotics program: Applied computational thinking for young children. *Early Childhood Research & Practice*; 12(2). Retrieved from http://ecrp.uiuc.edu/v12n2/bers.html.

Bers, M. U., González-González, C., Armas-Torres, M. B.(2019). Coding as a playground: Promoting positive learning experiences in childhood classrooms, *Computers & Education*; 138, 130-145. https://doi.org/10.1016/j.compedu.2019.04.013

Bers, M., Horn, M. (2010). *Tangible programming in early childhood: Revisiting developmental assumptions through new technologies*. In. Berson IR, Berson MJ (Eds), High-tech tots: Childhood in a digital world Greenwich, CT: Information Age Publishing; 2010. 49-70.

Bers, M. U., Ponte, I., Juelich, K., Viera, A., Schenker, J. (2002). Teachers as designers: Integrating robotics into early childhood education. *Information Technology in Childhood Education Annual*; 14, 123–145.

Hawes, Z., Moss, J., Caswell, B., Seo, J., Ansari, D.(2019). Relations between numerical, spatial and executive function skills and mathematics achievement: A latent-variable approach, *Cognitive Psychology*; 109, 68-90. https://doi.org/10.1016/j.cogpsych.2018.12.002

Kazakoff, E. R., Bers, M. U. (2014) Put your robot in, put your robot out: Sequencing through programming robots in early childhood, *Journal of Educational Computing Research*; *50*(4), 553-573. https://doi.org/10.2190%2FEC.50.4.f

Kazakoff, E. R., Sullivan, A., Bers, M. U. (2013). The effect of a classroom-based intensive robotics and programming workshop on sequencing ability in early childhood. *Early Childhood Education Journal*; 41(4), 245-255. https://doi.org/10.1007/s10643-012-0554-5

Manches, A., Plowman, L. (2017). Computing education in children's early years: A call for debate. *British Journal of Educational Technology*; 48(1), 191-201.

https://doi.org/10.1111/bjet.12355

Nulli, G. & Di Stasio, M. (2017). Coding alla scuola dell'infanzia con docente esperto della scuola primaria.*Italian Journal of Educational Technology, 25*(2), 59-65. Ortona, Italy: Edizioni Menabò - Menabò srl. Retrieved October 26, 2021 from https://www.learntechlib.org/p/183471/.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York, NY: Basic Books.

Paris, A. H., Paris, S. G.(2003) Assessing narrative comprehension in young children, *Reading Research Quarterly*; 38(1), 36-76. https://doi.org/10.1598/RRQ.38.1.3

Román-González, M., Pérez-González, J-C., Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test, *Computers in Human Behavior*; 678-691. http://dx.doi.org/10.1016/j.chb.2016.08.047

Rogers, C., Portsmore, M.(2004). Bringing engineering to elementary school. *Journal of STEM Education*; 5(3–4), 14–28.

Sung, W., Ahn, J, Black, J. B. (2017). Introducing Computational Thinking to Young Learners: Practicing Computational Perspective Trough Embodiment in Mathematics Education, *Technology, Knowledge, and Learning*; 1-28. https://doi.org/10.1007/s10758-017-9328-x

Wing, J. M. (2006). Computational Thinking. *Communications of the ACM*; 49(3), 33–35. https://doi.org/10.1145/1118178.1118215

Wing, J. M. (2008). Computational Thinking and Thinking about Computing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*; 366(1881), 3717–3725. https://doi.org/10.1098/rsta.2008.0118

Wing, J. M. (2010). Computational Thinking: What and Why? Link Magazine. Retrived from:https://www.cs.cmu.edu/link/research-notebook-computational-thinking-what-and-why

CHAPTER 6

General Discussion

Through this thesis I aimed to introduce and explore a fairly new field of psychological research specifically dealing with the child-robot interaction, eviscerating the multiple nuances that characterize it and that pertain to important psychological constructs, such as the Theory of Mind and trust, as well as the affective correlates of the relationship, investigated through the attachment relationship. This without neglecting the value of the specific design of the robots considered for the studies presented in this work. First of all, the theoretical framework showed that children's cognitive development has a significant relevance within the interactional process with robots. In particular, I observed that the development of mentalization skills and the understanding of false beliefs play a significant role. I had the opportunity to assess their relevant role within the first study about trust, in which the development of the understanding of false belief is shown to be important for the construction of future relationships based on trust. In particular, in the study designed to observe the processes of acquisition, loss and restoration of trust in two different agents (human and robot), the development of mentalization skills made the children able to reason more rationally about the fact that in the *Trusting Game*, in which children were asked to correctly identify the box under which the puppet was placed, the other player (human or robot) had the same probability of guessing as them, and also of making a mistake exactly like them, moderating the effect of the affective component of trust. This result is important as it can provide information on how the cognitive architectures of a robot, in particular, I'm referring to disciplines such as Developmental Robotics, can be modelled to make it trustful towards the human. Concerning the second study presented about children's understanding of the distinction between an intentionally false and an unintentionally false statement, I observed how the understanding of intentionality and the acquisition of a first-order false belief are crucial for children to understand the concepts of lies and mistakes and, therefore, to be able to distinguish more accurately between intentionally and unintentionally false statements. Moreover, by refining their mental abilities, children can make a more cogent moral assessment of lies and mistakes. An interesting finding of children who had acquired first-order ToM was that they were more accurate than children who had not yet acquired ToM in distinguishing between intentionally and unintentionally false statements when observed in the human. In contrast, regarding the distinction of the same statements in the robot, children who had acquired ToM responded randomly when attributing a lie and made the subsequent moral evaluation. What does this result indicate? I assumed that children who had acquired first-order ToM had a good understanding of intentionality in humans, and thus could correctly distinguish between an intentionally or unintentionally false statement, the answers given for the robot would indicate that these same children did not attribute fully intentional behaviour to the robot. Children who have acquired first-order ToM appear more likely to judge the robot’s behaviour as more acceptable because they do not consider it to be completely intentional. This result shows how important the acquisition or not of the ToM appears to be in understanding and attributing intentional behaviour to another non-human agent, such as the robot. The third study presented allowed me to continue this thread concerning the importance of understanding and attributing mental states to a robotic agent. Results from this work show how two humanoid social robots with different design characteristics, NAO robot and Robovie, can evoke different attributions of mental states by children. Here, I could observe how age, and the maturation of mentalistic competence, are elements to strongly consider when designing a robot. The mentalistic evaluation of robots represents an important indication to bear in mind regarding the effect of the design of robots on children, as well as taking into account the age of the subjects with whom the robots might interact, and also the different contexts of use.

These first three studies bring out interesting and stimulating reflections regarding the applicative implications of using and designing robotic agents conceived for a developing user, namely the child. The observed elements take on a specific value since they allow us to understand, or at least to start a more refined understanding process, on how the robot can be concretely and functionally inserted and acted within multiple contexts, from educational to clinical, according to its characteristics and what it elicits in the child at different stages of development and how these elements can be effectively functional from a developmental point of view. For example, I noted that the robotic partner is less susceptible to the dynamics of the attachment relationship and therefore appears to be a more stable partner of trust. Therefore, this could be used with children who have difficulties in attributing trust, due to difficult relational histories. In this way, the robot could be used, as a mediator, to play a supportive role in improving relationships between human subjects. Having in mind the mind of the child who interacts with the robotic agent, which carries features that are perceived with different meanings according to the age of the mind that is observing it, appears fundamental when proposing an activity within an educational context, not only for the child but also for the adult who is next to him/her. What is meant by this? I mean that the research field must be able to show the potentialities of the robotic tool inserted in a specific context, such as the educational one, presenting its multiple application nuances, i.e. showing how to use it to promote processes that have concrete effects on development and cognitive maturation, also to people who do not work in research contexts. In this way, therefore, the tool will be correctly literate by all subjects interacting with it, from children to adults, and used constructively. This reflection leads to the analysis of the latest research.

This finally work, based on the analysis of the effect of a coding activity related to Computational Thinking on the cognitive skills of 4-year-olds children, allowed me to observe the robotic artefact under another guise, namely within the framework of educational robotics. In this context, the children were able to manipulate the robotic object, explore it, understand it. They had their first exploratory encounter with it. I had the opportunity to observe the robot in a different role from the one observed in the first three studies, a more didactic-educational role, which assumes an important value as it allows us to understand how a robotic tool if it presents a well-thought-out activity and is adequate for the age of the subjects it is proposed to, could prove useful for cognitive development. The studies I have outlined represent the many facets of child-robot interaction. A kaleidoscope of pieces that make up the final image, which is the interaction between children and robots. Each of these components appears fundamental to understanding this type of interaction, to grasping its potential and richness in terms of knowledge. This is done through the exploration of trust, the attribution of mental states up to the direct programming of an educational robot. All this relates to child-robot interaction.

The exploration of the interaction between children and robots, precisely because of the characteristics of the interacting subjects, has proved and still is, fascinating, for me. Paraphrasing Dickinson[3], child-robot interaction, like the mind, is deeper than the sea because, putting them side by side, the child and the robot, one contains the other and brings with it an infinity of facets, marked above all by a subject that is maturing, that evolves year after year, that enriches the interaction with another agent precisely thanks to its maturational changes. While the other, the robotic agent, brings with it different elements, such as the design itself, which make it observable, and to which mental states can be attributed, from multiple eyes, minds that are growing. When a child meets a robot, worlds are generated, multiple interactional worlds. We should take on the role of Actaeon[4] seen, as it was described by Giordano Bruno in the myth telling of the encounter between Diana and Actaeon, as a metaphor for the human intellect moved by scientific curiosity, without fear, towards knowledge, to be increasingly moved by the desire for discovery and scientific knowledge to continue to explore these interactional worlds.

---

[3] Citation reference in Marchetti, A., & Massaro, D. (2012). *Capire la mente. La psicologia ingenua del bambino.* Carocci Editore, Roma

[4], Referring to the myth of Actaeon and Diana as revised by Giordano Bruno, in "*Giordano Bruno. Il mito di Atteone*" by Giulio Giorello (2013), Perle di Saggezza, AlboVersorio.

Publications not include in this thesis

1. Peretti, G., Villani, D., Marangi, M., Pellizzari, F., Di Bruno, S., Guida, I., Marchetti, A., Riva, G., Rivoltella P.C., Massaro, D. (Submitted). *Il viaggio di Cubetto: una proposta di coding per sviluppare il Pensiero Computazionale nella Scuola dell'Infanzia.* Pedagogika.
2. Manzi, F., Peretti, G., Cangelosi, A., Itakura, S., Kanda, T., Ishiguro, H., Di Dio, C. (Settembre, 2021). *I robot non sono tutti uguali: l'attribuzione di stati mentali a diversi robot umanoidi da parte dei bambini.* Paper presentato al XXXIII Congresso AIP Sezione di Psicologia dello Sviluppo e dell'Educazione, Bari, Settembre, 20-23, 2021.
3. Peretti, G., Villani, D., Marangi, Pelizzari, Di Bruno, Guida, Marchetti, A., Riva, G., Rivoltella, P.C., Massaro, D.(Settembre, 2021).*Cubetto's journey: attività e ricerca per sviluppare il pensiero computazionale a partire dalla scuola dell'infanzia.*Paper presentato al XXXIII Congresso AIP Sezione di Psicologia dello Sviluppo e dell'Educazione, Bari, Settembre, 20-23, 2021.
4. Peretti, G., Manzi, F., Di Dio, C., Cangelosi, A., Massaro, D., Marchetti, A. (Settembre, 2021). *Can a robot lie? A study on intentionality comprehension in 5- and 6-year-olds children.* Paper presentato a 25th Annual International CyberPsychology, CyberTherapy & Social Networking Conference (CYPSY25).
5. Di Dio, C., Peretti, G., Marchetti, A., Massaro, D. (Aprile, 2021). *Lo sviluppo della comprensione di Dio: una prospettiva cross-culturale.* Intervento presentato al seminario online "Religiosità, spiritualità e funzionamento psicologico positivo" – Università Cattolica del Sacro Cuore, 28 Aprile 2021.
6. Peretti, G., Manzi, F., Di Dio, C., Cangelosi, A., Massaro, D., Marchetti, A. (Aprile, 2021). *Un robot può mentire? Uno studio sulla comprensione dell'intenzionalità in bambini di 5 e 6 anni.* Paper presentato, come relatrice, al Convegno nazionale Interazione Bambini-Robot 2021 (IBR21) – Evento online, 13-14 Aprile 2021, Università Milano-Bicocca.
7. Pelizzari, F., Marangi, M., Rivoltella, P.C., Peretti, G., Villani, D., Massaro, D. *Coding e infanzia, tra gioco e apprendimento*. Paper presentato al Convegno nazionale Interazione Bambini-Robot 2021 (IBR21) – Evento online, 13-14 Aprile 2021, Università Milano-Bicocca.
8. Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P.L., Massaro, D., Marchetti, A. (2020). *Come i bambini pensano alla mente del robot: il ruolo dell'attaccamento e della*

Teoria della Mente nell'attribuzione di stati mentali ad un agente robotico. Sistemi Intelligenti.

9. Peretti, G., Rossini, G., Goldini, V., Marchetti, A., Di Dio C., Massaro, D.(Settembre, 2020). *La relazione adulto-bambino al tempo di Covid-19: la percezione della Didattica a Distanza.* Intervento presentato a Giornata di studio del Dipartimento: "La psicologia di fronte al COVID-19: ricerca e riflessioni applicative". Evento online, Università Cattolica del Sacro Cuore. 18 Febbraio 2021.
10. Manzi, F., Peretti, G., Di Dio, C., Cangelosi, A., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., Marchetti, A. (2020). *Children's attribution of mental state to humanoid robots: Nao vs Robovie.* Paper presentato alla 25th Annual CyberPsycology, CyberTherapy & Social Networking Conference (Milan).
11. Peretti, G., Rossini, G., Goldini, V., Marchetti, A., Di Dio C., Massaro, D.(Settembre, 2020). *La relazione adulto-bambino al tempo di Covid-19: la percezione della Didattica a Distanza.* Poster presentato alle Giornate di Studio Emergenza Covid-19. Ricadute evolutive ed educative. AIP della Sezione di Psicologia dello Sviluppo e dell'Educazione, 15-16 Settembre, Bari.
12. Di Dio, C., Manzi F., Peretti, G., Cangelosi, A., Harris, P., Massaro D., & Marchetti A (Settembre, 2019). *Mi posso fidare di te? Dall'interazione bambino-robot alle relazioni di fiducia.* Paper presentato al XXXII Congresso AIP Sezione di Psicologia dello Sviluppo e dell'Educazione, Napoli, 23-25 Settembre, 2019.
13. Di Dio, C., Manzi F., Peretti, G., Cangelosi, A., Harris, P., Massaro D., & Marchetti A (Settembre, 2019). *Esplorare fiducia e attaccamento nella relazione tra robot e bambini di 3 anni.* Paper presentato al XXXII Congresso AIP Sezione di Psicologia dello Sviluppo e dell'Educazione, Napoli, 23-25 Settembre, 2019.
14. Di Dio, C., Manzi F., Peretti, G., Cangelosi, A., Harris, P., Massaro D., & Marchetti A. (Giugno, 2019). *Exploring Trust In The Human-Robot Interaction In Preschool And School-Age Children.* Paper presentato alla 19th Conference on the European Association of Developmental Psychology, Athens, August, 29th– September 1st.
15. Di Dio, C., Manzi F., Peretti, G., Cangelosi, A., Harris, P., Massaro D., & Marchetti A. (Giugno, 2019). *Fiducia e attaccamento nella relazione tra bambini di 3 anni e un robot umanoide*. Paper presentato alla 1st Conferenza sull'Interazione bambino-robot, Milano, 12-13 Giugno, 2019.

16. S. Maggiolini, G. Peretti, *La Sindrome di Asperger al femminile. Problematiche ed abusi nelle relazioni sentimentali*, Autismo Erikson, Trento Gennaio 2017, Vol. 15, n.1 pp. 29-41.

Acknowledgments

L’inizio di questo percorso, per me, risulterà sempre associato al 31 agosto 2018. In quel giorno, dal cielo solo apparentemente grigio e dal clima elettrico, sono accaduti due eventi profondamente trasformativi: l’invio della domanda di ammissione al concorso di dottorato e un trapianto. Da quel giorno, nuovi passi per percorrere nuove vie e, su queste vie, numerose persone che hanno saputo accogliere dubbi, incertezze, perplessità, ansie e gioie. Per questo motivo, vorrei ringraziare chi ha fatto parte di questo mio percorso di dottorato.

Vorrei innanzitutto ringraziare la Professoressa Antonella Marchetti per la sua professionalità, per i suoi insegnamenti, per la sua capacità di saper cogliere in ognuno di noi il lato migliore e per essere una guida ferma e sicura anche nei momenti più difficili.

Un grande grazie a Cinzia, per essere stata un vero mentore in questo percorso e per avermi spronato a non guardarmi solo come un’allieva, ma elaborare un mio pensiero critico, essere autorevole, superare tutti i miei limiti. Grazie, inoltre, per aver saputo utilizzare le giuste parole in ogni occasione di confronto e lavoro.

Grazie a Davide, per la sua curiosità scientifica, per la sua passione al lavoro e per avermi dato la possibilità di sperimentarmi attraverso la realizzazione di materiale didattico, preparazione esami e lezioni.

Ringrazio Annalisa per la sua profonda passione verso l’insegnamento e il mondo accademico, passione che è giunta alla me studentessa magistrale. Il desiderio di proseguire è partito da quelle lezioni in Sant’Agnese.

Grazie a Silvia Maggiolini, per aver creduto in me.

Grazie a Daniela Villani, per la condivisione di esperienza e professionalità.

Ringrazio Teresa, per avermi accolta con un “Benvenuta!” che trasmetteva sicurezza, tranquillità, gioia sincera. Per aver condiviso caffè e ansie, incertezze sul futuro, viaggi mirabolanti su pulmini dalla tenuta incerta e lezioni di statistica.

Grazie a Federico, per la condivisione dell’esperienza pregressa di dottorando e per le numerose, rocambolesche, avventure affrontate durante le somministrazioni come, giusto per citarne qualcuna: la macchina a secco, le corse in metro con NAO, i viaggi in macchina dietro a pannelli per progetti di ricerca, l’acquisto disperato di figurine, i database matti e disperatissimi. Il tutto contrassegnato da ogni tipologia di condizione climatica (vento, neve, acqua, sole).

Un grazie a Betta, per il suo essere diretta, vera, professionale e di supporto nei momenti di bisogno che hanno contrassegnato questo percorso di dottorato.

Un grazie a Laura e Gisella, per avermi dato la possibilità di pensarmi in una veste nuova, da “senior”, e trasmettere loro quanto da me appreso.

Grazie ad Alessia, amica dai tempi della ginnastica, per avermi sostenuta in ogni momento di questo percorso e per continuare a condividere con me i momenti di crescita più importanti.

Grazie a Eleonora, per il suo continuo supporto e per la condivisione di sogni comuni.

Grazie a Monica, Sara e Serena per la profonda stima che ci lega dai banchi dell’università.

Grazie a Laura, Andrea, Damiano, Lucia, Daniele, Alessia, Federico e Martina per avermi accolta sospendendo ogni forma di giudizio, ma vedendomi semplicemente come un’amica.

Grazie a Simona, Davide, Lorenzo, Gaudenzio, Eleonora, Andrea, Andrea e Renato per le risate, i confronti e tutti i momenti di condivisione.

Grazie a Maria Vittoria, Valeria, Federico e Alberto per il loro esserci dai tempi della scuola, per aver visto ogni momento importante.

Grazie a Claudia, per avermi portata a fare il primo giro esplorativo della Cattolica. Non ho ancora smesso da allora!

Grazie a Maria Teresa e Sergio per avermi accolta all’interno della loro famiglia, insieme a Kira, così come Mirca e Gildo.

Ringrazio la mia famiglia, in particolare un grande grazie va ai miei nonni Pierluigia, Giuseppina e Augusto per lo sforzo impiegato per cercare di spiegare agli altri cosa io stessi facendo. Grazie per esserci ancora in questo mio cammino.

Un grande, immenso, grazie va ai miei genitori, Lorella e Giuseppe. Per il loro profondo amore, per il loro esserci sempre in ogni momento della mia vita. Per essere guide sicure in percorsi di vita non sempre facili, a volte tortuosi e incerti, ma sempre densi di affetto e sostegno. Per essere radici.

Non posso non ringraziare Tabata, fedele compagna dal pelo chiaro e morbido, sempre presente durante esami, conferenze, lezioni. Per i suoi occhi contenenti meraviglia e curiosità.

Infine, un grazie pieno di amore, stima e passione a Davide, il mio Davide. Sarebbe difficile condensare in poche righe tutto l’amore, il supporto, la condivisione e la cura che hai avuto nei miei confronti durante questi anni di dottorato. Sei stato base sicura nei momenti di incertezza, mi hai sostenuta e incoraggiata ad andare avanti. Hai metabolizzato i miei grovigli interiori e me li hai restituiti in una veste più comprensibile. La tua presenza mi ha consentito di comprendere meglio la mia essenza. Grazie per tutto quello che mi hai donato e per quello che costruiremo insieme.

Alle nuove avventure!

Curriculum Vitae

Giulia Peretti is a member of the Research Unit on Theory of Mind at Università Cattolica del Sacro Cuore in Milan, Italy.

In 2015 she graduated with a BSc in Science of Education at Università Cattolica del Sacro Cuore in Milan, with a thesis titled "La sindrome di Asperger al femminile. Problematiche e abusi nelle relazioni sentimentali", under the supervision of Professor Silvia Maggiolini

In 2017 she obtained an MSc in Pedagogy, writing a thesis titled "La relazione empatica medico-paziente: un approfondimento psicologico", under the supervision of Professor Annalisa Valle and Professor Davide Massaro.

Previous experience in research and writing her thesis work has developed a passion for research, and evolved into a PhD in the Unit directed by Professor Antonella Marchetti.

Her research interests focus on children robot-interaction and educational robotic.

She is a member of the International Society for the Study of Behavioral Development (ISSBD), the Italian Association of Psychology (AIP) and the Italian Association for Research and Intervention in Psychopathology of Learning (AIRIPA).